Co-expression and purification of the RadA recombinase with the RadB paralog from *Haloferax volcanii* yields heteromeric ring-like structures.

Authors:

Bushra B. Patoli[1, a,*], Jody A. Winter[1, b] Atif A. Patoli[1, a], Robin M. Delahay[2] and Karen A. Bunting[1, c]

Affiliations:
[1]School of Biology, Queens Medical Centre, University of Nottingham, Nottingham, UK, NG7 2UH.
[2]School of Medicine, University of Nottingham, Nottingham, UK, NG7 2UH. Email: rob.delahay@nottingham.ac.uk
[a]Institute of Microbiology, University of Sindh, Jamshoro, Pakistan (Present address). Email: bushrapatoli@gmail.com
[b]Nottingham Trent University, Nottingham, UK, NG11 8NS (Present address). Email: jody.winter@ntu.ac.uk
[c]Albumedix Ltd, Nottingham, UK, NG7 1FD (Present address). Email: karen.bunting@slidingclamp.com

[*]**Correspondence:** Bushra B. Patoli, Institute of Microbiology, University of Sindh, Jamshoro, Pakistan. Tel.: +92 3113005252. E.mail.bushrapatoli@gmail.com

Keywords: Homologous recombination, RadA, *Haloferax volcanii*, Co-expression, Halophilic protein purification.

Abbreviations: Immobilized metal affinity chromatography (IMAC); Discrete Optimised Protein Energy (DOPE); ATPase domain (AD); N-terminal domain (NTD); Size Exclusion Chromatography (SEC); Electron Microscopy (EM)

40 # Abstract

41

42 The study of archaeal proteins and the processes they contribute poses particular challenges due
43 to the often extreme environments in which they function. DNA recombination, replication and
44 repair proteins of the halophilic euryarchaeon, *Haloferax volcanii* (Hvo) are of particular interest
45 as they tend to resemble eukaryotic counterparts in both structure and activity and genetic tools
46 are available to facilitate their analysis. In the present study, we show using bioinformatics
47 approaches that the Hvo RecA-like protein RadA is structurally similar to other recombinases
48 although is distinguished by a unique acidic insertion loop. To facilitate expression of Hvo RadA
49 a co-expression approach was used, providing its lone paralog, RadB, as a binding partner. At
50 present, structural and biochemical characterization of Hvo RadA is lacking. Here, we describe
51 for the first time co-expression of Hvo RadA with RadB and purification of these proteins as a
52 complex under *in vitro* conditions. Purification procedures were performed under high salt
53 concentration (>1 M sodium chloride) to maintain the solubility of the proteins. Quantitative
54 densitometry analysis of the co-expressed and co-purified RadA-B complex estimated the ratio
55 of RadA to RadB to be 4:1, which suggests that the proteins interact with a specific
56 stoichiometry. Based on a combination of analyses, including size exclusion chromatography,
57 Western blot and electron microscopy observations we suggest that RadA multimerizes into a
58 ring-like structure in the absence of DNA and nucleoside cofactor.

59

60

61

62

Introduction

Homologous recombination is one of the principal pathways to repair deleterious DNA damage such as double-strand breaks. The basic mechanism of homologous recombination, comprising homologous base-pairing and a strand exchange reaction, is similar in all forms of life, however it varies in complexity and enzymology. Various enzymes participate in the recombination process, including the key recombinase which belongs to the RecA family of proteins (1). These enzymes are well conserved in all three domains of life and are termed RecA in bacteria (2), RadA in archaea (3), and Rad51 in eukaryotes (4). The strand exchange protein identified in archaeal RadA is more similar to the eukaryotic Rad51 protein (~40% amino acid identity) than to its prokaryotic counterpart RecA in eubacteria (~20% amino acid identity) (3, 5, 6). Crystallographic structure analysis of archaeal RadA has contributed significantly to our understanding of the structure and function of eukaryotic Rad51 (7, 8). Various accessory proteins also participate in the strand exchange reaction and these are usually termed recombination mediators. Many recombination mediators are encoded by genes that arise as a result of a duplication event within the genome, and are termed paralogs. In yeast, the paralogous proteins Rad55 and Rad57 share some similarity to Rad51 and function as a heterodimeric complex which may stabilize the Rad51-assembled nucleoprotein filament (9). Five Rad51 paralogs, XRCC2, XRCC3, RAD51B/RAD51L1, RAD51C/RAD51L2, and RAD51D/RAD51L3, are found in eukaryotes. These paralogs form various heteromeric complexes, mostly dimers such as RAD51B-RAD51C, RAD51D-XRCC2, RAD51C-XRCC3 or tetramers such as RAD51B-RAD51C-RAD51D-XRCC2 (10).

The identification and characterisation of RecA-like proteins in both crenarchaea and

86 euryarchaea shows that, in common with eukaryotic Rad51, archaeal RadA proteins exhibit a
87 high level of diversity. This may reflect that the archaea demonstrate an evolutionary mixture of
88 replication, repair and recombination functions in between simple bacterial and more complex
89 eukaryotic forms of life.

90 RadB was the first RadA paralog to be characterised in archaea and its presence is confined to
91 the Euryarchaeota. RadB lacks the N-terminal domain of RadA (11) and strand exchange activity
92 (12). Interaction of RadB with RadA has been demonstrated in *Pyrococcus furiosus* (Pfu) and it
93 was proposed that RadB functions in a manner analogous to the yeast Rad55-57 proteins in the
94 strand exchange reaction (12). Genetic analysis of *radB* from the euryarchaeon *Haloferax*
95 *volcanii* (Hvo) demonstrates that RadB functions in the homologous recombination pathways in
96 concert with RadA (13, 14).

