Supplemental Methods and Figures for

Structure of the periplasmic adaptor protein from a major facilitator superfamily (MFS) multidrug efflux pump

Philip Hinchliffe, Nicholas P. Greene, Neil G. Paterson¹, Allister Crow,

Colin Hughes, Vassilis Koronakis²

Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK

¹Diamond Light Source, Harwell Science and Innovation Campus, Didcot, OX11 0DE, UK ²Corresponding author: Vassilis Koronakis, Tel: +44 (0)1223 339766; E-mail: <u>vk103@cam.ac.uk</u>

Supplemental Methods

Cloning soluble EmrA homologues

To express soluble *Aquifex aeolicus* EmrA (*aa*EmrA) we removed the transmembrane (TM) helix by amplifying *emrA* codons 27-374 from *Aquifex aeolicus* genomic DNA with forward and reverse primers (5'-GCGCCATATGAAACACAGAATTGAGTATGCCATAACAAACG-3', 5'-GCGCTCGAGTCTCGTTCTCCTTATCTCAACCTCACCGC-3'). The product was digested with NdeI/XhoI and cloned into pET24 (Novagen). The resultant plasmid was designated pET24-*aa*EmrAΔTM. Plasmid pET24-*ec*EmrAΔTM expressing soluble *E.coli* EmrA (codons 48-390, lacking the TM domain) was produced by amplification of genomic DNA with forward (5'-GCGCCATATGCACTTCGAAGAAACCGATGACGCATACG-3') and reverse (5'-GCGCCTCGAGGCCAGCGTTAGCTTTTACGATATCG-3') primers, digestion and ligation as pET24-*aa*EmrAΔTM. We cloned EmrA homologues from *Rhodobacter capsulatus*, *Pseudomonas aeruginosa*, *Salmonella enterica* Typhimurium and *Thermodesulfobium narugense* with a similar strategy.

