BIOCHEMISTRY
including biophysical chemistry & molecular biology

Rapid Report

pubs.acs.org/biochemistry

The Solution Structures of Two Prophage Homologues of the Bacteriophage λ Ea8.5 Protein Reveal a Newly Discovered Hybrid Homeodomain/Zinc-Finger Fold

Jamie J. Kwan,[†] Ekaterina Smirnova,[†] Shahin Khazai,[†] Ferenc Evanics,[†,§] Karen L. Maxwell,[‡] and Logan W. Donaldson*[,†]

[†]Department of Biology, York University, 4700 Keele Street, Toronto, ON M3J1P3, Canada

[‡]Department of Molecular Genetics and Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada

Supporting Information

ABSTRACT: A cluster of genes in the *exoxis* region of bacteriophage λ are capable of inhibiting the initiation of DNA synthesis in *Escherichia coli*. The most indispensible gene in this region is *ea8.5*. Here, we report the nuclear magnetic resonance structures of two *ea8.5* orthologs from enteropathogenic *E. coli* and *Pseudomonas putida* prophages. Both proteins are characterized by a fused homeodomain/zinc-finger fold that escaped detection by primary sequence search methods. While these folds are both associated with a nucleic acid binding function, the amino acid composition suggests otherwise, leading to the possibility that Ea8.5 associates with other viral and host proteins.

Upon infection, a virus must subvert basic cellular processes of the host but not to the extent that the host becomes unviable. Bacteriophage λ has long served as a model for understanding virus−host relationships, yet many open reading frames (ORFs) remain relatively uncharacterized. Several of these ORFs cluster in the *bin* (blocks of inititation of DNA replication) region between the *exo* and *xis* genes of the p_L operon that is expressed during the early lytic stages of infection and silenced in the lysogenic stage (Figure 1). Experimental evidence suggests that effects of these genes are most beneficial to the virus when they are transient, tipping the balance of its diminished host's resources toward the production of viral DNA.[1] The effects of the *bin* genes on the inhibition of DNA synthesis is distinct from effects of the *kil* and *gam* gene products that block the bacterial cell cycle.[2]

Figure 1. ORFs comprising the *bin* region (blue) are located between the well-characterized *exo* and *xis* genes in the p_L operon.

The gene product of *ea8.5* appears to be least dispensable and, therefore, the most critical component of the *bin* region.[1] To understand the biochemical and structural basis of how host DNA replication may be deregulated by λ, we initially expressed *ea8.5* as a His_6-tagged protein under a strong T7 promoter. The resulting protein was expressed as insoluble inclusion bodies but could be refolded upon rapid dilution from denaturant into a physiological buffer. While an ^{15}N-edited HSQC nuclear magnetic resonance (NMR) spectrum of an isotopically labeled sample suggested a folded protein, a structural investigation could not be pursued as the protein polymerized into a clear gel at concentrations as low as 0.1 mM. As a result, a CS-BLAST search[3] was performed to identify homologues with potentially better solution properties. Two ORFs were identified in prophage-derived regions from enteropathogenic *Escherichia coli* (EC0111_1119) and *Pseudomonas putida* KT2440 (PP_3909), with sequences 41 and 16% identical, respectively, to that of λ ea8.5. A sequence alignment is presented in Figure 2A. A CxxH/CxxC motif was conspicuously conserved among the three sequences, suggesting zinc ion binding. In addition, a PSIPRED secondary structure prediction[4] of the sequences revealed the same $\alpha\beta\beta\alpha\alpha\alpha$ composition, suggesting EC0111_1119 and PP_3909 were structurally related to Ea8.5. As was observed previously for the λ tail protein, gpU, the only Ea8.5 homologues in the NCBI databases were other viral proteins, suggesting a considerable evolutionary divergence from other proteins that may share the same fold.[5]

Codon-optimized EC0111_1119 and PP_3909 DNA sequences were directly synthesized by DNA2.0 (Menlo Park, CA). While His_6-tagged variants were not expressed, sufficient protein was obtained for an NMR structural study in which glutathione *S*-tranferase fusion was used. As was observed in prior studies of the *bin* region genes, *E. coli* harboring EC0111_1119 or PP_3909 plasmids grew extremely slowly, suggesting the expressed fusion proteins were exerting a deleterious effect on bacterial cell division.