97 Hvo RadB has been successfully over-expressed in *Escherichia coli* (15) and both Hvo RadA
98 and RadB proteins have been purified after conditional over-expression in their native host (16,
99 17). A recent study has confirmed the *in vivo* interaction of Hvo RadA and RadB, exploiting
100 mass spectrometry to identify co-purifying proteins following IMAC chromatography (18).
101 However, both quantity and purity of the resulting proteins has been insufficient for structural
102 and biochemical analyses previously described for other DNA-associated halophilic proteins
103 such as Hvo PCNA and RPA3 (19, 20). Additionally, when over-expressed in *E. coli*,
104 recombinant Hvo RadA has also been found to co-purify with DNA, providing further
105 impediments to downstream analyses since conventional approaches to remove contaminating
106 DNA, such as denaturation and nuclease treatment, are likely to be detrimental to subsequent
107 characterization [17].

108 The study of recombination proteins in the hyperthermophilic crenarchaeon, *Sulfolobus tokodaii*
109 (St), has shown that over-expression and characterization of StRad55B, a paralog of StRadA,
110 was possible only when it was co-expressed with StRadA protein (21). We therefore explored
111 co-expression of Hvo RadA with RadB as both a novel strategy for maintenance and
112 optimisation of RadA protein stability in the non-halophilic host, *E. coli,* and as a potential
113 means to limit RadA interaction with host DNA to facilitate its purification. Using this approach
114 we successfully demonstrate soluble over-expression of the Hvo RadA-RadB complex, provide
115 clear evidence for RadA-RadB protein-protein interaction and present robust methodology to
116 enable their further characterization and study.
117

118 Materials and Methods

119

120 Bacterial strains, plasmids and growth conditions.

121 *E. coli* chemically competent strains DH5α™ (Invitrogen) and Rosetta 2 (DE3) (Novagen) were
122 used for cloning and gene expression, respectively. The PCR-amplified target gene was cloned
123 first into the Zero Blunt PCR vector (Invitrogen) and sequenced to exclude amplification errors.
124 The pET11 expression vector (Novagen) was used when single overexpression of *radA* or *radB*
125 was desired. For co-expression and purification of RadA and RadB, the *E. coli* Rosetta 2 (DE3)
126 strain was transformed with pET-Duet-1-based plasmid constructs (Novagen).
127 Bacterial cultures were grown at 37°C in Luria-Bertani (LB) broth in a shaking incubator or on
128 LB agar plates supplemented with ampicillin (100 μg ml^{-1}), chloramphenicol (34 μg ml^{-1}) or
129 kanamycin (34 μg ml^{-1}) as appropriate.
130

131 **Knock out of the *arnA* gene from *E. coli* Rosetta2 (DE3) expression strain**

132 A mutant of the Rosetta2 (DE3) expression strain carrying a kanamycin resistance cassette
133 (kan^R) in place of the *arnA* gene (arnA736 (del):: kan^R) was generated by P1 phage-mediated
134 transduction (22). Donor strain, obtained from the *E. coli* Genetic Stock center (CGSC), was
135 grown at 37°C to an OD_{650} of 0.3-0.4 to add the P1 phage stock (~10^7-10^8 PFU). After 3-4 h
136 incubation, purified lysate was used to transduce the kan^R marker into recipient Rosetta2 (DE3)
137 cells. Recipient strain was grown at 37°C to OD_{650} of 0.8, pelleted, re-suspended in 0.1 M
138 $MgSO_4$, 5 mM $CaCl_2$ and incubated with purified lysate for 25 minutes at 37°C. After addition
139 of sodium citrate to a final concentration of 0.5 M to chelate calcium ions and prevent further
140 total lysis of the recipient strain, transductants were plated and sub-cultured to purity on LB agar
141 containing kanamycin (34 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}).

142

143 **Molecular methods**

144 Hvo *radA* gene (Accession number: U45311) was PCR-amplified from the *H. volcanii* wild-type
145 DS2 strain (23). The forward primer (*radA*6HisF) contained an NdeI restriction site (underlined)
146 and a 6xHis tag
147 5'GACCT**CATATG****CATCACCATCACCATCAC**ATGGCAGAAGACGACCTC-3' and the
148 reverse primer (*radA*BamR) encoded a *Bam*HI restriction site (underlined) 5'-
149 GCAAT**GGATCC**TTATTACTCGGGCTTGAGACCGGCGTCCTG-3'. The His-tagged *radA*
150 gene was cloned first into the Zero Blunt PCR vector for sequencing and then sub-cloned for
151 overexpression either into pET11 (using NdeI and BamHI restriction sites) or pET-Duet1
152 plasmid at multiple cloning site 2 (MCS-2) (using the NdeI and EcoRV restriction sites). The
153 Hvo *radB* coding sequence (690bp) was excised from plasmid pCPG42 using the NdeI and

154 BamHI restriction sites (15) . After sequence verification in a sub-cloning vector, *radB* was
155 cloned into MCS-1 of pET-Duet1 using the NcoI and HindIII restriction sites to yield the
156 pBPRAD2 construct. The relevant properties of the strains are listed in Table 1.
157

158 **Over-expression, purification and analysis of RadA and RadB proteins**

159 Rosetta2 (DE3) or Rosetta2 (DE3) Δ*arnA E. coli* transformed with the pBPRAD2 construct were
160 used for over-expression and purification of co-expressed His-tagged RadA and un-tagged RadB
161 proteins. An overnight culture was inoculated into baffle flasks containing 2.4 L LB broth
162 supplemented with antibiotics as appropriate, and grown at 37°C. Over-expression of *radA-B*
163 was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.4 –
164 0.6 with incubation for a further 3 hrs in a shaking incubator at 30°C. The cell pellet was
165 harvested by centrifugation and lysed by sonication on ice in buffer A (50 mM HEPES, pH7.0,
166 10 mM imidazole and 1 M NaCl) containing EDTA-free protease inhibitor cocktail (Roche). The
167 cell lysate was clarified by centrifugation at 16,000 x *g* for 30 minutes at 4°C to remove
168 insoluble debris.

