

Half a Century After Their Discovery: Structural Insights into Exonuclease and Annealase Proteins Catalyzing Recombineering

Authors: Lucy J. Fitschen^{1,2}, Timothy P. Newing^{1,2}, Nikolas P. Johnston³, Charles E. Bell^{4,*}, Gökhan Tolun^{1,2,*}

¹ School of Chemistry and Molecular Bioscience, and Molecular Horizons, University of Wollongong, Wollongong, NSW, Australia.

² ARC Centre for Cryo-electron Microscopy of Membrane Proteins, University of Wollongong, Wollongong, NSW, Australia.

³ Faculty of Science, University of Technology Sydney, Ultimo, NSW, 2007, Australia

⁴ Department of Biological Chemistry and Pharmacology, The Ohio State University College of Medicine, Columbus, OH 43210.

* Corresponding Authors

Abstract

Recombineering is an essential tool for molecular biologists, allowing for the facile and efficient manipulation of bacterial genomes directly in cells without the need for costly and laborious *in vitro* manipulations involving restriction enzymes. The main workhorses behind recombineering are bacteriophage proteins that promote the single-strand annealing (SSA) homologous recombination pathway to repair double-stranded DNA breaks. While there have been several reviews examining recombineering methods and applications, comparatively few have focused on the mechanisms of the proteins that are the key players in the SSA pathway: a 5'→3' exonuclease and a single-strand annealing protein (SSAP or “annealase”). This review dives into the structures and functions of the two SSA recombination systems that were the first to be developed for recombineering in *E. coli*: the RecET system from *E. coli* Rac prophage and the λRed system from bacteriophage λ. By comparing the structures of the RecT and Redβ annealases, and the RecE and λExo exonucleases, we provide new insights into how the structures of these proteins dictate their function. Examining the sequence conservation of the λExo and RecE exonucleases gives more profound insights into their critical functional features. Ultimately, as recombineering accelerates and evolves in the laboratory, a better

30 understanding of the mechanisms of the proteins behind this powerful technique will drive the
31 development of improved and expanded capabilities in the future.

32 **Keywords**

33 Recombineering, Exonuclease, Red β , RecE, RecT, Protein Structure,

34 **Abbreviations and Nomenclature**

35 AFM – Atomic Force Microscopy

36 CTD – C-terminal domain

37 EATR – Exonuclease-Annealase Two-component Recombinase = TCR (Two Component
38 Recombinase)

39 ERF – Essential Recombination Function

40 HSV-1 – Herpes Simplex Virus 1

41 NS-EM – Negative Staining Electron Microscopy

42 NTD – N-terminal domain

43 SSA – Single-Strand Annealing

44 SSAP – Single-Strand Annealing Protein = Annealase

45 AFM – Atomic Force Microscopy

46 T2RE – Type 2 restriction endonuclease family of nucleases

47

1.1 What is Recombineering?

The ability to clone and edit genetic material is an essential component of life scientists' toolkit, allowing research in numerous fields, from molecular biology to biochemistry and cell biology to biophysics. One genome editing method known as recombineering, a portmanteau of recombination mediated genetic engineering, allows DNA manipulation without restriction enzymes or other *in vitro* enzymatic treatments [1]. Recombineering was initially developed for editing DNA within *Escherichia coli* using bacteriophage proteins, taking advantage of either the bacteriophage lambda (phage λ) Red (identified via recombination-deficient mutations) recombination system [2–4] or the λ prophage RecET system [5]. Both systems combine an exonuclease for resecting dsDNA ends in the 5'→3' direction, with an annealase for binding the resulting 3'-ssDNA overhang and annealing it to a homologous ssDNA molecule. Since first demonstrated in *E. coli*, recombineering has been successfully implemented in many other bacteria (Table 1), often using the exonuclease and annealase proteins from a host-specific bacteriophage. Recombineering is also the basis for related techniques such as multiplex automated genomic engineering (MAGE) [6] that can rapidly evolve new bacterial strains with enhanced functions.

The Red and RecET phage systems have been exploited for recombineering due to their simple, streamlined, and highly efficient pathway for homologous DNA recombination known as single strand annealing (SSA). SSA is one of three main pathways used in eukaryotic cells for the repair of dsDNA breaks, the other two being non-homologous end joining (NHEJ) and homologous recombination (HR) [REF]. SSA is one of the least studied DNA repair pathways, and its detailed mechanism of action is still largely unknown [7,8]. While numerous informative reviews covering different aspects of recombineering are currently available [1,9–11], most pay only limited attention to the structures and mechanisms of the proteins that are the key workhorses behind the method. As structural knowledge of a protein can dramatically improve our understanding of its function, this review will focus on the structures of the exonuclease and annealase proteins that have been determined to date, including the annealase structures reported during the past year. By digging deep into the structures of these proteins, we can understand not only how they function within their native bacterial hosts but also how we can continue to expand and improve recombineering in the future.

Table 1 – A list of the organisms in which recombineering has been reported. * Denotes a gene name rather than a protein. Multiple entries may exist for bacterial species where recombineering is reported with different EATR pairs. Similar tables have been created by others [1,12–14]

Bacterial Host Target	Exonuclease	Annealase	EATR Origin	Reference
<i>Acinetobacter baumannii</i>	ACINIS123_2462*	ACINIS123_2461*	<i>A. baumannii</i> strain IS-123	[15]
<i>Agrobacterium tumefaciens</i>	λ Exo	Red β	Bacteriophage λ	[16]
<i>Bacillus subtilis</i>	N/A	GP35	phage SPP1	[17]
<i>Burkholderia thailandensis</i>	λ Exo	Red β	Bacteriophage λ	[18]
<i>Burkholderia pseudomallei</i>	λ Exo	Red β	Bacteriophage λ	[18]
<i>Burkholderia</i> sp. DSM 7029	Red α 7029	Red β 7029	DSM 7029	[19]
<i>Caulobacter crescentus</i>	N/A	Red β	Bacteriophage λ	[20]
<i>Clostridium acetobutylicum</i>	N/A	CPF0939*	<i>C. perfringens</i>	[21]
<i>Collinsella stercoris</i>	N/A	CspRecT	<i>C. stercoris</i> phage	[22]
<i>Corynebacterium glutamicum</i>	RecT	RecE	Rac Prophage	[23]
<i>Corynebacterium glutamicum</i>	OrfC	OrfB	<i>L. pneumophila</i>	[23]
<i>Corynebacterium glutamicum</i>	GP61	GP60	Phage Che9c of <i>M. smegmatis</i>	[23]
<i>Escherichia coli</i>	N/A	CspRecT	<i>C. stercoris</i> phage	[22]
<i>Escherichia coli</i>	λ Exo	Red β	Bacteriophage λ	[4]
<i>Escherichia coli</i>	RecE	RecT	Rac Prophage	[5]
<i>Klebsiella pneumonia</i>	N/A	CspRecT	<i>C. stercoris</i> phage	[22]
<i>Lactobacillus brevis</i>	RecE homolog	RecT homolog	<i>L. brevis</i> KB290	[24]
<i>Lactobacillus casei</i>	LCABL_13060*	LCABL_13040*	prophage PLE3	[25]
<i>Lactobacillus plantarum</i>	lp_0642*	lp_0640*	prophage P1	[26]
<i>Lactobacillus reuteri</i>	N/A	RecT ₁	<i>L. reuteri</i>	[27]
<i>Lactobacillus rhamnosus</i>	N/A	LprRecT	<i>Lactobacillus reuteri</i> prophage	[20]
<i>Lactococcus lactis</i>	N/A	RecT ₁	<i>L. reuteri</i>	[27]
<i>Legionella pneumophila</i>	N/A	ORF C	<i>L. pneumophila</i>	[28]
<i>Mycoplasma pneumoniae</i>	N/A	GP35	phage SPP1	[29]
<i>Mycobacterium smegmatis</i>	Gp60	Gp61	Phage Che9c	[30]

<i>Mycobacterium tuberculosis</i>	Gp60	Gp61	Phage Che9c	[31]
<i>Photorhabdus luminescence</i>	Plu α	Plu β	<i>P. luminescence</i>	[13]
<i>Pseudomonas aeruginosa</i>	N/A	PapRecT	<i>P. aeruginosa</i> phage	[22]
<i>Pseudomonas aeruginosa</i>	λ Exo	Red β	Bacteriophage λ	[32]
<i>Pseudomonas putida</i>	N/A	Rec2	<i>P. putida</i>	[33]
<i>Pseudomonas syringae</i>	RecE _{Psy}	RecT _{Psy}	<i>P. syringae</i>	[34]
<i>Saccharomyces cerevisiae</i>	N/A	Red β	Bacteriophage λ	[35]
<i>Salmonella enterica</i>	λ Exo	Red β	Bacteriophage λ	[36]
<i>Shigella sonnei</i>	λ Exo	Red β	Bacteriophage λ	[37]
<i>Shigella flexneri</i>	λ Exo	Red β	Bacteriophage λ	[37]
<i>Shigella dysenteriae</i>	λ Exo	Red β	Bacteriophage λ	[37]
<i>Shewanella oneidensis</i>	N/A	W3 Beta	<i>Shewanella</i> sp. W3-18-1	[38]
<i>Sinorhizobium meliloti</i>	λ Exo	Red β	Bacteriophage λ	[39]
<i>Staphylococcus aureus</i>	N/A	EF2132*	<i>Enterococcus faecalis</i>	[40]
<i>Vibrio natriegens</i>	SXT-Exo	SXT-Beta	SXT mobile genetic element	[41]
<i>Xenorhabdus stockiae</i>	Plu α	Plu β	<i>P. luminescence</i>	[13]
<i>Xenorhabdus stockiae</i>	XBJ1_1172*	XBJ1_1171*	N/A	[42]
<i>Yersinia pseudotuberculosis</i>	λ Exo	Red β	Bacteriophage λ	[43]
<i>Zymomonas mobilis</i>	RecE	RecT	Rac Prophage	[44]

81

82 1.2 The Roles of EATR Proteins in Single-Strand Annealing

83 Recombineering utilizes bacteriophage proteins that catalyze homologous DNA
84 recombination. These proteins form an **Exonuclease-Annealase Two-component Recombinase**
85 system, or EATR. The term SynExo (**Synaptase Exonuclease pair**) has also been used, in which
86 case synaptase is synonymous with annealase. The terms annealase, synaptase, recombinase,
87 and single strand annealing protein (SSAP) have all been used to refer to the same group of
88 proteins. **Herein, we use the term annealase to describe proteins that bind to ssDNA and**

89 catalyze the annealing of two homologous ssDNA strands in an ATP-independent manner.
90 Prominent examples include RecT from *E. coli* [REF 61], Red β from phage λ [REF 100, 116],
91 Rad52 from yeast and humans [7], and ICP8 from Herpes Simplex Virus 1 (HSV1) [REF
92 127,128]. Annealases typically do not catalyze DNA strand-invasion reactions (insertion of a
93 ssDNA strand into a homologous dsDNA molecule) like the RecA and RAD51 recombinases
94 that are ATP-dependent [45]. Instead, the two EATR proteins work in concert to catalyze DNA
95 recombination by single strand annealing (SSA): the exonuclease binds to a dsDNA end and
96 carries out 5'→3' end-resection to form a long 3'-ssDNA overhang, to which the annealase binds
97 and anneals it to a homologous ssDNA. The two steps of the reaction, end-resection and
98 annealing are coupled to one another via a protein-protein interaction between the exonuclease
99 and annealased components [REF 108, 139].

100 Before discussing the mechanisms of EATR proteins in recombineering, it is worth
101 considering their natural roles in the propagation of the bacteriophage that encode them. In this
102 regard, the Red system from phage λ has been studied in the most detail. The Red genes are
103 not required for viability of phage λ , but they significantly enhance (by ~3-fold) the number of
104 phage particles produced upon lysis [REF 51]. Exactly how recombination promotes phage
105 propagation is not fully understood, but roles in replication [REF 139], generation of
106 concatemeric genomes for viral genome packaging [REF 141], repair of dsDNA breaks for
107 CRISPR evasion [REF 142], and generation of genetic diversity [REF 143,144] have been
108 proposed. Although the exact mechanisms by which EATR proteins promote recombination in
109 cells are still under investigation, several models have been proposed. Stahl *et al.* described
110 two models for phage λ recombination in *E. coli*, one that is dependent on the host RecA DNA
111 strand-exchange ATPase, and another that is RecA-independent [10]. The RecA-dependent
112 model, which is prevalent in non-replicating cells, starts with digestion of a dsDNA end by
113 λ Exo, which loads Red β onto the nascent 3' ssDNA overhang [10]. The host RecFOR proteins
114 then facilitate replacement of Red β with RecA on the 3'-overhang, which promotes invasion
115 of the 3'-overhang into a homologous dsDNA molecule, and recombination proceeds from
116 there via the normal host double-strand break repair system [10,46].

117 The RecA-independent model, also known as single strand annealing (SSA), is
118 prevalent in replicating cells. In this model, λ Exo and Red β again function in concert to form
119 a 3'-overhang bound by Red β , which in this case directly catalyzes annealing of the 3'-
120 overhang to a homologous ssDNA molecule. This pathway requires two DNA breaks at non-
121 allelic sites on separate phage λ chromosomes, which, upon end-resection, can produce 3'-

overhangs with complementary regions that can be directly annealed to one another. Following annealing, the excess non-homologous overhanging strands are removed, and any gaps formed are filled in by a polymerase. Finally, a ligase can seal the remaining nick to result in a fully repaired functional dsDNA molecule.

A similar type of SSA pathway can ensue when a dsDNA break occurs between two directly repeated sequences on the DNA. In such a scenario, the two 3'-overhangs formed by end-resection will have homologous sequences that can be directly annealed to one another. This results in repair of the dsDNA break, but with deletion of one of the two repeats along with the sequence between them. Although the phage λ chromosome doesn't have directly repeated regions, a very similar type of SSA pathway can be highly significant in eukaryotic cells, where such repeats are common. [REF 47, 144, 147]. Moreover, homologs of the RecET and Red EATR proteins encoded on the IncC conjugative plasmid that replicates in *Salmonella enterica* have been demonstrated to use this type of SSA to evade dsDNA breaks formed by a host CRISPR system [REF 144]. Thus, the EATR-promoted SSA pathway can be relevant to bacteriophage (or plasmid) propagation, and studies of the RecET and Red proteins have served as a model for understanding double-strand break repair in humans.

The recombination model that is most applicable to recombineering is the RecA-independent SSA model that relies on Red β annealase [47] [REF 146]. Recombineering can either employ synthetic ssDNA oligonucleotides as the input DNA electroporated into cells (typically in the range of 35-100 nucleotides), or a dsDNA cassette that can be much longer (5 - 10 KB). The dsDNA cassette is typically generated by PCR using primers containing terminal homologies to the recombination target site. In the case of oligonucleotides, the exonuclease component of the EATR is not required for recombination: the annealase binds the input oligonucleotide directly and anneals it to the target site exposed as ssDNA at the lagging strand of a replication fork [REF 71 Ellis et al.]. By contrast, recombineering with dsDNA as the input DNA requires end-resection by the exonuclease, and the resulting 3'-ssDNA overhang is bound by the annealase. Surprisingly, it appears that in dsDNA recombineering, the exonuclease typically digests one complete strand of the input dsDNA, and Red β anneals the intact opposing strand to the target site by the same mechanism as for short oligonucleotides, at the lagging strand of a replication fork [REF 50,146]. While recombination via the classical SSA pathway (i.e. end-to-end annealing) can occur during recombineering, the efficiency is much lower due to the requirement of appropriately positioned dsDNA breaks. Hence, most of the annealing events during recombineering occur at the replication fork.

As mentioned above, the classical RecA-independent SSA pathway in phage λ described by Stahl *et al.* was greatly stimulated in replicating cells, although the reason for this was unclear. It was suggested that the role of replication in these experiments could be to generate appropriately positioned dsDNA breaks. However, it is conceivable that some annealing events in a normal phage λ infection may occur at the replication fork, as seen in recombineering. Models for phage λ recombination involving replication have been proposed, including one involving a Replisome Invasion/Template Switch [48]. It seems unlikely that the full 50 kb phage λ chromosome would be digested for productive recombination as described above for dsDNA recombineering, but λ Exo is highly processive and can digest full dsDNA substrates of that length *in vitro* [REF 82,83]. In any case, the use of phage EATR proteins in recombineering, as well as structural and biophysical studies of them have shed new light on the possible mechanisms of phage λ recombination that have been studied for so many years by geneticists [REF 148].

