Post-translational environmental switch of RadA activity by extein–intein interactions in protein splicing

Natalya I. Topilina[1], Olga Novikova[1], Matthew Stanger[1], Nilesh K. Banavali[2] and Marlene Belfort[1,*]

[1]Department of Biological Sciences and RNA Institute, University at Albany, 1400 Washington Avenue, Albany, NY 12222, USA and [2]Laboratory of Computational and Structural Biology, Division of Genetics, Wadsworth Center, NYS Department of Health and Department of Biomedical Sciences, University at Albany, CMS 2008, Biggs Lab, Empire State Plaza, PO Box 509, Albany, NY 12201-2002, USA

ABSTRACT

Post-translational control based on an environmentally sensitive intervening intein sequence is described. Inteins are invasive genetic elements that self-splice at the protein level from the flanking host protein, the exteins. Here we show in *Escherichia coli* and *in vitro* that splicing of the RadA intein located in the ATPase domain of the hyperthermophilic archaeon *Pyrococcus horikoshii* is strongly regulated by the native exteins, which lock the intein in an inactive state. High temperature or solution conditions can unlock the intein for full activity, as can remote extein point mutations. Notably, this splicing trap occurs through interactions between distant residues in the native exteins and the intein, in three-dimensional space. The exteins might thereby serve as an environmental sensor, releasing the intein for full activity only at optimal growth conditions for the native organism, while sparing ATP consumption under conditions of cold-shock. This partnership between the intein and its exteins, which implies co-evolution of the parasitic intein and its host protein may provide a novel means of post-translational control.

INTRODUCTION

The existence of inteins, protein-splicing elements, at the active site of critical proteins suggests their regulatory potential (1). Inteins are mobile genetic elements that invade protein-coding genes at the DNA level, whereas at the protein level they catalyze self-removal from the host protein, the exteins. Although this process of protein splicing occurs spontaneously, a role for an extein–intein partnership in controlling activity of the host protein, suspected on the basis of enigmatic intein distribution, is largely unexplored.

Although inteins exist in proteins with diverse functions, proteins involved in DNA metabolism, such as polymerases, helicases, recombinases, topoisomerases and ribonucleotide reductases are the most common hosts for inteins (1). Adenosine triphosphatase (ATPase) domains are particularly common intein insertion points that are found in several classes of proteins including recombinases and helicases. Invasion of intein DNA often occurs into regions encoding conserved protein domains that are critical to protein function, such as a catalytic center, a ligand binding site or an interaction surface. This intein localization may be explained by the specificity of the intein's mobility apparatus, the homing endonuclease domain, for conserved sequences (2). An alternative explanation may be that inteins become more readily fixed in the population by occupying conserved sites (3,4). Alternatively, we proposed that the presence of inteins in some conserved motifs might be explained by an adaptive, regulatory role of inteins (1,5).

Protein splicing is a naturally occurring post-translational autoprocessing event, where the intein performs a series of autocatalytic peptide bond rearrangements. The mechanism of splicing for canonical inteins starts with two transesterification steps catalyzed by the first residue of the intein (Cys or Ser) and the first residue of the C-extein (Cys, Ser or Thr) (for review see (6–8)). A resulting branched intermediate, with the N-extein and the intein connected to the C-extein, is resolved by cyclization of the intein's C-terminal conserved Asn, with release of the intein. Finally, the thioester bond connecting the ligated exteins is converted to a peptide bond, leaving a scarless protein. Several conserved amino acid residues within the intein, including His residues, modulate activity of the catalytic residues. Although extein residues flanking the intein can also influence the rate of splicing, this extein

*To whom correspondence should be addressed. Tel: +1 518 437 4466; Fax: +1 518 442 4354; Email: mbelfort@albany.edu

effect was assumed to be limited to amino acid residues immediately proximal to the intein (9–11) and there is no known role for remote extein residues.

Conditional protein splicing (CPS) is consistent with the hypothesis that some inteins act as post-translational regulators of gene expression (1). CPS occurs in the presence of a particular environmental trigger, such as redox stress, temperature or pH (5,12–18). The existence of environmentally sensitive inteins hints at intein adaptation to the intracellular niche by development of a post-translational regulatory response. In these models an intein is thought to act as an environmental sensor and the role of exteins has been largely ignored (5,12–18). We are prompted to ask if exteins play a more substantive role and if three-dimensional (3D) extein–intein interactions might act in intein regulation, affecting intein splicing and thereby extein function, as a novel form of post-translational control.

To probe this hypothesis, we chose the intein-containing RadA protein from the hyperthermophilic archaeon *Pyrococcus horikoshii* (*Pho* RadA) for the following reasons. First, the intein is in the adenosine triphosphate (ATP) binding site (P-loop) of the RadA ATPase domain, which is the most common site for intein occupancy (1). Second, the RadA protein belongs to the conserved recombinase superfamily composed of bacterial RecA, archaeal RadA or Rad51, and eukaryal Rad51, all of which share a structurally conserved core ATPase domain with P-loop (19). Third, in contrast to strong transcriptional regulation of bacterial and eukaryal recombinase expression (20,21), hyperthermophilic RadA proteins were proposed to be regulated mostly post-translationally (22). Fourth, the RadA intein splices robustly in the context of different non-native exteins (23). Finally, the *Pho* RadA intein is a mini-intein of known structure, lacking the endonuclease domain, which simplifies RadA precursor modeling (24). In this work we discovered that the intein functions in a partnership with its native exteins in 3D space to regulate splicing in an environment-dependent manner, being responsive to temperature and solution conditions. Thus, protein splicing provides a thermal switch allowing full activity of the RadA protein only at the elevated temperature corresponding to the growth temperature of its native host. Thereby, superimposed on thermally regulated recombinase activity (25–30), protein splicing may impose an additional level of post-translational control, sparing ATP consumption at suboptimal temperatures and converting the inactive RadA zymogen to active RadA protein at the optimal growth temperature of the host.

MATERIALS AND METHODS

Computational analysis

The amino acid sequences used in this study were obtained from the Protein database at the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) and InterPro database (www.ebi.ac.uk/interpro/). Iterative and complementary searches were performed to identify and obtain sequences for the RadA/RecA inteins. We used the NCBI Protein database and the key-word search queries of the following composition: ('*Archaea*'[Organism] AND RadA[All Fields] AND intein[All Fields]) or ('*Bacteria*'[Organism] AND RecA[All Fields] AND intein[All Fields]), to search for RadA proteins with annotated inteins in archaea or all RecA proteins with annotated inteins in bacteria. Next, we performed a complementary search in the InterPro database which allows access to annotated proteins within a given protein family simultaneously. The family 'DNA recombination/repair protein RadA' (InterPro accession number: IPR011938) was represented by 291 proteins from archaea and eukaryota. By filtering species based on RadA domain organization, also through the InterPro database, the list of species with RadA proteins containing an intein-like Hint domain N-terminal (InterPro: IPR003587), Hint domain C-terminal (InterPro: IPR003586), and/or Homing endonuclease (InterPro: IPR027434) was obtained, and compared with results of the key-word search from the NCBI Protein database. The family 'DNA recombination and repair protein RecA' (InterPro: IPR013765) contained 22,046 proteins from bacteria. Using the same domain-identification strategy as for archaea, the bacterial RecA proteins with inteins were identified and compared. Finally, we performed a series of iterative blastp and tblastn analyses (both with default parameters) using the identified inteins as queries. This last step was especially useful in cases when genomes were not annotated or assembled. The identified inteins are listed in Table 1. Multiple protein alignments were performed by ClustalW (31) and edited manually. Phylogenetic analysis was performed using the Neighbor-Joining method in the MEGA5 program (32).

Bacterial strains and growth conditions

All strains and plasmids used in this study are listed in Table 2. *Escherichia coli* DH5α, BL21 Star (DE3) and ArcticExpress (DE3) strains were grown in Luria Broth (LB) medium with aeration. Where appropriate, the media contained ampicillin (100 μg/ml) or chloramphenicol (25 μg/ml). Electroporation of *E. coli* was performed with a Gene Pulser apparatus (BioRad). Transformants were recovered in SOC media (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) for 1 h at 37°C with aeration.

Plasmid methodology, enzymes and oligonucleotides

Plasmid DNA was isolated and purified using a QIAprep Spin Miniprep Kit (Qiagen). The enzyme digests and polymerase chain reaction (PCR) fragments were visualized by electrophoresis in 0.7% (w/v) agarose gels stained with ethidium bromide. DNA fragments were purified from agarose gels with QIAquick Gel Extraction Kit (Qiagen). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (NEB) and used as described in manufacturer protocols. The list of oligonucleotides used in this study is in Table 3. The sequences of all fragments generated by PCR were verified.

Construction of plasmids

Construction of the plasmid pMIG-RadAi, carrying the *Pho* RadA intein gene in foreign exteins, was accomplished

Table 1. List of archaea and bacteria with inteins in their RadA/RecA proteins

Kingdom/phylum	Species and strains (abbreviation used[a])	Acc. number[b]	Intein insertion (amino acid coordinates)
Archaea	*Thermococcus sibiricus* MM 739 **(Tsib)**	C6A058	c1 (154–321)
	Thermococcus litoralis DSM 5473 **(Tlit)**	H3ZR24	c1 (151–633)
	Thermococcus zilligii **(Tzill)**	WP_010479404[c]	c1 (148–629)
	Thermococcus kodakaraensis ATCC BAA-918 **(Tkod)**	Q5JET4	c1 (150–631)
	Pyrococcus sp. ST04 **(PspST04)**	I3RC67	c1 (149–568)
	Pyrococcus yayanosii CH1 / JCM 16557 **(PyayCH1)**	F8AHL2	c1 (150–579)
	Pyrococcus horikoshii ATCC 700860 **(Pho)**	O58001	c1 (153–324)
	Archaeoglobus profundus DSM 5631 **(Aprof)**	D2REA7	c1 (120–598)
	Nanoarchaeota archaeon SCGC AAA011-L22 **(Narch)**	WP_018203495[c]	e (278–774)
Bacteria			
Actinobacteria	*Verrucosispora maris* AB-18–032 **(Vmaris)**	F4FCR9	b (206–569)
	Saccharomonospora xinjiangensis XJ-54 **(SxinjiXJ-54)**	I0V674	b (206–572)
	Saccharomonospora paurometabolica YIM 90007 **(SpaurYIM-90)**	G4IYN5	b (206–572)
	Thermomonospora curvata ATCC 19995 **(Tcurvata)**	D1AAC0	b (205–568)
	Isoptericola variabilis 225 **(Ivar225)**	F6FSE8	b (206–569)
	Thermobifida fusca YX **(TfuscaYX)**	Q47RS6	b (650–1006), c2 (96–517)
	Mycobacterium rhodesiae JS60 **(MrhodJS60)**	G4I8P3	b (207–570)
	Mycobacterium thermoresistibile ATCC 19527 **(Mthermo)**	G7CID5	b (206–570)
	Mycobacterium sp. JDM601 **(MycspJDM601)**	F5YW72	b (206–569)
	Mycobacterium leprae TN **(MlepTN)**	P35901	b (206–570)
	Mycobacterium xenopi RIVM700367 **(MxenopiRIVM)**	I0RZP4	b (206–569)
	Mycobacterium orygis 112400015 **(Morygis)**	M8CJZ6	a (252–691)
	Mycobacterium tuberculosis H37Rv **(MtubH37Rv)**	I6YE90	a (252–691)
	Mycobacterium canettii CIPT 140060008 **(Mcan08)**	L0PY96; L0Q9X3	a (252–691)
	M. canettii CIPT 140010059 **(Mcan59)**	G0TQW8	a (252–691)
	Mycobacterium bovis BCG **(Mbovis)**	P0A5U5	a (252–691)
	M. canettii CIPT 140070017 **(Mcan17)**	L0QX80	b (211–573)
	Mycobacterium africanum GM041182 **(Mafric)**	F8M3J1	a (252–691)
Proteobacteria	*Thiovulum sp.* ES	J0LEE6	b (210–562)
	Acinetobacter baumannii ABNIH1 **(AbaumABMIH1)**	F9IG70	d (216–370)
	A. baumannii IS-143 **(AbaumIS-143)**	K1ED71	d (216–370)
	Enterobacter cloacae BWH 29 **(Ecloac)**	V3IU74	c2 (65–486)
	Enterobacter sp. MGH 26 **(EspMGH26)**	V3NDD1	c2 (65–486)
	Escherichia coli 2780750 **(Ecoli278075)**	M8ZI95	c2 (65–486)
	E. coli 2865200 **(Ecoli286520)**	M8W3V2	c2 (65–486)
	E. coli E1520 **(EcoliE1520)**	E9WNB4	c2 (65–486)
	E. coli HVH 139 (4–3192644) **(EcoliHVH139)**	T7BM04	c2 (65–486)
	E. coli HVH 43 (4–2173468) **(EcoliHVH43)**	T5ZLN3	c2 (65–486)
	E. coli KOEGE 10 (25a) **(EcoliKOEGE1)**	U0BGQ9	c2 (65–486)
	E. coli MP020980.1 **(Ecoli80.1)**	N3JUW8	c2 (65–486)
	E. coli MS 145–7 **(EcoliMS145)**	E1IXC8	c2 (65–486)
	Salmonella enterica 557928 **(Senter)**	V0ELC0	c2 (65–486)
Cyanobacteria	*Oscillatoriales cyanobacterium* JSC-12 **(OcyanJSC-12)**	K8GN85	b (196–550)

[a]The abbreviations used to indicate the corresponding species and strains in the phylogenetic tree in Figure 1A are in bold.
[b]Accession numbers for protein sequences deposited in InterPro Database (http://www.ebi.ac.uk/interpro/), unless indicated otherwise.
[c]Protein sequence accession number in NCBI GenBank Protein Database (http://www.ncbi.nlm.nih.gov/protein).

by inserting the RadA intein sequence with short native exteins (N-ext: EVFGEFGS and C-ext: TQLAHTLAVM) into ClaI and SphI sites between sequences coding for maltose binding protein (MBP) and green fluorescent protein (GFP) in a pACYCDuet-1 plasmid backbone (Table 2). The RadA intein fragment was generated by PCR with primer pair IDT3519/IDT3520 from *P. horikoshii* genomic DNA (a generous gift from K. Mills) using ExTaq DNA polymerase (TaKaRa).

