

# Sequencing and the alignment of structural genes in the *nqr* operon encoding the Na<sup>+</sup>-translocating NADH-quinone reductase from *Vibrio alginolyticus*

Maki Hayashi\*, Kouji Hirai, Tsutomu Unemoto

Laboratory of Membrane Biochemistry, Faculty of Pharmaceutical Sciences, Chiba University, 1–33 Yayoi-cho, Inage-ku, Chiba 263, Japan

Received 27 February 1995

**Abstract** We previously cloned a part of *nqr* operon encoding the Na<sup>+</sup>-translocating NADH-quinone reductase (NQR) from the marine *Vibrio alginolyticus* [Hayashi et al., FEBS Lett. 356 (1994) 330–332]. From its nucleotide sequences, four consecutive open reading frames (ORF) encoding the  $\gamma$ -subunit (27.7 kDa), two unidentified ORFs of 22.6 kDa and 21.5 kDa, and the  $\beta$ -subunit (45.3 kDa) were recognized. The gene encoding the  $\alpha$ -subunit was located upstream, and together with the recent report by Beattie et al. [FEBS Lett. 356 (1994) 333–338], the *nqr* operon was found to be constructed from six consecutive structural genes, where *nqr1*, *nqr3* and *nqr6* correspond to the  $\alpha$ -,  $\gamma$ -, and  $\beta$ -subunits, respectively, of the NQR complex.

**Key words:** Na<sup>+</sup> pump; NADH-quinone reductase; Nucleotide sequence; *nqr* operon; *Vibrio alginolyticus*

## 1. Introduction

Gram-negative marine and moderately halophilic bacteria have a unique Na<sup>+</sup>-translocating NADH-quinone reductase (NQR) complex as a first segment of the respiratory chain [1–5]. The enzyme complex purified from the marine *Vibrio alginolyticus* was composed of three subunits,  $\alpha$ ,  $\beta$ , and  $\gamma$  [6]. Recently we have cloned a part of *nqr* operon that contains the structural genes for the  $\gamma$ -subunit and  $\beta$ -subunit [7]. At the same time, Beattie et al. reported the presence of four consecutive open reading frames in the *nqr* operon containing the structural genes for the  $\alpha$ -subunit and  $\gamma$ -subunit [8]. Our sequence data overlapped with those reported by Beattie et al. [8] at the structural gene for the  $\gamma$ -subunit. Thus we could construct the alignment of structural genes in the *nqr* operon. This paper reports the sequencing of the latter half of the *nqr* operon and the complete alignment of 6 ORFs in the *nqr* operon.

## 2. Materials and methods

### 2.1. Materials

The subclone, p3Cl, was prepared as described in [7]. The probe DNA was labeled with digoxigenin-dUTP with the DIG DNA labeling kit (Boehringer-Mannheim Biochemica), and was used for the Southern hybridization.

### 2.2. Analytical methods

Subcloning, restriction mapping, the preparation of template DNA, and other standard methods were performed as described in [9]. DNA

was sequenced by the dideoxy method using Sequenase Version 2.0 DNA Sequencing Kit (Amersham). The primers used for sequencing were synthesized with Oligo1000 DNA Synthesizer (Beckman). A part of DNA was sequenced by an Automated Fluorescent DNA Sequencer (Shimazu DSQ-1000, Kyoto). Sequences were analyzed using the GEN-ETIX-Mac.