Lastly, while EATR proteins form a complex with one another, they also interact with host proteins to facilitate recombination. Most prominently, Red β from bacteriophage λ binds to *E. coli* single-stranded DNA binding protein (SSB), which coats the ssDNA at the lagging strand of the replication fork to protect it from nucleases and control access of numerous replication proteins [REF 20, 55]. This interaction with SSB is absolutely required for recombination *in vivo*, presumably to displace SSB and allow Red β to gain access to the lagging strand. Red β from phage λ also interacts with phage λ replication protein P, integrase, and antitermination protein [49], although the roles of these interactions are unclear. The interactions with host proteins are even more prevalent for EATR proteins from viruses that infect eukaryotic cells. For example, ICP8, the annealase from HSV-1, interacts with other proteins in the HSV-1 replisome, such as the UL9 origin-binding protein [50] and ICP27 that is essential for the regulation of viral gene expression [51]. [49][52]

In summary, structural information on EATR proteins and their interactions with other viral and host proteins is of broad interest for understanding multiple aspects of genome maintenance including replication, repair, and generation of genetic diversity. While the RecT and Red β annealases that have been predominantly employed for recombineering have been studied for over half a century, the key structural insights into these proteins have come relatively recently, as summarized in the historical timeline in Figure 1. The recent breakthroughs in annealase structures have put a new spotlight on these proteins and how

understanding their structure can help unravel their function and lead to improvements in recombineering in the future.

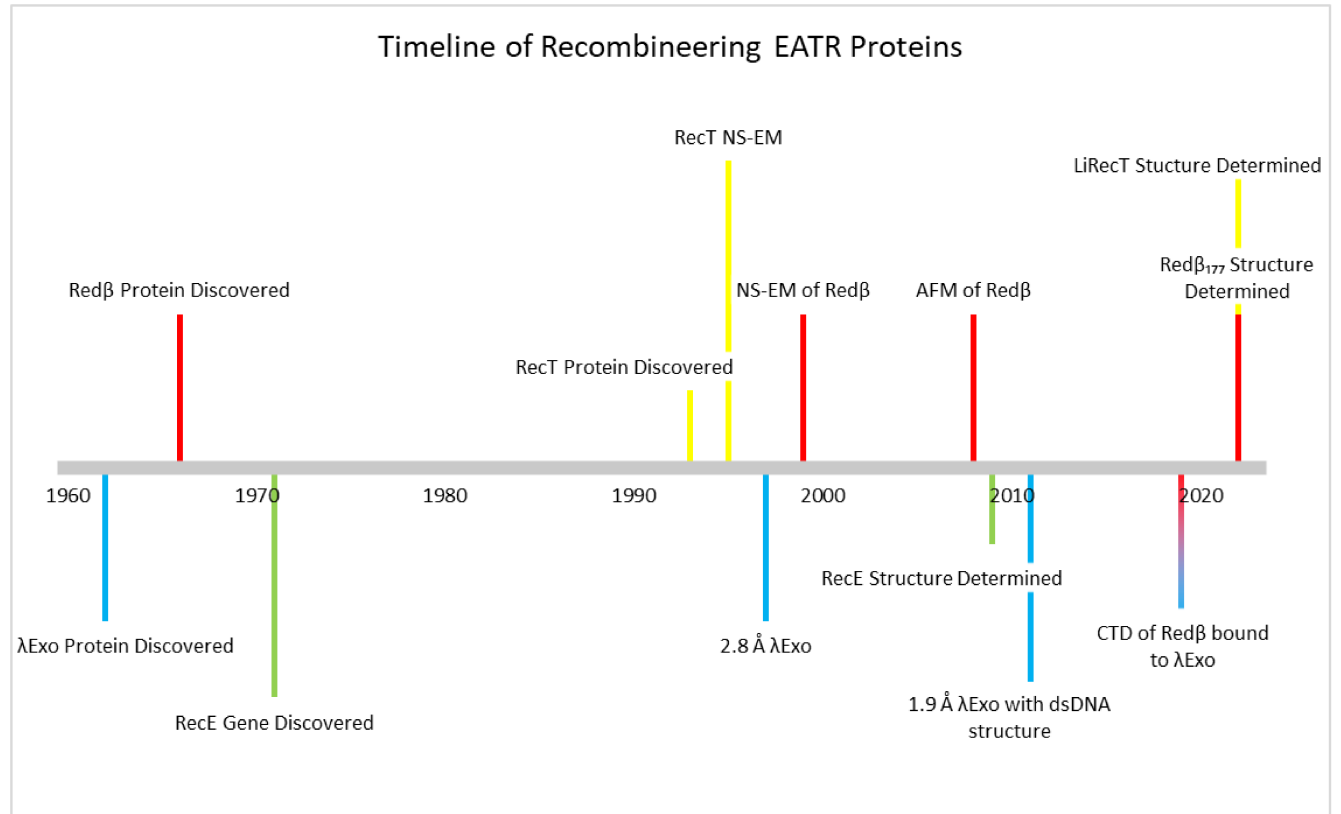


Figure 1: Timeline of the discovery and structural advances of EATR proteins used in recombineering [52–67]. Yellow indicates events about RecT, red indicates Redβ, blue λExo, and green RecE.

2.1 Exonucleases Structures

While recombineering with single-stranded oligonucleotides as the electroporated input DNA (often referred to as single-stranded oligonucleotide repair or ssOR) only requires the annealase [68], when the input DNA is double-stranded, both the annealase and exonuclease components of a specific EATR pair are needed. In addition to recombineering, exonucleases as stand-alone enzymes have been exploited for critical roles in other biotechnology applications such as generating ssDNA from dsDNA for PCR [69], CHIP-EXO protein-DNA footprinting [70], and generating ssDNA for several biosensor applications (a few examples include [71–74]). Despite these many uses, little is known about how these proteins are

evolutionarily related to one another, especially when compared to the work done for grouping annealases, discussed below [7,75].

The structures of the two main exonucleases used in *E. coli* recombineering, λ Exo from phage λ [52,60,65,66] and RecE from Rac prophage [63], have been determined by x-ray crystallography (Figure 2). Remarkably, despite having limited sequence identity, both exonucleases form ring-shaped oligomers with central funnel-shaped channels, although λ Exo forms a trimer and RecE a tetramer. In both structures, the dsDNA is thought to enter at the open end of the ring, such that the 5'-strand can feed into one of the active sites to be digested into mononucleotides. The 3'-overhang then exits out the back of the channel to tether the ring to the DNA as it moves forward digesting the 5'-strand [63,65,66]. This same oligomeric architecture has been seen for λ Exo, RecT, and for a third member of this family whose structure has been determined, the alkaline Exonuclease from *Laribacter hongkongensis* [76,77].

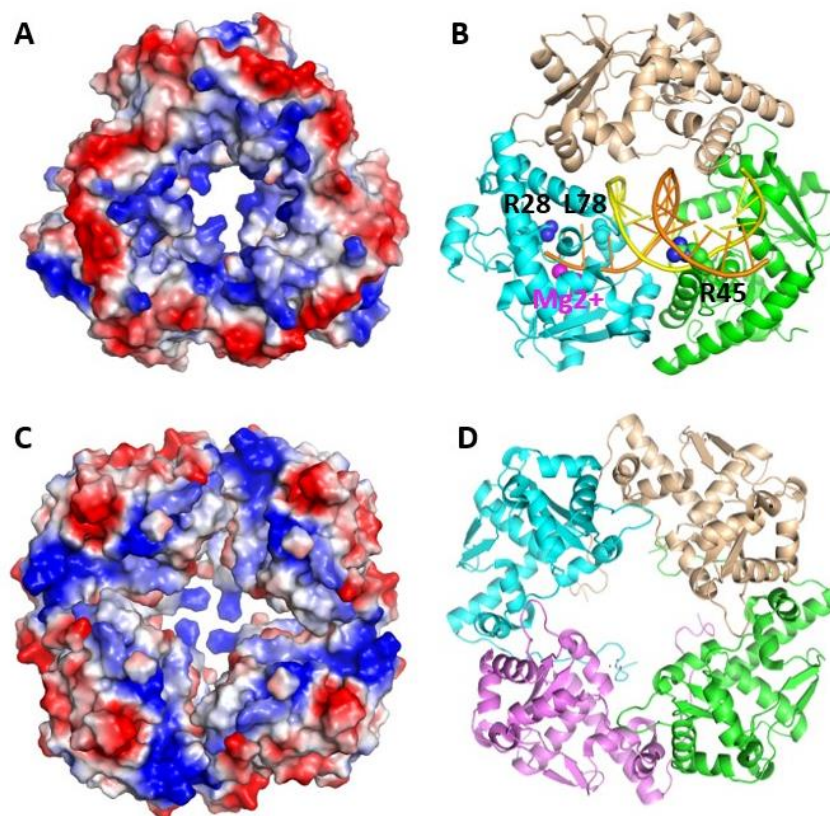


Figure 2: Structures of λ Exo and RecE. A. Surface electrostatic view of the λ Exo trimer (1AVQ). The wide end of the central channel to which the dsDNA substrate would enter is facing the viewer. B. Cartoon view of the λ Exo trimer complexed with dsDNA substrate (3SM4). The 5'-end of the orange DNA strand binds in the active site of the cyan subunit, near two Mg^{2+} ions. Side chains discussed in

the text are labeled. C, D. Surface electrostatics and cartoon view of the RecE nuclease domain tetramer (3H4R). The view is again looking down into the wide end of the central channel. Notice that the central channels in both structures narrow at the back end.

2.2 λ Exo Structure

λ Exo is a highly processive alkaline exonuclease that initiates digestion at dsDNA ends. The rate of dsDNA digestion is 5-40 nucleotides per second, as determined both at the single molecule level [78–80] and in bulk biochemical studies [65,81–83]. A peculiar feature of λ Exo is that it requires a 5'-phosphate on the dsDNA end for active digestion [56] yet binds to dsDNA with either 5'-OH or 5'-PO₄ ends with roughly equal affinity [84]. As the 5'-PO₄ is five covalent bonds removed from the phosphodiester bond that is cleaved in the reaction, its impact on catalytic activity but not on binding was perplexing. A clue as to the role of the 5'-phosphate came from mutagenesis studies indicating a pivotal role for Arg-28 in enzyme processivity, and an interaction of Arg-28 with the 5'-phosphate was suggested based on modeling [85].

λ Exo was the first recombineering protein to have its crystal structure determined. Although first crystallized in 1985, the crystals at that time only diffracted to a 6 Å resolution [86]. It was not until 12 years later that a crystal structure was determined at 2.8 Å resolution, without DNA [60]. The structure revealed a ring-shaped homotrimer [87] with a central channel of 30 Å at one end, enough to allow dsDNA to enter, but only 15 Å at the other end, allowing only ssDNA to exit [60]. The proposed DNA binding mode nicely explained the high processivity of λ Exo, as the ring-shaped trimer would be physically tethered to the DNA molecule as it moves along digesting it.

Over a decade later, the structure of λ Exo in complex with DNA substrate was determined [66]. The crystallized complex contained a 12-bp duplex with a 5'-phosphorylated 2-nt overhang at one end (a 14-mer/12-mer), the inactive K131A variant of λ Exo to prevent DNA digestion, and the Mg²⁺ ions that are required for nuclease activity. The structure showed that the DNA is indeed bound to the central channel, but significantly tilted to place the end of the DNA with the 2-nt overhang into one of the three active sites. The two nucleotides at the 5' end of the DNA are bent away from the duplex and inserted into an active site cleft, while the 3'-OH of the opposing strand is positioned to exit out the back of the trimer.

The unwinding of the DNA is mediated by apolar residues, including Leu-78 that wedge into the base pairs to separate them. The 5'-phosphate of the DNA is indeed bound at the end

of the active site to Arg-28, while the scissile bond is bound to two Mg^{2+} ions held in place by crucial acidic active site residues. The structure, which visualizes the nucleophilic water molecule that is poised for attack [65] supports a classic two-metal nuclease mechanism [88] characteristic of the type 2 restriction endonuclease (T2RE) family [87], also known as the PD-(D/E)XK family [89]. Three loops of λ Exo, one from each subunit, extend from the rim of the central channel to contact the downstream portion of the dsDNA substrate. The Arg-45 side chain from one of the three loops inserts into the minor groove of the DNA and is proposed to help the enzyme keep on track. In support of this role, mutation of Arg-45 to Ala completely disrupts activity [65,90].

Based on this structure, an “electrostatic ratchet” model for processive digestion was proposed in which the interaction of the 5'-phosphate on the DNA with Arg-28 at the end of the active site is key to moving the enzyme forward. As each mononucleotide is cleaved from the 5'-end and released with Mg^{2+} out the rear portal on the trimer, the newly generated 5'-phosphate of the next nucleotide on the DNA would be attracted to the positively charged pocket containing Arg-28. The hydrophobic wedge formed by Leu-78 is proposed to help unwind the base pairs as the enzyme moves along the DNA, and the Arg-45 side chain is thought to act as a rudder to help the trimer track along the minor groove of the downstream portion of the DNA.

The most recent crystal structure of λ Exo shows a trimer bound to three copies of the Red β CTD, resolved to 2.3 Å [52] (Figure 3). This structure provided the first direct insights into the architecture of the λ Exo-Red β EATR complex and is remarkably consistent with a model proposed by Tolun and Myers in 2007 [91] in which the role of the interaction is to load the Red β annealase directly onto the 3'-overhang that is formed by λ Exo during digestion. Further details of this interaction will be examined below in section 2.5.

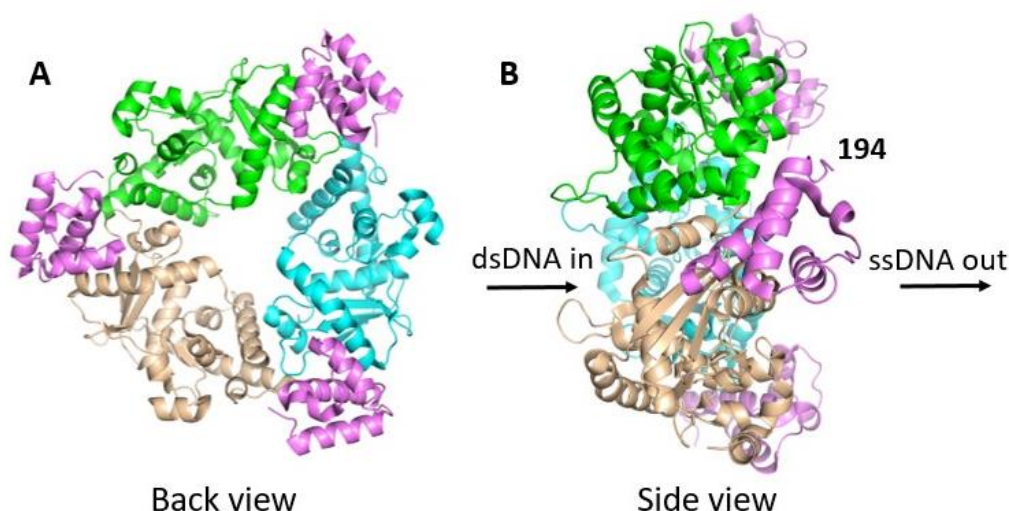


Figure 3. Structure of λ Exo in complex with the Red β CTD. A. view from the back of the λ Exo trimer (6M9K). The dsDNA substrate would enter the ring from the back and the 3'-overhang ssDNA would come out towards the viewer. The three Red β CTDs are colored magenta. B. Side view showing that the three Red β NTDs and linkers (residues 1-193) would lie on the face of the λ Exo trimer from which the 3'-overhang ssDNA is extruded. The structure thus suggests a mechanism in which Red β monomers are directly loaded onto the 3'-overhang as it is formed by λ Exo.