The plasmid carrying the full-length *radA* gene, pFL-RadAi, expressing the *Pho* RadA intein in native exteins, was constructed by amplifying the *radA* from *P. horikoshii* genomic DNA by PCR using primer pair

IDT3184/IDT3185 and inserting the resulting fragment into BamHI and XhoI sites of the expression vector pET45b. Plasmid pRadA, expressing the inteinless version of the RadA protein, was constructed by inverse PCR using primer pair IDT3962/IDT3964 and pFL-RadAi as a template. The CloneAmp HiFi PCR PreMix (Clontech) was used to ensure accuracy and efficiency of the amplification. The In-Fusion HD Cloning Plus kit (Clontech) was used to seal the ends of the amplified plasmid.

The pFL-RadAi-AA plasmid, expressing a splicing-inactive intein in the native exteins and plasmids pM1-pM12, (Table 2), carrying mutated extein sequences, were constructed from pFL-RadAi using site-directed muta-

Table 2. Bacterial strains and plasmids

Strain/plasmid	Relevant characteristics; Comments	Source
***E. coli* Strains**		
DH5α	F⁻ *endAI recAl hsdRl7* (rK⁻ mK⁻) *deoR supE44 thi-J gyrA96 relA*	Gibco-BRL
BL21 Star (DE3)	F⁻ *ompT* hsdS$_B$(r$_B^-$ m$_B^-$) *gal dcm rne131* (DE3)	Life Technologies
ArcticExpress	F⁻ *ompT* hsdS(r$_B^-$ m$_B^-$) *dcm*⁺ Tet^R *gal endA* Hte [*cpn10 cpn60* Gent^R]	Agilent Technologies
Plasmids		
pACYCDuet-1	expression vector, T7 promoter, Cam^R	Novagen
pET-45b(+)	expression vector, T7 promoter, Amp^R	Novagen
pMIG-RadAi	Pho RadA intein flanked by short extein sequences cloned into *Cla*I/*Sph*I sites between MBP and GFP coding sequences in pACYCDuet-1 backbone	Present study
pFL-RadAi	full-length intein-containing *radA*, cloned into *Bam*HI/*Xho*I sites of pET-45b(+)	Present study
pRadA	inteinless *radA* constructed based on pFL-Radi	Present study
pFL-RadAi-AA	full-length intein-containing *radA*, C153A, N324A splicing inactive intein mutant	Present study
pM1	full-length intein-containing *radA*, R358A, E360A, R361A, R363A, E364A helix 1 extein multiple mutant	Present study
pM2	R358A, R361A helix 1 extein multiple mutant	Present study
pM3	E360A, R363A, E364A helix 1 extein multiple mutant	Present study
pM4	R358A helix 1 extein single point mutant	Present study
pM5	E360A helix 1 extein single point mutant	Present study
pM6	R361A helix 1 extein single point mutant	Present study
pM7	E364A helix 1 extein single point mutant	Present study
pM8	R503A helix 1 single point extein mutant	Present study
pM9	E354A extein single point catalytic mutant, impaired ATPase activity	Present study
pM10	Q465A extein single point catalytic mutant, impaired ATPase activity	Present study
pM11	E77A, K79A, E80A extein multiple mutant	Present study
pM12	E57A, K58A, R60A, E61A extein multiple mutant	Present study

Table 3. List of oligonucleotides used in this study

Oligo ID	Sequence (5′ → 3′)	Application
IDT3519	ggggcatgcaacggaagtctttggggagttcggtagtggg	RadA intein amplification for MIG-RadAi
IDT3520	cccatcgatcattaccgcaagggtatgggctagctgagt	RadA intein amplification for MIG-RadAi
IDT3184	ggggatccggtgatgatcatggttaagaaaggg	FL-RadAi amplification
IDT3185	gggctcgagctaatcttcaattccttctccg	FL-RadAi amplification
IDT3208	agttcggtagtgggaaggcctttgctagggataccg	C153A mutagenesis
IDT3209	ccaatggacttgttctccatgctactcagctagcccatacc	N324A mutagenesis
IDT3962	actcagctagcccataccctttgcggtaatg	Intein removal form FL-RadAi
IDT3964	ggggagttcggtagtgggaagactcagctagcccataccttgcggtaatg	Intein removal form FL-RadAi
IDT3660	caggatctaatcccctattcttcgcaatcgctgcaatcgccgcaggtgcgaaggtgttttcagtgtcaatccaaatta	R358A, E360A, R361A, R363A, E364A mutagenesis
IDT3661	taatttggattgacactgaaaacaccttcgcacctgcggcgattgcagcgattgcgaagaatagggga ttagatcctg	R358A, E360A, R361A, R363A, E364A mutagenesis
IDT3662	gattgacactgaaaacaccttcgcacctgaggcgattagagagattgcgaagaat	R358A, R361A mutagenesis
IDT3663	attcttcgcaatctctctaatcgcctcaggtgcgaaggtgttttcagtgtcaatc	R358A, R361A mutagenesis
IDT3664	tgaaaacaccttcagacctgcgaggattgcagcgattgcgaagaataggggat	E360A, R363A, E364A mutagenesis
IDT3665	atcccctattcttcgcaatcgctgcaatcctcgcaggtctgaaggtgttttca	E360A, R363A, E364A mutagenesis
IDT3666	ctctctaatcctctcaggtgcgaaggtgtttttcagtgtca	R358A mutagenesis
IDT3667	tgacactgaaaacaccttcgcacctgagaggattagagag	R358A mutagenesis
IDT3668	gcaatctctctaatcctcgcaggtctgaaggtgtttt	E360A mutagenesis
IDT3669	aaaacaccttcagacctgcgaggattagagagattgc	E360A mutagenesis
IDT3670	aaacaccttcagacctgaggcgattagagagattgcgaag	R361A mutagenesis
IDT3671	cttcgcaatctctctaatcgcctcaggtctgaaggtgttt	R361A mutagenesis
IDT3672	ccctattcttcgcaatcgctctaatcctctcaggt	E364A mutagenesis
IDT3673	acctgagaggattagagcgattgcgaagaataggg	E364A mutagenesis
IDT3674	gaagggtaagggaggaaaggccggtagctaggttaatagat	R503A mutagenesis
IDT3675	atcattaacctagctaccgcctttcctcccttacccttc	R503A mutagenesis
IDT3956	ctcaggtctgaaggtgtttgcagtgtcaatccaaattac	E354A mutagenesis
IDT3957	gtaatttggattgacactgcaaacaccttcagacctgag	E354A mutagenesis
IDT3958	caggcctggcctgaactgcgttcgttacaaagacag	N465A mutagenesis
IDT3959	ctgtctttgtaacgaacgcagttcaggccaggcctg	N465A mutagenesis
IDT4120	cttcactgattcccgcaaccgctgcaagtgctatagggatgcaaccgct	E77A, K79A, E80A mutagenesis
IDT4121	cttcactgattcccgcaaccgctgcaagtgctatagggatgcaaccgct	E77A, K79A, E80A mutagenesis
IDT4122	tctagagtatcaaatcccgccgccgcaagcgctgcagcagtagcaggacccactc	E57A, K58A, R60A, E61A mutagenesis
IDT4123	gagtgggtcctgctactgctgcagcgcttgcggcggcgggatttgatactctaga	E57A, K58A, R60A, E61A mutagenesis

genesis with appropriate mutagenic primers (Table 3). All primers were designed using the QuikChange Primer Design Program (Agilent Technologies; www.genomics. agilent.com). The desired codon changes were introduced using QuikChange Lightning Site-Directed Mutagenesis Kit and QuikChange Lightning Multi Site-Directed Mutagenesis Kit from Agilent Technologies.

Protein expression and purification

Escherichia coli BL21 Star (DE3) (Life Technologies) carrying pMIG-RadAi was grown overnight at 37°C in LB media supplied with chloramphenicol, diluted 100-fold into fresh LB media and grown at 37°C to OD_{600} ~0.5. Protein expression was induced with 0.5 mM IPTG followed by 3 h incubation at 15 and 20°C. Induced cultures were collected by centrifugation and stored at −80°C. Pellets were thawed on ice, suspended in Lysis Buffer (50 mM Tris pH 8.0, 2 mM EDTA, 10% [v/v] glycerol) in one-fifth volume of the cell culture. Cells were lysed by sonication and debris was removed by centrifugation. Cleared lysates were loaded on a 12% Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel (33) without sample boiling. After electrophoresis, proteins of interest (MIG-RadAi precursor, 94 kDa; MG ligated exteins, 74 kDa) were visualized by in-gel GFP fluorescence using a Typhoon 9400 (GE Healthcare Life Sciences).

Preliminary experiments demonstrated that FL-RadAi precursor exists in two conformers when overexpressed in *E. coli* BL21 Star (DE3) at 30°C for 3 h, as confirmed by mass spectroscopy, (data not shown), similar to the RadA protein from thermophilic *Sulfolobus solfataricus* (34) and predominantly in one form when overexpressed in *E. coli* BL21 Star (DE3) at 46°C for 3 h. Thus, the FL-RadAi precursor protein expression in *E. coli* BL21 Star (DE3) was performed at 46°C in all subsequent experiments. *E. coli* ArcticExpress (DE3) cells were used for protein expression at 15 or 12°C for 18 h.

All His6-tagged proteins were purified using metal chelating chromatography under native conditions at 4°C. The 1 ml HisTrap HP (GE Healthcare Life Sciences) columns containing Ni Sepharose High Performance were used following the manufacturer instructions. Ni-NTA-purified proteins were separated by 12% SDS-PAGE gel and visualized by staining with Coomassie brilliant blue (FL-RadAi precursor, 62 kDa; ligated exteins, 42 kDa; intein, 20 kDa; N-extein, 19 kDa; C-extein, 23 kDa).

Splicing assays

To test a temperature effect on splicing *in vitro*, the Ni-NTA-purified FL-RadAi precursor protein and the extein mutant precursors were incubated at different temperatures (25–85°C with 10°C increments) for 15 min, 30 min, 1 h, or 2 h and then analyzed on 12% SDS-PAGE gel with Coomassie blue staining. To test solvent/solute effects, aliquots of the Ni-NTA-purified FL-RadAi precursor protein in a 96-well plate were supplemented with 96 reagents (1/20 dilution) from the Solubility and Stability Screen (Hampton Research), incubated overnight at 35 or 25°C and analyzed by 12% SDS-PAGE. Those compounds

that were found to facilitate intein splicing at 25°C were incubated with purified FL-RadAi precursor at 25°C for 30 min and then analyzed by 12% SDS-PAGE. To calculate the progression of intein splicing, protein bands on Coomassie blue-stained gels were quantified using ImageJ (www.imagej.nih.gov/ij/). The splicing efficiencies were calculated as follows: LE/(LE + PRE) × 100, where LE represents the amount of ligated exteins and PRE, the amount of precursor. Considering that intensity of Coomassie blue-stained bands depends on a protein size the signal from LE band was adjusted to accommodate the size difference between the ligated exteins and the precursor.

Circular dichroism measurements

Far-UV circular dichroism (CD) spectra were measured in a 0.1 cm temperature-controlled quartz cell using a Jasco J-720 spectropolarimeter (Tokyo, Japan) at various temperatures. Bandwidths of 1 nm and scan speed of 100 nm/min were utilized. Three spectra accumulations were averaged. The absorbance of the sample in the CD cell did not exceed 1.5, which was well within the recommended absorbance range (<3). The photomultiplier voltage was recorded in each run and did not exceed 600 V, which was also within the recommended range.