Using a suite of common triple-resonance experiments conducted at 600 MHz on uniformly ^{13}C- and ^{15}N-labeled samples of EC0111_1119 (0.6 mM, 16 °C) and PP_3909 (0.6 mM, 20 °C), final ensembles of 20 structures were determined using CYANA[6] with overall backbone atom precisions of 0.67 and 0.53 Å, respectively. Detailed protocols are provided as Supporting Information. A molecular model of Ea8.5 was produced from a PP_3909 template using SWISS-MODEL[7] and

Received: April 30, 2013
Revised: May 10, 2013
Published: May 13, 2013

   dx.doi.org/10.1021/bi400543w | *Biochemistry* 2013, 52, 3612−3614

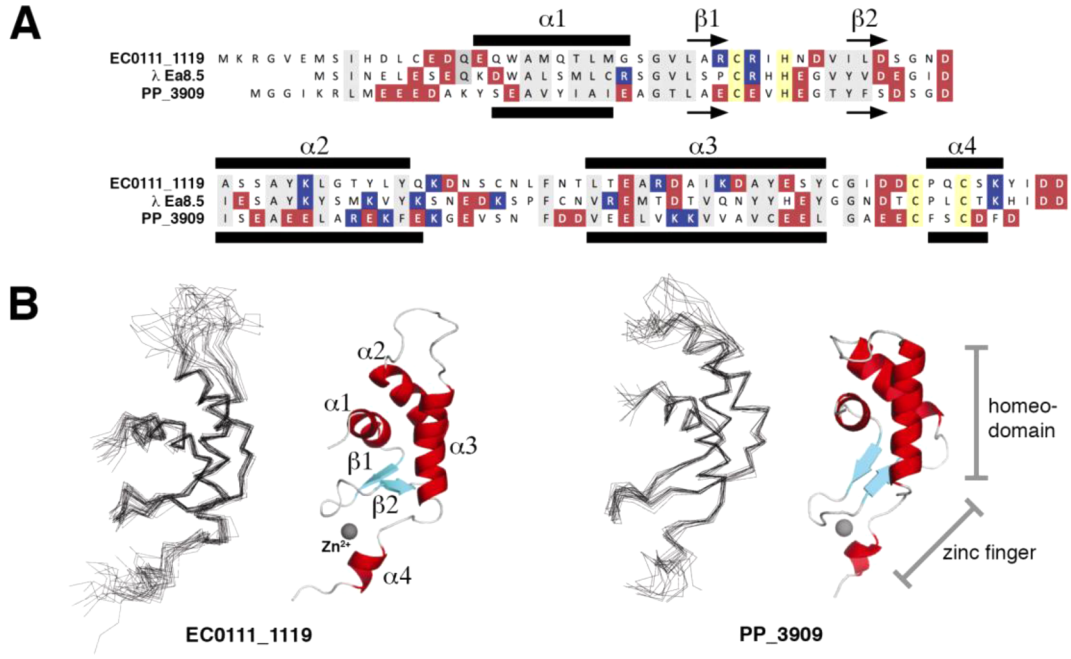

Figure 2. Sequence and structural highlights. (A) Sequence alignment of EC0111_1119, λ Ea8.5, and PP_3909. Gray boxes denote hydrophobic amino acids conserved at a given position; yellow boxes denote the amino acids that coordinate Zn(II), and red and blue boxes denote acidic and basic amino acids, respectively. Secondary structure elements for EC0111_1119 and PP_3909 are shown above and below the respective sequence. (B) Ensemble and ribbon diagrams of the two solution structures. The individual homeodomain and zinc binding motifs are indicated.