169 Soluble proteins in the supernatant were used for downstream purification of the complex using
170 Cobalt-based immobilized metal affinity chromatography (IMAC). A Liquid Chromatography
171 Column (2.5 cm x 10 cm, Sigma-Aldrich) containing 5 mL of Talon® metal affinity resin
172 (Clontech) was equilibrated with 20 column volumes of buffer A (50 mM HEPES pH 7.0, 1 M
173 NaCl and 30 mM imidazole). The clarified soluble fraction was loaded and incubated with
174 rolling for 10 minutes at room temperature. The flow-through and two 20 mL buffer A washes
175 were retained separately in fresh tubes for SDS-PAGE gel analysis. Bound proteins were eluted
176 with 13 mL of buffer B (50 mM HEPES pH 7.0, 1 M NaCl and 300 mM imidazole) following

177 rolling for 5 minutes at room temperature, prior to collection of the eluate. The eluate was loaded
178 onto a 26/60 Superdex 200 (S200) preparative column (GE Healthcare) using an ÄKTA Prime
179 Plus system. The column was pre-equilibrated and run using size exclusion chromatography
180 (SEC) buffer (50 mM HEPES pH 7.0 and 1 M NaCl). The proteins in each IMAC and SEC
181 fraction were analyzed by SDS-PAGE gel. SEC fractions were also analysed by agarose gel
182 electrophoresis to select fractions free from contaminating DNA. Samples containing low
183 concentrations of protein were first concentrated using StrataClean resin (5 µl ml^{-1}) (Stratagene)
184 as previously described (24) prior to gel loading. All samples were adjusted to maintain
185 equivalent loading in all gels.
186 The identity of RadA and RadB proteins was confirmed using MALDI-TOF-MS analysis of
187 excised bands by the Biopolymer Synthesis and Analysis Unit (BSAU), University of
188 Nottingham. Quantitative reflectance densitometric analysis of Coomassie Brilliant Blue-stained
189 proteins was performed using a BioRad GS-800 calibrated densitometer and Quantity One®
190 Software. Western blots were blocked with 5%, w/v, milk powder in phosphate buffered saline
191 (PBS)-Tween 20, then probed with 10 µL (1/1000 diluted) of alkaline phosphatase conjugated-
192 Mouse Anti-Hexa-His antibodies (Sigma) for 1 h and developed using BCIP/NBT substrate
193 according to manufacturer's instructions (Sigma).
194

195 **Electron microscopy**

196 Co-expressed, purified RadA-B proteins were concentrated to 1 mg ml^{-1} using a Vivapore 10/20
197 7500 Da cutoff (Vivascience) and were maintained in 1 M NaCl buffer to preserve the soluble
198 and native state of the proteins. Protein concentrations were determined using the Qubit protein
199 assay kit (Invitrogen). Protein samples were applied to a carbon-Formvar grid (Agar Scientific)

200 and allowed to settle for 10 minutes. The sample was then stained with either 0.5 or 1%, w/v,
201 phosphotungstic acid (PTA; pH 7.0) for 1 minute. The grids were then imaged using a JEOL
202 JEM1010 transmission electron microscope at 100, 000X or 200, 000X magnification.
203
204 **Homology Modelling**
205 Homology modelling of the Hvo RadA primary sequence was performed using the PyMod 2.0
206 plugin module for PyMol (25, 26) as a convenient interface to Modeller 9v4 (27). The Pfu RadA
207 structure (1PZN) chain A was used as the highest scoring template and aligned against the Hvo
208 RadA sequence using Clustal Omega. Modeller parameters were adjusted to accommodate
209 automated building of disulfide bridges and the highest level of optimization and refinement,
210 with additional energy minimisation performed on resulting models. Output models with
211 corresponding low scoring Discrete Optimised Protein Energy (DOPE) profiles, indicative of
212 limited modelling errors, were further inspected for agreement with secondary structural
213 elements and diversity in loop disposition prior to final model selection.
214
215
216 ## Results
217
218 **Conservation of the ATPase domain of RadA from both halophilic and non-halophilic**
219 **species**
220 Structural comparison of RadA from *P. furiosus* with RadA homologs in other domains of life
221 reveals similarity in the ATPase domain (AD) although differences are apparent at amino and
222 carboxyl terminal domains of the proteins (Fig. S1a). Conservation of amino acid residues at the

223 Walker A, Walker B and DNA binding (L1 and L2) motifs is also apparent in a protein sequence
224 alignment of Hvo RadA with RadA homologs (Fig. S1b), and the Hvo RadB protein. Aside from
225 this conserved motif organization, Hvo RadA and RadB proteins are otherwise dissimilar,
226 sharing only 18.5%/34.1% identity/similarity at the amino acid level. Hvo RadB is also a
227 somewhat smaller protein, due to lack of a RadA-equivalent N-terminal domain (NTD) (Fig.
228 S1c).