## 3. Results

### 3.1. Sequencing of *nqr* genes

Fig. 1 shows the restriction maps and the strategies of nucleotide sequencing of p3Cl. The structural genes were renamed using numbers as discussed below. The nucleotide and deduced amino acid sequences of the cloned *nqr* genes are shown in Fig. 2, where 4,586 bps were sequenced. Four open reading frames (ORF) were detected. The first ORF (*nqr3*) starting from the nucleotide sequence of 965 had the predicted N-terminal and intermediate amino acid sequences determined from the  $\gamma$ -subunit [7]. Its sequence has been reported in [10] and registered under Accession Number D43958. It was composed of 256 amino acid residues with the  $M_r$  value of 27,703 Da, which was close to the experimental value of 32 kDa [6] determined by SDS-PAGE gel electrophoresis. The nucleotide sequence of the first ORF was in agreement with that of the *nqrC* gene reported by Beattie et al. [8] except for one point. In our nucleotide sequence, 1,513 and 1,514 were read as AT (Fig. 2), and thus the predicted amino acid residue was converted from R to W. The second ORF (*nqr4*) encoded 210 amino acid residues, which was in complete agreement with that of *nqrD*. Undoubtedly, our sequence data overlapped with that of Beattie et al. [8] at the first and the second ORFs. The third ORF (*nqr5*) encoded 198 amino acid residues. This protein was not identified in the purified NQR complex.

The fourth ORF (*nqr6*) encoded 407 amino acid residues with 45,274 Da, which was close to the value of 46 kDa for the  $\beta$ -subunit [6]. We previously prepared the probe B for the  $\beta$ -subunit, but it was inappropriate for the cloning of *nqr* operon [7]. By cloning a gene from the genome of *V. alginolyticus* with the probe B and sequencing its subclone, it was found that the amino acid sequence predicted an outer membrane maltoporin-like protein (unpublished results). Thus, the amino acid sequence for the  $\beta$ -subunit was incorrect. The region of the fourth ORF was confirmed to be indispensable for the expression of the  $\beta$ -subunit [7]. Moreover, the deduced amino acid sequence contained a common structure for NADH- or FAD-binding site. Thus, this ORF was assigned to be the structural gene for the  $\beta$ -subunit. Downstream of this ORF was a region of dyad symmetry followed by a T-rich region (Fig. 2), suggesting the end of *nqr* operon.

\*Corresponding author. Fax: (81)(43) 255 1574.

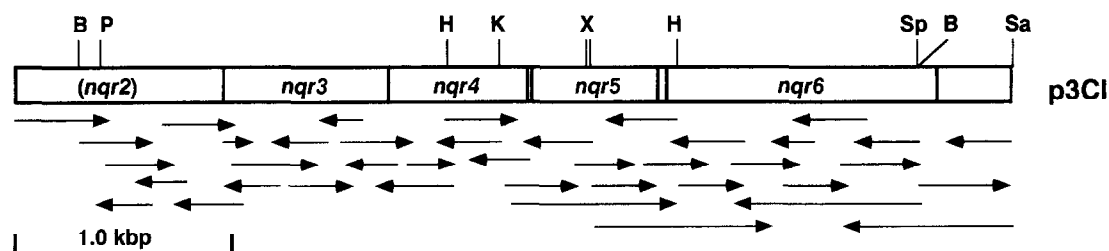


Fig. 1. Restriction maps and sequencing strategies of p3Cl. B, *Bgl*II; H, *Hind*III; K, *Kpn*I; P, *Pst*I; Sa, *Sal*I; Sp, *Sph*I; X, *Xba*I.

Table 1  
Properties of four structural genes in the *nqr* operon

	<i>nqr3</i>	<i>nqr4</i>	<i>nqr5</i>	<i>nqr6</i>
Nucleotide residue	965–1,735	1,735–2,367	2,375–2,971	3,005–4,228
Amino acid residue	256	210	198	407
$M_r$ (Da)	27,703	22,602	21,540	45,274
pI	4.38	9.17	7.00	4.55
Corresponding subunit	$\gamma$	unknown	unknown	$\beta$
Membrane-spanning helix	1	6–7	6	1–2