refined with CNS_SOLVE.[8] As shown in Figure 2B, the NMR structures are both characterized by an $\alpha1\alpha2\alpha3$ homeodomain-like fold tightly integrated with a $\beta1\beta2\alpha4$ zinc-finger fold in the original configuration. The Cα root-mean-square deviation of 2.5 Å between the two structures highlights the differences in the relative positioning of the secondary structural elements and the fact that only 14 amino acids are similar between the two proteins. Supported by missing resonances, the N-terminal 15 amino acids are unstructured, as well as the $\alpha2-\alpha3$ loop. The last few amino acids leading to the C-terminus were assigned in their entirety but were relatively sharper in line width and demonstrated no long-range nuclear Overhauser effects indicative of being unstructured. Evidence of zinc ion binding was corroborated by significant changes in the ^{15}N HSQC spectra of each protein upon addition of ethylenediaminetetraacetic acid (Figure S1 of the Supporting Information). As shown in Figure 3, a structural search of the Protein Data Bank (PDB) using PDBeFOLD[9] identified a similar CCCC-type zinc figure in the protein ZNF483 (PDB entry 2CTU).

A structure search of the EC0111_1119 and PP_3909 three-helix motifs identified a number of homeodomain-like DNA binding domains such as Aristaless[10] among the majority of hits as well as the RNA binding domain of ribosomal protein L11.[11] The major nucleic acid binding site on a homeodomain fold is localized to the third helix. A molecular surface comparison centered on the third helix is presented in Figure 4. Typically,

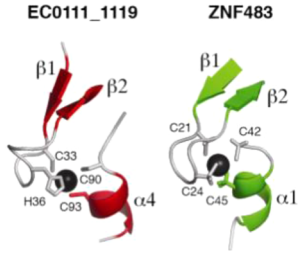

Figure 3. CHCC zinc finger of the Ea8.5 family that shares a similar geometry with the CCCC zinc finger of ZNF483.

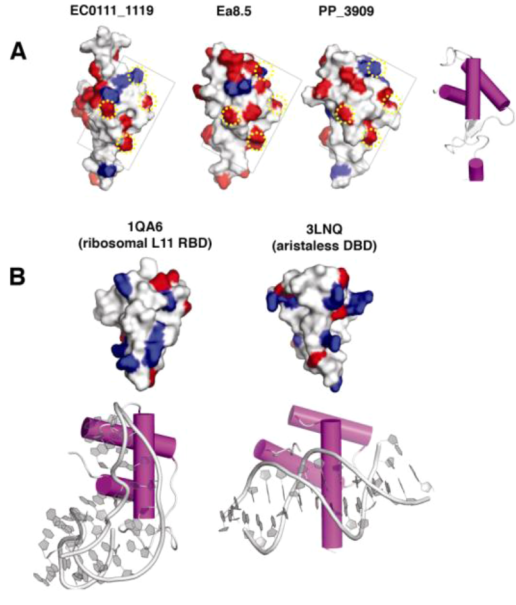

Figure 4. Surface comparisons. (A) Acidic and basic amino acids are colored red and blue, respectively. λ Ea8.5 is a molecular model. PDB entries are shown, where appropriate. The orientation of the surfaces places $\alpha3$ in front of the viewer, a cartoon diagram is presented as a visual aid. Conserved and similarly positioned charged amino acids are identified by yellow circles, and the perimeter of the surface is denoted with a box. (B) Two structurally similar RNA and DNA binding proteins demonstrate a more positively charged surface that is required for base-specific recognition.

dx.doi.org/10.1021/bi400543w | *Biochemistry* 2013, 52, 3612−3614

nucleic acid binding motifs derive their sequence specificity from basic amino acids. Among the Ea8.5 proteins, these amino acids are absent. Titration of ^{15}N-labeled PP_3909 with a 24 bp oligonucleotide from an unrelated study produced no apparent chemical shift changes with an up to 20-fold excess of DNA, suggesting that it does not bind DNA. It is worth noting that even nonspecific oligonucleotides can still bind some homeodomains with a reduced affinity.[12]

A closer inspection of the protein surfaces identifies some spatially conserved, charged amino acids (Figure 4). Together, they may define the boundary of a protein interaction site as opposed to a nucleic acid binding site. Indeed, a yeast two-hybrid screen of a λ protein library against itself identified Ea8.5 as its own protein partner in addition to the Int/NinB recombinase proteins, the Q antiterminator, and the Tfa/Stf tail fiber proteins.[13] With respect to Ea8.5 as a partner to itself, this protein did demonstrate concentration-dependent oligomerization. Resonance line broadening was also observed for EC0111_1119 and PP_3909 above 0.6 mM or at elevated temperatures, indicating that they also had the propensity to self-associate, albeit more weakly.