		2000000000000000	
aaEmrA ecEmrA ecEmrK vcVceA smEmrA rcEmrA stEmrA paEmrA tnEmrA bpEmrA hiEmrA ngFarA	MKKYTI MSA NAETQTPQQPVKKSGKRKRLL MEQINSNKKSNRRKYF MN.SNNSNTEFDGVRARQSRKKGE MSQ.TQD.TAAPAAPNRRGNLL MT.AE.TQAPKKS.PKKLLM MT.AE.TQAPKKS.PKKLLM MSA.NAEIQTPQQPAKKKGKRKTAL MSDQLKENNIKAKFKLPQPKKPSKKGE MTTN.PNPTDNPKAASGAPRTRKRLL MSDQLKENNIKAKFKLPQPKKPSKKGE MTTTN.PNPTDNPKAASGAPRTRKRLL MSDQLKENNIKAKFKLPQKKLSKKKRK MTQ.IATENPSTKSVSNKTDRKKGL MKSGNSEPNLMET.HTD.ETKLQNTQVKRKRRL	SVFTVVALTVGFGTYAFFF LLLTLLFIIIAVAIGIYWFLV SLLAVVLFIAFSGAYAYWSME GGAAATVVAGGSYALYWHFI GGLAAATVAGSYALYWHFI GGLFVIVVLLAALALWYFMF LLTLLFVIIAVAYGIYWFLV LLLTLFVIIAVAYGIYWFLV LILLAMIVLATLASVAWEFFY LIFLLVVVVIILSPAYQKIVY ALLGLGALVAAAGYGAYYYSV TLLVVVVIAIAAIAYGLYYLVSV TLLVVVVIAIAAIAYGLYYLVSV TLLVVVVIAIAAIAYGLYYLVSV TLLVVVVIAIAAIAYGLYYLVSV TLLVVVVIAIAAIAYGLYYLVSV TLLVVVVIAIAAIAYGLYYLVSV	IKHRIEYAITNAVFVKADELSYLSFRÜ LRHFEETDDAYVA.GNQIQIMSQV LEDMISTDDAYVA.GNQIDJISAQV GSRYISTDDAYVA.GNQVQITPAV GRWFEETDDAYVA.GNQVQITPLV SGRYESTENAVQ.QARISVASDI LRHIEETDDAYVA.GNQVQIMAQV GRWHEDTDDAYVA.GNQVQITPQI SLTHQ.STDDAYVA.GNLVQLTPQV ARFHETDDAYVN.GNLVQLTPQV ARFHETDDAYVN.GNLVQLTPQV ARFHETDDAYVN.GNLVQLTPQV ARFHETDDAYVN.GNLVQLTPQV ARFHETDDAYVN.GNLVQLTPQV ARFHETDDAYVA.GNVVQITPQV QUHEEETEDAYVA.GNVVQVTPQK
aaEmrA ecEmrA vcVceA smEmrA stEmrA paEmrA tnEmrA bpEmrA hiEmrA ngFarA	SGKVI I EVY KDLGD YV KRG EALAKLDP TY YELEK RTLEI SGSVI TKVWADNTDFV KEGDVLVTLDP TDARQAFE KAK SGSVI TVVNHKDTNYV RQGDILVSLDKTDATIALLQAE GRDYTASECRWI PSTSTGDVLVQLDDTDARALLQAE AGTVVAINADDGMRVERGQLLVQLDP SDTAVALQQAE SGSVI TVVAINADDGMRVERGQLLVQLDP SDTAVALQQAE SGSVI TKVWADNTDFV KEGDVLVTLDQTDAKQAFE KAK VGTVV3I GADDGDLVRKGQE VLWRLDP SDADIALQAE SGSVI TKVWADNTDFV KEGDVLVVLDR DP SDADIALQAE SGSVI TKVWADNTDFV KEGDVLVVLDR DP SDADIALQAE SGSVI TKVWADNTDFV KEGDVLVVLDR DP SDADIALQAE SGKVI VAVNVER YQEV KKGQU VLAELTS SDADIALALANAE SGKVI VAVNVER VQEV KKGQD LIVRLDP SDADIALQAE SGKVI VAVNVER VQEV KKGQL VLAELTS SDADIALANAE SGKVI VAVNVER VQEV KKGQD LIVLDP ADS QVALQAE GGTVR KULHDDTDAVKKGDVLAVLDDDNDVLAYERAK	90 11 100 10 100 100 10 100	0 120 130 KG[]HISLSAKKLKVESLKKKREALREK ANIE AEVA AMVEA AQEQ ANID AQEQ ANID AQEQ ANTO AVEA ANTO AVEA ANTO ANTO ANTO APUA ANTO ANTO ANTO ANTO ANTO ANTO AQUA AQUA AQVA AQVA
aaEmrA ecEmrA ecEmrK vcVceA smEmrA stEmrA stEmrA tnEmrA tnEmrA hiEmrA ngFarA	00000000000000000000000000000000000000	180 190 180 190 NILKVILHEREYLEKSIQEINT AVTSAQAQLDVAIQQYN. ISIAQAQANLNAAKAAMAQ SFAQAQANLNAAKAAMAQ LAAABAAVAGSETVE. JASSALBSLAAKQGVA. AVASAQAQLDVAIQQYN. ALDSAKASLTSSQUN. AVASAQAQLDVAIQQYN. ALDSAKASLTSSQQLA. VKAAQASLDAAKQQLA. VKAAQASLDAAGQQLA. AVKASKQLAANLNASKNQLA.	0000000000000000000000000000000000