#### 4. Discussion

We previously used *nqrA*, *B* and *C* to call the structural genes for the  $\alpha$ -,  $\beta$ - and  $\gamma$ -subunits [7]. Beattie et al. reported the sequencing of four consecutive structural genes of *nqr* operon from *V. alginolyticus* and these genes were named as *nqrA*, *B*, *C* and *D* [8]. Among them, *nqrA* and *C* were identified as the structural genes for the  $\alpha$ - and  $\gamma$ -subunits, respectively. The predicted N-terminal amino acid sequences coincided with our previous data [7]. Moreover, the nucleotide sequences of *nqrC* and *D* reported by Beattie et al. [8] coincided with those of the first and the second ORFs in the present paper. It is clear that the latter half of the *nqr* operon is sequenced in this paper. To avoid confusion, we recommend renaming the structural genes of the *nqr* operon with numbers starting from *nqrA*. Table 1 summarizes the properties of four structural genes in the *nqr* operon. In combination with the data of Beattie et al. [8], the *nqr* operon is apparently constructed from 6 ORFs, where *nqr1*, *nqr3* and *nqr6* encode the  $\alpha$ -,  $\gamma$ - and  $\beta$ -subunits, respectively. The proteins encoded by *nqr4* and *nqr5* are very hydrophobic and have several predicted membrane-spanning helices. The amino acid sequences for proteins Nqr3–6 are not closely related to any subunits of  $H^+$ -translocating NADH-quinone reductase from prokaryotes [11,12] and eukaryotes [13,14], suggesting a unique structure of the  $Na^+$ -translocating NQR complex.

Our purified NQR complex was composed of three subunits,  $\alpha$ ,  $\beta$  and  $\gamma$  [6]. Using ubiquinone-1 as a substrate, the three subunits were essential for the  $Na^+$ -dependent NADH-quinone reductase activity. Therefore, the presence of three subunits is a minimum requirement for the catalytic activity. The role of subunits encoded by *nqr2*, *nqr4* and *nqr5* is unknown at present, and it is necessary to confirm their roles by reconstitution experiments.

**Acknowledgements:** A part of this work was supported by a Grant-in-Aid from the Japanese Ministry of Education, Science and Culture.

#### References

- [1] Hayashi, M. and Unemoto, T. (1984) *Biochim. Biophys. Acta* 767, 470–478.
- [2] Unemoto, T. and Hayashi, M. (1989) *J. Bioenerg. Biomembr.* 21, 649–662.
- [3] Unemoto, T. and Hayashi, M. (1993) *J. Bioenerg. Biomembr.* 25, 385–391.
- [4] Unemoto, T., Tokuda, H. and Hayashi, M. (1990) in: *The Bacteria*, Vol. XII, Bacterial Energetics (Krulwich, T.A., Ed.) pp. 33–54, Academic Press, New York.
- [5] Unemoto, T., Akagawa, A. and Hayashi, M. (1993) *J. Gen. Microbiol.* 139, 2779–2782.
- [6] Hayashi, M. and Unemoto, T. (1987) *Biochim. Biophys. Acta* 890, 47–54.
- [7] Hayashi, M., Hirai, K. and Unemoto, T. (1994) *FEBS Lett.* 356, 330–332.
- [8] Beattie, P., Tan, K., Bourne, R.M., Leach, D., Rich, P.R. and Ward, F.B. (1994) *FEBS Lett.* 356, 333–338.
- [9] Sambrook, J., Fritsch, F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [10] Hayashi, M. and Unemoto, T. (1994) *EBEC Short Reports*, vol. 8, pp. 77, Valencia, Spain.
- [11] Xu, X., Matsuno-Yagi, A. and Yagi, T. (1991) *Biochemistry* 30, 6422–6428.
- [12] Weidner, U., Geier, S., Ptock, A., Friedrich, T., Leif, H. and Weiss, H. (1993) *J. Mol. Biol.* 233, 109–122.
- [13] Pilkington, S.J., Skehel, J.M., Gennis, R.B. and Walker, J.E. (1991) *Biochemistry* 30, 2166–2175.
- [14] Preis, D., Weidner, U., Conzen, C., Azevedo, J.E., Nehls, U., Röhlen, D., Van der Pas, J., Sackmann, U., Schneider, R., Werner, S. and Weiss, H. (1991) *Biochim. Biophys. Acta* 1090, 133–138.

→  
Fig. 2. The nucleotide and deduced amino acid sequences of the *nqr* genes. Each open reading frame was named as *nqr3*, *4*, *5* and *6* as shown in Table 1. Boxed amino acid sequences indicate the N-terminal and intermediate sequences determined for the  $\gamma$ -subunit. Underlined nucleotide sequence denotes putative Shine–Dalgarno sequences. Arrows represent a region of dyad symmetry adjacent to a poly(dT) region. The data are submitted under the Accession Number D49364.