2.3 RecE Structure

While λ Exo is a 226 amino acid protein, RecE is a much larger 866 amino acid protein that contains a C-terminal nuclease domain (residues 564-866) and an N-terminal domain of unknown function. The nuclease domain can substitute genetically for the full-length protein [92], although full-length RecE has enhanced activity for some recombineering applications *in vivo* [93]. The crystal structure of the nuclease domain was determined at 2.8 Å in the absence of DNA in 2009, between the time the two λ Exo structures without and with DNA were published [63]. The RecE fold has a core topology similar to λ Exo and a common set of conserved active site residues. Intriguingly, the RecE monomers pack into the tetramer in essentially opposite orientations as λ Exo, relative to the end of the channel at which the DNA would enter. This suggests that although RecE and λ Exo are evolutionarily related at the tertiary structure level, their similar quaternary structures (RecE tetramer and λ Exo trimer) likely evolved independently from a monomeric ancestor. Clearly, a ring-shaped structure with a tapered central channel is a fundamental architectural feature for this processive 5'-3' exonuclease enzyme family.

Each subunit of RecE has a channel that contains an active site that connects to a positively charged portal that could allow for the release of mononucleotides as they are cleaved (Figure 2c). The structure was determined without DNA in the presence of Ca^{2+} , which

supports DNA binding but not cleavage. Although only one Ca^{2+} ion is bound per active site, two Mg^{2+} ions are presumably needed for cleavage. The set of critical active site residues is primarily conserved between RecE and λExo , with one notable exception: Glu-85 of λExo is replaced by His-652 in RecE. This residue is also histidine in the C-terminal nuclease domain of RecB of the *E. coli* RecBCD complex, another member of the T2RE family. The role of this residue in catalysis is not yet clear, but it could help to stabilize the 3'-OH leaving the group after hydrolysis.

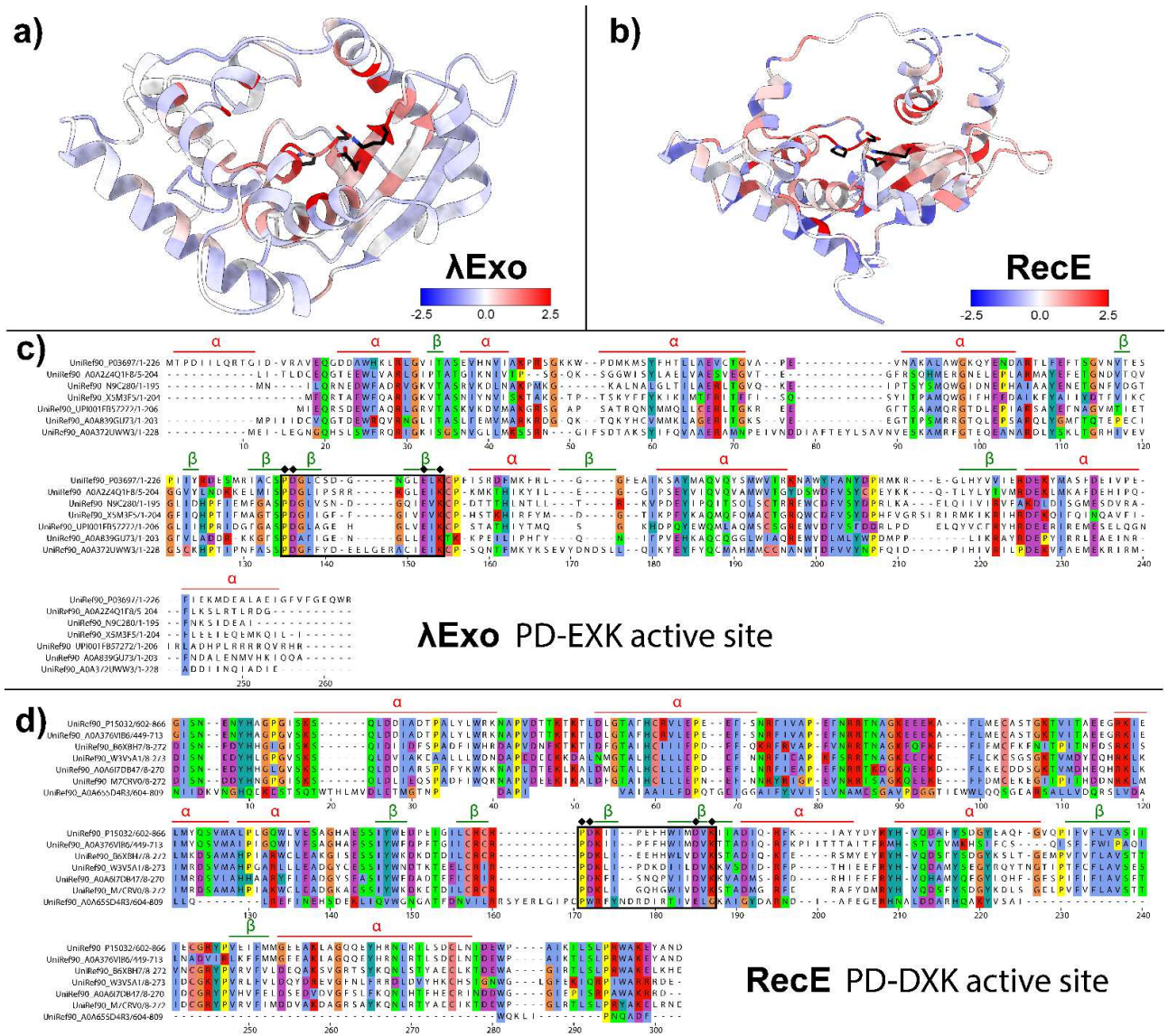
Another difference between RecE and λExo is that RecE contains much longer loops projecting out from the rim of the central channel, presumably to capture the dsDNA substrate. These loops, formed by residues 665-698 of RecE, are largely disordered in the crystal structure and are not part of the final refined model. One of our laboratories (Bell) successfully crystallized the RecE nuclease domain in a complex with different lengths of DNA. However, the DNA could never be visualized, presumably because it did not sit down in a unique orientation relative to the crystal packing interactions. The loops did, however, become partially visualized in these structures.

In summary, the RecE and λExo structures show several common features that appear fundamental to the processive nuclease activity required for 5'→3' end-resection. These features also appear to be conserved for the additional structures of related exonuclease proteins of the phage recombination systems that have been determined.

2.4 Evolutionary Analysis of Recombineering Exonucleases by Sequence Alignments

We analyzed sequence conservation in both λExo and RecE using the 2000 hit blast search results against the UniProt ref90 database. Following multiple sequence alignment and quality control of these datasets, the final MSAs consisted of 1347 sequences for λExo and 183 for RecE. Many of the RecE sequences were eliminated once sequences with greater than 90% similarity were clustered. Both λExo and RecE belong to the PD-(D/E)XK phosphodiesterase superfamily, a highly diverse group of proteins with homologs present in all domains of life. The superfamily consists primarily of nucleases, including processive exonucleases such as λExo , RecE, *E. coli* RecB, and the herpesvirus alkaline nuclease UL12, as well as many restriction endonucleases including those used in traditional cloning techniques. As with most members of this superfamily, λExo and RecE have a conserved core fold consisting of a four-stranded, mixed β -sheet flanked by α -helices, with $\alpha\beta\beta\beta\alpha\beta$ topology [89]. Embedded within

334 this fold are the conserved aspartate, glutamate, and lysine residues that give the PD-(D/E)XK
 335 family its name.



336
 337 **Figure 4: Sequence conservation of λExo and RecE:** Sequence conservation of a) λExo
 338 (PDBID:1AVQ) 1347 sequences, and b) RecE (PDBID:3H4R) 183 sequences. Scale values represent
 339 the entropy-based AL2CO conservation index, where blue and red correspond to the lowest and highest
 340 sequence conservation, respectively. Active site residues are indicated in black stick representation.
 341 Representative MSA is shown for c) λExo and d) RecE that summarise the full 2000 hit analysis.
 342 Secondary structural features, mapped from atomic structures of the reference sequences, are indicated
 343 above the MSA, with α-helices in orange and β-sheets in green. Active site residues are indicated above
 344 the MSA by a black diamond, and the active site region is bound by a black box. The consensus active
 345 site for each family is provided at the end of each respective MSA.

346
 347 In our analysis, the active site residues of both proteins appear to be highly conserved,
 348 and in most cases identical across all constituent sequences. In contrast, other regions of the

protein are more variable (Figure 4a & b, Supplementary Figures 1 & 2). Notably, λ Exo and RecE differ in the composition of their active site residues. λ Exo displays a highly conserved PD-EXK active site structure (Figure 4c), whereas RecE has the alternate structure PD-DXK. In both alignments, conserved positively charged residues flank the active site and are thought to facilitate binding to the DNA substrate.

Of particular interest is the great disparity in length between the λ Exo and RecE families. While the PD-(D/E)XK-like domain spans the entire length of the λ Exo sequences, it comprises only the C-terminal segment of RecE. A review of the hits retrieved by RecE reveals a heterogeneity in sequence length, with some showing homology across the entire length of RecE including its large N-terminal domain. In contrast, others like λ Exo consist of only a single exonuclease domain (Supplementary Figure 1). As expected, the C-terminal PD-(D/E)XK domain showed the greatest conservation across all alignment regions prior to truncation, possibly explaining the higher number of sequences eliminated during the clustering step.

A review of the hits retrieved by BLAST search of λ Exo identified many sequences identified as homologs of YqaJ, a domain from a known EATR pair found in the skin element of *Bacillus subtilis* [94]. Other homologs of YqaJ include Chu exonuclease of the *B. subtilis* phage SPP1, which forms an EATR pair with its partner annealase GP35 [17,94].

Also of note, the elements that encode λ Exo and RecE have different reproductive methods. λ Exo is encoded within a phage that can undergo lytic reproduction, whereas RecE is present in a defective prophage replicating with the host. This difference in reproduction method could have a marked effect on sequence evolution. Phage λ has a generation time of ~ 7.7 phages min^{-1} [95], which is $>20\times$ faster than that of *E. coli*, which is estimated at ~ 0.3 bacterium min^{-1} [96]. The more rapid evolution of λ Exo could account for the higher level of sequence similarity for the RecE family in our analysis. Alternatively, the differences in similarity could reflect the limited number of RecE sequences available in the current UniProt Database.

2.5 The Lambda Phage EATR Complex: λ Exo+Red β

The interaction between the two phage EATR proteins has been known for over half a century. In fact, the Red β protein was discovered during the purification of λ Exo, as the two proteins were seen to co-purify with an apparent 1:1 stoichiometry [REF 139]. While the

nuclease activity of λ Exo had been well known [55], the function of Red β was not established until nearly a decade later when it was discovered that Red β could promote the annealing of homologous ssDNA strands [REF 100]. Although Red β can function independently, somewhat higher annealing activity was observed in the presence of λ Exo [97]. The reason for this is still not apparent. There is actually a third protein in the Red system known as γ -protein (also referred to as Gam), encoded by the *gam* gene [REF 103-105]. The γ -protein is often not present in genomes encoding the typical EATR pair, including in the *E. coli* Rac prophage encoding RecE and RecT, and it appears to take on a more supplementary role. In phage λ the *gam* gene is required to transition from the early to the late stage of viral infection [98], but the γ -protein does not appear to interact with any other phage λ recombination proteins [99] [DIFFERENT REF? 101 is a review]. Instead, it binds to the host RecBCD helicase/exonuclease complex to prevent it from digesting dsDNA ends [REF 103,105], which are present on the linear form of the λ phage genome [100]. [Best REF??] The γ -protein can efficiently inhibit both nuclease activities of RecBCD, including its exonuclease activity on dsDNA and its ssDNA and endonuclease activity on ssDNA [101,102]. Crystal structures of γ -protein reveal a small alpha-helical dimer, and a cryo-EM structure of γ -protein in complex with RecBCD has been determined [103,104]. Red-mediated recombination with linear dsDNA can occur without γ -protein *in vivo* [12, BEST REF?]. Still, γ -protein is typically included in recombineering strains with active RecBCD to prevent the destruction of linear duplex DNA [12, BEST REF?].

While there is currently no structure of a complete EATR complex, there have been attempts to model what the λ phage EATR complex could look like (Figure 5). One of the first models, proposed by Tolun in 2007, considered the available biochemical and stoichiometric data [91] and was comprised of four λ Exo trimers bound to a dodecameric ring of Red β (Figure 5a). This complex would presumably load onto a dsDNA end through one of the λ Exo trimers. According to the model, as the 5'-strand is digested, the exposed ssDNA would be bound to the N-terminal domains of the associated Red β subunits[91]. Although the purified EATR complex has a 1:1 stoichiometry, the concentration of Red β in cells is higher than that of λ Exo [REF? IS THIS TRUE?]. Mechanistically, λ Exo should only need to be stoichiometric with dsDNA ends, whereas higher levels of Red β would be required to form the larger oligomeric complexes on DNA. Presumably, Red β monomers would detach from the EATR complex to form the complex with the nascent 3'-overhang ssDNA for recombination [64]. Indeed, expression of Red β at higher levels than λ Exo leads to a significant improvement of recombination efficiency,

whereas an excess of λ Exo over Red β decreases recombination levels [105]. A similar relationship was observed for RecE and RecT (in the same publication), suggesting that the two EATR pairs work by similar mechanisms [105].

While the first EATR model was primarily based on biochemical data, Newing *et al.* [64] proposed a model based on newly available structural data (Figure 5b). This model incorporated the structures of λ Exo bound to DNA [65] and to the CTD of Red β [52], combined with the Red β_{177} cryo-EM structure [64]. AlphaFold 2 was used to predict a structure for the linker region of Red β (residues 178-193), which has not yet been resolved experimentally. This model also assumed higher levels of Red β than λ Exo as the long 3'-ssDNA overhang generated by λ Exo trimer digestion would require multiple Red β monomers to form a continuous protein-DNA filament seen in the structure.

Here we propose a third possible model for the phage λ EATR complex, generated using AlphaFold 2 (Figure 5c). The model contains three λ Exo and three Red β subunits, and retains the signature 1:1 ratio, with λ Exo forming the characteristic trimer, three Red β N-terminal domains interacting with one another as in the Red β_{177} structure, and three Red β C-terminal domains bound to λ Exo as in the crystal structure of the complex [REF 55]. While this model is asymmetric in how the N-terminal domains are positioned, the Red β linker region is likely to be flexible enough to allow for the conformational differences. Interestingly, the cleft on the Red β N-terminal domain that contains DNA in the cryo-EM structure is occupied by a new α -helix from the linker region that is generated by AlphaFold 2 [64]. This α -helix would block DNA binding and could conceivably control how Red β monomers assemble on the nascent ssDNA that is generated by λ Exo. While there is still no experimentally determined structure of the full phage λ EATR complex, the insights we gain from each new structure and model help to assemble the pieces.