2 aaEmrA ecEmrA ecEmrK vcVceA smEmrA stEmrA paEmrA tnEmrA bpEmrA hiEmrA ngFarA	20 20 23 240 250 LSS LEEETKSLKERIKTAEQKIKDTVLIAPFDGVVAK KLED QPAVQQAATEVRNAWLALERTRIISPMTGYVSR PLNRQPQVVEAADATKEAWLALKRTDIKSPVTGYIAQ TVD SNPEVLLAKARYEQAKID LERTVIRAPISGIVAK VIATQPDVQAAAQLRQAFLNNARAGIVAPVTGYVAR ETDTHFTVMAALTARDKAAYNLGLTEVKAPADGVVYK, NLED QPAVQQAATEVRNAWLALERTRIVSPMTGYVAR KVTALXAIXQAKANLDQALDDARSTIVAPVTGYVAK SVADHPSVQAAATEVRNAWLALERTRIVSPMTGYVAK SVADHPSVQAAATEVRNAWLALERTRIVSPMTGYVAK SVADHPSVQAAATEVRNAWLALERTRIVSPMTGYVAK SVADHPSVQAAATEVRDAYLANARNTLPAPVTGYVAK SLREQPEVQTAIGRLKDAWLNLRRTQVRAPADGQVAK	260 270 RFTISR DVVRAGOPAFALVN RAVOPGAOISPTTPLMAVVP RSVQVGERVQVGMPLMTVVP RSVQVGERVQVGMPLMTVVP DSFKPGOPVAACAALFSVE RAVOPGAOISPTTPLMAVVP RSVQVGERVQPGNALMAVVP RSVQVGOPCAALFSVE RAVOPGAOISPTTPLMAVVP RSVQVGOPVAPGNALMAVVP RSVQVGOPVAPGNLTIV RSVQVGORVAPGNLTVNP RSVQVGORVAPGNLTVNP RSVQVGORVAPGNPLMAVVP RSVQVGORVAPGNPLMAVVP RSVQVGORVAPGNPLMAVVP RSVQVGORVAPGNPLMAVVP RSVQVGORVAPGNPLMAVVP	280 290 280 290 300 PESFYVEVLLEETKLKGVKVGNKAYVR ATNMWVDANFKETOIANMRIGOPVTII TDHIYVDANFKEVELRDVKVGOPVTII TDHIYVDANFKEVELRDVKVGOPVTLL TGDTWIEANFKETOIGKMAPGOKAEVE ATDLWVDANFKETOIGKMAPGOKAEVE KNDLWVTANFKETOIDKMRIGOPVEI LDGVWVDANFKEVOLKHIRIGOPVELH LNAVWVDANFKEVOLKHIRIGOPVELH LNAVWVDANFKEVOLKHIRIGOPVELH LNAVWVDANFKETOITNRIGOPVELH LNAVWVDANFKETOITNRIGOPVELH LSDVWVDANFKETOITNRIGOPVELH LSDVWVDANFKETOITNRIGOPVELH
2 aaEmrA ecEmrK vvVceA smEmrA stemrA paEmrA hjEmrA hjEmrA hjEmrA ecEmrA ecEmrA ecEmrA stemrA stemrA stemrA stemrA ngFarA	20 20 20 23 240 250 250 250 250 250 250 250 25	260 270 RFTSRGDVVRAGQPAFALVN RAVOPGAQISPTTPLMAVVP RSVQVGCRVQVGMPLMTVVP RSVQVGCRVQPGNALMAVVP RSVQVGCRVQPGNALMAVVP RSVQVGCRVQPGNALMAVVP RSVQVGCQVQFGNALMAVVP RSVQVGCQVQFGNALMAVVP RSVQVGCQVAPGNALMAVVP RSVQVGQRVQPGNALMAVVP RSVQVGQVAPGNLMAVVP RSVQVGQVAPGNLMAVVP RSVQVGQVAAGAVLMAVVP RSVQVGQVAAGAVLMAVVP 1000 350 11 12 14 15 16 17 18 18 18 18 18 19 18 19 100 11 12 12 13 14 15 15 16 17 10 10 10 10 10 10 10 <th>280 280 280 280 290 300 PESFYVEVLLESTKLKGVKVGNKAVVR ATNMWVDANFKETCILANMRIGOPVTILT ARQMWVNANFKETCILTDVRIGOSVNII TDHIYVDANFKEVELRDVKVGOPVTLT TEQMWVEANFKETCILTANMRIGOPVTLI TEQMWVEANFKETCILTANMRIGOPVTLI LDQIWIDANFKETCILTANMRIGOPVEIR KNDLWVTANFKETCILTANMRIGOPVEIR KNDLWVDANFKETCILTANMRIGOPVEIR LDQIWIDANFKETCILTANMRIGOPVEIR LDQVWVDANFKETCILTANMRIGOPVEIR LDAVWVDANFKETCILTANMRIGOPVEIR SOUWVDANFKETCILTANMRIGOPVEIR NEQMWLEANFKETCILTANMRIGOPVEIR NEQMWLEANFKETCILTANMRIGOPVEIR NEQMWLEANFKETCILTANMRIGOPVEIR SOUWVDANFKETCILTANT PLRIGLSTLVSVNTTNRDGQVLANKVR PLRIGLSMTATIDTAGTDPQTLVQYR PLRIGLSMATIDTARTDQULAQVE PLRIGLSMATIDTAGTDQULAQVE PLRIGLSMATAVVDTARKTNLDALVMRV PLRIGLSMVAVVDTARKTNLDALVMRV PLRIGLSMVVDVDTGF PLQGLSMQVDVDTGF PLRIGLSMQVDVDTGF PLRIGLSMQVDVDTRAKTNLDALVMRV PLRIGLSMQVDVDTRAKTNLDALVMPT</th>	280 280 280 280 290 300 PESFYVEVLLESTKLKGVKVGNKAVVR ATNMWVDANFKETCILANMRIGOPVTILT ARQMWVNANFKETCILTDVRIGOSVNII TDHIYVDANFKEVELRDVKVGOPVTLT TEQMWVEANFKETCILTANMRIGOPVTLI TEQMWVEANFKETCILTANMRIGOPVTLI LDQIWIDANFKETCILTANMRIGOPVEIR KNDLWVTANFKETCILTANMRIGOPVEIR KNDLWVDANFKETCILTANMRIGOPVEIR LDQIWIDANFKETCILTANMRIGOPVEIR LDQVWVDANFKETCILTANMRIGOPVEIR LDAVWVDANFKETCILTANMRIGOPVEIR SOUWVDANFKETCILTANMRIGOPVEIR NEQMWLEANFKETCILTANMRIGOPVEIR NEQMWLEANFKETCILTANMRIGOPVEIR NEQMWLEANFKETCILTANMRIGOPVEIR SOUWVDANFKETCILTANT PLRIGLSTLVSVNTTNRDGQVLANKVR PLRIGLSMTATIDTAGTDPQTLVQYR PLRIGLSMATIDTARTDQULAQVE PLRIGLSMATIDTAGTDQULAQVE PLRIGLSMATAVVDTARKTNLDALVMRV PLRIGLSMVAVVDTARKTNLDALVMRV PLRIGLSMVVDVDTGF PLQGLSMQVDVDTGF PLRIGLSMQVDVDTGF PLRIGLSMQVDVDTRAKTNLDALVMRV PLRIGLSMQVDVDTRAKTNLDALVMPT