10 20 30 40 50 60 70 80 90 100 110 120  
CAACTAGCGACGATTATCTCGTAACTGGCACTACTGGTAACTGAATCTGGCGGAACGATTGCCGTGATGCAGGGGTTGGCAGTAAGTGTCTAGGTGGCACTACTCTCTA  
130 140 150 160 170 180 190 200 210 220 230 240  
CCTATCTACGCAACGATTCTCTGTTGGTGGTTCGGGAAGTCTTTCTGATGGTGGTGAAGCAGGAAGAAAGGCTCTTTGTTACTCTATCTTGTTCGACCTTACTGTT  
250 260 270 280 290 300 310 320 330 340 350 360  
CCACCAACGCTACCTCTATGGCAAGGGCACTAGGTATCACTATGGTGTGTTGTTGGCAAGGAGATCTCGGTGGTACAGGTCTTAACTCTTAACTCTGACCTTGCAGGTCTGCA  
370 380 390 400 410 420 430 440 450 460 470 480  
TTCCTATTCTTCGCTACCGCAGCAAACTCTCGGTGACGTAGTGAAGTGCAGCTGACGGCTTCTGGTGAACCTCTTACCAAGGGCTCAAGGTGGTAAAGGCTGCTAGTT  
490 500 510 520 530 540 550 560 570 580 590 600  
AACACAGTAACTGGTCTCTCTACTCTGGTGAAGCTTCTCGGTAACTCCCTGGCTCTATTGGTGAAGTATCACTCTAGCTCTAATGATCGGTGCAGCGATGCTGTTACATG  
610 620 630 640 650 660 670 680 690 700 710 720  
CGAATCGCTTCATGGCGCATTCATGGCGGCGTAATGATCGGTATGATTCAGTATCACTCTGTTAACTGATCGGTCTGACATAACCAATGTTCAACATGCCATGGCAGTGGAC  
730 740 750 760 770 780 790 800 810 820 830 840  
CTAGTCTAGCTGGTTCATTGCGGTATCTCTCATGGCAACGACCTAGTATCAGCTCATTTACTAAACAAAGGTAAGTGGTGTACGGCATCTGATGGTGAATGTTGTGTGATG  
850 860 870 880 890 900 910 920 930 940 950 960  
ATCCGTGTAGTAAACCGGCTACCAAGAGTATGATGCTGGCGATTCTATTGGCACTATTGCACTCTGTTGCACATGTAGTGATTGAGAAGAACATCAAGCGAGACTAGCA

970 980 990 1000 1010 1020 1030 1040 1050 1060 1070 1080  
CGCTATGGCAATAAAGCAGCAGCATTAAAGAGCGCTGGTCTTGTATCGGGTGGGCTTGTCTTCAATCATCGTATCAACAGCAGCAGTGGTCTGCTGATAAGCAAAAGC  
nqr3 M A S N N D S I K K T L G V V I G L S L V C S I I V S T A A V G L R D K Q K A  
1090 1100 1110 1120 1130 1140 1150 1160 1170 1180 1190 1200  
TAACGCTGTCTAGATAAGCAATCAAGATCTTGAAGTTCAGGCAATGAGCGAAGCGTAAGAAAGTACCAAGCATTGCTGAGTACGAACTCTGCTGGTAGCCTTGAAC  
N A V L D K V E V A G I D A N G K K V P E L F A E Y I E P R L V D L E T  
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 1310 1320  
AGTAACCTTCTAGGCGTATGCATCTAGCTGATCAGCGTGAAGCATCAAAAGTGCAGAAAGTCAATCGATTGACACCAAGAAAGTGTGCTGATATCCGTCTGCTGGCAG  
G N F T E G N A S T Y D Q R E A S K D A E R S I A L T P E E D V A D I R R R A N  
1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440  
TACAGCATGGTGTACTAGTAAAGACCAAGTGAAGTCAAAAGTATCTTGCCTATGCAGCGTAAGGACTTGGTGTGATGATGCTTGTAGCCGTGAAATCATGATGATA  
T A V V V L V K D Q D E V K V I L P M H G K G L W S M M Y A F V A V E T D G N  
1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560  
CACTGTGCTGTCTATTACTAGCAGCAAGGTGAACTCTGCTGACTTGGTGAAGTGAAGAACTCTTATGGCGCATCAATTCATTGGCAAGAAATGTCAATGAAGTCAATCA  
T V S A I T Y Y E Q G E T P G L G G E V E N P S W R D Q F I G K K L Y N E D H Q  
1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680  
GCCAGCGATTAAAGTGTGAAGCGCGCGCAGCGTCTGAGCAGCGTGTGATGGCTATCAGGAGCAAGCTCACTAGTAAGTGGTGTCAACACACATTTGACTCTGGTAGG  
P A I K V V K G G A P Q G S E H G V D G L S G A T L T S N G V O H T F D W L G  
1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800  
TGACAGGGCTTGGTCTTCTAGCAAAAGTCTGACGGAGAACTTAATGATGTCAGTGCACAAAGCTTAAAGGAGCATTTAGCGCCAGTATTGGATAACCAACCAATCCGG  
D K G F G P F L A K V R D G E L N \*