229 Along with a characteristic negative charge distribution on the protein surface, a distinct
230 preference for certain amino acids is also a hallmark of halophilic proteins. This is apparent in
231 Hvo RadA as a comparative increase in the acidic residue, aspartate, which is the most common
232 adaptation relative to mesophilic and thermophilic counterparts (Fig. S2). However, whereas a
233 marked reduction in the lysine content and increases in serine and alanine have been reported in
234 other halophilic proteins (28) this is not similarly the case for Hvo RadA. This disparity reflects
235 the lack of any currently known universal form of halophilic adaptation.

236 It is of particular interest to understand how halophilic proteins have adapted to interact with
237 DNA, given the involvement of basic residues in DNA binding of the phosphate backbone. Guy
238 and others (15) demonstrated that residues in a conserved basic patch (the KHR triplet) were
239 crucial for DNA binding in Hvo RadB. To gain some insight into how preference for particular
240 amino acid usage in Hvo RadA, would likely affect the surface character of Hvo RadA, we built
241 a structural model in Modeller 9v4 for comparative analysis. Superposition of the Hvo RadA
242 model with RadA homologs shows the expected correspondence with the principal structural
243 elements in the N-terminal and ATPase domains (Fig. 1a) although, unlike other RadA proteins,
244 an extensive insertion loop found only in the Hvo RadA sequence (159-182) is also apparent
245 (Fig. 1a and Fig. S1b).

246 Fig. 1(b) shows the conserved elements of RadA in the Hvo RadA homology model in relation to
247 lysine residues within the molecule. The L1 and L2 loops within the ATPase domain are
248 involved in ssDNA binding. Chen and others (29) identified a number of key positively charged
249 residues in Sso RadA L1 (R217, R223 and R229) that are also conserved in Hvo RadA.
250 Conversely, of the two lysine residues in the N-terminal domain of Sso RadA involved in
251 dsDNA binding, only the equivalent of K27 is retained in Hvo RadA, with K60 being substituted
252 with as aspartate residue (D52). However, Hvo RadA possesses an alternative lysine at position
253 42. The majority of the lysine residues within Hvo RadA are located on the face of the molecule
254 containing the NTD, L1 and L2 motifs, suggesting their retention may be related to involvement
255 in DNA-binding.
256 The large insertion sequence seemingly unique to Hvo RadA (residues 159-182) is of potential
257 interest (Fig. 1 and Fig S1b). Given the lack of homology to other family members the
258 conformation of the modelled loop remains speculative but it can be assumed that it is positioned
259 on the face of the molecule as indicated in Fig. 1(b). The loop region comprises a large number
260 of acidic residues, both aspartate and glutamate, which, in the context of the uncommonly higher
261 proportion of lysines in Hvo RadA relative to other halophilic proteins, may function to maintain
262 solubility of the molecule or promote/stabilize RadA interactions with other proteins.
263
264 **Differential co-expression and purification of soluble RadA-RadB proteins**
265 His-tagged Hvo *radA* and un-tagged *radB* were co-expressed from pBPRAD2 in the Rosetta 2
266 (DE3) *E. coli* strain in order to minimise issues with codon usage that can occur when expressing
267 archaeal proteins in a heterologous host (30). SDS-PAGE gel analysis of clarified lysates of
268 induced cultures identified overexpressed proteins at the anticipated molecular mass for 6xHis-

269 RadA (~ 40 kDa) and RadB (~25 kDa) (Fig. 2). However, by comparison with the broadly
270 equivalent amounts of RadA and RadB observed when each protein was expressed from pET11
271 individually (Fig 2. lanes 2 & 4 respectively), RadB, when co-expressed with RadA, appeared
272 substantially diminished in abundance. This may suggest a possible issue with bicistronic
273 expression, although DUET vectors are engineered for independent promoter control of both
274 multiple cloning sites to avoid such problems. Alternatively, it may be the case that RadA affects
275 RadB expression or stability; a previous analysis of the RadA:RadB ratio in *P. furiosus* cell
276 extract showed the cellular amount of RadB protein to be ~200 times lower than that of RadA
277 (12). Consequently, the lack of equimolar expression observed may better reflect a more
278 physiological ratio of RadA and RadB proteins and, as such, was not considered to be an
279 impediment to further analysis.

280 *H. volcanii* requires 2 M sodium chloride to maintain internal osmotic balance in artificial media
281 (23). Consideration is therefore required around buffer selection to maintain salt levels to
282 promote solubility of halophilic proteins and avoid aggregation. Structural characterization of
283 Hvo PCNA at 2 Å resolution demonstrated the correct folding of the trimeric protein with a root
284 mean square deviation of 1.3-1.7 Å compared with other archaeal and eukaryotic PCNA
285 molecules (19, 20). Similarly, the Hvo RPA3 DNA-binding protein was shown to be competent
286 to bind DNA in buffers containing 1 M KCl, but not at the lower concentration tested (0.2 M
287 KCl) (19, 20). Both proteins were expressed in *E. coli* and were principally purified in buffers
288 containing 1M NaCl, therefore this concentration was maintained in this study.

289 IMAC was used as a first step for the purification of His-tagged RadA and un-tagged RadB using
290 a cobalt-based metal affinity resin. The cobalt-based resin was observed to bind His-tagged
291 halophilic RadA protein with a higher specificity compared to nickel-based resins, reducing non-

292 specific binding of host proteins (data not shown).

