Gene, 152 (1995) 135–136

GENE 08475

The deduced *Vibrio cholerae* RecA amino-acid sequence*

(Cholera; nucleotide sequence; RecA protein; recombination)

Rebecca L. Margraf, Alberto I. Roca and Michael M. Cox

Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706-1569, USA

Received by R.E. Yasbin: 29 July 1994; Revised/Accepted: 20 August/23 August 1994; Received at publishers: 27 October 1994

SUMMARY

The nucleotide sequence of the *recA* gene of *Vibrio cholerae* (*Vc*) has been determined. The amino acid (aa) sequence of the protein product is very similar to other known RecA aa sequences. However, this sequence does not agree with a previously reported *Vc* RecA aa sequence [Ghosh et al., Nucleic Acids Res. 20 (1992) 372].

The sequence of the *recA* gene of *Vibrio cholerae* (*Vc*) strain 2740-80 was ascertained from a clone (pGP69A) that had previously been shown to complement a *recA E. coli* strain when tested for ultraviolet light resistance and P1 transduction (Goldberg and Mekalanos, 1986a). The deduced aa sequence is 354-aa long (38 254 Da). The *Vc* RecA aa sequence is 87% identical to *V. anguillarum* (*Va*) RecA (Tolmasky et al., 1992) and 79% identical to *E. coli* RecA (Horii et al., 1980). Most of the divergence was in the last 50 aa. This was expected because the C terminus of the RecA protein is the least well-conserved region among different bacterial species (Roca and Cox, 1990). The ATP-binding site (Walker et al., 1982) and the RecA signature sequence (Bairoch, 1991) are completely conserved. This work will aid the elucidation of RecA protein's role (Goldberg and Mekalanos, 1986b; Ketley et al., 1990; Kumar et al., 1994) in cholera, a diarrheal disease of humans.

A *Vc* 'RecA' aa sequence has been previously reported (Ghosh et al., 1992). We do not believe that the sequence reported by Ghosh et al. represents a RecA aa sequence for the following reasons: First, their reported *Vc* 'RecA' aa sequence is only 10% identical to the *E. coli* RecA aa sequence. Second, the *E. coli* 'RecA' aa sequence that was published as a comparison in their paper is not the *E. coli* RecA aa sequence from the GenBank database (J01672) taken from the literature (Horii et al., 1980). For example, their *E. coli* 'RecA' sequence had 26 additional aa at the C terminus. Third, the highly conserved ATP-binding domain and the RecA signature sequence were not present in either their *E. coli* or *Vc* 'RecA' aa sequences. Last, when we used their *Vc* 'RecA' aa sequence in a BLAST (Altschul et al., 1990) search to find related sequences in the GenBank database, the closest match was to a hypothetical ORF (L10328 o223) in the *E. coli* genome. The similarity of the Ghosh et al. (1992) 'RecA' aa sequence to this hypothetical ORF was also noted by Fujita et al. (1994). Therefore, we conclude that the previously reported sequence of the *Vc* 'RecA' protein (Ghosh et al., 1992) is something other than a RecA protein. Other groups have come to the same conclusion (De Mot et al., 1993; Lloyd and Sharp, 1993).

After this paper was submitted, it came to our attention that another *Vc* RecA aa sequence has been determined (Stroeher et al., 1994). There are 9 nt conflicts between the sequence in Fig. 1A and the Stroeher et al. (1994) sequence of the *Vc* strain O17 *recA* gene and flanking regions. We have checked and confirmed the *Vc recA* nt sequence as reported in Fig. 1A. Six of the nt conflicts reside within the ORF, resulting in 5 aa conflicts at aa positions 27, 49, 52, 64 and 94 as shown in Fig. 1B (one nt conflict is silent). At these 5 aa positions, the Stroeher et al. (1994) *Vc* RecA aa sequence is not consistent with the *Va* RecA aa sequence or a multiple sequence alignment of 15 other bacterial RecA aa sequences (Roca and Cox, 1990). After the 6 nt conflicts within the ORF of the Stroeher et al. (1994) *recA* nt sequence were reexamined, the first 5 (corresponding to aa positions 27, 49, 52, 56 and 64) were corrected to agree with the nt sequence

Correspondence to: Dr. M.M. Cox, Department of Biochemistry, University of Wisconsin-Madison, 420 Henry Mall, Madison, WI 53706-1569, USA. Tel. (1-608) 262-1181 or 262-27982; Fax (1-608) 265-2603; e-mail: coxlab@macc.wisc.edu

*On request, the authors will supply detailed experimental evidence for the conclusions reached in this Brief Note.