nqr4 M S S A Q N V K K S I L A P V L D N N P I A  
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920  
CTACAGTCTTGGTGTATGTTCTGCTCTGAGTAACAACTAACTAGAAACAGCTTTGTAATGAGCTAGCGGTAACTTGTAACTGGCTGTCTAACTCTCTGTTTCATTGATC  
L Q V L G V C S A L A V T T K L E T A F V M T L A V T F V T A L S N F S V L I  
1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040  
CGTAACCATCTCTAAGCGTGGTATCATCGTTCAGTGGCAATTCGCTAGTAATCGTGGTAGCAGGTGCTAAAGCTTACTATACGATATCTCTAAACAGTATCT  
R N H I P N S V R I I V Q M A I I A S L V I V V D Q V L K A Y L Y D I S K Q L S  
2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160  
GTATTCGTTGGTGTGATTACTCACTAATCTGATGTAAGTGGTCTGCTGAAGCATTGCAATGAAGTCTGCGCAGTACCATCTCTTATTGACGGTATGGTAACGGTCTAGGTACGGT  
V F V G L I I T N C I V M G R A E A F A M K S A P V P S L I D G I G N G L Y G  
2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280  
TTCGTTCTTCTCACTGAGGTTTCTCGGTGAGTATTTGGCTCAGGCAAGCTATTTGGCTAGAAAGTCTACCTCTAGTGAGCAACGGTGGTGTGACCAAGCTTAAGCGCTTAAGCTA  
F V L I T V G F F R E L F G S G K L F G L E V L P L V S N G G W Y Q P N G L M L  
2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400  
CTAGCAGCATCAGGATCTTCTAATCGGCTCTTATCTGGGTAACTCGTATCTGAAACGAGAAAGTAGAAGCGAAGGAGTAAAGGACGTGATGGAACATTACATTAGTCTGTAGT  
L A P S A F F L I G F L I W V I R I L K P E Q V E A K E \*