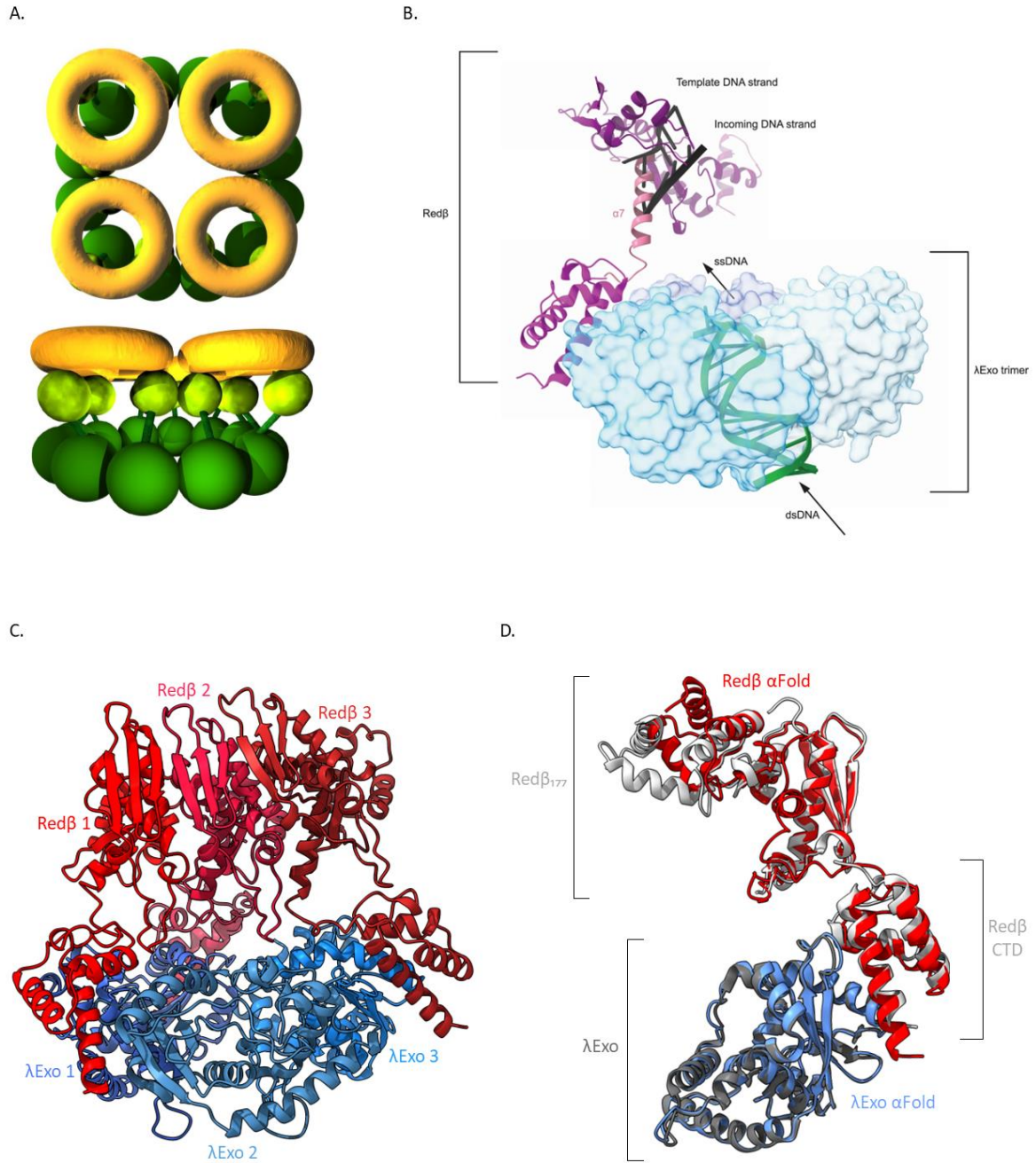


Figure 5: Proposed models of the EATR complex of Redβ and λExo. (A) is an early model, reproduced with author permission from Tolun [91], showing four λExo trimers (yellow) bound to the C-terminal domains of Redβ (lime) that form a larger ring with their N-terminal domains (darker green). (B) is a reproduction with author permission of the model proposed by Newing *et al.* [64] showing a composite structural model of the complex, with Redβ highlighted in pink and λExo in blue. AlphaFold 2 was used to predict the structure of Redβ and λExo (C), showing a λExo trimer (blue) bound to three Redβ monomers (red). (D) highlights the interaction of single λExo and Redβ monomers with the experimentally determined structures of λExo (PDB ID = 6M9K) in dark grey and of the C-terminal domain (PDB ID= 6M9K) and the first 177 amino acids (PDB ID = 7UJL) in light grey.

3.1 Annealase Proteins

There are many distinct types of proteins with annealase activity found in nature, as first mapped out by Iyer *et al.* who proposed three distinct superfamilies grouped around ERF (essential recombination function), RecT/Red β , and Rad52. Each family was predicted to have a different core fold and a distinct pattern of sequence conservation [7]. Lopes *et al.* later proposed a different grouping based on Rad52-like, Gp2.5-like, and Rad51-like sequences [75]. Most recently, seven annealase families were proposed including Sak3, Sak4, Rad52/22, ERF, RecT/Red β , Gp2.5, and RecA [106]. The latter two groupings included Rad51/RecA family proteins that have annealase activity, but primarily function in ATP-dependent DNA strand exchange for homologous recombination [45]. Similarly, Gp2.5 is a single-stranded DNA binding protein from bacteriophage T7 that presumably has annealase activity as a side effect of ssDNA-binding [107]. Due in part to the diversity of annealase proteins, we have yet to arrive at a consensus mechanism for how they catalyze DNA annealing. Based on their distinct core folds and presumably different evolutionary origins, the different types of annealase proteins could indeed operate by different mechanisms. Of the 7 families described most recently, only three have representative high-resolution structures available, namely the Rad52, RecA/RAD51, and RecT/Red β families.

Recombineering has primarily been developed and optimized for use in *E. coli*, and the RecET and phage λ Red proteins have evolved to function in *E. coli*. The annealase activity is now known to depend on an interaction with the host SSB protein [REF 20,55], which will vary in sequence in different hosts. Therefore, it can be challenging to predict if λ Red or RecET will be functional in a given bacteria of interest. However, recombineering can be expanded to new organisms by mining for EATR proteins from a bacteriophage (or prophage) that infect them (Table 1). Moreover, Red β from phage λ functions efficiently as an annealase for recombineering in close relatives of *E. coli* including *Salmonella enterica* [36]. The interaction between annealases and host SSB proteins has recently been shown to largely involve the last ~9 residues of SSB [REF 55], which is the site for interaction of numerous *E. coli* host proteins [REF 149]. Altering this sequence has allowed the portability of a given annealase into a new bacterium of interest [20] increasing the efficiency of recombineering in new bacterial hosts. Knowledge of how RecT and Red β operate can also benefit our understanding of the annealase mechanism, for which there is a general lack of structural information, particularly for the relevant protein-DNA complexes. While structures with ssDNA substrate have been available for eukaryotic annealases including Rad52 and ICP8, the recent structures of RecT and Red β

[64,67] have been determined in complex with a duplex intermediate of annealing, and therefore provide important new insights into the possible annealing mechanism, as will now be described.

3.2 Structure of Red β

While Red β was discovered over half a century ago [55,108], structural investigations only began approximately 20 years ago. The first structures reported by Passy *et al.* used negative staining electron microscopy (NS-EM) [62] and revealed oligomeric rings in the absence of DNA and larger rings with ssDNA. Left-handed helical filaments were observed when Red β was mixed with heat-denatured double stranded DNA, which was the first indication of a structural transition upon annealing [62]. Almost ten years after Passy *et al.*'s findings, further investigations using atomic force microscopy (AFM) [109] revealed similar helical filaments in the presence of two complementary ssDNA sequences, but disperse monomers bound to a single ssDNA sequence. A model for annealing was proposed in which a clamped dimer of Red β stabilizes a nucleus of complementarity from which annealing can propagate [109].

While the work of Erler *et al.* confirmed that Red β forms ring-like structures in the absence of DNA, it more clearly showed structures resembling a split-lock washer, with a gap or a slight overlap between monomers at one end of each ring [109]. Unlike NS-EM, which gives 2D projection, AFM imaging is sensitive to height, which explains how split-lock washers were detected. As there were some differences in the oligomeric complexes seen by different low resolution imaging methods, a high-resolution structure of Red β was clearly needed to resolve the discrepancies.

The first atomic structure of Red β was of its C-terminal domain, determined in complex with a λ Exo trimer. The overall architecture of the complex supported a model in which Red β is loaded onto ssDNA during DNA end-resection by λ Exo [52]. From a mutational analysis of the λ Exo-CTD interface, a second role for the CTD in binding to the host SSB protein was discovered. The two interactions were found to use an overlapping site and are thus likely to be mutually exclusive. A 'hand-off' model was proposed in which the interaction with λ Exo loads Red β onto the first ssDNA (the 3'-overhang formed by λ Exo). In contrast, the subsequent interaction with SSB localizes the initial Red β -ssDNA complex to the lagging strand of the replication fork, where it can scan the lagging strand for a sequence that is complementary to the first ssDNA [52]. The structure of the Red β CTD is also significant because the other

available structure of Red β that would become available only includes its N-terminal DNA-binding domain [64].

Most recently, cryo-EM revealed the structure of the Red β N-terminal domain (NTD) that is responsible for DNA binding and oligomerization (Figure 6b). Rather dramatically, the structure captured a helical filament of Red β in complex with a novel intermediate of DNA annealing that has an unusual conformation of duplex DNA [64]. The structure used a truncated form of Red β that only included its first 177 amino acids (out of 261 in native Red β). The cryo-EM 2D class averages showed 1- and 2- start helical filaments, **with the start of a helix denoting the number of threads that are found per turn of a helix** [64]. While 2-start filaments have also been observed for the ICP8 annealase from HSV-1, their functional role is not clear [110]. The 1-start filaments of Red β_{177} on the other hand suggested a compelling mechanism for annealing.

The Red β_{177} structure demonstrates the molecular mechanism of how this protein, and most likely its homologs in the Red β /RecT family, anneal DNA. The cryo-EM sample was created by sequentially adding complementary 27 nt ssDNA oligonucleotides, a common approach for forming annealed DNA intermediates [62,109,111,112]. For the first time, a protein was visualized in complex with a conformation of DNA with a ribbon-like planar structure, as opposed to the typical B-form double helix [64]. While this conformation was first seen for Red β , it was quickly confirmed for RecT, lending strong support for functional significance and a conserved mechanism of annealing [67]. The structure clearly revealed a site size of 4 base pairs of DNA for each Red β monomer. However, each Red β monomer makes contact with nucleotides from 6 consecutive base pairs of the bound DNA [64].

This structure also revealed the mechanism of oligomerization of Red β subunits. While it had been well known that Red β forms oligomeric complexes [62,109] the Red β_{177} structure showed that oligomerization is stabilized mainly through electrostatic interactions between monomers, with patches of positive and negative charges that mirror each other on opposite faces [64]. It was also noted that some of the residues involved in oligomerization help to stabilize the DNA-binding pocket, reinforcing the idea that oligomerization is inherently coupled to DNA-binding [64]. Caldwell *et al.* [113] examined the oligomeric structures formed by Red β with different DNA substrates. They found that while the binding of Red β to ssDNA gave rise to a wide distribution of different species, the sequential addition of two complementary oligonucleotides formed a much more distinct and stable complex with Red β

[111]. While the relationship between oligomerization and DNA binding is still not fully understood, the new structural data support the conclusion that they are coupled.

Although not resolved in the cryo-EM map, residues 130–139 of Red β form a so-called “finger loop.” Molecular Dynamics simulations suggested that this loop is highly flexible and moves between open and closed conformations to control access of the DNA strands to the binding groove on Red β . By holding the second (incoming) ssDNA strand within the groove to sample the more deeply bound template ssDNA, while allowing it to disengage if the match is not perfect, this loop could play a crucial role in decreasing the sequence homology search from 3D (i.e., diffusing into and out of the DNA-binding site from the solution) to 1D (i.e., sliding in the DNA-binding site, along the template ssDNA strand). Further experiments will be needed to test this model more rigorously.

3.3 The Structure of RecT from a prophage of *Listeria innocua*

For many years, one of our laboratories (Bell) had been attempting to crystallize RecT/Red β homologs (and fragments thereof) alone and in complex with DNA. Although several crystals with DNA were obtained, none diffracted x-rays well enough to permit structure determination. With the advent of methods for high-resolution cryo-EM, we prepared grids of full-length Red β alone and in complex with different lengths of ssDNA and sequentially added complementary DNAs. These complexes invariably gave spiral C-shaped structures that sat in the plane of the ice (Figure 6a). Although the 3D structures of the complexes could not be determined due to strong preferred orientation, wedge-shaped subunits were readily apparent. Curiously, the structures contained many more subunits (>20) than expected from earlier ring-shaped complexes of λ Red β seen by negative stain EM, with previous EM studies estimating between 11 and 18 subunits per ring [62,109]. In hindsight, these appear very similar to the cryo-EM structure solved at high resolution for the Red β_{177} NTD, in terms of the approximate number of subunits per turn, overall diameter, and shallow helical pitch [64]. When compared to LiRecT, the Red β_{177} NTD structure is much less tightly wound, has a larger outer diameter and a larger inner diameter and a larger hole down the center, and thus a much shallower helical rise per subunit. Between Red β_{177} and LiRecT, full length Red β is a much closer approximation to the Red β_{177} structure. This suggests that the complexes formed by the NTD as seen by cryo-EM are likely to be very similar to the complexes formed by the full-length Red β . However, since the complexes of full-length Red β

576 did not stack end to end as for the NTD, they did not form the longer filaments that sat in the
577 plane of the ice to be amenable to helical reconstruction.

578 To approach this problem, we took advantage of several full-length RecT/Red β proteins
579 we had purified over the years to screen for crystallization. While three of these gave similar
580 C-shaped structures as λ Red β , one of them, RecT from the A118 prophage of *Listeria innocua*,
581 gave distinct helical filaments when mixed with two complementary 83-mer oligonucleotides
582 that were added to the protein sequentially [67]. The filaments appeared strikingly like those
583 of full-length Red β when mixed with long heat-denatured dsDNA as seen by negative stain
584 EM [REF 65]. Single particle analysis of these filaments resulted in the 3.4 Å 3D
585 reconstruction of a single 1.5-turn filament containing 18 LiRecT subunits and 83 bp of duplex
586 DNA (Figure 6c). The DNA was bound in a highly extended, unwound, and flattened
587 conformation to a narrow, positively charged groove that runs along the outer surface of the
588 filament. Although the 2D class averages and 3D reconstruction converged on a single 1.5-turn
589 filament with 10 subunits per turn, considerably longer filaments were observed in the raw
590 cryo-EM images, suggesting that the filaments can stack end to end. While these longer
591 filaments were not used in the single particle analysis, they would likely be amenable to helical
592 reconstruction, although this has not yet been attempted.

593 The LiRecT monomer shares a common core fold with RAD52 and Red β . LiRecT
594 contains three inserted segments: an N-terminal helical bundle that forms inter-subunit
595 interactions at the upper rim of the filament, a β_A - β_B hairpin inserted between β_3 and β_4 that
596 forms inter-subunit interactions at the lower rim of the filament, and a pair of helices (α_D - α_E)
597 inserted after β_5 that forms the lower rim of the DNA binding groove. Only residues 34 to 244
598 of each LiRecT monomer were visible in the reconstruction. The C-terminal residues (245-
599 271) could form a small domain analogous to the CTD of Red β for interacting with its
600 exonuclease partner and host SSB protein. However, at only 26 residues (instead of ~70), it
601 would be considerably smaller.

602 While the LiRecT mixed with two complementary 83-mer DNA strands formed a stable
603 complex that was solved to 3.4 Å resolution, a lower resolution structure of LiRecT was
604 determined in a complex with a single 83-mer oligonucleotide. This complex also forms left-
605 handed helical filaments, although the filaments were less well-organized and did not stack
606 end-to-end. Although the bound DNA could not be resolved clearly enough to model, there was
607 a considerable amount of density in the groove. One clear and informative feature of this

complex was that the N-terminal lobes of each LiRecT subunit were apparently disordered, as their corresponding density was weak. This observation was intriguing as it suggested a mechanism in which the N-terminal lobes clamp down on the duplex only when a fully complementary strand of ssDNA is incorporated into the complex.

Native mass spectrometry of LiRecT-DNA complexes largely confirmed the numbers of subunits in the LiRecT-ssDNA and LiRecT-dsDNA complexes seen by cryo-EM. Interestingly, a mixture of two different types of LiRecT-ssDNA complexes was observed by native MS: some with one copy of 83-mer ssDNA and 10-12 LiRecT subunits, and some with two copies of the same 83-mer ssDNA and 17-18 LiRecT subunits. While the functional significance of these different types of complexes is yet to be established, we hypothesize that the smaller complexes with a single copy of ssDNA represent initial LiRecT-ssDNA complexes, while the larger complexes represent attempts at annealing at sites of partial complementarity along the 83-mer. In agreement with this hypothesis, the larger complexes are similar in composition to the helical filaments formed with two-complementary DNAs and were likely the species resolved by single-particle analysis of the complex formed when only one 83-mer ssDNA was added to LiRecT.

Based on the LiRecT structures, a model for annealing was proposed in which a cluster of ~10 LiRecT subunits (one for each 5 nt of ssDNA) forms on ssDNA with the ssDNA bound to the inner site. This initial LiRecT-ssDNA complex can sample additional segments of ssDNA and bring them into the groove to attempt to pair with the initially bound template ssDNA. When a complementary match is found, the two DNAs can form a closely paired duplex, and the N-terminal lobe of LiRecT can clamp down on the duplex to stabilize and consolidate annealing. This forms an additional and extensive amount of inter-subunit interactions at the N-terminal lobes (above the DNA binding groove), to form the highly stable complex seen by gel shift and single-molecule experiments.