Abbreviations: aa, amino acid(s); BLAST, basic local alignment search tool; *E.*, *Escherichia*; nt, nucleotide(s); *recA*, gene encoding RecA; *V.*, *Vibrio*; *Va*, *V. anguillarum*; *Vc*, *V. cholerae*.

SSDI 0378-1119(94)00686-5

A

```
TTCTATACAGGGGTGGACACTGTACAAATCAACAGTATAATGAACCCCAACATGACAAAT   60
           -35                    -10
                LexA box
CAGCGTCGACCCGATTTCGACACAATTACCTATCACCTGAGAGTGACTATCCGGAGAGAG   120
                                                     RBS
TAATGGACGAGAATAAACAGAAGGCACTTGCGGCCGCCCTCGGTCAAATTGAAAAGCAAT   180
   M  D  E  N  K  Q  K  A  L  A  A  A  L  G  Q  I  E  K  Q     19
TTGGTAAAGGCTCCATCATGCGCTTAGGCGACAACCGCGCGATGGATGTAGAAACCATTT   240
F  G  K  G  S  I  M  R  L  G  D  N  R  A  M  D  V  E  T  I     39
CGACCGGTTCTCTGTCTCTGGATATCGCGTTGGGTGCTGGTGGTCTACCAATGGGACGTA   300
S  T  G  S  L  S  L  D  I  A  L  G  A  G  G  L  P  M  G  R     59
TCGTTGAGATCTTCGGCCCTGAATCTTCGGGTAAAACGACACTGACTCTGGAACTGATTG   360
I  V  E  I  F  G  P  E  S  S  G  K  T  T  L  T  L  E  L  I     79
CTGCTGCACAACGTGAAGGTAAAACCTGTGCGTTTATCGATGCCGAGCACGCACTGGATC   420
A  A  A  Q  R  E  G  K  T  C  A  F  I  D  A  E  H  A  L  D     99
CGGTTTATGCCAAAAAATTGGGCGTGAATATCGATGAGCTACTGGTTTCTCAGCCAGATA   480
P  V  Y  A  K  K  L  G  V  N  I  D  E  L  L  V  S  Q  P  D     119
CCGGCGAGCAAGCACTGGAAATTTGTGATGCACTGGCTCGCTCTGGTGCTGTGGATGTCA   540
T  G  E  Q  A  L  E  I  C  D  A  L  A  R  S  G  A  V  D  V     139
TCGTTGTTGACTCAGTAGCGGCTTTGACACCAAAAGCGGAAATTGAAGGCGAAATGGGCG   600
I  V  V  D  S  V  A  A  L  T  P  K  A  E  I  E  G  E  M  G     159
ATAGCCACATGGGTCTGCAAGCGCGTATGTTGTCGCAAGCAATGCGTAAACTGACGGGTA   660
D  S  H  M  G  L  Q  A  R  M  L  S  Q  A  M  R  K  L  T  G     179
ACCTCAAGCAATCCAACTGTATGTGTATCTTCATCAACCAAATTCGTATGAAGATTGGTG   720
N  L  K  Q  S  N  C  M  C  I  F  I  N  Q  I  R  M  K  I  G     199
TGATGTTTGGTAACCCAGAAACTACCACTGGCGGTAACGCACTGAAATTCTACGCTTCTG   780
V  M  F  G  N  P  E  T  T  T  G  G  N  A  L  K  F  Y  A  S     219
TTCGTTTGGATATTCGCCGTACTGGCGCAATCAAAGAAGGCGAAGAAGTGGTGGGTAACG   840
V  R  L  D  I  R  R  T  G  A  I  K  E  G  E  E  V  V  G  N     239
AAACCCGTATCAAGGTTGTGAAGAACAAGATTGCTGCGCCGTTTAAAGAAGCGAACACTC   900
E  T  R  I  K  V  V  K  N  K  I  A  A  P  F  K  E  A  N  T     259
AGATCATGTATGGCCAAGGTTTCAATCGTGAAGGTGAATTGATCGATCTTGGTGTGAAAC   960
Q  I  M  Y  G  Q  G  F  N  R  E  G  E  L  I  D  L  G  V  K     279
ACAAAATGGTTGAGAAGTCAGGCGCTTGGTACAGCTATAACGGCGATAAAATCGGCCAAG   1020
H  K  M  V  E  K  S  G  A  W  Y  S  Y  N  G  D  K  I  G  Q     299
GTAAAGCGAATGCCTGCAAGTATCTGAAAGAAAACCCAGAAATTGCGAAAACTCTAGATA   1080
G  K  A  N  A  C  K  Y  L  K  E  N  P  E  I  A  K  T  L  D     319
AAAAACTGCGTGAAATGCTGTTAAACCCAGAAAACATGCAGCTGATCGCTGAAACATCTT   1140
K  K  L  R  E  M  L  L  N  P  E  N  M  Q  L  I  A  E  T  S     339
CTGCGGCGGATGACGTCGAATTTGGCGCGGTGCCAGAAGAGTTTTAATCGGCAGGCTGAA   1200
S  A  A  D  D  V  E  F  G  A  V  P  E  E  F  .                 354
```