ATPase activity of RadA

The inteinless RadA protein and the splicing inactive FL-RadAi-AA protein precursor were Ni-NTA-purified as described above, exchanged into assay buffer (25 mM Tris pH 7.5, 10 mM $MgCl_2$) pre-treated with PiBind resin (Innova Biosciences) to remove contaminating inorganic phosphate. ATPase activities of RadA and FL-RadAi-AA (2 μM) were analyzed using the High Throughput Colorimetric ATPase Assay kit (Innova Biosciences) following the manufacturer instructions. Enzymatic reactions were performed at different temperatures (25–85°C with 10°C increments) for 30 min. Different concentrations of the substrate, ATP, was added in the presence of 0.2 μg/ml of M13mp18 ssDNA (NEB). To eliminate the impact of temperature-dependent non-enzymatic ATP hydrolysis, protein-free controls were used for each data point. All experiments were performed in triplicate.

RadA precursor modeling

Homology models for the RadA extein and intein were generated separately based on the closest templates for the extein PDB ID: 2ZUB (35) and for the intein PDB ID: 4E2T (24). The initial sequence alignment required for homology modeling was performed using ClustalW (31) (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and was manually corrected for proper alignment of conserved intein sequences and correcting for artificial mutations by also taking into consideration a structural alignment performed using Phyre (36). Homology models were generated using MODELLER (37), with the DOPE (38) and GA341 (39) energy functions, and the best scoring model was chosen for further optimization. The intein homology model was

manually positioned such that its ends were as close as possible to its connection with the extein homology model using the molecular visualization program Visual Molecular Dynamic (VMD) (40). The two loops connecting the intein to the N-extein and the C-extein were then remodeled using the loop prediction program Loopy (41) to get an initial continuous precursor structural model. This model showed the first residue of the intein and the first residue of the C-extein to be far apart, which is not a feasible conformation for the splicing reaction. Therefore, they were brought closer using the following protocol, which used restrained energy minimizations and molecular dynamics (MD) simulations. A harmonic distance restraint was applied between the centers-of-mass of the two separated residues with a force constant of 10 kcal/mol/Å^2. The distance minimum of the restraint was gradually reduced to 6 Å in 2 Å decrements to achieve the optimized precursor model. At each distance minimum, the structure was optimized *in vacuo* using 500 steps of Steepest Descent (SD) minimization (42), followed by 100 steps of Adopted Basis Newton Raphson (ABNR) minimization (43) and then by Langevin dynamics for 2 ps with a friction coefficient of 60 ps at a temperature of 200 K, followed by another round of 500 steps of SD minimization and 100 steps of ABNR minimization. These minimizations and MD simulations were performed using the program CHARMM (44,45) and the CHARMM22 protein force field (46).

The optimized precursor models were analyzed for vacuum interaction energies between catalytically involved intein amino acid residues (C153, H245, H312, H323 and N324) and all other residues in the precursor. This was done by looping over all the residues and calculating the extein sidechain:intein sidechain and extein sidechain:intein backbone components of these interaction energies to get an understanding of the approximate influence of each extein amino acid sidechain on the intein catalytic residues.

RadA model assessment

A few different Model Quality Assessment Programs were used to assess the RadA precursor wild-type and mutant models with the scores shown in Supplementary Table S1. The Swiss-Model Qmean server (http://swissmodel.expasy.org/qmean/cgi/index.cgi) was used to obtain the Qmean and Qmean derived Z-score (47). QMEAN score is a global score for each model that scores its reliability from values ranging between 0 and 1, with higher values indicative of greater accuracy. The associated Z-score relates this QMEAN score of each model to the scores of a non-redundant set of high-resolution X-ray structures of similar size with ideal values being close to 0 (i.e. showing least deviation from the X-ray set scores). ProSA-Web Z-scores (48) were obtained using the ProSA-Web server (https://prosa.services.came.sbg.ac.at/prosa.php). These Z-scores are mostly within the range −7 to −12 for X-ray structures of proteins of length close to 500 residues. The Molprobity score (49) of each model was obtained using the Joint Center for Structural Genomics' Quality Control Check version 3.1 (http://smb.slac.stanford.edu/jcsg/QC/). The Molprobity score reflects the X-ray resolution at which its individual scores would be the expected values such that lower

values indicated better models. The Protein Structure Validation Server (PSVS1.5) (50) was used to obtain Z-scores for Verify3D (51) and Procheck G-factor phi/psi (52) assessments. These Z-scores are derived from comparison of the Verify3D and Procheck G-factor phi/psi scores of each model to the scores of a set of 252 X-ray crystal structures with resolution ≤1.80 Ang., R-factor ≤0.25 and R-free ≤0.28. As with other Z-scores, values closer to 0 indicate models that better agree with the high quality X-ray structure dataset.

RESULTS

Diverse intein insertion sites in functional domains of the RadA/RecA proteins

To characterize the diversity and distribution of inteins in the archaeal RadA and bacterial RecA proteins, we performed computational mining (Figure 1 and Table 1). In addition to 72 inteins reported in the intein database In-Base (53), we identified 26 new inteins. An interesting aspect of intein distribution is the exceptionally wide span of organisms containing RadA/RecA inteins, given only a few other examples of inteins present in orthologous genes in both archaea and bacteria (1). There are five distinct intein insertion points in RadA/RecA proteins, a–e, including two newly identified ones (54,55). The RadA/RecA inteins vary greatly between archaeal and bacterial species, based on sequence similarity and their insertion points in RadA/RecA proteins, suggestive of independent events of intein invasion (Figure 1 and Table 1).

The intein insertion points are clustered in important functional domains, at the monomer–monomer interaction interface (sites a, b and d) and the ATP-binding site (sites c and e) (Figure 1B). Interestingly, insertion point c exists in both archaeal RadA and bacterial RecA proteins (designated c1 and c2, respectively). This insertion point is located in the conserved active site P-loop of the ATP-binding domain (Figure 1B), which is a hot-spot for inteins in different ATPase-containing proteins in various archaeal and bacterial species (Figure 1A) (1). Although, the comparative analysis demonstrated a high degree of RadA and RecA extein sequence correspondence, there is little similarity between the RadA and RecA inteins at position c, suggesting independent invasion at the identical insertion point in archaea and bacteria (Figure 1C). Given that this site appears to have been repeatedly targeted by different inteins in two domains of life, the small, well-characterized intein in the *Pho* RadA protein was selected as the focus of the present study.

Native exteins inhibit splicing as post-translational thermal switch of RadA activity

To test a potential relationship between intein catalysis and the nature of the extein, we made two precursors with the *Pho* RadA intein with foreign and native exteins. The construct with foreign exteins contains MBP and GFP flanking the intein, to form the MPB-Intein-GFP fusion (MIG-RadAi) (Figure 2A). The MIG-RadAi construct contains short native exteins (8–10 amino acid residues) as extein residues at the splice junctions can modulate splicing by
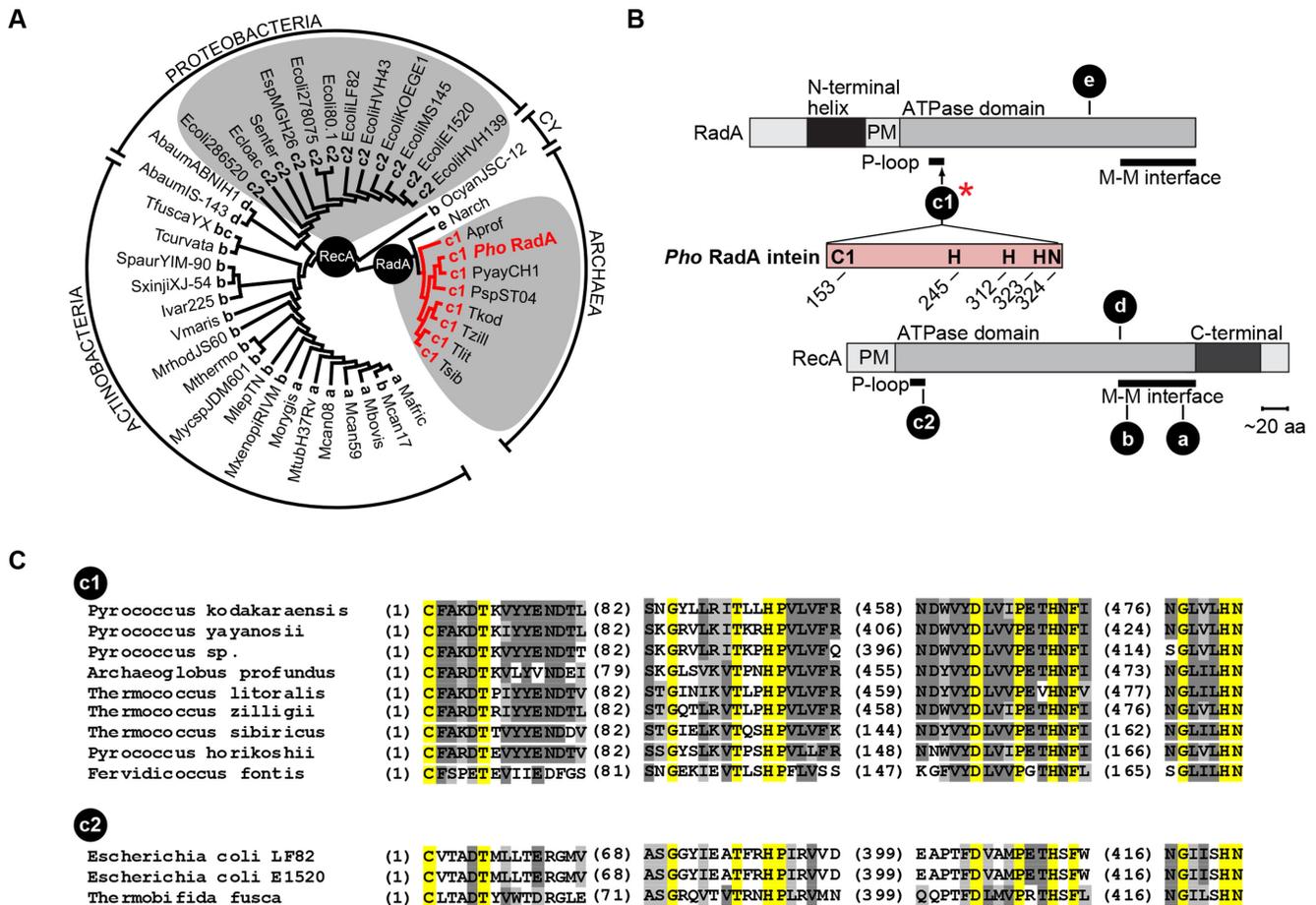
Figure 1. Distribution and diversity of RadA/RecA inteins among archaea and bacteria. (**A**) Neighbor-Joining tree. The tree was constructed based on the multiple alignment of 42 RadA/RecA extein amino acid sequences from archaea and bacteria (proteobacteria, actinobacteria and cyanobacteria—CY). The insertion points (a–e) are indicated together with abbreviated species names, listed in full in Table 1. Letters a–e reflect the chronology of intein discovery (54,55). The species containing an intein in insertion point c (either c1 or c2) are on a gray background. The precursor of the protein of interest, *Pho* RadA (bold, red), carries an intein in insertion point c, subpopulation c1. (**B**) Intein insertion points. Intein locations are shown along RadA (c1 and e) and RecA (a–d) relative to structural and functional domains. The intein of interest, *Pho* RadA, is in insertion point c1 (red asterisk), which is located at the end of the RadA P-loop in the ATPase domain. Conserved, catalytically important amino acid residues and their positions are shown in single-letter code in the intein, as further identified in Figure 4A. Abbreviations: PM, polymerization motif; M–M interface, monomer–monomer interface. (**C**) Multiple sequence alignment of the inteins (splicing domains) from insertion point c. Comparative analysis of the splicing domains from inteins occupying insertion point c in RadA and RecA proteins revealed substantial differences in amino acid sequences between archaeal (insertion c1) and bacterial (insertion c2) inteins indicating independent acquisition of inteins by RadA and RecA in insertion point c.

affecting the electronic properties of intein active site (9–11). The construct with the native exteins contains the full-length RadA protein (FL-RadAi) with the intein at its P-loop (Figures 1B and 2B). Splicing of the RadA intein from both precursors was examined.