293 Co-eluted His-tagged RadA and un-tagged RadB proteins were observed at a similar ratio as
294 before at their expected molecular mass positions by SDS-PAGE analysis (Fig. 2b), indicating
295 specific interaction. Eluted fractions were subsequently subjected to SEC for further purification
296 and to provide an initial indication of the stoichiometry of RadA-RadB interaction. Both proteins
297 eluted together as a complex of ~370 kDa across the elution range of 120-170 mL (Fig. 3a). No
298 difference in protein ratio was apparent in any elution fraction suggesting that complex
299 formation may be constrained by a particular stoichiometric configuration (Fig. 3b). Absence of
300 contaminating DNA was confirmed by agarose gel electrophoresis.

301

302 **Optimisation of the heterologous expression system**

303 Whereas previous experiments demonstrated successful purification of a stable RadA-RadB
304 complex, subsequent analysis was hampered by a limitation of the purification protocol to
305 remove all non-specific host proteins. One major persistent contaminant (~64 kDa) that could not
306 be removed by standard optimisation approaches without significant concomitant decrease in
307 yield of the RadA-RadB complex was identified by MALDI analysis to be the product of the
308 *arnA* gene (bi-functional polymyxin resistance protein, ArnA). The primary sequence of ArnA
309 demonstrates an abundance of histidine residues (31) likely accounting for its capture by IMAC
310 resins. Similarly, the ~74 kDa ArnA protein is reported to form a hexameric structure consisting
311 of a dimer of trimers (31, 32) with a predicted molecular mass of ~440 kDa, close to the
312 observed ~370 kDa RadA-B complex.

313 In view of this, bacteriophage-mediated P1 transduction was performed to delete the *arnA* gene
314 from the Rosetta 2 (DE3) expression strain. The resulting Rosetta 2 (DE3) Δ*arnA* mutant strain

315 was found to be equivalent to the parent strain in respect of growth characteristics and over-
316 expression of RadA-RadB. Subsequently, RadA-RadB complex expressed from the modified
317 strain was purified to ~95% homogeneity following the previous two-step purification protocol
318 (data not shown).
319

320 **RadA oligomerizes into ring like structures**

321 Recombinase proteins characteristically oligomerise into ring structures or helical nucleoprotein
322 filaments. Electron microscopy of *P. furiosus* RadA revealed that the protein forms dimers of
323 heptameric ring structures in solution in the absence of DNA (8) whereas, in *S. solfataricus*,
324 RadA forms an octameric ring bound to DNA (33). Since Hvo RadA-RadB is consistently
325 purified as a large ~370 kDa stable complex, some form of RadA oligomerization is clearly
326 indicated. Preliminary densitometric analysis of co-purified RadA and RadB protein bands
327 estimates the ratio of RadA co-purified with RadB to be 4:1. Therefore the observed ~370 kDa
328 RadA-RadB SEC complex would most closely correspond to eight molecules of RadA in
329 complex with two molecules of RadB although this estimation is limited by the resolution of the
330 S200 SEC column employed in this study.

331 Subsequent electron microscopy observation of our purified Hvo RadA-RadB protein complex
332 confirms that Hvo RadA-RadB complex also possesses a tendency to self-associate into
333 multimeric structures in the absence of DNA (Fig. 4). In order to distinguish protein structures
334 from electron microscopy (EM) artifacts resulting as a consequence of high salt levels, several
335 control grids were prepared and imaged with either buffer only or buffer plus stain at 100,000X
336 magnification. The images captured showed ring-like structures throughout the field which
337 suggests that Hvo RadA-RadB complex in its native state similarly exists as a ring-like structure

338 (Fig. 4).

339

340

341 Discussion

342 The expression and purification of halophilic proteins, particularly from heterologous host
343 overexpression systems presents several challenges, principally due to the particular
344 physicochemical properties of these proteins which constrain conventional approaches.
345 Halophilic proteins are adapted to maintain solubility and function in hyper saline (1-4 M)
346 conditions (34). The presence of high salt in halophilic organisms affects the gross molecular
347 conformation of both proteins (35) and DNA (36). Mesophilic DNA-binding proteins typically
348 make heavy use of positively charged residues to mediate electrostatic interactions with DNA
349 (15, 29, 37). In the majority of organisms the presence of higher concentrations of salt interferes
350 with DNA-binding and leads to aggregate formation in mesophilic proteins. The situation is
351 reversed in halophilic counterparts. Halophilic proteins are generally decorated with negatively
352 charged amino acid residues which allow the binding of surplus water and salt to build up a
353 hydrated solvent network on the surface of proteins. Structural studies have shown that binding
354 of hydrated cations (provided by excess salt and water molecules) around the negatively charged
355 residues on the protein surface reduces the electrostatic repulsive forces between polyanionic
356 DNA and protein molecules (19, 38). At lower concentrations of salt, the protective effect is lost
357 and the repulsive forces between the acidic residues lead to unfolding and inactivation of the
358 protein (28).

359 Consistently, *H. volcanii* RadA has an overall abundance of acidic amino acids, particularly
360 aspartate which, based on homology modelling of RadA, contributes to the overall predicted
361 negative surface charge of the protein. Accounting for this, we successfully adapted a protocol
362 for purification of soluble RadA under moderate salt conditions. Further protocol modifications
363 were introduced to enable purification from persistent contaminants including both DNA and
364 particular host proteins. Most notably, we overcame the previously observed deleterious effect of
365 RadA over-expression in *E. coli* by co-expression of RadA with RadB. These adaptations
366 enabled RadA purification as a DNA-free complex with RadB using a relatively facile two-step
367 purification protocol. Based on our observations we speculate that RadA interacts with a small
368 proportion of RadB in a particular stoichiometry to stabilize it as a binding partner.

