Expression and nucleotide sequence analysis of the *Legionella pneumophila* recA gene

Xun Zhao and Lawrence A. Dreyfus

Department of Microbiology, The University of Texas Medical Branch, Galveston, TX, U.S.A.

Received 23 March 1990
Revision received and accepted 4 April 1990

Key words: *Legionella pneumophila* recA gene; Nucleotide sequence

1. SUMMARY

The nucleotide sequence of the *L. pneumophila* recA gene was determined. The coding region was 1044 nucleotides (348 codons), specifying a 37,934 Da protein. Preceding the recA gene was a tandem set of transcription regulatory sequences and putative LexA binding sites. When expressed in *E. coli*, the cloned recA gene yielded two proteins with molecular weights of approximately 38,000 and 35,500 Da. The larger of these two proteins shared 70.4% and 74.6% identity with the *E. coli* and *Pseudomonas aeruginosa* RecA proteins, respectively. The 35,500 Da recA encoded protein was presumed to be the product of translation from Met26 which was preceded by an alternate ribosomal binding site.

2. INTRODUCTION

The bacterial family *Legionellaceae*, is a closely related group of organisms unlike other Gram-negative bacteria [1, 2]. The *Legionellaceae* type-species, *Legionella pneumophila*, is the etiologic agent of the serious and often fatal pneumonia termed Legionnaires disease [3]. To date, a growing list of *Legionella* species are placed in at least 33 recognized serogroups [4]. Recently, our laboratory reported the cloning and characterization of the *L. pneumophila* recA gene [5]. Like that of *E. coli*, the *L. pneumophila* recA gene was induced by UV treatment, promoted homologous recombination, and cross-reacted with antibodies raised against *E. coli* RecA protein [5]. Our previous findings prompted the present study to determine the sequence of the *L. pneumophila* recA gene and ascertain the molecular basis for the high degree of functional conservation between homologous genes from such widely divergent bacteria.

3. MATERIALS AND METHODS

3.1. Bacterial strains, bacteriophage, and plasmids

*E. coli* HB101 [6] was used as a host strain for recombinant plasmid DNAs. *E. coli* JM105 [7] was used as a host for M13-mp18 and M13-mp19 [8]. The previously described plasmid, pLAD111 [5] was the source of the *L. pneumophila* recA.

---

Correspondence to: Lawrence A. Dreyfus, Department of Microbiology, The University of Texas Medical Branch, Galveston, TX 77550, U.S.A.

0378-1097/90/$03.50 © 1990 Federation of European Microbiological Societies
gene. The plasmid derivatives of pT7-1 [9], pT7-5 and pT7-6, and pGP1-2 were obtained from Stan Tabor.

3.2. Media and growth conditions

LB medium [10] was used as a routine maintenance medium for *E. coli* strains. JM105 was maintained on M9 agar medium [10] supplemented with thiamine (1 µg/ml). JM105 was cultivated on 2X-YT medium [10] prior to the isolation of single stranded sequencing template DNA. HB101 strains harboring pGP1-2 and the appropriate pT7 vectors were grown at 30°C. All other bacterial cultivations were at 37°C. When appropriate, antibiotics were added to culture media at the following concentrations in µg/ml: carbenicillin (Cb), 100; kanamycin (Km), 25; rifampicin (Rif), 300.

3.3. Expression and metabolic labeling of the recA gene encoded products

A 1.7 kb BglII-EcoRI fragment of pLAD111, previously determined to encode functional recA activity [5], was subcloned into pT7-5 and pT7-6. The resulting plasmids, pT7-5rec and pT7-6rec, were transformed individually into *E. coli* HB101 (pGP1-2) for analysis by T7 RNA polymerase mediated expression [9]. During expression, T7-promoted gene products were labeled with [35S]methionine (> 1000 Ci/mmol (Trans35S-i-label, ICN Biochemicals, Inc., Costa Mesa, CA)). Expression samples were fractionated by SDS-PAGE on 10% (w/v) gels [11]. Separated proteins were electrotransferred to nitrocellulose [12] and exposed to X - OMAT AR film for 2 h at -70°C.