MIG-RadAi was overexpressed in *E. coli* and the cellular lysates were used to analyze the extent of MIG-RadAi processing *in vivo*. The protein bands of the MIG-RadAi precursor and the MG (MBP-GFP) splice product were visualized by GFP fluorescence in an SDS-PAGE gel, without boiling the extracts (Figure 2A). The MIG-RadAi precursor processed effectively *in vivo* upon expression at 15 and 25°C for 3 h, with splicing in the 59–81% range. *In vitro*, the MIG-RadAi precursor recovered after induction at 15°C was 94% spliced within 1 h at 25°C (Figure 2A). These experiments indicate high splicing activity of the RadA intein from foreign exteins, consistent with a previous report (23).

In contrast to the MIG-RadAi precursor, the FL-RadAi precursor demonstrated compromised splicing activity *in vivo* (Figure 2B), with no detectable splicing *in vitro* at 25°C over several days (data not shown). Interestingly, catalytic activity of the RadA intein in the context of native exteins was strongly dependent on temperature, showing an increase in splicing at 55°C and reaching a maximum at 75–85°C (Figure 2B; plotted Figure 3B). The different activity of the RadA intein in the context of its native and foreign exteins suggests that the native exteins modulate performance of the intein in a temperature-dependent manner. Moreover, given that the amino acid residues immediately flanking the intein are identical in MIG-RadAi and FL-RadAi, remote residues must be mediating the regulatory extein effect, likely via extein–intein interactions that occur in 3D space.
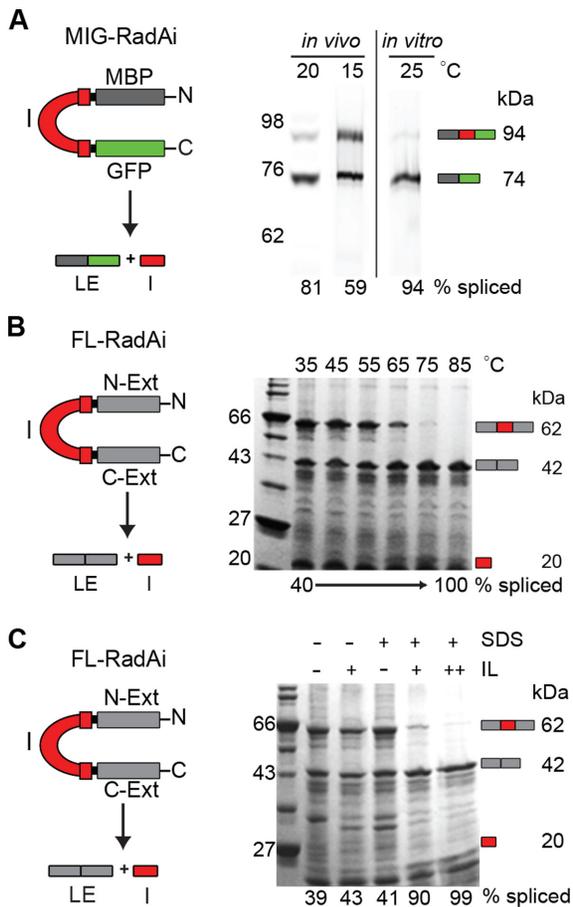
Figure 2. RadA intein splicing is regulated by its native exteins in temperature and solution-dependent manner. (**A**) RadA intein splicing in foreign exteins is efficient even at low temperatures. The RadA intein was cloned into the non-native extein context (pMIG-RadAi; Table 2) and the recombinant protein was overexpressed in BL21 Star (DE3). The MIG-RadAi precursor (left) consists of the RadA intein (I; red) with short native exteins fused to MBP (dark gray) and GFP (green). The precursor (94 kDa) and the ligated exteins (LE; 74 kDa) were visualized by in-gel GFP fluorescence. More than 50% of precursor was spliced *in vivo* at 15°C and >80% was spliced at 20°C. The MIG-RadAi precursor recovered from a 15°C induction spliced efficiently (94% splicing) *in vitro* at 25°C within 1 h. (**B**) RadA intein splicing in the native exteins is inefficient at low temperatures, but efficient at high temperature. In the FL-RadAi precursor (left) the RadA intein (red) is flanked by its native exteins (N-Ext and C-Ext; gray). RadA intein splicing of the Ni-NTA purified FL-RadAi precursor was visualized in a Coomassie blue-stained gel. Accumulation of the spliced intein (I; 20 kDa) and the ligated exteins (LE; 42 kDa) and disappearance of the FL-RadAi precursor (62 kDa) were observed at high temperatures. A plot of the data from Figure 2B is shown in Figure 3B (middle panel). (**C**) Solution effects on RadA intein splicing in the native exteins. RadA intein splicing of FL-RadAi is shown in the presence 1.25% (+) and 2.5% (++) of ionic liquid (IL) 1-butyl-3-methylimidazolium chloride and 0.5% (+) SDS for samples incubated at 25°C for 30 min.

Extein-imposed inhibition of splicing can be modulated by solution environment

If thermal regulation of the RadA intein is mediated by contacts between the exteins and the intein, agents other than temperature might disrupt these extein–intein interactions to facilitate splicing. Denaturants, detergents and solvents are known to activate some enzymes by relaxing

the rigidity of interactions at the active site (56). To test whether this might be the case, we subjected FL-RadAi to a Hampton Solubility and Stability Screen (see 'Materials and Methods' section) which includes a panel of 96 compounds that modulate protein solubility and stability. We determined that several compounds had modest effects (data not shown). Although 1.25% 1-butyl-3-methylimidazolium chloride, a water-miscible ionic liquid (57) and 0.5% SDS, a detergent, had little effects on splicing of FL-RadAi upon incubation at 25°C for 30 min, the two reagents together resulted in a sharp increase in splicing, to 90% (Figure 2C). Doubling the concentration of the ionic liquid allowed splicing to go almost to completion. We hypothesized that the ionic liquid in combination with the detergent acts by reducing the interactions between the native exteins and the intein. Thereby solution environment is capable of releasing intein inhibition in the FL-RadAi precursor at low temperature, indicating that zymogen activation by splicing is also sensitive to non-thermal cues.

Thermal responses of structure and activities of FL-RadAi and RadA

To test if the FL-RadAi precursor might undergo a secondary structure transition that could allow protein splicing, folding of FL-RadAi precursor at different temperatures was characterized by CD spectroscopy. To eliminate changes that might reflect rearrangements during protein splicing, a catalytically inactive FL-RadAi-AA precursor was utilized, where AA designates C153A and N324A mutations at the N- and C- termini of the intein (Figure 1B). The inteinless RadA protein was used as a control to show secondary structure transitions in the ligated exteins and RadA protein plus RadA intein in a 1:1 mixture (RadA + Intein) were used to mimic FL-RadAi in composition, alongside RadA intein alone. Far-UV spectra of the proteins showed temperature-dependent changes in the 25–85°C interval in extein-containing species, FL-RadAi-AA, RadA and the RadA + Intein. However, there were minimal changes associated with RadA intein spectra. CD spectra of the samples containing the RadA exteins showed common temperature-dependent secondary structure rearrangements, similar to those reported for the *Pyrobaculum islandium* RadA (25). Given that the exteins have predominantly α-helical content (35) and the intein comprises mainly β strands (24), we followed changes in ellipticity around the middle of the spectrum in the 217–223 nm interval, which covers characteristic wavelengths for α-helices (222 nm) and β-sheets (218 nm) (Figure 3A and B, top panel). We observed an increase in slope at 55°C for both the FL-RadAi-AA precursor and for the RadA protein, albeit less pronounced for RadA (Figure 3B). This similarity suggests that temperature-dependent rearrangements in the FL-RadAi precursor are related to its extein component.

Considering the temperature-dependence of *Pho* RadA intein splicing, it is interesting that catalytic activity of hyperthermophilic RadA/Rad51 proteins are also regulated by temperature (25–28). We therefore tested the temperature dependence of ATPase activity of the *Pho* RadA protein. Similar to other RecA-like proteins, the inteinless *Pho* RadA shows ssDNA-dependent ATPase activity (25) (Fig-
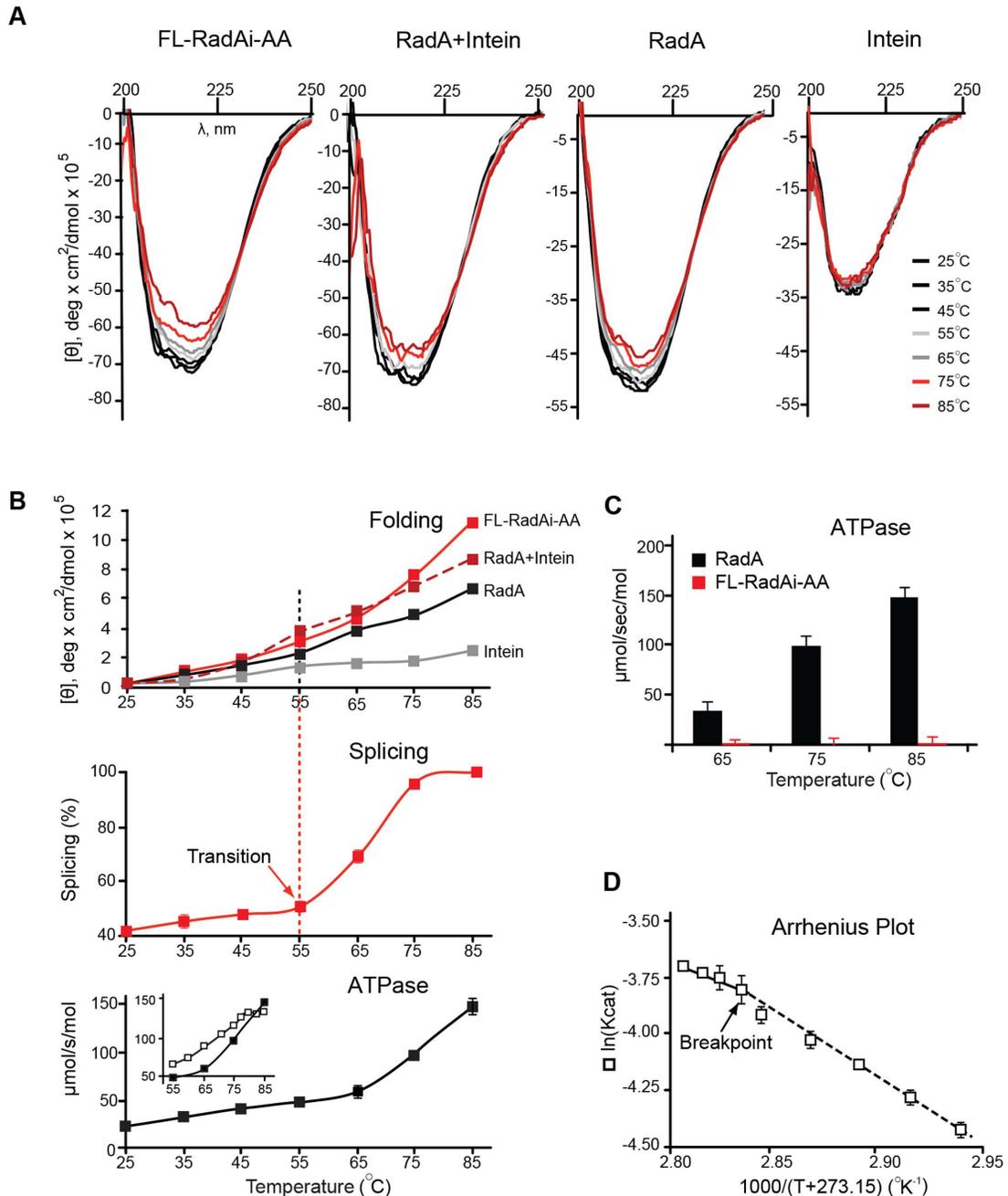
Figure 3. Temperature-dependent structure transition and ATPase activity of the *Pho* RadA protein and the precursor. (**A**) Far UV circular dichroism (CD) spectra. CD spectra of the splicing-inactive FL-RadAi-AA precursor, the RadA protein, RadA + Intein and Intein alone were recorded from 25–85°C with 10°C increments. (**B**) Temperature dependence of secondary structure transitions, splicing and ATPase activity. Top: Ellipticity at 217–223 nm measured for the four protein combinations in panel A. Middle: FL-RadAi splicing plot derived from data in Figure 2B. Splicing increases sharply at the structural transition, above 55°C. Bottom: ATPase activity of RadA. Inset corresponds to data used in panel D. CD data are from the experiment represented in panel A. Splicing and ATPase activity measurements in all panels were performed in triplicate at the temperatures indicated and error bars represent standard deviation. (**C**) ATPase activity of RadA and FL-RadAi at different temperatures. ATPase activity of inteinless RadA (black) increased with temperature while FL-RadAi (red) containing the inactive intein had no detectable ATPase activity. The experiment was performed in triplicate and error bars represent standard deviation. (**D**) *Pho* RadA exhibits biphasic ATPase activity. The Arrhenius plot demonstrates two distinct catalytic modes (55–76 and 76–83°C) with a slope change (Breakpoint) close to 76°C. The experiment was performed in triplicate and error bars represent standard deviation.

ure 3B, bottom). The rate of ATP hydrolysis by RadA rises above temperatures of 65°C. Unlike RadA, the splicing inactive FL-RadAi-AA precursor showed no catalytic activity at elevated temperatures (Figure 3C), consistent with the assumption that intein splicing is required for functional activity of the host protein.