B

```
Vcho  --MDENKQKALAAALGQIEKQFGKGSIMRLGDNRAMDVETISTGSLSLDIALGAGGLPMG  58
Vc94  --...........................H......................T..S......  58
Vang  --...............................T.........................  58
Ecol  MAI.........................ED.S...........................  60
Vc92  -MLLVRRP.R..WWILPVCDAI.LAVFVGI.VEK.LIYQDSALIAIIMGVIT.C..GIIR  59

            ATP-binding
Vcho  RIVEIFGPESSGKTTLTLEL--------IAAAQREGKTCAFIDAEHALDPVYAKKLGVNI 110
Vc94  .....L..............--------..............T................. 110
Vang  ....VY..............--------......V...............I......... 110
Ecol  .....Y.............QV--------....................I..R...D. 112
Vc92  DVLAREI.MVLRSEVYATAVLWWHFNYAL.MGYSSSRGT.LLGGFFHALNSSRSDPVAPF 119

Vcho  DELLVSQPDTGEQALEICDALARSGAVDVIVVDSVAALTPKAEIEGEMGDSHMGLQARML 170
Vc94  ............................................................ 170
Vang  .........................I....I............................. 170
Ecol  .N..C.................................I.......A...M 172
Vc92  SAN.CPDQI..SAPNAP.EVPLSQKSPSYLKLS.LCE.NARPRLRFHITST.-------- 172

                                               signature
Vcho  SQAMRKLTGNLKQSNCMCIFINQIRMKIGVMFGNPETTTGGNALKFYASVRLDIRRTGAI 230
Vc94  ............................................................ 230
Vang  .........................................................S. 230
Ecol  .......A......TLL.....................................I..V 232
Vc92  ---LFPFVI.TCY.KALKPYWCSAKE.TMSRT.SQLYISQSAVS.RI.NLEKKLSKKLIV 228

Vcho  KEGEEVVGNETRIKVVKNKIAAPFKEANTQIMYGQGFNREGELIDLGVKHKMVEKSGAWY 290
Vc94  ............................................................ 290
Vang  ...D.A...................Q.D...L...........V.......L...A.... 290
Ecol  ....N...S...V............Q.EF..L..E.I.FY...V.....E.LI..A.... 292
Vc92  PA.RHIKLTADAEQIIASIGPTFNELHGLIFEQQTLEDNTLLGM.CSETLVAGYLGTCRH 288

Vcho  SYNGDKIGQGKANACKYLKENPEIAKTLDKKLREMLLNPENMQLIAETSSAADDVEFGAV 350
Vc94  ............................................................ 350
Vang  ................F.R...AA.MA..T.........AELIVEEPIL.EMPQE.EL   348
Ecol  ..K.E.........TAW..D...T..EIE..V..L..SNP.STPDFSVDDSEGVA.TNED 352
Vc92  ASLALARRW.NIARSQARSHWVITTNHRNASI.TLRVITT.HTPRIVENVQSGKATL.FC 348

Vcho  PEEF                                                         354
Vc94  ....                                                         354
Vang                                                               348
Ecol  F                                                            353
Vc92  AGYLPANPFTVLAKRRCTHCQVSLSPMI                                 376
```