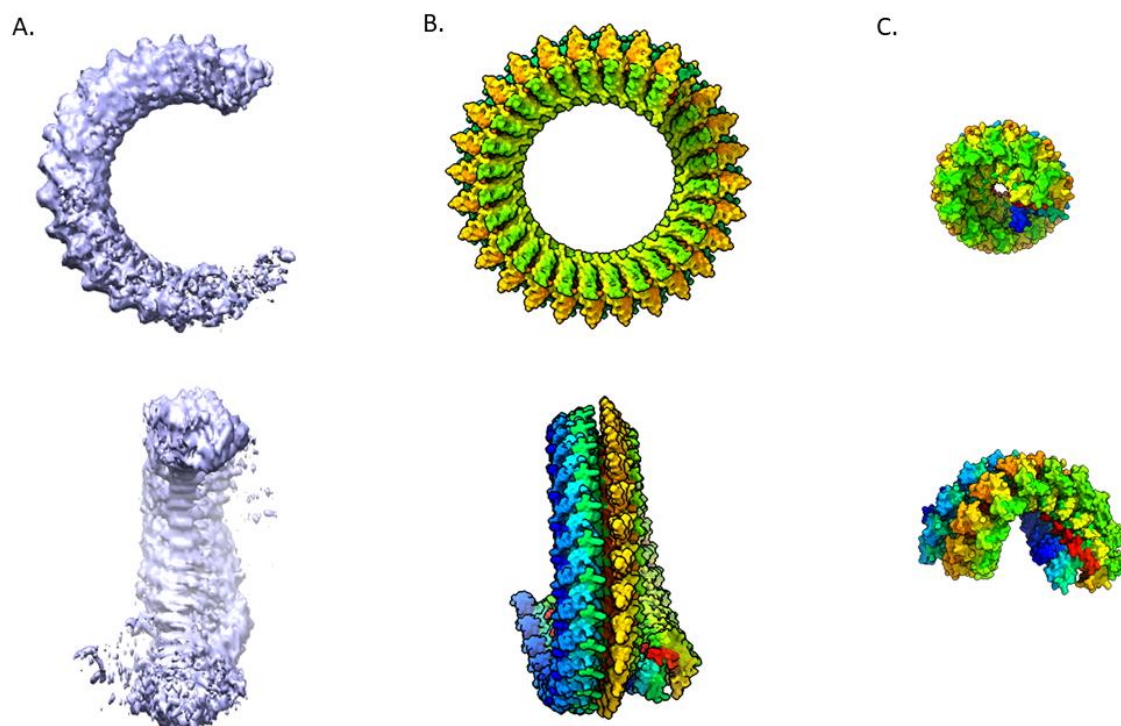


Figure 6: Comparison of Redβ and LiRecT cryo-EM structures. (A) shows a preliminary cryo-EM map of full-length Redβ, showing the incomplete ring shape. This is compared to the Redβ₁₇₇ structure (B), which has been assembled into the ‘split-lock washer’ conformation. A comparison of these two structures highlights the extra density in (A), which matches up with the overlapping parts of the helix in (B). LiRecT (C) structure is also shown, with both a top and side view.

3.4 Comparison of Redβ and RecT

The recent boom in annealase structures creates an opportunity to discover common features that may be important for function. While both Redβ and RecT have been studied since their discovery in the 1960s and 1990s, respectively, it was only in 2022 that their structures became available [64,67]. Despite their high structural similarity, RecT and Redβ have only ~10% sequence identity, and the phage annealase family is extremely divergent[75]. While RecT and Redβ are the defining members of the RecT/Redβ annealase superfamily [7] and were expected to have similar structures, visualization of their common core fold has cemented this relationship. This β-β-α-β-β-α fold is seen in both LiRecT and Redβ₁₇₇, with the most conserved portion being the β-β-β-α core comprised of an α-helix crossing diagonally over three antiparallel β-strands. This β-β-β-α section of the fold was first observed in the structure

of the N-terminal domain (NTD) of Rad52 [114], which solidifies the notion that this entire group of annealases is evolutionarily related from bacteriophages to humans [75].

A notable structural divergence in phage annealase proteins is in their C-terminal domains (CTD), which are believed to bind the exonuclease partner and host SSB protein. The first interaction was seen in a crystal structure of the CTD of Red β bound to λ Exo [52]. Whether other phage annealases including RecT use a similar CTD to bind to their respective partner proteins is unknown. The interaction has been biochemically and functionally confirmed for the *E. coli* RecT/RecE EATR, with RecT binding to RecE but not to λ Exo, demonstrating specificity [105]. While there have been attempts to model what a complete EATR complex containing λ Exo, Red β , and DNA could look like [64], exactly how this complex forms and its mechanistic details will remain unknown until its structure is determined. Such a structure would be expected to explain how the full-length proteins assemble with a 1:1 stoichiometry, despite λ Exo forming trimers and Red β forming rings, split-lock washers, or continuous helical filaments.

The bound DNA conformation is another important similarity between the Red β and RecT structures. Both complexes have the DNA duplex held in a highly extended, flattened, and completely unwound conformation. However, there are some notable differences. First, each LiRecT monomer binds to 5 bp of DNA duplex, whereas Red β (and RAD52) bind to 4 bp (or nt) per monomer. Second, the Red β filaments contain 27 subunits (108 nt) per helical turn, whereas the LiRecT filaments are much more tightly wound with 10 subunits (50 bp) per helical turn. Both have a 5-nt (or 4-nt) repeating pattern of extended DNA, where the protein wedges into the DNA every 5 (or 4) nt to form an extended gap where the base pairs are completely un-stacked.

The details of the protein-DNA and protein-protein interactions that build up the filaments of LiRecT and Red β differ greatly, due to the extremely low sequence similarity. Nonetheless, a striking similarity is that the inner template DNA strand, added first to each protein to form the annealing intermediate, forms considerably more extensive interactions with the protein than the outer (incoming) DNA strand that was added second. The outer (incoming) strand is mainly held in place through normal Watson-Crick base pairs with the inner (template) strand. In the case of Red β ₁₇₇ there could be additional interactions involving the disordered finger loop **that was not visualized** in the structure. Indeed, the finger loop harbors three basic residues (K132, R134, R137) that could potentially interact with the

incoming strand, to supplement its Watson-Crick base pair interactions with the template strand.

4.1 Recombineering annealases and human Rad52

While Red β and RecT are essential to *E. coli* recombineering with both dsDNA and ssDNA, there are a multitude of other annealases found in nature from organisms other than bacteriophages. The idea of recombineering in eukaryotes is not new, and by characterizing the annealases from other organisms, there may be new insights into how this can be achieved. Unfortunately, these proteins are poorly understood, and many annealases such as Erf and Sak4, do not have available structures. Despite the dearth of phage annealase structures, several structures are available for human Rad52 [114–117]. Al-Fatlawi *et al.* [118] compared the structures of Rad52, Red β_{177} , and LiRecT, and described a common annealase fold as the β - β - β - α core. When Al-Fatlawi *et al.* compared LiRecT and Red β_{177} to the crystal structure of Rad52, the conserved region of phage annealases was expanded to encompass a β - β - α - β - β - α fold as the DNA binding region [118]. Not only was this fold visible in both structures, but it had previously been predicted to be conserved among annealases [75]. This β - β - α - β - β - α fold was also seen within the LiRecT structure, again showing homology between all the current annealase structures. Figure 7 demonstrates the conserved nature of this fold, highlighting the distinctive secondary structures in LiRecT, Red β_{177} , and Rad52. These findings demonstrate a structural link between eukaryotic and phage annealases, and show that they bind to ssDNA substrate, and possibly duplex intermediate, via similar mechanisms.

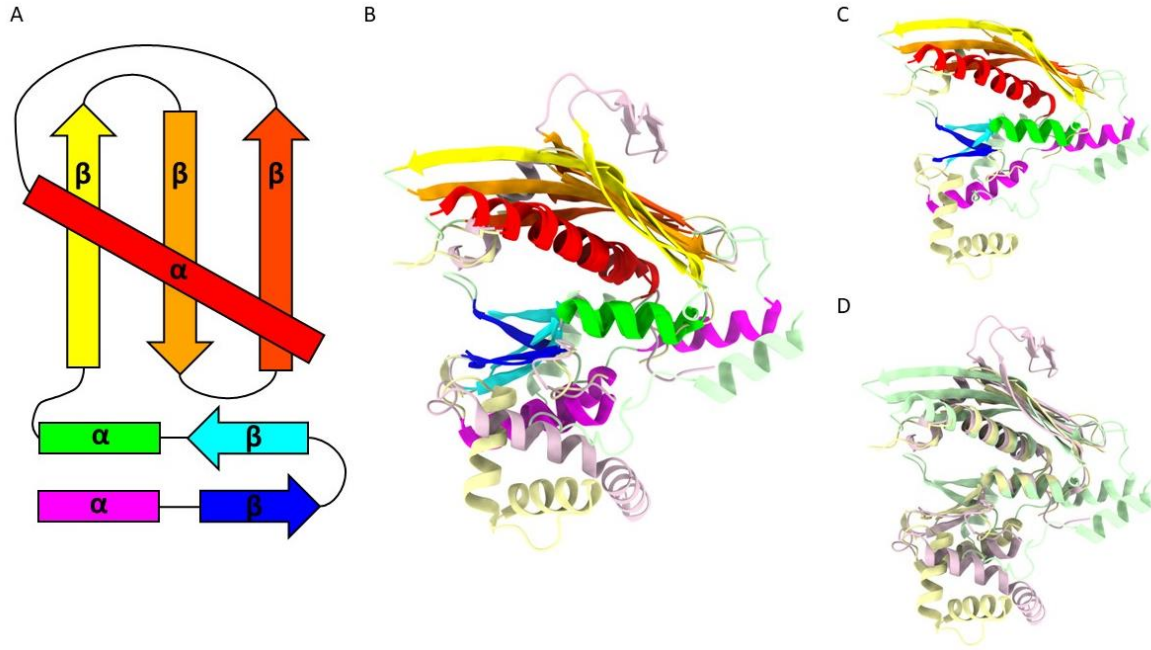


Figure 7: Visualization of the comparison between Red β_{177} , LiRecT, and Rad52 structures. (A) shows a topographical representation of the α - β - β - α - β - β - α fold, inspired by [118]. Each of the secondary structures is highlighted and corresponds to the appropriate structure in (B), which superimposes Red β_{177} (yellow, PDB: 7UJL), LiRecT (pink, PDB: 7UB2), and Rad52 (green, PDB: 8H1P)[64,67,117]. These structures were also represented without the highlighted secondary structures (D) and without Rad52 (C).

The available crystal structures of the NTD of Rad52, including a structure with ssDNA substrate, reveal planar undecamermeric rings [114–117]. By contrast, negative stain EM of full length Rad52 showed heptamermeric rings [119]. Helical filaments have never been observed for Rad52, despite the functional and now structural similarities with Red β and LiRecT. We envision two possibilities to account for this. First, as proposed by Al-Fatlawi *et al.*, the model for Rad52 involving ring-to-ring annealing *in trans* may need to be re-examined [118]. It is conceivable that Rad52 could form filaments under conditions yet to be identified, such as at lower concentrations that are likely to be more relevant *in vivo*. Along these lines, Rad52 has only been visualized at high resolution when bound to ssDNA substrate, while RecT and Red β have only been visualized bound to duplex intermediate. RecT and Red β could indeed function as oligomeric rings at earlier stages of their annealing reactions, such as when bound to an ssDNA substrate. While there is no high-resolution structure of Red β or RecT in the ring form, AFM [REF 113] and NS-EM [REF 65] have shown that Red β can form rings or split-lock washers. The extensive range of oligomeric complexes observed for Red β (summarized in Table 2) indicates a dynamic oligomerization yet to be completely unraveled. It is clear that the

relationship between the annealase fold and the multimeric complexes formed by annealases has much to still be explored.

Table 2 shows the number of subunits in rings formed by Red β observed using different techniques.

Number of subunits	Technique Used	DNA Substrate	Reference
15 - 16	TEM	30 nt oligonucleotide ¹	[62]
18	TEM	Heat-denatured calf thymus DNA	[62]
11 – 12 ²	TEM	No DNA	[62]
14 $\pm$ 3	AFM	83 nt, 123 nt, and 163 nt oligonucleotide ¹	[109]
11 - 12	AFM	No DNA	[109]
14	SEC-MALS	38 nt ssDNA	[111]
21.1	SEC-MALS	83 nt ssDNA	[111]
11.3	SEC-MALS	38 nt oligonucleotide ¹	[111]
18.2	SEC-MALS	83 nt oligonucleotide ¹	[111]
7.7 - 9.1	SEC-MALS	No DNA	[111]
11.5	AUC	38 nt oligonucleotide ¹	[111]
12.5	AUC	38 nt ssDNA	[111]
9 - 13	AUC	No DNA	[111]
11	Native-MS	38 nt oligonucleotide ¹	[111]
22	Native-MS	83 nt oligonucleotide ¹	[111]
9	Native-MS	38 nt ssDNA	[111]
11	Native-MS	83 nt ssDNA	[111]
$\approx$ 12 ¹	Native-MS	No DNA	[111]

¹ Complementary oligonucleotide used.

² Range where multiple rings subunit numbers were determined, with the bold value being the majority value in these cases.

4.2 Recombineering Annealases and other DNA binding proteins

While Rad52 is one of the most well-studied eukaryotic annealases, there are many more annealases found throughout nature, including those from viruses that infect humans, such as HSV1. While Rad52 is the best-characterized eukaryotic annealase, viral annealases could also be utilized for eukaryotic recombineering. The ICP8 annealase from HSV-1 is one of the few eukaryotic annealases with an available structure [120]. ICP8 exhibits annealase functions similar to Red β and is able to stimulate gene targeting in human cells [121]. Interestingly, although ICP8 is not homologous to the RecT and Red β annealases from bacteriophage, it also functions with an exonuclease partner called UL12 to form an EATR complex. ICP8 increases the efficiency of SSA in human cell lines [122] and thus could be key to expanding recombineering for therapeutic purposes.

The most noticeable difference between ICP8 and the prokaryotic recombineering annealases is its size, being over three times the size of the bacteriophage annealases [PLEASE CLARIFY, LENGTH OR PROTEIN, OR SIZE OF OLIGOMER??]. One of the main reasons behind this is likely to be the involvement of ICP8 in more than just SSA. ICP8 is involved in DNA replication [123] and DNA binding [124]. Another difference is the shape of the filament formed. While both RecT and Red β form helical filaments upon DNA binding and annealing [64,67], the ICP8 filament was formed without any DNA [125,126]. The former filaments were single, while ICP8 forms double (bipolar) filaments. Another difference is that the ICP8 DNA binding groove does not appear to be on the outside of the filament, as seen for Red β and LiRecT, but rather is located between the two filaments, going through the center of the double filament. This would indicate a different annealing mechanism, if in fact, the bipolar filaments of ICP8 are relevant. While the ICP8 crystal structure was not of a filament form, NS-EM shows that it forms filaments [127]. All of these indicate a larger oligomer, and more elaborate oligomerization behavior than for the phage annealases currently used for recombineering. This may be one of the reasons why eukaryotic recombineering has proven to be more challenging than bacterial recombineering. While there have been attempts to examine human cell ssDNA recombineering with ICP8, these tests required replicating cells and had rates of recombination less than 0.5 % [121]. This rate is much lower than bacterial systems and demonstrates that further research and optimization is still needed before eukaryotic recombineering becomes feasible. In addition, potential additional interactions with the target cell's proteome may result in further complications. Ultimately, while ICP8 is a promising candidate for the advancement of eukaryotic recombineering, the differences in structure from

the currently used recombineering annealases suggest that we will need to address various challenges before it can be used for recombineering effectively.

Differences between ICP8 and Red β /RecT family are not limited to size or annealing mechanism. For ICP8, oligomerization is achieved by a CTD of one monomer docking onto the NTD of another monomer [120]. The folds at the DNA binding site of ICP8 are also different from the common core seen in Red β /RecT and Rad52. Although ICP8 has a similar fold to the β - β - α - β - β - α fold discussed above, the protein sequence forming this pocket is not continuous, and the packing of the α -helices with the β -strands is also different. The details of herpes virus annealases are outside the scope of this review and will be presented in detail in an upcoming article from the Tolun group.