To further analyze the effect of temperature on the rate of ATP hydrolysis by the RadA protein, an Arrhenius plot was generated by measuring ATPase activity at temperatures in the 55 to 80°C interval (Figure 3B, bottom, inset; Figure 3D). *Pho* RadA exhibits biphasic ATPase activity and an Arrhenius plot with a breakpoint at 76°C, with two activation energies, of 12.7 and 22.7 kJ/mol (Figure 3D). This breakpoint in activation energy is 20°C higher than the increase in splicing activity. The breakpoint may therefore correspond to a local rearrangement of amino acid residues at the ATPase active site similar to *P. islandium* RadA (25) rather than to global conformational changes. Given that *P. horikoshii* lives in the temperature range of 70–103°C, the two distinct catalytic modes of RadA below and above the 76°C breakpoint may reflect the state of the protein *in vivo* (25).

FL-RadAi precursor models suggest inhibitory extein–intein interactions

Considering the innate ability of the RadA intein to splice at low temperatures (23) (Figure 2A) and its temperature-dependent splicing from the native precursor, together with the similarity in FL-RadAi-AA and RadA temperature-dependent structure transitions (Figure 3A and B, top panel), we suggest that the temperature sensitivity in the FL-RadAi precursor is attributable to its extein component. We also speculate that intein splicing is blocked by interactions with extein residues and that these interactions can be disrupted via conformational changes in the exteins induced by temperature or by the solvent environment, thereby releasing the intein from a locked, inactive state. We therefore wished to gain insight into these inhibitory interactions.

In the absence of structural data on the full-length intein-containing protein precursor, the *Pho* FL-RadAi precursor (Figure 4A) was modeled using the available structures of the *Pho* RadA intein (24) and the *S. solfataricus* RadA protein (35) (Figure 4B). Two assumptions were made in generating the model: first, the individual intein and extein fragments maintain their overall 3D structures in the precursor, consistent with the minimal differences in the CD spectra of the intact FL-RadAi precursor and its compositional mimic of RadA + Intein (Figure 3B); and second, the covalent connectivity between the RadA intein and extein sequences imposes 3D constraints that are sufficient for a reasonable prediction of the intein and extein regions with respect to one another. A representative structure of the FL-RadAi precursor illustrates multiple features pertinent to splicing (Figure 4B). First, a large amount of the internal structure can be maintained for both the RadA exteins and the intein in the precursor without steric clashes. Second, the RadA intein and the exteins share a large interaction surface and the intein is in direct contact with the C-terminal extein, which constitutes the main part of the ATPase core domain of the RadA protein. Finally, the catalytic residues of the in-

tein are located on the extein–intein interface (Figure 4C), revealing the possibility for 3D extein–intein interactions affecting intein catalysis. Conserved residues of the intein C153, H245, H312, H323 and N324 are oriented toward the RadA exteins (Figure 4C). C153 corresponds to the first residue of intein (C1 in Figure 4A), which initiates splicing, whereas N324 is the terminal Asn of intein which cyclizes to release the intein from the C-terminal extein (6–8). The His residues are in conserved sequence blocks B, F and G, respectively (Figure 4A), and they modulate the activity of the catalytic residues of the intein.

The proximity of the intein active site to the extein predicts which particular extein sidechains could affect splicing. For this analysis, the intein residues C153, H245, H312, H323 and N324 were used. Ten optimized precursor structures were generated to analyze the sensitivity of these interaction energies to model variations. Identification of extein–intein contacts was based on calculation of absolute values of interaction energies between each extein sidechain and the sidechains or the backbone of the individual intein catalytic residues. The extein–intein interactions with values >1 kcal/mol (positive or negative) were considered significant and shown in Figure 4D. Such interactions were observed with groups of extein residues that are remote from the splice sites in the precursor sequence, in helix 1, loop 1 and loop 2 (Figure 4A), but close to catalytic residues of the intein in 3D space in the FL-RadAi precursor structure model (Figure 4C). These residues belong to highly electrostatic secondary structure elements of the ATPase domain of RadA: helix 1 (D352-K367), loop 1 (R496-R503) and loop 2 (E524-D529) (Figure 4A, C and D). Helix 1 residues interact with sidechains and backbones of H312, H323 and N324; especially strong interactions were identified between R358 and R361 and the sidechain of N324. Loop 1 residues K497, K498 and R503 interact with the sidechain of H245 and R503 has additional interactions with the sidechain of H323 and the backbone of N324. Loop 2 residues interact only with the sidechain of H323. Interestingly, most of the identified 3D extein–intein interactions involve residues H312, H323 and N324, responsible for C-terminal cleavage activity of the intein, while only weak interactions were observed with H245 and none with C153, the residues involved in N-terminal activity of intein. The strongest interactions identified in FL-RadAi were taken into account in extein mutant design to probe the basis of thermoregulation of protein splicing.

Design of extein mutants to establish interactions that control protein splicing

To test whether identified 3D interactions between the extein residues and the catalytic intein residues are responsible for the observed temperature dependence of RadA intein splicing from its native exteins, specific extein residues of the FL-RadAi precursor were mutated to alanine (Figure 5A and B). These mutations can be classified into three categories: extein-intein interaction mutants designed based on the interaction energies, RadA functional mutants based on residues involved in ATPase activity of RadA and control mutants based on the distance from the intein active site
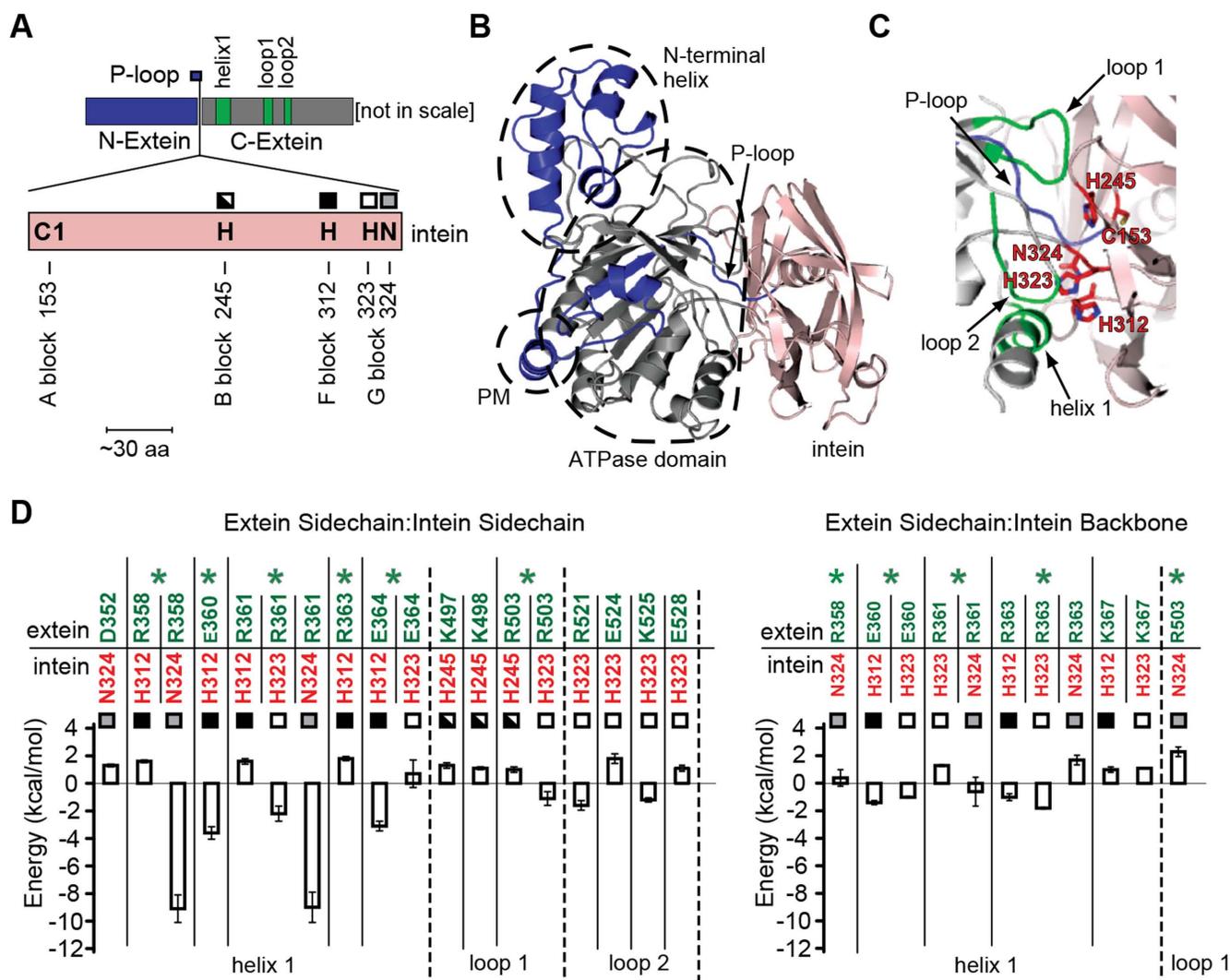
Figure 4. Model of interactions between the RadA exteins and the intein. (**A**) Linear representation of the RadA exteins and the intein. The intein (pink) is in the P-loop of the RadA ATPase domain. The N-extein and the C-extein are shown in blue and gray, respectively, as in panel **B** and **C**. Computational analysis predicted that residues within helix 1, loop 1 and loop 2 (green), located in the C-extein, interact strongly with the intein catalytic site. The catalytic residues of the intein are indicated by the one-letter amino acid code and their positions and conserved sequence blocks are shown. These residues, H245, H312, H323 and N324 are marked by different squares also used in panel **D** and Figure 5C. (B) Putative 3D structure of the *Pho* FL-RadAi precursor was generated by computer modeling based on avaliable structures of the *Pho* RadA intein (PDB ID: 4E2T) and *Sulfolobus solfatarius* RadA protein (PDB ID: 2ZUB). The large contact surface can be seen between the intein (pink) and C-extein (gray). Functionally important domains of the RadA protein: N-terminal helix, polymerization domain (PM) and ATPase domain, are demarcated by dashed lines. (C) The intein active site is at the extein–intein interface and has contacts with the C-extein. Catalytically important residues of the intein (pink) C153, H245, H312, H323 and N324 are shown as red sticks alongside the C-extein (gray). Computational analysis predicted that residues within helix 1, loop 1 and loop 2 (all green) in the C-extein interact strongly with the catalytic residues of the intein. (D) Calculated interaction energies. Interaction energies between the extein residues (green) and the intein active-site residues (red) are shown. Only amino acid residue pairs with interaction energy values >1 kcal/mol are shown. 'Extein Sidechain:Intein Sidechain' (left) and 'Extein Sidechain:Intein Backbone' (right) interactions are plotted separately. The intein catalytic amino acid residues (H245, H312, H323 and N324) are marked by squares as in panel A. The error bars in the figure are the standard deviations obtained from energy calculations for 10 modeled precursor structures. The extein amino acid residues which show the strongest interactions with the intein (green asterisks) were chosen for experimental studies of the extein–intein interactions.

in the models. Twelve mutants were generated, as justified in Table 4 and illustrated in Figure 5A and B.

Among the extein–intein interaction mutants, the residues from the electrostatic helix 1, which show the strongest interactions with the intein, were given the most attention (Figure 4D, marked with * and Figure 5C). Mutant 1 (M1) has R358, E360, R361, R363 and E364 residues of helix 1 changed to alanine; M2 has mutations in the residues R358 and R361 that have strong interactions

with intein residue N324 (Figure 5C); M3 has mutations in residues E360, R363 and E364 that interact with H312 (Figure 5C); M4-M7 have single point mutations in R358, E360, R361 and E364 (Figure 5 and Table 2). Additionally, R503 from loop 1 interacts with H245, H323 and N324 residues of the intein and was mutated to alanine in M8 (Figure 5C).