Supplemental Figure 1. Sequence alignment of EmrA homologues.

Amino acid sequence alignment of EmrA from *Aquifex aeolicus* (aaEmrA), *Escherichia coli* (ecEmrA and ecEmrK), *Vibrio cholerae* (vcVceA), *Stenotrophomonas maltophilia* (smEmrA), *Rhodobacter capsulatus* (rcEmrA), *Salmonella* Typhimurium (stEmrA), *Pseuodomonas aeruginosa* (paEmrA), *Thermodesulfobium narugense* (tnEmrA), *Cupriavidus metallidurans* (cmEmrA), *Burkholderia pseudomallei* (bpEmrA), *Haemophilus influenza* (hiEmrA), *Neisseria gonorrhoeae* (ngEmrA). Strictly conserved residues are boxed in white on a red background and highly conserved residues are boxed in red on a white background. Secondary structure is indicated above and coloured according to domain (purple, predicted transmembrane helix; yellow, β -barrel; green, lipoyl; blue, α -helical coiled-coil; red, β -barrel disordered loop).



Supplemental Figure 2. Disorder and electrostatic distribution over the EmrA structure.

(A) B-factor distribution over the EmrA structure. The crystal structure of Aquifex EmrA is coloured according to C_{α} B-factors, from blue (low B-factor) to red (high B-factor).

(B) EmrA electrostatic surface. The EmrA surface is shown with positively and negatively charged surface areas coloured blue and red, respectively. Electrostatic surface potentials were calculated using the program APBS and contoured at $\pm 9 \text{ kT/}e$.



Supplemental Figure 3. Alignment of structurally characterised periplasmic adaptors. EmrA structure compared to structurally characterised drug efflux pump adaptors BesA (*B.burgdorferi*), MexA (*P.aeruginosa*), AcrA (*E.coli*) and MacA (*E.coli*) and the metal efflux pump adaptors CusB (*E.coli*) and ZneB (*C.metallidurans*). Adaptors are coloured by domain, blue (α-helical coiled-coil), green (lipoyl), yellow (β-barrel) and orange (membrane proximal, MP). Dotted orange lines indicate unobserved adaptor C- and N- terminal regions.



Supplemental Figure 4. Sequence conservation over EmrA.

Left, structure of EmrA coloured by domain as in main text Figure 1. The disordered residues 322-342 have been modelled as a loop (grey) projecting out from the β -barrel. *Centre*, surface representation of EmrA with conserved amino acids coloured from highest conserved residues (pink) to lowest conserved residues (yellow). *Right*, a close up of the β -barrel, rotated 180° from the centre view.



Supplemental Figure 5. Isothermal titration calorimetry properties of EmrA titrated with known EmrAB substrates.

Soluble constructs of 0.1mM *E.coli* and *A.aeolicus* EmrA (*ec*EmrA and *aa*EmrA, respectively) were exchanged into 20mM HEPES pH8.0, 150mM NaCl on a superdex 200 gel filtration column and concentrated to 0.1mM in an Amicon 10kDA molecular weight cut-off concentrator (Millipore). They were then titrated with 10mM nalidixic acid (NA) or 1mM DNP in the same buffer. ITC measurements were carried out at 25°C using a VP-ITC MicroCalorimeter (Microcal). Protein (0.1mM) was placed in the 1400µl calorimeter cell and 8µl of titrant was injected over 16s with 200s intervals between each injection. Titrant into buffer alone was used as a negative control. As a positive control, binding of 10mM NiCl₂ to the his-tags of *ec*EmrA and *aa*EmrA was analysed using the same parameters. Data were analysed using Origin software (Microcal). In each case, the upper panel shows raw energy changes during the titration (time), while the lower panel presents the derived integrated total energy changes as a function of the molar ratio of the interactants.