369 *H. volcanii* RadA is similar to other characterised RadA homologues. Consistent with its
370 function as a recombinase, ATP and DNA binding motifs are apparent which have a well-
371 conserved sequence composition despite the common skews in overall amino acid usage
372 typically observed in halophilic proteins (Fig. 1 and Fig. S1). The abundance of acidic amino
373 acids such as aspartate in Hvo RadA could potentially result in electrostatic repulsion with the
374 negatively charged DNA backbone under *in vitro* conditions (Fig. 1). The crystal structure
375 analysis of Hvo PCNA indicates that this protein compensates for the reduction in positively
376 charged surface residues by employing cation binding (19). However, the interaction in this
377 instance is largely topological in nature rather than via interaction of specific residues. A
378 conserved basic patch on the surface of RadB has been shown to be crucial in DNA binding (15)
379 and number of similarly conserved positively charged residues are retained in Hvo RadA. The
380 presence of a highly negatively charged loop uniquely present in Hvo RadA may compensate for

381 the retention of lysine residues that presumably are maintained to enable direct DNA-binding
382 (Fig. 1a).
383 The failure to observe DNA binding activity in Hvo RadA across a range of methods (refer to
384 supplementary text, Fig. S3 and Table S1) suggests that, for this protein, the protein-DNA
385 interaction may require additional, as yet unidentified, factors. It is also possible that under the
386 conditions of our assays, RadA and RadB interaction occluded the DNA binding site or
387 prevented an active conformational state of the proteins required for DNA binding and ATPase
388 activity. It is also possible that additional co-factors are required for enzymatic activity of the
389 complex.
390 A recently published study exploring the role of Hvo RadB in homologous recombination also
391 confirmed interaction of RadA with RadB when the proteins were individually over-expressed in
392 *H. volcanii*, identifying co-purifying proteins by mass spectrometry [39]. Genetic analyses
393 identified point mutations in RadA that suppress the Δ*radB* phenotype, supporting the hypothesis
394 that RadB functions as a recombination mediator, potentially by inducing conformational
395 changes within RadA. Additionally, the authors noted that *in vitro* assays performed did not
396 demonstrate functional DNA binding, as found in this study, also noting that additional co-
397 factors may be required.
398 Recombinases have been generally observed to self-assemble into ring structures (six to eight
399 protomers) in their native state and monomerize to interact with DNA to form the helical
400 nucleoprotein filaments. So far the precise function of the ring structures is not clear in the
401 process of homologous base pairing. The toroidal form in recombinases has been suggested to
402 function as an inactive storage form of protein, which is likely utilized to occlude the
403 polymerization motif, preventing unwanted interaction of the protein with DNA. EM analysis of

the purified protein complex showed that RadA-RadB also multimerizes into ring-like structures, which we consider may be of the order of a heptamer or octamer (Fig. 4), similar to its RadA/Rad51 homologs. On the basis of our collective observations, further studies to understand the dynamics of halophilic protein-protein and protein-DNA interactions in *H. volcanii*, including gross structural determination and robust biophysical characterization can now be anticipated.

Funding

This work was supported by the University of Sindh, Jamshoro and the Higher Education Commission of Pakistan. K.B. and J.W. were supported by a Wellcome Trust RCDF award to K.B. (Grant no. 076556/Z/05/Z).

Conflicts of interest

None.

References

1. **Kowalczykowski SC, Eggleston AK.** Homologous pairing and DNA strand-exchange proteins. *Annu Rev Biochem.* 1994;63:991-1043.

2. **Story RM, Weber IT, Steitz TA.** The structure of the E. coli recA protein monomer and polymer. *Nature.* 1992;355(6358):318-25.

3. **Sandler SJ, Satin LH, Samra HS, Clark AJ.** recA-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast Saccharomyces cerevisiae. *Nucleic Acids Res.* 1996;24(11):2125-32.