Since the intein insertion is in the ATPase active site of the RadA protein, we also tested the effect of mutating residues
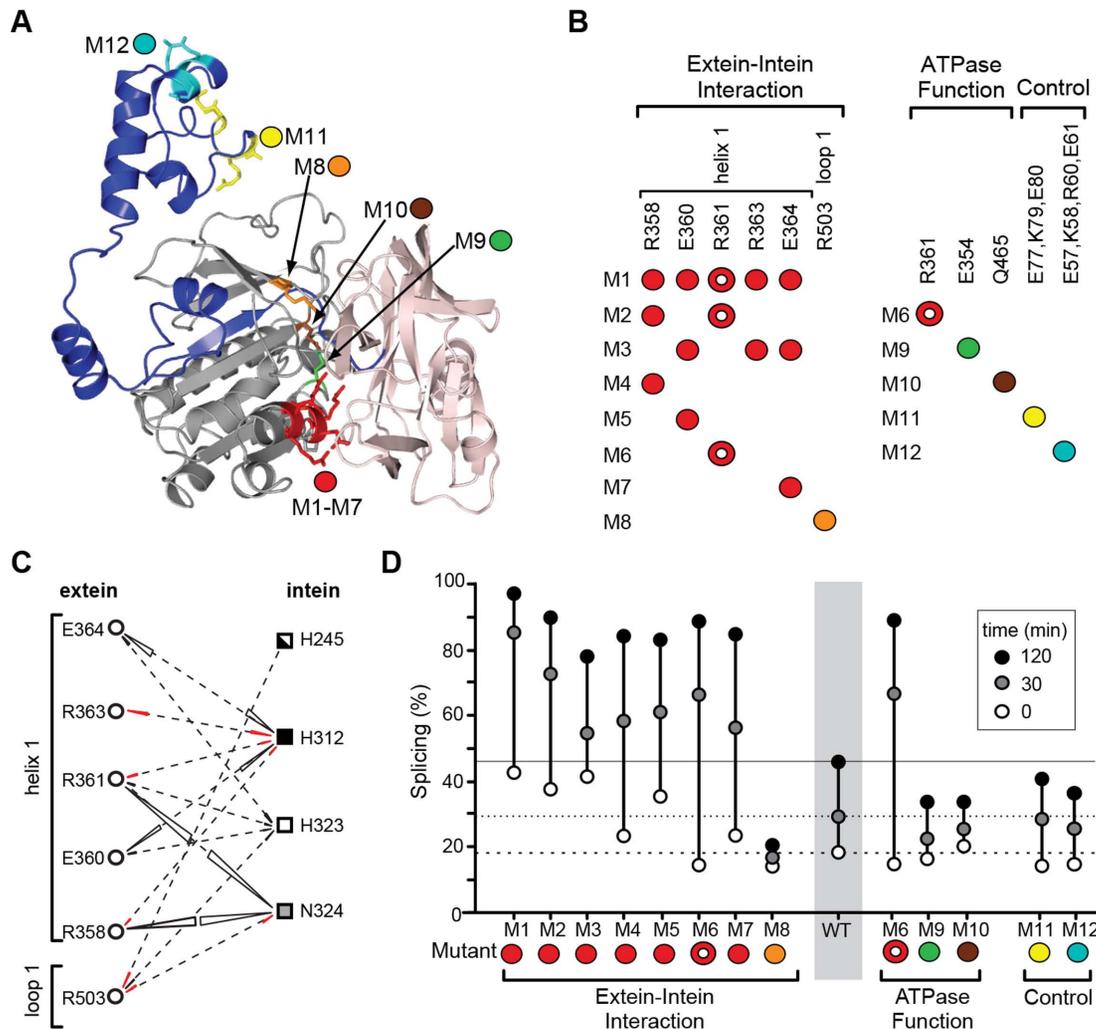
Figure 5. The extein–intein interactions have a profound effect on the intein splicing. (**A**) Locations of the extein mutations. Mutations M1–M12 are highlighted on the structure model of the *Pho* FL-RadAi precursor. (**B**) Mutant design. The extein–intein interaction mutants (M1–M8) were designed based on the interaction energy analysis (Figures 4D and 5C); the ATPase function mutants (M6, M9, M10) carry mutations which disrupt ATPase activity of the RadA protein; control mutants (M11 and M12) carry mutations at amino acid residues that are distant from the intein and are unlikely to interact with the intein active site. R361 in M6 (circle with white center) is a residue that emerged from the computational screen as an interacting residue and is also a catalytic residue of the RadA protein. The color-coding corresponds to the model in (A). (**C**) The extein–intein interactions. The extein amino acid residues selected for experimental studies (derived from Figure 4D, asterisks) are shown with interacting intein residues by dashed lines. The interaction energies between amino acid residue pairs are shown with triangles, the area which corresponds to the value of the interaction energy (negative, black; positive, red). (**D**) Characterization of the FL-RadAi extein mutants. Splicing was monitored *in vitro* at 55°C in comparison with the wild-type FL-RadAi precursor (WT). The percentages of splicing at three time points: 0, 30 and 120 min, are plotted. The 0 min time-point reflects *in vivo* splicing during induction at 46°C.

that are involved in the ATPase activity of RadA. The effect of mutations in highly conserved residues E354 and Q465 that are proposed to coordinate the nucleophilic water molecule (58) were tested in mutants M9 and M10, respectively. The effect of mutation in K152, the residue directly involved in ATP hydrolysis, could not be studied here as K152 is also the terminal residue of the N-extein, a site that perturbs intein splicing (9,10). Interestingly, the ATP base-stacking residue, R361 (58), came up in our computational screen as having one of the strongest interactions with the intein and is mutated in M1, M2 and M6. As controls, M11 and M12 have mutations in charged residues that are not expected to be in proximity to the catalytic intein residues. This range of mutants will probe the veracity of the intein–

extein structure model and the basis of the environmental control of protein splicing.

Splicing of extein mutants supports three-dimensional model for extein–intein interactions

To probe extein–intein interactions intein splicing activity of the FL-RadAi extein mutants and the isogenic wild-type FL-RadAi precursor (WT) were analyzed, after over-expression and Ni-NTA purification. The single and multiple points mutants M6 and M1, respectively, have folding and temperature-dependent secondary structure transitions similar to the ones observed in WT FL-RadAi, as confirmed by CD analysis (data not shown). Modulation

Table 4. Justification for amino acid selection in mutant design

Category	Mutant	Mutation to A	Rationale
Exstein–intein interaction	M1	R358, E360, R361[a], R363, E364	helix 1, all residues
	M2	R358, R361[a]	helix 1, grouped by interactions with N324
	M3	E360, R363, E364	helix 1, grouped by interactions with H312
	M4	R358	helix 1, single point mutant
	M5	E360	helix 1, single point mutant
	M6	R361[a]	helix 1, single point mutant
	M7	E364	helix 1, single point mutant
	M8	R503	loop 1, single point mutant
ATPase function	M9	E354	Coordination of the nucleophilic water molecule
	M10	Q465	Coordination of the nucleophilic water molecule
Control	M11	E77, K79, E80	Charged residues that are not expected to be in proximity to intein catalytic residues
	M12	E57, K58, R60, E61	Charged residues that are not expected to be in proximity to intein catalytic residues

[a]R361 is the ATP base-stacking residue.

of intein splicing by the extein mutations was analyzed *in vivo* and *in vitro* (Figure 5D). The extent of *in vivo* splicing activity is reflected by the amount of spliced precursor at 0 min corresponding to ∼20% for WT during induction at 46°C (Figure 5D). Mutants M1, M2, M3 and M5 showed an increased level of *in vivo* splicing, exceeding 30%.

In vitro analysis of intein splicing was performed by incubation of the samples at elevated temperatures. The RadA intein spliced efficiently from all samples at 75°C (data not shown). In contrast, at 55°C, while WT splicing increased from 20 to 30% after 30 min and to >40% after 120 min, the extein mutations were found to produce dramatic effects on intein splicing at 55°C, especially in the case of the extein–intein interaction mutants (Figure 5D, *cf* M1–M8 with WT). M1, with multiple electrostatic residues in helix 1 mutated to alanine, showed the strongest phenotype, with 85 and 98% of precursor spliced after 30 min and 120 min of incubation, respectively. Although less dramatic, other multiple mutants, M2 and M3, and single point mutants M5–M7 also showed increased splicing activity compared to WT (Figure 5D). Interestingly, M8, which is a loop 1 extein mutant, has inhibited splicing activity compared to WT, opposite to all helix 1 mutants. M6, with mutation of the R361 residue involved in ATP binding, was identified computationally by virtue of extein–intein interactions, and has the strongest splicing activation phenotypes among the single mutants. Other functional mutants in the ATPase domain, which were not identified by computational screen, such as M9 and M10 have only small deviations in splicing compared to WT. The control mutants, M11 and M12, showed negligible modulation of intein splicing, which is consistent with the charged residues in these mutants not interacting directly with the intein catalytic residues.

Mutants M1, M6 and M8 were selected for more detailed characterization of the kinetics of splicing at 55°C (Figure 6A). Consistent with previous data, M1 had a significant amount of splicing *in vivo*, and spliced *in vitro* at least three times faster than WT, resulting in 98% splicing after 2 h at 55°C. Also, in corroboration with previous data, single point mutations in M6 (helix 1) and M8 (loop 1) spliced >80% and <20% after 2 h at 55°C resulting in 2.7-fold acceleration and 5-fold inhibition of intein splicing, respectively.

To further investigate the temperature dependence of M1, this mutant and WT were overexpressed at 12°C to prevent M1 splicing *in vivo*. With recovery of more unspliced precursors (Figure 6B, T = 0) characterization of splicing was performed *in vitro* at 55, 45 and 35°C (Figure 6B). At 55°C, M1 spliced 10 times faster than WT with almost complete precursor conversion after 120 min. At 45°C, M1 spliced 19 times faster than WT with >85% of precursor spliced within 120 min while WT had virtually no activity. Even at 35°C with extremely low levels of activity M1 spliced at five times the rate of WT.

Together these data show unequivocally that the RadA exteins can modulate intein splicing. Significant stimulation and inhibition of protein splicing were observed even with single point mutations in extein sequences remote from the intein in primary sequence, but predicted to be close to the intein catalytic center in 3D space. The most dramatic effects were found in the predicted extein–intein interaction mutants. These results support our 3D precursor model for extein–intein interactions that regulate splicing, while providing the basis for a post-translational environmental switch of RadA activity.

DISCUSSION

This work reports the discovery that splicing of the *Pho* RadA intein, located in the P-loop of the ATP-binding domain of the hyperthermophilic RadA protein, is regulated by its native exteins in a manner dependent upon temperature and solution environment. The extein effect on the *Pho* RadA intein stands out from previous cases of extein-derived modulation of splicing for three reasons. First, regulation is observed only in the native exteins. Second, extein–intein interactions occur via residues that are remote in primary sequence but proximal in 3D space. Finally, exteins serve as an environmental sensor to control splicing, providing a new form of post-translational control. Thereby the exteins impose a lock on splicing, which can be viewed as a cold-shock response that is released at high temperature, such that RadA is most active under the optimal growth conditions of the native organism (Figure 7).
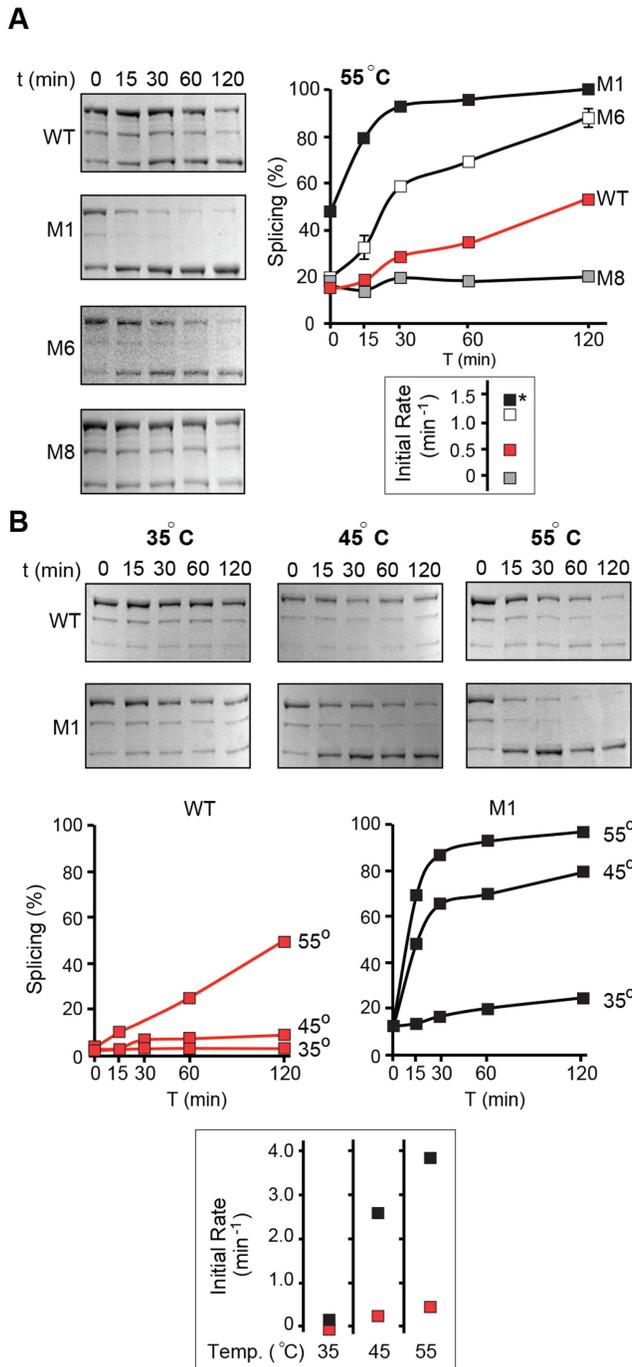
Figure 6. Extein–intein interactions affect temperature dependence of *Pho* RadA intein splicing. (**A**) Kinetics of splicing of extein mutants M1, M6 and M8. Mutants are compared with the wild-type FL-RadAi precursor (WT) at 55°C. Proteins were overexpressed in BL21 Star (DE3) cells at 46°C and Ni-NTA purified. The experiment was performed in triplicate and error bars represent standard deviation. All splicing assays have very small deviations, obscuring error bars in the plot. Initial rates of splicing are shown on the right. * indicates that the initial rate for M1 was calculated for the sample that had 50% of the precursor converted *in vivo* and therefore is likely to be an underestimate. (**B**) Splicing of extein mutant M1. WT and M1 mutant proteins were overexpressed in ArcticExpress (DE3) cells at 12°C. Mutant M1 and the wild-type FL-RadAi precursor (WT) were compared at 55, 45 and 35°C. All splicing assays have very small deviations, obscuring error bars in the plot. The experiment was performed in triplicate and error bars represent standard deviation. Initial rates of splicing at different temperatures are shown on the right.
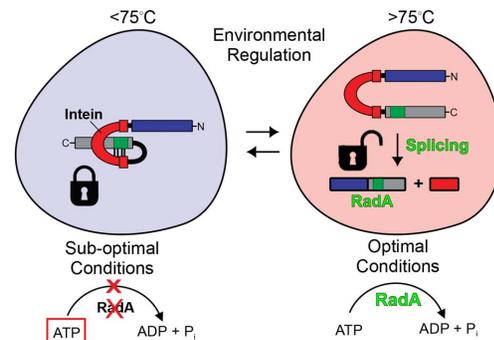