4.3 Future Directions for Recombineering and Annealases

Knowledge of how the proteins from the phage λ and the Rac prophage EATR systems work is essential for understanding the SSA pathway and improving recombineering and related biotechnologies. The recent structures of phage EATR proteins has provided at least one structure for Red β , λ Exo, RecT, and RecE. While this has highlighted key similarities among these proteins, many questions remain. How do the exonuclease and annealase proteins complex together to form a full EATR? During SSA, does the EATR function as a stable, intact complex, or do the Exonuclease and annealase components continually associate and dissociate more dynamically? While the exonucleases are needed at a stoichiometric ratio to the DNA ends available, annealases are presumably needed at a stoichiometric ratio to the length of the nascent ssDNA generated by the exonucleolytic digestion, which could be thousands of nucleotides. Yet, the EATR complexes appear to have an equal number of exonuclease and annealase monomers (i.e., 1:1 monomer: monomer stoichiometry). How do these different stoichiometric ratios factor in during an SSA reaction? While we now have structures for each of the four main EATR proteins independently, no structures are available for a full EATR complex. There have been attempts to visualize what this complex could look like [64], but only limited structural data supports the proposed model [52]. Therefore, one of the next big frontiers will be determining the structures of complete EATR complexes.

An exciting goal is to identify EATRs from viruses infecting organisms from different kingdoms of life. This would facilitate the development and expansion of recombineering to new organisms. While hundreds of exonucleases and annealases have been identified by various bioinformatics and structure prediction approaches [REF??], most of them function

only in bacteria [REF??] and are typically very distantly related to one another at the sequence level. Outside of the four main proteins discussed here, few other EATR proteins have been extensively characterized *in vitro*. While eukaryotic recombineering is an exciting goal for the future, further structural characterization of the EATRs of HSV1 and related viruses infecting humans and animals will be needed to develop it.

4.4 Conclusion

This review summarized the literature on EATRs catalyzing SSA used in recombineering from a structural biology perspective. The recent annealase structures determined by our groups revealed striking molecular mechanistic details and evolutionary connections. These structures can be exploited to improve recombineering efficiency in the future. Yet, the many questions listed above remain. Future structures of EATR complexes bound to their DNA substrates will unravel further mechanistic details about how an SSA reaction is catalyzed during recombineering. The potential of developing methods for recombineering in eukaryotic cells using the EATRs of HSV1 and related viruses is an enticing goal for the future. Given the lack of a requirement for a specific type of DNA sequence, recombineering by SSA could potentially be even more impactful than CRISPR, though further development is needed.

5.0 Methods

5.1 Comparison of sequence conservation between λ Exo and RecE

A multiple sequence alignment (MSA) was prepared based on 2000 hit BLASTsearch against the Uniref90 protein database. The subsequent hits were then initially aligned using the FFT-NS-2 algorithm in MAFFT v7.52, and the alignments truncated to the region of the query sequence containing the PD-(D/E)XK endonuclease-like domain, as identified in InterPro (1-226 for Exo, 602-866 for RecE) [128,129]. Truncated sequences were clustered by 90% global sequence identity using CD-HIT [130]. Further quality control was completed, including removing sequences less than 150 aa and removing outliers based on a PCA within Jalview v2.11 [131]. The final MSA were re-aligned using the L-INS-i algorithm in MAFFT v7.52 [129]. The resultant MSA were then used to project sequence conservation onto the protein structures in ChimeraX v1.61 using the entropy-based AL2CO method [132,133]. To visualize the sequence variation, secondary structure location, and identity of the active site residues, a small subsection of the original MSA was created containing the query sequence and 6

additional sequences representing the variation retrieved from the full 2000 hit analysis. MSA figures were generated using the pyMSAviz package (available from: <https://moshi4.github.io/pyMSAviz/>). Structural predictions were performed using the ColabFold [134] implementation of AlphaFold2 multimer version 3.

6.0 Bibliography

- [1] T.M. Wannier, P.N. Ciaccia, A.D. Ellington, G.T. Filsinger, F.J. Isaacs, K. Javanmardi, M.A. Jones, A.M. Kunjapur, A. Nyerges, C. Pal, M.G. Schubert, G.M. Church, Recombineering and MAGE, *Nature Reviews Methods Primers*. 1 (2021). <https://doi.org/10.1038/s43586-020-00006-x>.
- [2] K.C. Murphy, Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*, *J Bacteriol*. 180 (1998) 2063–2071. <https://doi.org/10.1128/JB.180.8.2063-2071.1998>.
- [3] K.A. Datsenko, B.L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proceedings of the National Academy of Sciences*. 97 (2000) 6640–6645. <https://doi.org/10.1073/PNAS.120163297>.
- [4] D. Yu, H.M. Ellis, E.C. Lee, N.A. Jenkins, N.G. Copeland, D.L. Court, An efficient recombination system for chromosome engineering in *Escherichia coli*, *Proc Natl Acad Sci U S A*. 97 (2000) 5978–5983. <https://doi.org/10.1073/pnas.100127597>.
- [5] Y. Zhang, F. Buchholz, J.P.P. Muirers, A. Francis Stewart, A new logic for DNA engineering using recombination in *Escherichia coli*, *Nat Genet*. 20 (1998) 123–128. <https://doi.org/10.1038/2417>.
- [6] H.H. Wang, F.J. Isaacs, P.A. Carr, Z.Z. Sun, G. Xu, C.R. Forest, G.M. Church, Programming cells by multiplex genome engineering and accelerated evolution, *Nature*. 460 (2009) 894–898. <https://doi.org/10.1038/nature08187>.
- [7] L.M. Iyer, E. V Koonin, L. Aravind, Classification and evolutionary history of the single-strand, *BMC Genomics*. 3 (2002). <https://doi.org/https://doi.org/10.1186/1471-2164-3-8>.
- [8] S. Ramakrishnan, Z. Kockler, R. Evans, B.D. Downing, A. Malkova, Single-strand annealing between inverted DNA repeats: Pathway choice, participating proteins, and genome destabilizing consequences, *PLoS Genet*. 14 (2018) 1–29. <https://doi.org/10.1371/journal.pgen.1007543>.
- [9] L.J. Marinelli, G.F. Hatfull, M. Piuri, Recombineering A powerful tool for modification of bacteriophage genomes, 2 (2012) 5–14. <https://doi.org/10.4161/bact.18778>.
- [10] K.C. Murphy, λ Recombination and Recombineering, *EcoSal Plus*. 7 (2016). <https://doi.org/10.1128/ECOSALPLUS.ESP-0011-2015>.
- [11] S.K. Sharan, L.C. Thomason, S.G. Kuznetsov, D.L. Court, Recombineering: A homologous recombination-based method of genetic engineering, *Nat Protoc*. 4 (2009) 206–223. <https://doi.org/10.1038/nprot.2008.227>.

- 865 [12] S. Datta, N. Costantino, X. Zhou, D.L. Court, Identification and analysis of recombineering
866 functions from Gram-negative and Gram-positive bacteria and their phages, *Proc Natl Acad*
867 *Sci U S A.* 105 (2008) 1626–1631. <https://doi.org/10.1073/pnas.0709089105>.
- 868 [13] J. Yin, H. Zhu, L. Xia, X. Ding, T. Hoffmann, M. Hoffmann, X. Bian, R. Müller, J. Fu, A.F. Stewart,
869 Y. Zhang, A new recombineering system for *Photorhabdus* and *Xenorhabdus*, *Nucleic Acids*
870 *Res.* 43 (2015) e36. <https://doi.org/10.1093/nar/gku1336>.
- 871 [14] B.J. Caldwell, C.E. Bell, Structure and mechanism of the Red recombination system of
872 bacteriophage λ , *Prog Biophys Mol Biol.* 147 (2019) 33–46.
873 <https://doi.org/10.1016/j.pbiomolbio.2019.03.005>.
- 874 [15] A.T. Tucker, E.M. Nowicki, J.M. Boll, G.A. Knauf, N.C. Burdis, M. Stephen Trent, B.W. Davies,
875 Defining gene-phenotype relationships in *acinetobacter baumannii* through one-step
876 chromosomal gene inactivation, *MBio.* 5 (2014) 1–9. [https://doi.org/10.1128/MBIO.01313-](https://doi.org/10.1128/MBIO.01313-14)
877 14.
- 878 [16] S. Hu, J. Fu, F. Huang, X. Ding, A.F. Stewart, L. Xia, Y. Zhang, Genome engineering of
879 *Agrobacterium tumefaciens* using the lambda Red recombination system, *Appl Microbiol*
880 *Biotechnol.* 98 (2014) 2165–2172. <https://doi.org/10.1007/s00253-013-5412-x>.
- 881 [17] Z. Sun, A. Deng, T. Hu, J. Wu, Q. Sun, H. Bai, G. Zhang, T. Wen, A high-efficiency
882 recombineering system with PCR-based ssDNA in *Bacillus subtilis* mediated by the native
883 phage recombinase GP35, *Appl Microbiol Biotechnol.* 99 (2015) 5151–5162.
884 <https://doi.org/10.1007/s00253-015-6485-5>.
- 885 [18] Y. Kang, M.H. Norris, B.A. Wilcox, A. Tuanyok, P.S. Keim, T.T. Hoang, Knockout and pullout
886 recombineering for naturally transformable *Burkholderia thailandensis* and *Burkholderia*
887 *pseudomallei*, *Nature Protocols* 2011 6:8. 6 (2011) 1085–1104.
888 <https://doi.org/10.1038/nprot.2011.346>.
- 889 [19] X. Wang, H. Zhou, H. Chen, X. Jing, W. Zheng, R. Li, T. Sun, J. Liu, J. Fu, L. Huo, Y. zhong Li, Y.
890 Shen, X. Ding, R. Müller, X. Bian, Y. Zhang, Discovery of recombinases enables genome mining
891 of cryptic biosynthetic gene clusters in *Burkholderiales* species, *Proc Natl Acad Sci U S A.* 115
892 (2018) E4255–E4263. <https://doi.org/10.1073/PNAS.1720941115>.
- 893 [20] G.T. Filsinger, T.M. Wannier, F.B. Pedersen, I.D. Lutz, J. Zhang, D.A. Stork, A. Debnath, K. Gozzi,
894 H. Kuchwara, V. Volf, S. Wang, X. Rios, C.J. Gregg, M.J. Lajoie, S.L. Shipman, J. Aach, M.T. Laub,
895 G.M. Church, Characterizing the portability of phage-encoded homologous recombination
896 proteins, *Nat Chem Biol.* 17 (2021) 394–402. <https://doi.org/10.1038/s41589-020-00710-5>.
- 897 [21] H. Dong, W. Tao, F. Gong, Y. Li, Y. Zhang, A functional *recT* gene for recombineering of
898 *Clostridium*, *J Biotechnol.* 173 (2014) 65–67. <https://doi.org/10.1016/j.jbiotec.2013.12.011>.
- 899 [22] T.M. Wannier, A. Nyerges, H.M. Kuchwara, M. Czikkely, D. Balogh, G.T. Filsinger, N.C. Borders,
900 C.J. Gregg, M.J. Lajoie, X. Rios, C. Pál, G.M. Church, Improved bacterial recombineering by
901 parallelized protein discovery, *Proc Natl Acad Sci U S A.* 117 (2020) 13689–13698.
902 <https://doi.org/10.1073/pnas.2001588117>.

903 [23] Y. Chang, Q. Wang, T. Su, Q. Qi, Identification of phage recombinase function unit in genus
904 *Corynebacterium*, *Appl Microbiol Biotechnol.* 105 (2021) 5067–5075.
905 <https://doi.org/10.1007/s00253-021-11384-x>.

906 [24] H. Huang, X. Song, S. Yang, Development of a RecE/T-Assisted CRISPR–Cas9 Toolbox for
907 *Lactobacillus*, *Biotechnol J.* 14 (2019) 1800690. <https://doi.org/10.1002/biot.201800690>.

908 [25] Y. Xin, T. Guo, Y. Mu, J. Kong, Identification and functional analysis of potential prophage-
909 derived recombinases for genome editing in *Lactobacillus casei*, *FEMS Microbiol Lett.* 364
910 (2017) 243. <https://doi.org/10.1093/femsle/fnx243>.

911 [26] P. Yang, J. Wang, Q. Qi, Prophage recombinases-mediated genome engineering in
912 *Lactobacillus plantarum*, *Microb Cell Fact.* 14 (2015) 1–11. <https://doi.org/10.1186/s12934-015-0344-z>.

914 [27] J.P. Van Pijkeren, R.A. Britton, High efficiency recombineering in lactic acid bacteria, *Nucleic
915 Acids Res.* 40 (2012) e76. <https://doi.org/10.1093/nar/gks147>.

916 [28] A. Bryan, M.S. Swanson, Oligonucleotides stimulate genomic alterations of *Legionella
917 pneumophila*, *Mol Microbiol.* 80 (2011) 231–247. <https://doi.org/10.1111/j.1365-2958.2011.07573.x>.

919 [29] C. Piñero-Lambea, E. Garcia-Ramallo, S. Martinez, J. Delgado, L. Serrano, M. Lluch-Senar,
920 *Mycoplasma pneumoniae* Genome Editing Based on Oligo Recombineering and Cas9-
921 Mediated Counterselection, *ACS Synth Biol.* 9 (2020) 1693–1704.
922 <https://doi.org/10.1021/acssynbio.0C00022>.

923 [30] J.C. Van Kessel, G.F. Hatfull, Efficient point mutagenesis in mycobacteria using single-stranded
924 DNA recombineering: Characterization of antimycobacterial drug targets, *Mol Microbiol.* 67
925 (2008) 1094–1107. <https://doi.org/10.1111/J.1365-2958.2008.06109.X>.

926 [31] J.C. van Kessel, G.F. Hatfull, Recombineering in *Mycobacterium tuberculosis*, *Nat Methods.* 4
927 (2007) 147–152. <https://doi.org/10.1038/nmeth996>.

928 [32] B.-2199-9-20. pdf Lesic, L.G. Rahme, Use of the lambda Red recombinase system to rapidly
929 generate mutants in *Pseudomonas aeruginosa*, *BMC Mol Biol.* 9 (2008) 1–9.
930 <https://doi.org/10.1186/1471-2199-9-20>.

931 [33] T. Aparicio, A. Nyerges, I. Nagy, C. Pal, E. Martínez-García, V. de Lorenzo, Mismatch repair
932 hierarchy of *Pseudomonas putida* revealed by mutagenic ssDNA recombineering of the *pyrF*
933 gene, *Environ Microbiol.* 22 (2020) 45–58. <https://doi.org/10.1111/1462-2920.14814>.

934 [34] B. Swingle, Z. Bao, E. Markel, A. Chambers, S. Cartinhour, Recombineering Using RecTE from
935 *Pseudomonas syringae*, *Appl Environ Microbiol.* 76 (2010) 4960–4968.
936 <https://doi.org/10.1128/AEM.00911-10>.

937 [35] E.M. Barbieri, P. Muir, B.O. Akhuetie-Oni, C.M. Yellman, F.J. Isaacs, Precise Editing at DNA
938 Replication Forks Enables Multiplex Genome Engineering in Eukaryotes, *Cell.* 171 (2017)
939 1453–1467. <https://doi.org/10.1016/j.cell.2017.10.034>.

940 [36] K. Bunny, J. Liu, J. Roth, Phenotypes of *lexA* Mutations in *Salmonella enterica*: Evidence for a
 941 Lethal *lexA* Null Phenotype Due to the Fels-2 Prophage, *J Bacteriol.* 184 (2002) 6235–6249.
 942 <https://doi.org/10.1128/JB.184.22.6235-6249.2002>.

943 [37] R.T. Ranallo, S. Barnoy, S. Thakkar, T. Urick, M.M. Venkatesan, Developing live *Shigella* vaccines
 944 using λ Red recombineering, *FEMS Immunol Med Microbiol.* 47 (2006) 462–469.
 945 <https://doi.org/10.1111/j.1574-695X.2006.00118.x>.

946 [38] A.D. Corts, L.C. Thomason, R.T. Gill, J.A. Gralnick, A new recombineering system for precise
 947 genome-editing in *Shewanella oneidensis* strain MR-1 using single-stranded oligonucleotides,
 948 *Sci Rep.* 9 (2019) 1–10. <https://doi.org/10.1038/s41598-018-37025-4>.

949 [39] J. Yang, Q. Zhang, G. Zhang, G. Shang, Recombineering-Mediated *Sinorhizobium meliloti*
 950 Rm1021 Gene Deletion, *Curr Microbiol.* 80 (2023) 76. [https://doi.org/10.1007/s00284-023-](https://doi.org/10.1007/s00284-023-03188-1)
 951 [03188-1](https://doi.org/10.1007/s00284-023-03188-1).