3.4. DNA sequencing and data analysis

DNA sequencing was performed on both strands by the dideoxynucleotide chain termination method of Sanger et al. [13]. Modified T7 DNA polymerase (Sequenase kit) were purchased from United States Biochemical Corp. (Cleveland, OH). Deoxyadenosine 5'-[α-35S]-thiotriphosphate (> 400 Ci/mmol) was purchased from Amersham Corp. (Arlington Heights, IL.). Synthetic oligonucleotides which primed synthesis within the subcloned *L. pneumophila* DNA were purchased from Bio-Synthesis Inc. (Denton, TX).

4. RESULTS

4.1. T7-Expression of the *L. pneumophila* recA gene

The results of T7 expression indicated that the 1.7 kb *L. pneumophila* insert encoded two protein products (Fig. 1C). The higher molecular weight form (38000 Da) corresponded almost exactly in size to the *E. coli* recA gene product; the lower molecular weight species (35500 Da) was assumed to correspond to a specific proteolytic cleavage of RecA [5]. This analysis also indicated that the recA gene was transcribed from the direction of the EcoRI site toward the BglII site, or as shown with pT7-5rec (Fig. 1A,C).

4.2. DNA sequence analysis

The *L. pneumophila* recA gene sequence and the encoded protein are shown in Fig. 2. The coding region of the recA gene was 1044 nucleotides from the amino terminal ATG triplet to the TAA termination codon. A search for transcription initiation sites similar to those of *E. coli* indicated that two complete sets of (-35)- and (-10)-like sequences preceded the coding region of the *L. pneumophila* recA gene. The 5′-most set of sequences began with a (-35)-like sequence at nucleotide -158 (TTGAAC) followed by a Pribnow heptamer [-14] at nucleotide -131 (TATAATC). In addition, a sequence homologous to a LexA binding site (SOS box) [-15], was located between the -35 and -10 sequences beginning at nucleotide -155 (Figure 2). A second (-35)-like sequence, which read GTGATA, was found at nucleotide -43 to -38. Downstream from the second -35 site was located a second Pribnow heptamer which read TATAATA. A second sequence homologous to a LexA binding site (SOS box) [-15], was located between the -35 and -10 sequences beginning at nucleotide -155 (Figure 2). A second (-35)-like sequence, which read GTGATA, was found at nucleotide -43 to -38. Downstream from the second -35 site was located a second Pribnow heptamer which read TATAATA. A second SOS box was located between the second set of promoter sequences at nucleotides -36 to -17. This second LexA binding site contained the highly conserved ..TTGT.........CAT.. motif found in several SOS genes of *E. coli* [-15]. A proposed Shine-Dalgarno sequence [-14] which was located immediately after the 3′-most Pribnow heptamer. A second possible ribosomal binding site was located within the coding region of the gene, just upstream of Met26 (Fig. 2).

The *L. pneumophila* recA gene terminated with a TAA ochre codon. The termination codon was
A B

Fig. 1. T7-expression of recA encoded proteins. (A) Schematic representation of plasmids pT7-5rec and pT7-6rec. The subcloned L pneumophila DNA fragment (EcoRI-BglII) is shown as a boxed area. The portion of the boxed area corresponding to the approximate borders of the recA gene are shown by the black area. (B) Coomassie blue stained gel following SDS-PAGE analysis of T7 expression analysis. Molecular mass markers (lane M) were: myosin H-chain, 219600; phosphorylase B, 100420; bovine serum albumin, 68030; ovalbumin, 42730; carbonic anhydrase, 27420; β-lactoglobulin; 18080; lysozyme, 14880. Lanes 1, 2, and 3 are the T7 expressed products from pT7-5rec, pT7-6rec and pT7-5, respectively. Panel C is an autoradiogram of an identical set of lanes as in panel B following electrotransfer to nitrocellulose membrane.

followed by a proposed stemmed loop composed of a 10 base 5′ stem beginning at nucleotide 1063, a 5 base loop, and a 3′-stem beginning at nucleotide 1078 and terminating at position 1087 (Fig. 2). Though not completely base paired (8/10 correct pairs), the hyphenated stem was G·C rich and had a free energy of formation of ΔG = −8 kcal [16]. The proposed termination loop was followed by a 10 base palindrome (ATAGCGCTAT) positioned one base pair 3′ to the loop end (Fig. 2).