Figure 7. Extein–intein partnership regulates splicing. At suboptimal conditions of temperature and solution environment, the exteins place a lock on splicing through interactions with the intein in 3D space and ATP (boxed) accumulates (left). Under optimal conditions of high temperature and/or solution environment (right) the lock imposed by extein–intein interactions is released and splicing ensues to yield an active RadA protein (green).

P-loop is a hot spot for intein invasion

Recently, detailed characterization of genome organization and gene expression has resulted in paradigm shifts from viewing mobile elements as purely selfish and parasitic entities to considering their dynamic role in the evolution of species (59). Increased frequencies of intein invasion of particular types of proteins and especially of common sites in different proteins may inform us of potential benefits to the organism of such intein localization. Our attention was drawn to a previously reported hot-spot for intein invasion found in the conserved motif in a phosphate-binding loop of the ATPase domain, called the P-loop. Seven of 16 of the most common proteins with inteins have insertions in the P-loop (1). The attractiveness of the P-loop for intein insertions is poorly understood. It is possible, that the bias toward insertion into the P-loop arose from its sequence conservation and specificity of the homing endonucleases from different inteins (2). Another possibility is that intein invasion is quasi-random and inteins are retained in P-loops for some adaptive advantage to the organism (1), including their partnership with exteins. Given the diverse ATPases invaded by inteins, we propose below that this arrangement might allow modulation of ATP consumption under various conditions of stress.

Among the intein-containing proteins, the RecA family stands out as a favorite niche. Strikingly, the newly identified bacterial RecA intein, the first intein reported for *E. coli*, is at precisely the same insertion point as an archaeal RadA counterpart, in the P-loop of the ATP-binding domain (Figure 1). Comparative analysis of the RadA and RecA extein and intein sequences suggested independent invasions at this insertion point in archaea and bacteria, given the conserved nature of the exteins containing the disparate inteins (Figure 1) (1). In the present study we show that the activity of the RadA intein is regulated by its native exteins, suggesting that the intein insertion in the P-loop of the ATP-binding domain of some proteins can be of functional importance. Although we cannot rule out the possibility that the extein-intein partnership is a secondary adaptation, it is clearly valuable to view the intein as part of a complex sys-

tem and to consider the nature of its host protein, the intein insertion site, the host species and its environment.

Post-translational regulation of RadA by superimposed mechanisms

RecA/RadA/Rad51 orthologs in bacteria, archaea and eukarya are ATPases that facilitate DNA strand exchange during homologous recombination, repair of double-strand DNA breaks and restart of stalled replication forks (19). In striking contrast to bacterial and eukaryotic proteins, the archaeal RadA gene is not induced by DNA damage caused by γ and UV irradiation and heat shock, suggesting that it might be constitutively expressed (60–62). It was proposed that the RadA protein might be regulated post-translationally (22).

Although the cellular environment in *E. coli* and *P. horikoshii* is different, it is widely accepted that hyperthermophilic proteins expressed in *E. coli* retain their native folding and biochemical properties (56). Considering the observed temperature dependence of RadA intein splicing, it is important to recognize that ATPase activity of the RadA protein itself is strongly regulated by temperature (Figure 3) (25–28). Most hyperthermophilic enzymes have optimal activity within the range of growth temperatures of the host organism, typically 70–125°C. Thermal control of hyperthermophilic enzymes is often related to their higher rigidity at moderate temperatures and, rarely, with temperature induced conformational changes (56). An Arrhenius plot, which measures the effect of temperature on reaction rate, is linear for the majority of hyperthermophilic proteins, suggesting a uniform functional conformation with changes in temperature (56). However, a biphasic Arrhenius plot for some hyperthermophilic proteins (63–67) suggests functionally significant conformational changes (56), such as reported for RadA of some hyperthermophiles. These RadA proteins exhibit two catalytic modes above 70°C related to rearrangement of hydrophobic amino acid residues near the ATPase active site (25,68). *Pho* RadA has similar temperature dependence of ATPase activity including the increase of ATPase activity at elevated temperatures (Figure 3B) and the breakpoint in Arrhenius plot at 76°C (Figure 3D). Considering that below the optimal growth temperature archaea shut down replication and recombination, the existence of two catalytic modes of RadA was proposed to represent *in vivo* temperature dependent regulation of RadA function (25,68). The presence of the intein in *Pho* RadA and the independent control of splicing from ATPase activity by temperature and solution environment represent superimposed regulatory mechanisms on ATPase function that may safeguard against futile ATP consumption or recombination at unphysiologically low temperatures.

Native exteins as sensors that modulate intein splicing

This work shows that splicing of the RadA intein in the context of its native exteins is temperature dependent. In contrast, several hyperthermophilic inteins show thermally sensitive splicing in foreign exteins (5,12–14,16–18), a phenomenon that may be related to high intein rigidity at suboptimal temperature. Indeed, structural characterization of

the hyperthermophilic temperature-dependent *Pyrococcus abyssi* PolII intein revealed that this intein has a significantly more rigid structure than that found in mesophilic inteins (69). In sharp contrast to these inteins with innate temperature dependence, the *Pho* RadA intein can readily splice from foreign exteins even at low temperature (Figure 2A) (23), showing that RadA intein splicing *per se* is insensitive to temperature. However, splicing of the *Pho* RadA intein from its native precursor has pronounced temperature dependence, suggestive of a direct involvement of the exteins in the thermal properties of RadA intein splicing.

Regulation of RadA intein splicing is related to the temperature dependent changes in the secondary structure of the RadA exteins (Figure 3A and B). Splicing begins at 55°C, coincident with secondary structure rearrangements of RadA exteins of the FL-RadAi precursor. Splicing accelerates with increased temperature and plateaus above 75°C. This leveling off is possibly due to rearrangements in the ATPase active site of RadA responsible for the *Pho* RadA transition between the two catalytic modes (Figure 3D). Such modulation of FL-RadAi splicing observed only in the native extein context suggests that the RadA exteins can serve as a sensor that regulates intein activity through temperature- and solution-sensitive extein–intein interactions.

Extein–intein partnership in three-dimensional space for post-translational control of RadA activity

To understand the nature of the interactions involved in the extein-dependent modulation of intein activity, atomic details of the FL-RadAi precursor are needed. In the absence of a full extein–intein precursor structure, it was possible to derive a model using the independently determined structures of the intein and the extein, as the prediction is reduced to mostly the relative orientation of the two structures and the limited extein distortion to accommodate the intein insertion. This approach is legitimized by the similarity in secondary structure rearrangements of FL-RadAi precursor and RadA + Intein (Figure 3A and B). We generated 10 *Pho* RadA precursor models and identified extein sidechains that could potentially affect splicing through their electrostatic and van der Waal's interactions with the intein catalytic residues, leading to a mutagenesis study that supported the model, as discussed further below. Interestingly, when we performed similar modeling for the newly identified bacterial *E. coli* RecA precursor, with intein insertion in its P-loop, we found that precursor models with similar extein–intein orientations as *Pho* RadA are not feasible due to steric clashes between the *E. coli* RecA intein and its C-extein (data not shown). Orientation of the disparate inteins within the highly conserved ATPase domains could be different for the archaeal hyperthermophilic RadA precursor and the bacterial mesophilic RecA precursors, suggestive of coevolution of intein and extein within each precursor, and precursor-specific functional adaptations.

Mutational analysis strongly supported the modeled interactions between the exteins and the intein. Reduced electrostatic interactions between helix 1 and the intein active site in M1, allow splicing in the temperature interval of 35–55°C (Figure 6B), suggesting that these helix 1-intein active

site interactions are the primary source of the inhibition of splicing above 55°C. Intein splicing has a profound sensitivity toward its exteins as even single point mutations, such as R361A (M6) and R503A (M8), cause strong stimulation and inhibition of protein splicing, respectively. Interestingly, all helix 1 mutations facilitate intein splicing, whereas the loop 1 mutation leads to complete inhibition of splicing, suggesting that different secondary structure elements of the RadA exteins may have divergent effects on intein activity. Such differential modulation of intein splicing by the exteins suggests the potential for extein–intein interaction to toggle intein activity in response to different stimuli.

The conserved residue R361 from helix 1, directly involved in ATP base stacking (58), has the strongest predicted interaction with intein catalytic residues and the strongest phenotype among the single point mutants. Importantly, the interactions between R361 and the intein may affect activity of both the intein and the exteins, by affecting splicing catalysis on one hand and preventing ATP binding to R361 in the functionally inactive FL-RadAi precursor on the other. A similar form of thermal modulation of a hyperthermophilic enzyme favoring activity at high temperatures was reported for the *Pho* acylphosphatase, where formation of a salt bridge with catalytic residues inhibits protein activity at low temperature and the removal of these interactions by mutagenesis releases the inhibition resulting in enzymatic activity of the protein at low temperature (70).

In addition to temperature and mutations, the inhibitory extein–intein interactions within FL-RadAi could be released in a defined solution environment. This environment was formed by the combination of the detergent SDS and the ionic liquid 1-butyl-3-methylimidazolium chloride, which together allow FL-RadAi splicing at low temperature. Similarly, detergents, denaturants and solvents can activate some enzymes by relaxing the rigidity of interactions at the active site (56). However, ATPase activity of the RadA protein at low temperature is not stimulated by the solution composition mentioned above (data not shown), suggesting independent modulation of splicing activity of the RadA precursor and ATPase activity of the RadA protein. Thus, the extein–intein interactions are sensitive to the nature of extein amino-acid side chains and environmental conditions, including temperature and solution composition, and allow independent regulation of intein splicing and ATPase activity *per se*. It would also be interesting to probe the effect of pressure on these interactions given that *P. horikoshii* is hyperbaric.

In conclusion, this work demonstrates a partnership between the RadA intein and the exteins in the *P. horikoshii* RadA precursor. This interaction between the two serves as a sensor for temperature and solution composition, modulating intein splicing. We propose that the intein serves as a transducer that permits extein function in response to environmentally induced interruption of extein–intein interactions. During this modulation, the intein transforms the inactive exteins into the active RadA protein, providing a novel mechanism of post-translational regulation of RadA function.