952 [40] K. Penewit, E.A. Holmes, K. McLean, M. Ren, A. Waalkes, S.J. Salipante, Efficient and scalable
 953 precision genome editing in *Staphylococcus aureus* through conditional recombineering and
 954 CRISPR/Cas9-mediated counterselection, *MBio.* 9 (2018) e00067-18.
 955 <https://doi.org/10.1128/mBio.00067-18>.

956 [41] H.H. Lee, N. Ostrov, M.A. Gold, G.M. Church, Recombineering in *Vibrio natriegens*, *BioRxiv.*
 957 (2017). <https://doi.org/10.1101/130088>.

958 [42] X. Huang, Y. Sun, S. Liu, Y. Li, C. Li, Y. Sun, X. Ding, L. Xia, Y. Hu, S. Hu, Recombineering using
 959 RecET-like recombinases from *Xenorhabdus* and its application in mining of natural products,
 960 *Appl Microbiol Biotechnol.* 106 (2022) 7857–7866. [https://doi.org/10.1007/s00253-022-](https://doi.org/10.1007/s00253-022-12258-6)
 961 [12258-6](https://doi.org/10.1007/s00253-022-12258-6).

962 [43] A. Derbise, B. Lesic, D. Dacheux, J.M. Ghigo, E. Carniel, A rapid and simple method for
 963 inactivating chromosomal genes in *Yersinia*, *FEMS Immunol Med Microbiol.* 38 (2003) 113–
 964 116. [https://doi.org/10.1016/S0928-8244\(03\)00181-0](https://doi.org/10.1016/S0928-8244(03)00181-0).

965 [44] Y. Wu, T. Li, Q. Cao, X. Li, Y. Zhang, X. Tan, RecET recombination system driving chromosomal
 966 target gene replacement in *Zymomonas mobilis*, *Electronic Journal of Biotechnology.* 30
 967 (2017) 118–124. <https://doi.org/10.1016/j.ejbt.2017.10.005>.

968 [45] G.M. Weinstock, K. McEntee, I.R. Lehman, ATP-dependent renaturation of DNA catalyzed by
 969 the *recA* protein of *Escherichia coli*, *Proceedings of the National Academy of Sciences.* 76
 970 (1979) 126–130. <https://doi.org/10.1073/PNAS.76.1.126>.

971 [46] M.M. Stahl, L. Thomason, A.R. Poteete, T. Tarkowski, A. Kuzminov, F.W. Stahl, Annealing vs.
 972 invasion in phage λ recombination, *Genetics.* 147 (1997) 961–977.
 973 <https://doi.org/10.1093/GENETICS/147.3.961>.

974 [47] M. Maresca, A. Erler, J. Fu, A. Friedrich, Y. Zhang, A.F. Stewart, Single-stranded heteroduplex
 975 intermediates in λ Red homologous recombination, *BMC Mol Biol.* 11 (2010).
 976 <https://doi.org/10.1186/1471-2199-11-54>.

977 [48] A.R. Poteete, Involvement of DNA replication in phage lambda Red-mediated homologous
978 recombination, *Mol Microbiol.* 68 (2008) 66–74. [https://doi.org/10.1111/J.1365-](https://doi.org/10.1111/J.1365-2958.2008.06133.X)
979 2958.2008.06133.X.

980 [49] S. V. Rajagopala, S. Casjens, P. Uetz, The protein interaction map of bacteriophage lambda,
981 *BMC Microbiol.* 11 (2011) 213. <https://doi.org/10.1186/1471-2180-11-213>.

982 [50] P.E. Boehmer, I.R. Lehman, Physical interaction between the herpes simplex virus 1 origin-
983 binding protein and single-stranded DNA-binding protein ICP8., *Proceedings of the National*
984 *Academy of Sciences.* 90 (1993) 8444–8448. <https://doi.org/10.1073/pnas.90.18.8444>.

985 [51] M. Olesky, E.E. McNamee, C. Zhou, T.J. Taylor, D.M. Knipe, Evidence for a direct interaction
986 between HSV-1 ICP27 and ICP8 proteins, *Virology.* 331 (2005) 94–105.
987 <https://doi.org/10.1016/j.virol.2004.10.003>.

988 [52] B.J. Caldwell, E. Zakharova, G.T. Filsinger, T.M. Wannier, J.P. Hempfling, L. Chun-Der, D. Pei,
989 G.M. Church, C.E. Bell, Crystal structure of the Red C-terminal domain in complex with
990 Exonuclease reveals an unexpected homology with Orf and an interaction with *Escherichia*
991 *coli* single stranded DNA binding protein, *Nucleic Acids Res.* 47 (2019) 1950–1963.
992 <https://doi.org/10.1093/nar/gky1309>.

993 [53] D. Korn, A. Weissbach, The effect of lysogenic induction on the deoxyribonucleases of
994 *Escherichia coli* K12(λ): II. The kinetics of formation of a new exonuclease and its relation to
995 phage development, *Virology.* 22 (1964) 91–102. [https://doi.org/10.1016/0042-](https://doi.org/10.1016/0042-6822(64)90051-0)
996 6822(64)90051-0.

997 [54] A. Weissbach, D. Korn, The Effect of Lysogenic Induction on the Deoxyribonucleases of
998 *Escherichia coli* K12λ, *Journal of Biological Chemistry.* 237 (1962) PC3312–PC3314.
999 [https://doi.org/10.1016/S0021-9258\(18\)50170-8](https://doi.org/10.1016/S0021-9258(18)50170-8).

1000 [55] C.M. Radding, D.C. Shreffler, Regulation of λ exonuclease: II. Joint regulation of exonuclease
1001 and a new λ antigen, *J Mol Biol.* 18 (1966) 251–261. [https://doi.org/10.1016/S0022-](https://doi.org/10.1016/S0022-2836(66)80244-9)
1002 2836(66)80244-9.

1003 [56] J.W. Little, An Exonuclease Induced by Bacteriophage λ II. Nature of the Enzymatic Reaction, *J*
1004 *Biol Chem.* 242 (1967) 679–686. [https://doi.org/10.1016/S0021-9258\(18\)96258-7](https://doi.org/10.1016/S0021-9258(18)96258-7).

1005 [57] A.J. Clark, Toward a Metabolic Interpretation of Genetic Recombination of *E. coli* and its
1006 Phages, 25 (1971) 437–464. <https://doi.org/10.1146/annurev.mi.25.100171.002253>.

1007 [58] S.D. Hall, M.F. Kane, R.D. Kolodner, Identification and characterization of the *Escherichia coli*
1008 RecT protein, a protein encoded by the RecE region that promotes renaturation of
1009 homologous single-stranded DNA, *J Bacteriol.* 175 (1993) 277–287.
1010 <https://doi.org/10.1128/jb.175.1.277-287.1993>.

1011 [59] R.J. Thresher, A.M. Makhov, S.D. Hall, R. Kolodner, J.D. Griffith, Electron microscopic
1012 visualization of RecT protein and its complexes with DNA, *J Mol Biol.* 254 (1995) 364–371.
1013 <https://doi.org/10.1006/JMBI.1995.0623>.

1014 [60] R. Kovall, B.W. Matthews, Toroidal structure of λ-exonuclease, *Science* (1979). 277 (1997)
1015 1824–1827. <https://doi.org/10.1126/science.277.5333.1824>.

- 1016 [61] A. Erler, S. Wegmann, C. Elie-Caille, C.R. Bradshaw, M. Maresca, R. Seidel, B. Habermann, D.J.
1017 Muller, A.F. Stewart, Conformational Adaptability of Red β during DNA Annealing and
1018 Implications for Its Structural Relationship with Rad52, *J Mol Biol.* 391 (2009) 586–598.
1019 <https://doi.org/10.1016/j.jmb.2009.06.030>.
- 1020 [62] S.I. Passy, X. Yu, Z. Li, C.M. Radding, E.H. Egelman, Rings and filaments of β protein from
1021 bacteriophage λ suggest a superfamily of recombination proteins, *Proc Natl Acad Sci U S A.* 96
1022 (1999) 4279–4284. <https://doi.org/10.1073/pnas.96.8.4279>.
- 1023 [63] J. Zhang, X. Xing, A.B. Herr, C.E. Bell, Crystal Structure of *E. coli* RecE Protein Reveals a Toroidal
1024 Tetramer for Processing Double-Stranded DNA Breaks, *Structure.* 17 (2009) 690–702.
1025 <https://doi.org/10.1016/J.STR.2009.03.008>.
- 1026 [64] T.P. Newing, J.L. Brewster, L.J. Fitschen, J.C. Bouwer, N.P. Johnston, H. Yu, G. Tolun, Red β 177
1027 annealase structure reveals details of oligomerization and λ Red-mediated homologous DNA
1028 recombination, *Nat Commun.* 13 (2022) 5649. <https://doi.org/10.1038/s41467-022-33090-6>.
- 1029 [65] J. Zhang, K.A. McCabe, C.E. Bella, Crystal structures of λ exonuclease in complex with DNA
1030 suggest an electrostatic ratchet mechanism for processivity, *Proc Natl Acad Sci U S A.* 108
1031 (2011) 11872–11877. <https://doi.org/10.1073/pnas.1103467108>.
- 1032 [66] J. Zhang, X. Pan, C.E. Bell, Crystal structure of λ exonuclease in complex with DNA and Ca²⁺,
1033 *Biochemistry.* 53 (2014) 7415–7425. <https://doi.org/10.1021/bi501155q>.
- 1034 [67] B.J. Caldwell, A.S. Norris, C.F. Karbowski, A.M. Wiegand, V.H. Wysocki, C.E. Bell, Structure of a
1035 RecT/Red β family recombinase in complex with a duplex intermediate of DNA annealing, *Nat*
1036 *Commun.* 13 (2022) 1–14. <https://doi.org/10.1038/s41467-022-35572-z>.
- 1037 [68] H.M. Ellis, D. Yu, T. DiTizio, D.L. Court, High efficiency mutagenesis, repair, and engineering
1038 of chromosomal DNA using single-stranded oligonucleotides, *Proceedings of the National*
1039 *Academy of Sciences.* 98 (2001) 6742–6746. <https://doi.org/10.1073/PNAS.121164898>.
- 1040 [69] R.G. Higuchi, H. Ochman¹, Production of single-stranded DNA templates by exonuclease
1041 digestion following the polymerase Production of single-stranded DNA templates by
1042 exonuclease digestion following the polymerase chain reaction, *Nucleic Acids Res.* 17 (1989)
1043 5865. <https://doi.org/10.1093/nar/17.14.5865>.
- 1044 [70] H.S. Rhee, B.F. Pugh, Comprehensive Genome-wide Protein-DNA Interactions Detected at
1045 Single-Nucleotide Resolution, *Cell.* 147 (2011) 1408–1419.
1046 <https://doi.org/10.1016/j.cell.2011.11.013>.
- 1047 [71] Z.J. Liu, L.Y. Yang, T.C. Lu, Y.Q. Liang, M.M. Liu, G.X. Zhong, X.H. Lin, P.F. Huang, J.Y. Chen, A
1048 zero-background electrochemical DNA sensor coupling ligase chain reaction with lambda
1049 exonuclease digestion for CYP2C19*2 allele genotyping in clinical samples, *Sens Actuators B*
1050 *Chem.* 368 (2022) 132096. <https://doi.org/10.1016/j.snb.2022.132096>.
- 1051 [72] L.L. Xu, W. Zhao, J. Pu, S. Wang, S. Liu, H. Li, R. Yu, A Pax-5a gene analysis approach enabled by
1052 selective digestion with lambda exonuclease, *Analytical Methods.* 14 (2022) 2415–2422.
1053 <https://doi.org/10.1039/D2AY00469K>.

- 1054 [73] L. Cui, Y. Li, M. Lu, B. Tang, C. yang Zhang, An ultrasensitive electrochemical biosensor for
1055 polynucleotide kinase assay based on gold nanoparticle-mediated lambda exonuclease
1056 cleavage-induced signal amplification, *Biosens Bioelectron.* 99 (2018) 1–7.
1057 <https://doi.org/10.1016/j.bios.2017.07.028>.
- 1058 [74] J. Sun, C. Li, Y. Hu, Y. Ding, T. Wu, A structure change-induced fluorescent biosensor for uracil-
1059 DNA glycosylase activity detection based on the substrate preference of Lambda exonuclease,
1060 *Talanta.* 243 (2022) 123350. <https://doi.org/10.1016/j.talanta.2022.123350>.
- 1061 [75] A. Lopes, J. Amarir-Bouhram, G. Faure, M.A. Petit, R. Guerois, Detection of novel
1062 recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs,
1063 *Nucleic Acids Res.* 38 (2010) 3952–3962. <https://doi.org/10.1093/nar/gkq096>.
- 1064 [76] W. Yang, W.Y. Chen, H. Wang, J.W.S. Ho, J.D. Huang, P.C.Y. Woo, S.K.P. Lau, K.Y. Yuen, Q. Zhang,
1065 W. Zhou, M. Bartlam, R.M. Watt, Z. Rao, Structural and functional insight into the mechanism
1066 of an alkaline exonuclease from *Laribacter hongkongensis*, *Nucleic Acids Res.* 39 (2011) 9803–
1067 9819. <https://doi.org/10.1093/nar/gkr660>.
- 1068 [77] I. V. Shevelev, U. Hübscher, The 3'–5' exonucleases, *Nature Reviews Molecular Cell Biology*
1069 2002 3:5. 3 (2002) 364–376. <https://doi.org/10.1038/nrm804>.
- 1070 [78] S.H. Mueller, L.J. Fitschen, A. Shirbini, S.M. Hamdan, L.M. Spenkelink, A.M. van Oijen, Rapid
1071 single-molecule characterisation of enzymes involved in nucleic-acid metabolism, *Nucleic*
1072 *Acids Res.* 1 (2022) e5. <https://doi.org/10.1093/nar/gkac949>.
- 1073 [79] J. Dapprich, Single-Molecule DNA Digestion by Lambda-Exonuclease, *Cytometry.* 36 (1999)
1074 163–168. [https://doi.org/10.1002/\(SICI\)1097-0320\(19990701\)36:3](https://doi.org/10.1002/(SICI)1097-0320(19990701)36:3).
- 1075 [80] A.M. Van Oijen, P.C. Blainey, D.J. Crampton, C.C. Richardson, T. Ellenberger, X.S. Xie, Single-
1076 molecule kinetics of λ exonuclease reveal base dependence and dynamic disorder, *Science*
1077 (1979). 301 (2003) 1235–1238. <https://doi.org/10.1126/science.1084387>.
- 1078 [81] X. Pan, C.E. Smith, J. Zhang, K.A. McCabe, J. Fu, C.E. Bell, A Structure-Activity Analysis for
1079 Probing the Mechanism of Processive Double-Stranded DNA Digestion by λ Exonuclease
1080 Trimers, *Biochemistry.* 54 (2015) 6139–6148. <https://doi.org/10.1021/acs.biochem.5b00707>.
- 1081 [82] X. Pan, J. Yan, A. Patel, V.H. Wysocki, C.E. Bell, Mutant poisoning demonstrates a
1082 nonsequential mechanism for digestion of double-stranded DNA by λ exonuclease trimers,
1083 *Biochemistry.* 54 (2015) 942–951. <https://doi.org/10.1021/bi501431w>.
- 1084 [83] G. Tolun, R.S. Myers, A real-time DNase assay (ReDA) based on PicoGreen[®] fluorescence,
1085 *Nucleic Acids Res.* 31 (2003) e111. <https://doi.org/10.1093/nar/gng111>.
- 1086 [84] P.G. Mitsis, J.G. Kwagh, Characterization of the interaction of lambda exonuclease with the
1087 ends of DNA, *Nucleic Acids Res.* 27 (1999) 3057–3063.
1088 <https://doi.org/10.1093/NAR/27.15.3057>.
- 1089 [85] K. Subramanian, W. Rutvisuttinunt, W. Scott, R.S. Myers, The enzymatic basis of processivity in
1090 λ exonuclease, *Nucleic Acids Res.* 31 (2003) 1585–1596.
1091 <https://doi.org/10.1093/NAR/GKG266>.

1092 [86] J. van Oostrum, J.L. White, R.M. Burnett, Isolation and crystallization of λ exonuclease, Arch
1093 Biochem Biophys. 243 (1985) 332–337. [https://doi.org/10.1016/0003-9861\(85\)90510-7](https://doi.org/10.1016/0003-9861(85)90510-7).