431
432 4. **Ogawa T, Shinohara A, Nabetani A, Ikeya T, Yu X, et al**. RecA-like recombination
433 proteins in eukaryotes: functions and structures of RAD51 genes. *Cold Spring Harb Symp Quant*
434 *Biol*. 1993;58:567-76.
435
436 5. **Shinohara A, Ogawa H, Ogawa T**. Rad51 protein involved in repair and recombination
437 in S. cerevisiae is a RecA-like protein. *Cell.* 1992;69(3):457-70.
438
439 6. **Komori K, Miyata T, Daiyasu H, Toh H, Shinagawa H, et al.** Domain analysis of an
440 archaeal RadA protein for the strand exchange activity. *J Biol Chem*. 2000;275(43):33791-7.
441
442 7. **Ariza A, Richard DJ, White MF, Bond CS.** Conformational flexibility revealed by the
443 crystal structure of a crenarchaeal RadA. *Nucleic Acids Res.* 2005;33(5):1465-73.
444
445 8. **Shin DS, Pellegrini L, Daniels DS, Yelent B, Craig L, et al.** Full-length archaeal Rad51
446 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2. *EMBO J.*
447 2003;22(17):4566-76.
448
449 9. **Sung P, Klein H.** Mechanism of homologous recombination: mediators and helicases
450 take on regulatory functions. *Nat Rev Mol Cell Biol*. 2006;7(10):739-50.
451
452 10. **Symington LS.** Role of RAD52 epistasis group genes in homologous recombination and
453 double-strand break repair. *Microbiol Mol Biol Rev.* 2002;66(4):630-70.
454
455 11. **Sandler SJ, Hugenholtz P, Schleper C, DeLong EF, Pace NR, et al.** Diversity of radA
456 genes from cultured and uncultured archaea: comparative analysis of putative RadA proteins and
457 their use as a phylogenetic marker. *J Bacteriol.* 1999;181(3):907-15.
458
459 12. **Komori K, Miyata T, DiRuggiero J, Holley-Shanks R, Hayashi I, et al.** Both RadA
460 and RadB are involved in homologous recombination in Pyrococcus furiosus. *J Biol Chem.*
461 2000;275(43):33782-90.
462
463 13. **Allers T, Ngo HP.** Genetic analysis of homologous recombination in Archaea: Haloferax
464 volcanii as a model organism. *Biochem Soc Trans.* 2003;31(Pt 3):706-10.
465
466 14. **Haldenby S, White MF, Allers T.** RecA family proteins in archaea: RadA and its
467 cousins. *Biochem Soc Trans*. 2009;37(Pt 1):102-7.
468
469 15. **Guy CP, Haldenby S, Brindley A, Walsh DA, Briggs GS, et al.** Interactions of RadB, a
470 DNA repair protein in archaea, with DNA and ATP. *J Mol Biol.* 2006;358(1):46-56.
471
472 16. **Allers T.** Overexpression and purification of halophilic proteins in Haloferax volcanii.
473 *Bioeng Bugs.* 2010;1(4):288-90.
474

475 17. **Allers T, Barak S, Liddell S, Wardell K, Mevarech M.** Improved strains and plasmid
476 vectors for conditional overexpression of His-tagged proteins in Haloferax volcanii. *Appl*
477 *Environ Microbiol.* 2010;76(6):1759-69.
478 18. **Wardell K, Haldenby S, Jones N, Liddell S, Ngo GHP, et al.** RadB acts in homologous
479 recombination in the archaeon Haloferax volcanii, consistent with a role as recombination
480 mediator. *DNA Repair* (Amst). 2017;55:7-16.
481
482 19. **Winter JA, Christofi P, Morroll S, Bunting KA.** The crystal structure of Haloferax
483 volcanii proliferating cell nuclear antigen reveals unique surface charge characteristics due to
484 halophilic adaptation. *BMC Struct Biol.* 2009;9:55.
485
486 20. **Winter JA, Patoli B, Bunting KA.** DNA binding in high salt: analysing the salt
487 dependence of replication protein A3 from the halophile Haloferax volcanii. *Archaea.*
488 2012;2012:719092.
489
490 21. **Sheng D, Li M, Jiao J, Ni J, Shen Y.** Co-expression with RadA and the characterization
491 of stRad55B, a RadA paralog from the hyperthermophilic crenarchaea Sulfolobus tokodaii. *Sci*
492 *China C Life Sci.* 2008;51(1):60-5.
493
494 22. **Sternberg N.** Bacteriophage P1 cloning system for the isolation, amplification, and
495 recovery of DNA fragments as large as 100 kilobase pairs. *Proc Natl Acad Sci U S A*.
496 1990;87(1):103-7.
497
498 23. **Mullakhanbhai MF, Larsen H.** Halobacterium volcanii spec. nov., a Dead Sea
499 halobacterium with a moderate salt requirement. *Arch Microbiol.* 1975;104(3):207-14.
500
501 24. **Ziegler J, Vogt T, Miersch O, Strack D.** Concentration of dilute protein solutions prior
502 to sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal Biochem.* 1997;250(2):257-
503 60.
504 25. **Janson G, Zhang C, Prado MG, Paiardini A**. PyMod 2.0: improvements in protein
505 sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics*. 2016.
506
507 26. **DeLano WL.** The PyMol Molecular Graphics System. *DeLano Scientific LLC, Palo*
508 *Alto, Calif, USA*. 2008.
509
510 27. **Webb B, Sali A.** Comparative Protein Structure Modeling Using MODELLER. Curr
511 Protoc *Bioinformatics.* 2014;47:5 6 1-32.
512
513 28. **Lanyi JK.** Salt-dependent properties of proteins from extremely halophilic bacteria.
514 *Bacteriol Rev.* 1974;38(3):272-90.
515
516 29. **Chen LT, Ko TP, Chang YW, Lin KA, Wang AH, et al.** Structural and functional
517 analyses of five conserved positively charged residues in the L1 and N-terminal DNA binding
518 motifs of archaeal RADA protein. *PLoS One.* 2007;2(9):e858.
519