REFERENCES

1. Novikova,O., Topilina,N. and Belfort,M. (2014) Enigmatic distribution, evolution, and function of inteins. *J. Biol. Chem.*, **289**, 14490–14497.
2. Liu,X.Q. (2000) Protein-splicing intein: genetic mobility, origin, and evolution. *Annu. Rev. Genet.*, **34**, 61–76.
3. Gogarten,J.P., Senejani,A.G., Zhaxybayeva,O., Olendzenski,L. and Hilario,E. (2002) Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.*, **56**, 263–287.
4. Swithers,K.S., Senejani,A.G., Fournier,G.P. and Gogarten,J.P. (2009) Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol. Biol.*, **9**, 303.
5. Callahan,B.P., Topilina,N.I., Stanger,M.J, Van Roey,P. and 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.
6. Paulus,H. (2001) Inteins as enzymes. *Bioorg. Chem.*, **29**, 119–129.
7. Volkmann,G. and Mootz,H.D. (2013) Recent progress in intein research: from mechanism to directed evolution and applications. *Cell. Mol. Life. Sci.*, **70**, 1185–1206.
8. Mills,K.V., Johnson,M.A. and Perler,F.B. (2014) Protein splicing: how inteins escape from precursor proteins. *J. Biol. Chem.*, **289**, 14498–14505.
9. Iwai,H., Zuger,S., Jin,J. and Tam,P.H. (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett.*, **580**, 1853–1858.
10. Amitai,G., Callahan,B.P., Stanger,M.J., Belfort,G. and Belfort,M. (2009) Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 11005–11010.
11. Shemella,P.T., Topilina,N.I., Soga,I., Pereira,B., Belfort,G., Belfort,M. and Nayak,S.K. (2011) Electronic structure of neighboring extein residue modulates intein C-terminal cleavage activity. *Biophys. J.*, **100**, 2217–2225.
12. Shao,Y., Xu,M.Q. and Paulus,H. (1995) Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence. *Biochemistry.*, **34**, 10844–10850.
13. Xu,M.Q. and Perler,F.B. (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J.*, **15**, 5146–5153.
14. Shao,Y. and Paulus,H. (1997) Protein splicing: estimation of the rate of O-N and S-N acyl rearrangements, the last step of the splicing process. *J. Pept. Res.*, **50**, 193–198.
15. Wood,D.W., Wu,W., Belfort,G., Derbyshire,V. and Belfort,M. (1999) A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotechnol.*, **17**, 889–892.
16. Cambon-Bonavita,M.A., Schmitt,P., Zieger,M., Flaman,J.M., Lesongeur,F., Raguenes,G., Bindel,D., Frisch,N., Lakkis,Z., Dupret,D. *et al.* (2000) Cloning, expression, and characterization of DNA polymerase I from the hyperthermophilic archaea Thermococcus fumicolans. *Extremophiles.*, **4**, 215–225.
17. Mills,K.V., Manning,J.S., Garcia,A.M. and Wuerdeman,L.A. (2004) Protein splicing of a Pyrococcus abyssi intein with a C-terminal glutamine. *J. Biol. Chem.*, **279**, 20685–20691.

18. Mills,K.V., Dorval,D.M. and Lewandowski,K.T. (2005) Kinetic analysis of the individual steps of protein splicing for the Pyrococcus abyssi PolII intein. *J. Biol. Chem.*, **280**, 2714–2720.

19. Haldenby,S., White,M.F. and Allers,T. (2009) RecA family proteins in archaea: RadA and its cousins. *Biochem. Soc. Trans.*, **37**, 102–107.

20. Krejci,L., Altmannova,V., Spirek,M. and Zhao,X. (2012) Homologous recombination and its regulation. *Nucleic Acids Res.*, **40**, 5795–5818.

21. Daley,J.M., Kwon,Y., Niu,H. and Sung,P. (2013) Investigations of homologous recombination pathways and their regulation. *Yale J. Biol. Med.*, **86**, 453–461.

22. Sheng,D., Zhu,S., Wei,T., Ni,J. and Shen,Y. (2008) The in vitro activity of a Rad55 homologue from Sulfolobus tokodaii, a candidate mediator in RadA-catalyzed homologous recombination. *Extremophiles.*, **12**, 147–157.

23. Ellila,S., Jurvansuu,J.M. and Iwai,H. (2011) Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in Escherichia coli. *FEBS Lett.*, **585**, 3471–3477.

24. Oeemig,J.S., Zhou,D., Kajander,T., Wlodawer,A. and Iwai,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.

25. Spies,M., Kil,Y., Masui,R., Kato,R., Kujo,C., Ohshima,T., Kuramitsu,S. and Lanzov,V. (2000) The RadA protein from a hyperthermophilic archaeon Pyrobaculum islandicum is a DNA-dependent ATPase that exhibits two disparate catalytic modes, with a transition temperature at 75 degrees C. *Eur. J. Biochem.*, **267**, 1125–1137.

26. Shalguev,V.I., Kil,Y.V., Yurchenko,L.V., Namsaraev,E.A. and Lanzov,V.A. (2004) Rad51 protein from the thermotolerant yeast Pichia angusta as a typical but thermodependent member of the Rad51 family. *Eukaryot. Cell.*, **3**, 1567–1573.

27. McIlwraith,M.J., Hall,D.R., Stasiak,A.Z., Stasiak,A., Wigley,D.B. and West,S.C. (2001) RadA protein from Archaeoglobus fulgidus forms rings, nucleoprotein filaments and catalyses homologous recombination. *Nucleic Acids Res.*, **29**, 4509–4517.

28. Chen,L.T., Ko,T.P., Chang,Y.C., Lin,K.A., Chang,C.S., Wang,A.H. and Wang,T.F. (2007) Crystal structure of the left-handed archaeal RadA helical filament: identification of a functional motif for controlling quaternary structures and enzymatic functions of RecA family proteins. *Nucleic Acids Res.*, **35**, 1787–1801.

29. Seitz,E.M., Brockman,J.P., Sandler,S.J., Clark,A.J. and Kowalczykowski,S.C. (1998) RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev.*, **12**, 1248–1253.

30. Kil,Y.V., Baitin,D.M., Masui,R., Bonch-Osmolovskaya,E.A., Kuramitsu,S. and Lanzov,V.A. (2000) Efficient strand transfer by the RadA recombinase from the hyperthermophilic archaeon Desulfurococcus amylolyticus. *J. Bacteriol.*, **182**, 130–134.

31. Thompson,J.D., Gibson,T. and Higgins,D.G. (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics*, doi:10.1002/0471250953.bi0203s00.

32. Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M. and Kumar,S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**, 2731–2739.

33. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.*, **227**, 680–685.

34. Ariza,A., Richard,D.J., White,M.F. and Bond,C.S. (2005) Conformational flexibility revealed by the crystal structure of a crenarchaeal RadA. *Nucleic Acids Res.*, **33**, 1465–1473.

35. Chang,Y.W., Ko,T.P., Lee,C.D., Chang,Y.C., Lin,K.A., Chang,C.S., Wang,A.H. and Wang,T.F. (2009) Three new structures of left-handed RADA helical filaments: structural flexibility of N-terminal domain is critical for recombinase activity. *PLOS One*, **4**, e4890.

36. Bennett-Lovsey,R.M., Herbert,A.D., Sternberg,M.J.E. and Kelley,L.A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, **70**, 611–625.

37. Sali,A. and Blundell,T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

38. Shen,M. and Sali,A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, **15**, 2507–2524.

39. Melo,F., Sanchez,R. and Sali,A. (2002) Statistical potentials for fold assessment. *Protein Sci.*, **11**, 430–448.

40. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graph.*, **14**, 33–38.

41. Soto,C.S., Fasnacht,M., Zhu,J., Forrest,L. and Honig,B. (2008) Loop modeling: sampling, filtering, and scoring. *Proteins.*, **70**, 834–843.

42. Curry,H.B. (1994) The method of steepest descent for nonlinear minimization problems. *Quart. Appl. Math*, **2**, 250–261.

43. Chu,J.W., Trout,B.L. and Brooks,B.R. (2003) A super-linear minimization scheme for the nudged elastic band method. *J. Chem. Phys.*, **119**, 12708.

44. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., Swaminathan,S. and Karplus,M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.

45. Brooks,B.R., Brooks,C.L. III, Mackerell,A.D. Jr, Nilsson,L., Petrella,R.J., Roux,B., Won,Y., Archontis,G., Bartels,C., Boresch,S. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.

46. MacKerell,A.D., Bashford,D., Bellott,M., Dunbrack,R.L., Evanseck,J.D., Field,M.J., Fischer,S., Gao,J., Guo,H., Ha,S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.*, **102**, 3586–3616.

47. Benkert,P., Kunzli,M. and Schwede,T. (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Res.*, **37**, W510–W514.

48. Wiederstein,M. and Sippl,M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, **35**, W407–W410.

49. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

50. Bhattacharya,A., Tejero,R. and Montelione,G.T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins*, **66**, 778–795.

51. Luthy,R., Bowie,J.U. and Eisenberg,D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83–85.

52. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **27**, 283–291.

53. Perler,F.B. (2002) InBase: the Intein Database. *Nucleic Acids Res.*, **30**, 383–384.

54. Davis,E.O., Sedgwick,S.G. and Colston,M.J. (1991) Novel structure of the recA locus of Mycobacterium tuberculosis implies processing of the gene product. *J. Bacteriol.*, **173**, 5653–5662.

55. Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T. (1993) Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae. *Mol. Microbiol.*, **7**, 197–206.

56. Vieille,C. and Zeikus,G.J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.*, **65**, 1–43.

57. Heller,W.T., O'Neill,H.M., Zhang,Q. and Baker,G.A. (2010) Characterization of the influence of the ionic liquid 1-butyl-3-methylimidazolium chloride on the structure and thermal stability of green fluorescent protein. *J. Phys. Chem. B.*, **114**, 13866–13871.

58. Wu,Y., Qian,X., He,Y., Moya,I.A. and Luo,Y. (2005) Crystal structure of an ATPase-active form of Rad51 homolog from Methanococcus voltae. Insights into potassium dependence. *J. Biol. Chem.*, **280**, 722–728.

59. Faulkner,G.J. and Carninci,P. (2009) Altruistic functions for selfish DNA. *Cell Cycle*, **8**, 2895–2900.

60. Komori,K., Miyata,T., DiRuggiero,J., Holley-Shanks,R., Hayashi,I., Cann,I.K., Mayanagi,K., Shinagawa,H. and Ishino,Y. (2000) Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus. *J. Biol. Chem.*, **275**, 33782–33790.

61. McRobbie,A.M., Carter,L.G., Kerou,M., Liu,H., McMahon,S.A., Johnson,K.A., Oke,M., Naismith,J.H. and White,M.F. (2009) Structural and functional characterisation of a conserved archaeal RadA paralog with antirecombinase activity. *J. Mol. Biol.*, **389**, 661–673.

62. Frols,S., Gordon,P.M., Panlilio,M.A., Duggin,I.G., Bell,S.D., Sensen,C.W. and Schleper,C. (2007) Response of the hyperthermophilic archaeon Sulfolobus solfataricus to UV damage. *J. Bacteriol.*, **189**, 8708–8718.

63. Cannio,R., Fiorentino,G., Carpinelli,P., Rossi,M. and Bartolucci,S. (1996) Cloning and overexpression in Escherichia coli of the genes encoding NAD-dependent alcohol dehydrogenase from two Sulfolobus species. *J. Bacteriol.*, **178**, 301–305.

64. Fabry,S. and Hensel,R. (1987) Purification and characterization of D-glyceraldehyde-3-phosphate dehydrogenase from the thermophilic archaebacterium Methanothermus fervidus. *Eur. J. Biochem.*, **165**, 147–155.

65. Faraone-Mennella,M.R., Gambacorta,A., Nicolaus,B. and Farina,B. (1998) Purification and biochemical characterization of a poly(ADP-ribose) polymerase-like enzyme from the thermophilic archaeon Sulfolobus solfataricus. *Biochem. J.*, **335**, 441–447.

66. Hensel,R., Laumann,S., Lang,J., Heumann,H. and Lottspeich,F. (1987) Characterization of two D-glyceraldehyde-3-phosphate dehydrogenases from the extremely thermophilic archaebacterium Thermoproteus tenax. *Eur. J. Biochem.*, **170**, 325–333.

67. Wrba,A., Schweiger,A., Schultes,V., Jaenicke,R. and Zavodszky,P. (1990) Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium Thermotoga maritima. *Biochemistry.*, **29**, 7584–7592.

68. Kil,Y.V., Glazunov,E.A. and Lanzov,V.A. (2005) Characteristic thermodependence of the RadA recombinase from the hyperthermophilic archaeon Desulfurococcus amylolyticus. *J. Bacteriol.*, **187**, 2555–2557.

69. Du,Z., Liu,J., Albracht,C.D., Hsu,A., Chen,W., Marieni,M.D., Colelli,K.M., Williams,J.E., Reitter,J.N., Mills,K.V. *et al.* (2011) Structural and mutational studies of a hyperthermophilic intein from DNA polymerase II of Pyrococcus abyssi. *J. Biol. Chem.*, **286**, 38638–38648.

70. Lam,S.Y., Yeung,R.C., Yu,T.H., Sze,K.H. and Wong,K.B. (2011) A rigidifying salt-bridge favors the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity. *PLOS Biol.*, **9**, e1001027.