1094 [87] A. Pingoud, M. Fuxreiter, V. Pingoud, W. Wende, Type II restriction endonucleases: Structure
1095 and mechanism, Cellular and Molecular Life Sciences. 62 (2005) 685–707.
1096 <https://doi.org/10.1007/S00018-004-4513-1/METRICS>.

1097 [88] W. Yang, J.Y. Lee, M. Nowotny, Making and Breaking Nucleic Acids: Two-Mg²⁺-Ion Catalysis
1098 and Substrate Specificity, Mol Cell. 22 (2006) 5–13.
1099 <https://doi.org/10.1016/J.MOLCEL.2006.03.013>.

1100 [89] K. Steczkiewicz, A. Muszewska, L. Knizewski, L. Rychlewski, K. Ginalski, Sequence, structure
1101 and functional diversity of PD-(D/E)XK phosphodiesterase superfamily, Nucleic Acids Res. 40
1102 (2012) 7016–7045. <https://doi.org/10.1093/NAR/GKS382>.

1103 [90] X. Pan, C.E. Smith, J. Zhang, K.A. McCabe, J. Fu, C.E. Bell, A Structure-Activity Analysis for
1104 Probing the Mechanism of Processive Double-Stranded DNA Digestion by λ Exonuclease
1105 Trimers, Biochemistry. 54 (2015) 6139–6148.
1106 <https://doi.org/10.1021/ACS.BIOCHEM.5B00707>.

1107 [91] G. Tolun, More than the sum of its parts: physical and mechanistic coupling in the phage
1108 lambda red recombinase, (2007).

1109 [92] S.D. Hall, R.D. Kolodner, Homologous pairing and strand exchange promoted by the
1110 Escherichia coli RecT protein, Proc Natl Acad Sci U S A. 91 (1994) 3205–3209.
1111 <https://doi.org/10.1073/pnas.91.8.3205>.

1112 [93] J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P.M. Seibert, A. Plaza, L. Xia, R. Müller, A.F. Stewart, Y.
1113 Zhang, Full-length RecE enhances linear-linear homologous recombination and facilitates
1114 direct cloning for bioprospecting, Nat Biotechnol. 30 (2012) 440–446.
1115 <https://doi.org/10.1038/nbt.2183>.

1116 [94] T.S. Vellani, R.S. Myers, Bacteriophage SPP1 Chu is an alkaline exonuclease in the SynExo
1117 family of viral two-component recombinases, J Bacteriol. 185 (2003) 2465–2474.
1118 <https://doi.org/10.1128/JB.185.8.2465-2474.2003>.

1119 [95] I.N. Wang, Lysis Timing and Bacteriophage Fitness, Genetics. 172 (2006) 17–26.
1120 <https://doi.org/10.1534/genetics.105.045922>.

1121 [96] B. Gibson, D.J. Wilson, E. Feil, A. Eyre-Walker, The distribution of bacterial doubling times in
1122 the wild, Proceedings of the Royal Society B. 285 (2018).
1123 <https://doi.org/10.1098/rspb.2018.0789>.

1124 [97] E. Kmiec, W.K. Holloman, Beta protein of bacteriophage lambda promotes renaturation of
1125 DNA., Journal of Biological Chemistry. 256 (1981) 12636–12639.
1126 [https://doi.org/10.1016/s0021-9258\(18\)42938-9](https://doi.org/10.1016/s0021-9258(18)42938-9).

1127 [98] L.W. Enquist, A.M. Skalka, Replication of Bacteriophage λ DNA Dependent on the Function of
1128 Host and Viral Genes: I. Interaction of red, gam and rec, J Mol Biol. 75 (1972) 185–212.
1129 [https://doi.org/https://doi.org/10.1016/0022-2836\(73\)90016-8](https://doi.org/https://doi.org/10.1016/0022-2836(73)90016-8).

1130 [99] J.L. Brewster, G. Tolun, Half a century of bacteriophage lambda recombinase: In vitro studies
1131 of lambda exonuclease and Red-beta annealase, *IUBMB Life*. 72 (2020) 1622–1633.
1132 <https://doi.org/10.1002/iub.2343>.

1133 [100] K.C. Murphy, The λ Gam Protein Inhibits RecBCD Binding to dsDNA Ends, *J Mol Biol*. 371
1134 (2007) 19–24. <https://doi.org/10.1016/J.JMB.2007.05.085>.

1135 [101] A.E. Karu, Y. Sakaki, H. Echols, S. Linn, The gamma protein specified by bacteriophage lambda.
1136 Structure and inhibitory activity for the recBC enzyme of *Escherichia coli*, *J Biol Chem*. 250
1137 (1975) 7377–7387. [https://doi.org/https://doi.org/10.1016/S0021-9258\(19\)40955-1](https://doi.org/https://doi.org/10.1016/S0021-9258(19)40955-1).

1138 [102] K.C. Murphy, Lambda Gam protein inhibits the helicase and chi-stimulated recombination
1139 activities of *Escherichia coli* RecBCD enzyme, *J Bacteriol*. 173 (1991) 5808–5821.
1140 <https://doi.org/10.1128/JB.173.18.5808-5821.1991>.

1141 [103] M. Wilkinson, L. Troman, W.A. Wan Nur Ismah, Y. Chaban, M.B. Avison, M.S. Dillingham, D.B.
1142 Wigley, Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial
1143 effect with quinolones, *Elife*. 5 (2016). <https://doi.org/10.7554/ELIFE.22963>.

1144 [104] R. Court, N. Cook, K. Saikrishnan, D. Wigley, The Crystal Structure of λ -Gam Protein Suggests a
1145 Model for RecBCD Inhibition, 271 (2007) 25–33. <https://doi.org/10.1016/j.jmb.2007.05.037>.

1146 [105] J.P.P. Muyrers, Y. Zhang, F. Buchholz, A.F. Stewart, RecE/RecT and Red α /Red β initiate double-
1147 stranded break repair by specifically interacting with their respective partners, *Genes Dev*. 14
1148 (2000) 1971–1982. <https://doi.org/10.1101/gad.14.15.1971>.

1149 [106] K. Steczkiewicz, E. Prestel, E. Bidnenko, A.K. Szczepankowska, Expanding Diversity of
1150 Firmicutes Single-Strand Annealing Proteins: A Putative Role of Bacteriophage-Host Arms
1151 Race, *Front Microbiol*. 12 (2021) 644622. <https://doi.org/10.3389/fmicb.2021.644622>.

1152 [107] A.J. Hernandez, C.C. Richardson, Gp2.5, the multifunctional bacteriophage T7 single-stranded
1153 DNA binding protein, *Semin Cell Dev Biol*. 86 (2019) 92–101.
1154 <https://doi.org/10.1016/J.SEMCDB.2018.03.018>.

1155 [108] C.M. Radding, Regulation of λ exonuclease: I. Properties of λ exonuclease purified from
1156 lysogens of λ T11 and wild type, *J Mol Biol*. 18 (1966) 235–250.
1157 [https://doi.org/10.1016/S0022-2836\(66\)80243-7](https://doi.org/10.1016/S0022-2836(66)80243-7).

1158 [109] A. Erler, S. Wegmann, C. Elie-Caille, C.R. Bradshaw, M. Maresca, R. Seidel, B. Habermann, D.J.
1159 Muller, A.F. Stewart, Conformational Adaptability of Red β during DNA Annealing and
1160 Implications for Its Structural Relationship with Rad52, *J Mol Biol*. 391 (2009) 586–598.
1161 <https://doi.org/10.1016/j.jmb.2009.06.030>.

1162 [110] A.S. Darwish, L.M. Grady, P. Bai, S.K. Weller, ICP8 Filament Formation Is Essential for
1163 Replication Compartment Formation during Herpes Simplex Virus Infection, *J Virol*. 90 (2016)
1164 2561–2570. <https://doi.org/10.1128/JVI.02854-15>.

1165 [111] B.J. Caldwell, A. Norris, E. Zakharova, C.E. Smith, C.T. Wheat, D. Choudhary, M. Sotomayor,
1166 V.H. Wysocki, C.E. Bell, Oligomeric complexes formed by Red β single strand annealing protein
1167 in its different DNA bound states, *Nucleic Acids Res*. 49 (2021) 3441–3460.
1168 <https://doi.org/10.1093/nar/gkab125>.

1169 [112] G. Karakousis, N. Ye, Z. Li, S.K. Chiu, G. Reddy, C.M. Radding, The beta protein of phage λ binds
1170 preferentially to an intermediate in DNA renaturation, *J Mol Biol.* 276 (1998) 721–731.
1171 <https://doi.org/10.1006/jmbi.1997.1573>.

1172 [113] K. Zakharova, B.J. Caldwell, S. Ta, C.T. Wheat, C.E. Bell, Mutational analysis of red β single
1173 strand annealing protein: Roles of the 14 lysine residues in dna binding and recombination in
1174 vivo, *Int J Mol Sci.* 22 (2021). <https://doi.org/10.3390/IJMS22147758>.

1175 [114] W. Kagawa, H. Kurumizaka, R. Ishitani, S. Fukai, O. Nureki, T. Shibata, S. Yokoyama, Crystal
1176 Structure of the Homologous-Pairing Domain from the Human Rad52 Recombinase in the
1177 Undecameric Form, *Mol Cell.* 10 (2002) 359–371. [https://doi.org/10.1016/S1097-](https://doi.org/10.1016/S1097-2765(02)00587-7)
1178 [2765\(02\)00587-7](https://doi.org/10.1016/S1097-2765(02)00587-7).

1179 [115] M.R. Singleton, L.M. Wentzell, Y. Liu, S.C. West, D.B. Wigley, Structure of the single-strand
1180 annealing domain of human RAD52 protein, *Proc Natl Acad Sci U S A.* 99 (2002) 13492–13497.
1181 <https://doi.org/10.1073/pnas.212449899>.

1182 [116] M. Saotome, K. Saito, K. Onodera, H. Kurumizaka, W. Kagawa, Structure of the human DNA-
1183 repair protein RAD52 containing surface mutations, *Acta Crystallogr F Struct Biol Commun.* 72
1184 (2016) 598–603. <https://doi.org/10.1107/S2053230X1601027X>.

1185 [117] C. Kinoshita, Y. Takizawa, M. Saotome, S. Ogino, H. Kurumizaka, W. Kagawa, C.W. Kagawa, H.
1186 Kurumizaka, The cryo-EM structure of full-length RAD52 protein contains an undecameric
1187 ring, *FEBS Open Bio.* 13 (2023) 408–418. <https://doi.org/10.1002/2211-5463.13565>.

1188 [118] A. Al-Fatlawi, M. Schroeder, A.F. Stewart, The Rad52 SSAP superfamily and new insight into
1189 homologous recombination, *Commun Biol.* 6 (2023) 87. [https://doi.org/10.1038/s42003-023-](https://doi.org/10.1038/s42003-023-04476-z)
1190 [04476-z](https://doi.org/10.1038/s42003-023-04476-z).

1191 [119] A.Z. Stasiak, E. Larquet, A. Stasiak, S. Müller, A. Engel, E. Van Dyck, S.C. West, E.H. Egelman,
1192 The human Rad52 protein exists as a heptameric ring, *Current Biology.* 10 (2000) 337–340.
1193 [https://doi.org/10.1016/S0960-9822\(00\)00385-7](https://doi.org/10.1016/S0960-9822(00)00385-7).

1194 [120] M. Mapelli, S. Panjkar, P.A. Tucker, The Crystal Structure of the Herpes Simplex Virus 1 ssDNA-
1195 binding Protein Suggests the Structural Basis for Flexible, Cooperative Single-stranded DNA
1196 Binding, *Journal of Biological Chemistry.* 280 (2005) 2990–2997.
1197 <https://doi.org/10.1074/jbc.M406780200>.

1198 [121] M. Valledor, R.S. Myers, P.C. Schiller, Herpes ICP8 protein stimulates homologous
1199 recombination in human cells, *PLoS One.* 13 (2018) e0200955.
1200 <https://doi.org/https://doi.org/10.1371/journal.pone.0200955>.

1201 [122] A.J. Schumacher, K.N. Mohni, Y. Kan, E.A. Hendrickson, J.M. Stark, S.K. Weller, The HSV-1
1202 Exonuclease, UL12, Stimulates Recombination by a Single Strand Annealing Mechanism, *PLoS*
1203 *Pathog.* 8 (2012) e1002862. <https://doi.org/10.1371/journal.ppat.1002862>.

1204 [123] M. Gao, D.M. Knipe, Potential role for herpes simplex virus ICP8 DNA replication protein in
1205 stimulation of late gene expression, *J Virol.* 65 (1991) 2666–2675.
1206 <https://doi.org/10.1128/JVI.65.5.2666-2675.1991>.

1207 [124] M. Gao, D.M. Knipe, Genetic evidence for multiple nuclear functions of the herpes simplex
1208 virus ICP8 DNA-binding protein, *J Virol.* 63 (1989) 5258–5267.
1209 <https://doi.org/10.1128/JVI.63.12.5258-5267.1989>.

1210 [125] A.M. Makhov, A. Sen, X. Yu, M.N. Simon, J.D. Griffith, E.H. Egelman, The Bipolar Filaments
1211 Formed by Herpes Simplex Virus Type 1 SSB/Recombination Protein (ICP8) Suggest a
1212 Mechanism for DNA Annealing, *J Mol Biol.* 386 (2009) 273–279.
1213 <https://doi.org/10.1016/J.JMB.2008.12.059>.

1214 [126] M.E. O'Donnell, P. Elias, B.E. Funnell, I.R. Lehman, Interaction between the DNA polymerase
1215 and single-stranded DNA-binding protein (infected cell protein 8) of herpes simplex virus 1.,
1216 *Journal of Biological Chemistry.* 262 (1987) 4260–4266. [https://doi.org/10.1016/S0021-](https://doi.org/10.1016/S0021-9258(18)61341-9)
1217 [9258\(18\)61341-9](https://doi.org/10.1016/S0021-9258(18)61341-9).

1218 [127] S. Weerasooriya, K.A. DiScipio, A.S. Darwish, P. Bai, S.K. Weller, Herpes simplex virus 1 ICP8
1219 mutant lacking annealing activity is deficient for viral DNA replication, *Proc Natl Acad Sci U S*
1220 *A.* 116 (2019) 1033–1042. <https://doi.org/10.1073/pnas.1817642116>.

1221 [128] T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B.L. Pinto, G.A. Salazar, M.L. Bileschi, P.
1222 Bork, A. Bridge, L. Colwell, J. Gough, D.H. Haft, I. Letunić, A. Marchler-Bauer, H. Mi, D.A.
1223 Natale, C.A. Orengo, A.P. Pandurangan, C. Rivoire, C.J.A. Sigrist, I. Sillitoe, N. Thanki, P.D.
1224 Thomas, S.C.E. Tosatto, C.H. Wu, A. Bateman, InterPro in 2022, *Nucleic Acids Res.* 51 (2023)
1225 D418–D427. <https://doi.org/10.1093/nar/gkac993>.

1226 [129] K. Katoh, D.M. Standley, MAFFT Multiple Sequence Alignment Software Version 7:
1227 Improvements in Performance and Usability, *Mol Biol Evol.* 30 (2013) 772–780.
1228 <https://doi.org/10.1093/molbev/mst010>.

1229 [130] L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation
1230 sequencing data, *Bioinformatics.* 28 (2012) 3150–3152.
1231 <https://doi.org/10.1093/bioinformatics/bts565>.

1232 [131] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton, Jalview Version 2—a
1233 multiple sequence alignment editor and analysis workbench, *Bioinformatics.* 25 (2009) 1189–
1234 1191. <https://doi.org/10.1093/bioinformatics/btp033>.

1235 [132] E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris, T.E.
1236 Ferrin, UCSF ChimeraX: Structure visualization for researchers, educators, and developers,
1237 *Protein Science.* 30 (2021) 70–82. <https://doi.org/10.1002/PRO.3943>.

1238 [133] J. Pei, N. V. Grishin, AL2CO: calculation of positional conservation in a protein sequence
1239 alignment, *Bioinformatics.* 17 (2001) 700–712.
1240 <https://doi.org/10.1093/bioinformatics/17.8.700>.

1241 [134] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making
1242 protein folding accessible to all, *Nat Methods.* 19 (2022) 679–682.
1243 <https://doi.org/10.1038/s41592-022-01488-1>.

1244