Overexpression of *ea8.5* produces fuzzy plaques that may be a morphological consequence of a sustained block on bacterial DNA replication.[1] The inability of the bacteria to initiate DNA replication may provide an opportunity for the bacteriophage to utilize resources and coordinate genome integration if the lysogenization program is favored. Possible bacterial targets of Ea8.5 may include DnaA and factors such as DiaA or Hda that regulate it, or nucleoid-associated proteins such as HU, Fis, and IHF that affect the architecture of the bacterial chromosome,[14] and transcription factors that act upon cII-dependent promoters invoved in the lysis versus lysogenization decision.

An investigation of the Ea8.5 protein rather than its orthologous representatives was hampered at the level of expression by the exclusive production of inclusion bodies and, upon refolding in zinc-containing buffers, limited solubility and a strong tendency to oligomerize. Using the sequence and structural homology as a guide, a variant of Ea8.5 (C19A, I39A, C62A, and Y74A) not only was expressed in the soluble fraction but also demonstrated improved solution characteristics, although not to the extent that an NMR solution structure could be pursued. Nevertheless, the variant provides a means of embarking on a proteomics survey to identify protein partners.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Detailed cloning, expression, and structure determination methods; structural statistics (Table S1); and supporting NMR spectra (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

Chemical shifts and coordinates of EC0111_0119 and PP_3909 were deposited in the BioMagResBank as entries 19178 and 19179, respectively, and in the PDB as entries 2M7A and 2M7B, respectively.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: logand@yorku.ca. Telephone: (416) 736-2100. Fax: (416) 736-5698.

Present Address

[§]F.E.: Agilent Technlogies, Mississauga, ON L5N 5M4, Canada.

Author Contributions

K.L.M. and L.W.D. conceived the work. J.J.K., E.S., S.K., F.E., and L.W.D. conducted and analyzed the experiments. L.W.D. wrote the manuscript.

Funding

L.W.D. is supported by a Discovery Grant and a Discovery Accelerator Award from the Natural Sciences and Engineering Research Council (NSERC). K.L.M. is supported by a Canadian Institutes for Health Research Operating Grant (FRN 62796).

Notes

The authors declare no competing financial interests.

■ ACKNOWLEDGMENTS

We thank David Kleinman for technical assisstance.

■ REFERENCES

(1) Łoś, J. M., Łoś, M., Wegrzyn, A., and Wegrzyn, G. (2008) *Folia Microbiol. (Dordrecht, Neth.) 53*, 443−450.

(2) Sergueev, K., Court, D., Reaves, L., and Austin, S. (2002) *J. Mol. Biol. 324*, 297−307.

(3) Biegert, A., and Söding, J. (2009) *Proc. Natl. Acad. Sci. U.S.A. 106*, 3770−3775.

(4) McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) *Bioinformatics 16*, 404−405.

(5) Edmonds, L., Liu, A., Kwan, J. J., Avanessy, A., Caracoglia, M., Yang, I., Maxwell, K. L., Rubenstein, J., Davidson, A. R., and Donaldson, L. W. (2007) *J. Mol. Biol. 365*, 175−186.

(6) Güntert, P. (2004) *Methods Mol. Biol. 278*, 353−378.

(7) Bordoli, L., and Schwede, T. (2012) *Methods Mol. Biol. 857*, 107−136.

(8) Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.

(9) Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. D60*, 2256−2268.

(10) Miyazono, K.-I., Zhi, Y., Takamura, Y., Nagata, K., Saigo, K., Kojima, T., and Tanokura, M. (2010) *EMBO J. 29*, 1613−1623.

(11) Conn, G. L. (1999) *Science 284*, 1171−1174.

(12) Iwahara, J., Zweckstetter, M., and Clore, G. M. (2006) *Proc. Natl. Acad. Sci. U.S.A. 103*, 15062−15067.

(13) Rajagopala, S. V., Casjens, S., and Uetz, P. (2011) *BMC Microbiol. 11*, 213.

(14) Mott, M. L., and Berger, J. M. (2007) *Nat. Rev. Microbiol. 5*, 343−354.