Fig. 1. Nucleotide sequence of the *V. cholerae recA* gene and a comparison of the deduced aa sequence with other RecA proteins. (**A**) The sequence was determined by the dideoxy technique using [^{35}S]dATP on double-stranded plasmids pGP69A and pRLM2 (a deletion derivative of pGP69A). Forward and reverse universal primers and designed sequence-specific primers were used. Both strands of the gene were sequenced. The putative *lexA* operator (repressor-binding site), −10 and −35 regions, and ribosome-binding site (RBS) are indicated. GenBank accession No. U10162. (**B**) Comparison of the RecA proteins of *V. cholerae* (Vcho), *V. cholerae* of Stroeher et al. (1994) (Vc94 X71969), *V. anguillarum* (Vang M80525), *E. coli* (Ecol J01672) and 'RecA' of *V. cholerae* (Vc92 X61384, Ghosh et al., 1992). Residues identical to Vcho are represented by a dot (·) and gaps are depicted by a dash (-). The ATP-binding site and signature sequence are shaded.

of Fig. 1A. The remaining conflict (corresponding to aa^{94}) remained uncertain (P.A. Manning, personal communication).

The *Vc recA* clone (pGP69A) was a generous gift from John J. Mekalanos. This work was supported by NIH Grant GM32335 to M.M.C. and a Mary Shine Peterson undergraduate scholarship to R.L.M.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.: Basic local alignment search tool. J. Mol. Biol. 215 (1990) 403–410.

Bairoch, A.: PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. 19 (1991) 2241–2245.

De Mot, R., Laeremans, T., Schoofs, G. and Vanderleyden, J.: Characterization of the *recA* gene from *Pseudomonas fluorescens* OE 28.3 and construction of a *recA* mutant. J. Gen. Microbiol. 139 (1993) 49–57.

Fujita, N., Mori, H., Yura, T. and Ishihama, A.: Systematic sequencing of the *Escherichia coli* genome: analysis of the 2.4–4.1 min (110,917–193,643 bp) region. Nucleic Acids Res. 22 (1994) 1637–1639.

Ghosh, S.K., Biswas, S.K., Paul, K. and Das, J.: Nucleotide and deduced amino acid sequence of the *recA* gene of *Vibrio cholerae*. Nucleic Acids Res. 20 (1992) 372.

Goldberg, I. and Mekalanos, J.J.: Cloning of the *Vibrio cholerae recA* gene and construction of a *Vibrio cholerae recA* mutant. J. Bacteriol. 165 (1986a) 715–722.

Goldberg, I.S. and Mekalanos, J.J.: Effect of a *recA* mutation on cholera toxin gene amplification and deletion events. J. Bacteriol. 165 (1986b) 723–731.

Horii, T., Ogawa, T. and Ogawa, H.: Organization of the *recA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 77 (1980) 313–317.

Ketley, J.M., Kaper, J.B., Herrington, D.A., Losonsky, G. and Levine, M.M.: Diminished immunogenicity of a recombination-deficient derivative of *Vibrio cholerae* vaccine strain CVD103. Infect. Immun. 58 (1990) 1481–1484.

Kumar, K.K., Srivastava, R., Sinha, V.B., Michalski, J., Kaper, J.B. and Srivastava, B.S.: *recA* mutations reduce adherence and colonization by classical and El Tor strains of *Vibrio cholerae*. Microbiology 140 (1994) 1217–1222.

Lloyd, A.T. and Sharp, P.M.: Evolution of the *recA* gene and the molecular phylogeny of bacteria. J. Mol. Evol. 37 (1993) 399–407.

Roca, A.I. and Cox, M.M.: The RecA protein: structure and function. Crit. Rev. Biochem. Mol. Biol. 25 (1990) 415–456.

Stroeher, U.H., Lech, A.J. and Manning, P.A.: Gene sequence of *recA*$^{+}$ and construction of *recA* mutants of *Vibrio cholerae*. Mol. Gen. Genet. 244 (1994) 295–302.

Tolmasky, M.E., Gammie, A.E. and Crosa, J.H.: Characterization of the *recA* gene of *Vibrio anguillarum*. Gene 110 (1992) 41–48.

Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J.: Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1 (1982) 945–951.