4.3. Analysis of the encoded L. pneumophila recA protein

The L pneumophila recA gene encoded 348 amino acids which yielded a calculated molecular mass of 37934 Da. This predicted mass was consistent with the 38000 Da value obtained by SDS-PAGE of T7 expression products. An aligned comparison of the L pneumophila RecA protein and those of E. coli [17,18] and Pseudomonas aeruginosa [19] indicated that L pneumophila RecA shared 70.4% and 74.6% homology with these two proteins, respectively (Fig. 3). A consensus ATP binding site proposed for ATP binding proteins (R·····G··SGKT) [20] was present in all three RecA sequences. The L pneumophila ATP binding sequence began with Arg<sup>60</sup> and ended with Thr<sup>73</sup> (Fig. 3). Other conserved functional residues were Gly<sup>204</sup> which is involved
Fig. 2. Nucleotide sequence of the L. pneumophila recA gene and the deduced amino acid sequence. Transcriptional and translational initiation sites and proposed transcription termination structure are underlined. The proposed LexA binding sites (SOS-box) are indicated as double underlined. The nucleotide numbering begins with the first base of the ATG start codon.

in SOS induction activities, Glu^38 which is thought to be critical for recognition and binding of single stranded DNA, and Gly^160, mutation of which results in a totally non-functional RecA protein [21].

5. DISCUSSION

Work by a number of investigators has resulted in the cloning of recA genes from several bacteria including Rhizobium meliloti [22], P. aeruginosa [23], Vibrio cholera [24,25], Methyphilus methylotrophus [26], Thiobacillus ferroxidans [27], Shigella flexneri [28], Erwinia carotovora [28], Proteus vulgaris [28], Haemophil influenzae [29], Legionella pneumophila [5], Neisseria gonorrhoeae [30], Agrobacterium tumefaciens [31], and Vibrio anguillarum [32]. Complementation of the E. coli recA defect by such a wide variety of apparent gene homologues is an impressive example of functional conservation across a widely divergent group of bacteria. Of the cloned recA genes, relatively few have been sequenced [17,19,27,33], but the available data support the notion that the observed functional homology is a direct consequence of considerable DNA sequence relatedness. The L. pneumophila recA gene is no exception to this observation, in that it shares 70.4% and 74.6% protein sequence identity with E. coli and P. aeruginosa, respectively.

The previous finding [5] that the L. pneumophila recA gene was induced following UV treatment was corroborated by the identification of a LexA binding sequence 5' to the coding region of this gene. These data suggest that L. pneumophila possesses an inducible DNA repair mechanism similar to that of E. coli.

The identification of two proteins as products of T7 expression was not totally surprising in view
of previous minicell data [5]. The interpretation of this finding, however, must be reconsidered based on sequence data presented in this study. Previous results supporting the fact that the 38000 Da and 35 500 Da proteins are both products of the recA gene include the elimination of both products by a single transposon insertion, coordinate induction by UV treatment of L. pneumophila, and cross-reactivity of both products with antibody to E. coli RecA. Previously [5], it was assumed that the smaller species was a product of specific proteolytic degradation of the RecA protein. The proposed tandem transcription initiation signals and a potential translational start at Met26 suggest, however, that the 35 000 Da product may result from alternate translation of the recA message or, that a second mRNA molecule may be transcribed using the 3'-most promoter sequences with translation beginning at Met26. The resulting protein would have a molecular mass very close to that of the observed 35 500 Da protein. Based on known functional regions of RecA [21], there is no reason to assume that the truncated protein would not possess full RecA activity. Nonetheless, there is no apparent rationale for two recA encoded proteins nor is there currently any evidence of a functional role for the 35 500 Da species. Experiments are currently being planned to resolve the issue.

ACKNOWLEDGEMENTS

X. Z. is a recipient of a James W. McLaughlin Graduate Fellowship. This work was supported in part by Public Health Services grant 5R29 AI24684 from the National Institutes of Health.

REFERENCES