520 30. **Poidevin L, MacNeill SA.** Biochemical characterisation of LigN, an NAD+-dependent
521 DNA ligase from the halophilic euryarchaeon Haloferax volcanii that displays maximal in vitro
522 activity at high salt concentrations. *BMC Mol Biol.* 2006;7:44.
523
524 31. **Gatzeva-Topalova PZ, May AP, Sousa MC.** Crystal structure and mechanism of the
525 Escherichia coli ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with
526 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry.* 2005;44(14):5328-38.
527
528 32. **Breazeale SD, Ribeiro AA, McClerren AL, Raetz CR.** A formyltransferase required
529 for polymyxin resistance in Escherichia coli and the modification of lipid A with 4-Amino-4-
530 deoxy-L-arabinose. Identification and function oF UDP-4-deoxy-4-formamido-L-arabinose. *J*
531 *Biol Chem.* 2005;280(14):14154-67.
532
533 33. **Yang S, Yu X, Seitz EM, Kowalczykowski SC, Egelman EH.** Archaeal RadA protein
534 binds DNA as both helical filaments and octameric rings. *J Mol Biol.* 2001;314(5):1077-85.
535
536 34. D**ym O, Mevarech M, Sussman JL.** Structural features that stabilize halophilic malate
537 dehydrogenase from an archaebacterium. *Science.* 1995;267(5202):1344-6.
538
539 35. **Madern D, Ebel C, Zaccai G.** Halophilic adaptation of enzymes. *Extremophiles.*
540 2000;4(2):91-8.
541
542 36. **Schlick T, Li B, Olson WK**. The influence of salt on the structure and energetics of
543 supercoiled DNA. *Biophys J.* 1994;67(6):2146-66.
544
545 37. **Lee CD, Wang TF.** The N-terminal domain of Escherichia coli RecA have multiple
546 functions in promoting homologous recombination. *J Biomed Sci.* 2009;16:37.
547
548 38. **Bergqvist S, Williams MA, O'Brien R, Ladbury JE.** Halophilic adaptation of protein-
549 DNA interactions. *Biochem Soc Trans*. 2003;31(Pt 3):677-80.
550
551 39. **Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, et al.** The complete
552 genome sequence of Haloferax volcanii DS2, a model archaeon. *PLoS One*. 2010;5(3):e9605.
553

554

555

556

557

558

559 # Table

560

561 **Table 1. List of strains and plasmids.**

Strains and plasmids	Genotype/Description	Source
DH5α	*F-ϕ80lacZ ΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK+) phoA supE44 λ– thi-1 gyrA96 relA1*	Novagen
Rosetta 2 (DE3)	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE23 (CamR)*	Novagen
Rosetta 2 (DE3)Δ*arnA*	Rosetta 2 (DE3) with *arnA* inactivating mutation	This study
DS2	*H. volcanii* wild-type strain	(39) & (23)
pZero Blunt	Cloning vector	Invitrogen
pCPG42	Hvo *radB* encoding plasmid	(1515)
pET11-*radA*	*radA* cloned into NdeI/BamHI sites for overexpression of His-	This study
pETDUET-*radA*	*radA* cloned into NdeI/EcoRV sites of MCS2 for overexpression	This study
pBPRAD2	*radB* cloned into NcoI/HindIII sites of pETDUET-*radA* MCS1 for co-expression of His-tagged RadA and untagged RadB	This study

562
563
564
565
566
567
568
569
570
571
572
573
574
575
576

Figure legends

Fig. 1(a). Superposition of the Hvo RadA homology model with the Pfu RadA crystal structure. The Hvo RadA homology model is shown in grey and the Pfu RadA structure (1PZN:Chain A) in red (cartoon representation). Good correspondence is apparent between major structural elements, although Hvo RadA can be distinguished by an additional large loop insertion. (**b). Cartoon representation of the backbone of the modelled Hvo RadA structure.** The Walker A motif (green), Walker B motif (red) and DNA binding loops (blue) are indicated. The Cα position of the lysine residues in Hvo RadA are indicated with orange spheres. The Cα position of acidic residues located in an insertion loop in Hvo RadA relative to related sequences (see supplementary Fig. S1) are indicated with red spheres. NTD, AD and the insertion loop are labelled.

Fig. 2(a). Coomassie Brilliant-Blue stained SDS-PAGE gel demonstrating co-over-expression of Hvo RadA and RadB in *E. coli.* Broadly equivalent levels of recombinant RadA and RadB were observed in whole cell lysates when expressed independently (lanes 3 & 4, respectively) compared with reduced levels of RadB observed when co-expressed with RadA (lane 2, arrow). Lane 1 shows the pre-induction sample. RadA and RadB positions are indicated in (b). Molecular mass markers are indicated (kDa). **(b) Purification of co-expressed RadA-RadB proteins by cobalt-based immobilized metal affinity chromatography.** Untagged RadB is observed to co-purify with His-tagged RadA.

599

600 **Fig. 3(a). Purification of RadA-RadB complex by Size Exclusion Chromatography.**
601 Chromatogram monitoring UV absorbance at 280 nm indicating the elution of the RadA-RadB
602 complex at 130-170mL volume (arrow). Fractions as collected are indicated on the *x*-axis. The
603 relative elution peaks for a series of molecular mass standards are shown with arrows. **(b).**
604 **Coomassie Brilliant-Blue stained SDS-PAGE gel representing the purified proteins after**
605 **SEC.** Lane 1 showing 10 µl Talon purified load prior to purification on a S200 column. All S200
606 purified fractions corresponding to the UV peak at 100-180 mL (lanes 2-9) show RadA and
607 RadB co-purification at their estimated MW markers. An ~64kDa contaminating host protein
608 (ArnA) consistently co-eluted with the RadA-B complex as indicated in lanes 5-8.

609

610 **Fig. 4. Transmission Electron Microscopy (TEM) images of co-expressed RadA-B complex.**
611 TEM images of co-expressed RadA-B complex at 100,000X magnification. All of the protein
612 samples were negatively stained with either (a) 1% or (b) 0.5% phosphotungstic acid (PTA).
613 Consistently sized (~13nm diameter) ring-like RadA-RadB complexes are indicated with arrows.
614 Boxed ring-like structures are shown enlarged in inset images.