nqr5 M E H Y I S L L V  
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520  
TAAATGATTTTCATCAAAACATGGCTCTGTTCTTCTAGTATGTGACTTCTTGGCTATCAAGAAAGTTAAGACCTTTTGGCTAGGTGTTGAGTGTGAGTGTG  
K S I F I E N M A L S F F L G M C T F L A V S K K V K T S F G L G V A V V V V L  
2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640  
GACTATCGTCTTCTGGAACCACTAGTATACCACTAGTTCGAGAGAGAAACGCTTGTAGTGAGGCGGTGACCTAGCTTCTTAACTTATCATCACCTTATCGGTGAATGTCAGC  
T I A V P N N L V Y N L V R E N A L V E G V D L S F L N F I T F I G V I A A  
2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760  
ACTTGTACAGTCTAGAGATGTTCTGACCGTCTTCCACCTTTGTACACGCGCTAGGCACTTCTCTACCGCTGATCAGATAAAGTGTGCAATCTCGGTGGTGTATCTTTCTAT  
L V Q I L E M V L D R F F P P L Y N A L G I F L P L I T V N C A I F G G V S F M  
2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880  
GGTACCAAGTGTACTCAAACTTCTGTAATCTATGTTTACGGTTCGGCTCTGGTGTGGCTGGATGCTAGCAATCGTCTGCTTCTGAGTATCCGTGAGAAGTGAAGTACTGTGAGCT  
V Q R D Y N F A E S I V Y G F G S G W M L A T V A L A G I R E K M V S D V  
2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000  
ACCTCAAGTCTAGCTGGTCTGGTATCACATTTACTGAGGTCTTATGGCGTTAGGCTTTATGTTCTTCTGCTGTTCACTGTAAGTGGGTAAACCGCAACATTAAGGGAATA  
P P G L R G L G I T F I T V G L M A L G F M S F S G V Q L \*

nqr6 M D I I L G V V M F T L I V L A L V L V I L F A K S K L V P T G D I T I S V N  
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120  
GTCAATGGCAATTTCTTGGTGTAGTGTCTTACTCTGATGCTACTTGCAGTACTAGTATTCTTTTCCGTAAGCTTACCAACAGGTGACATTAACAATTTCTGTGAA  
3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240  
CGATGACCTTCTCTGGGATCGTTACACCAACAGGTGGTAAGCTACTGAGTGTCTAGCGGTGCTGGCTATTGATCTTCTGCTGTGGTGGCGGTGGCTCATGTGGTCACTGCTG  
D D P S L A I V T Q P G G K L L S A L A G A G V F V S S A C G G G G S C G Q C R  
3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360  
CGTAAAGTTAAATCAGGTGGTGGGACATCTCAACCAACGAGCTTGAACATATTAAAGGTGAAGCAGTGAAGTGAAGCTTCTAGCTGTGCTAGTGTGCTATGAAACCTGACATGGA  
T I A V P N N L V Y N L V R E N A L V E G V D L S F L N F I T F I G V I A A  
3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480  
CATCGAGCTTCTGAGGAATCTTCCGCGTTAAGAGTGGGAATGTACCTTATCTCTAACGATAACAAAGCGACATTCATCAAGAGCTTAAAGCTACAAATTCAGATGGCAATCAGT  
I E L P E E I F G V K K W E C T V I S N D N K A T F I K E L K L Q I P D G E S V  
3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600  
ACCTTTCGCTGCTGGTGTACATCAAGATTGAAGCAGCTGCTCACACGTGAATACGACGACTAGATATTCTGAGGAATACCGTGAAGACTGGGAAGTTCAACCTTTTCCGTTA  
P F R A G G Y I Q I E A P A H H V K Y A D Y D I P E E Y R E D W E K F N L F R Y  
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720  
CGAGTCTAAAGTTAAAGAGACAATCCGTGATCTCAATGGCTAACTACCAGAAGACGAGTATCATCATGCTTAAAGCTTCTGATCGCAACTCCGCGCCGAACCAACGAGACGT  
E S K V N E E T I R A Y S M A N Y P E E H G I I M L N V R I A T P P P N N P D V  
3730 3740 3750 3760 3770 3780 3790 3800 3810 3820 3830 3840  
ACCACAGGCATCATGCTCTGATCATCTGGTCTCTAAAGAGGCGGACAAAGTGTACTTTCTGGTCAATTTGGTGAAGTCTTCTGCAAGAGATACAGATGCAAAATGGTATCGTAGG  
P P G I M S S Y I W S L K E G D K C T I S G P F G E F F A K D T D A E M V F V G  
3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960  
TGGTGTGGTGTGATGGCTCAATCGCTCACACATCTTCAACCACTTAAAGCGTCTACACTCTAAGCTAAGATGTCTTCTGGTACGGCGACGTTCTAAGCTGAAATGTTCTACAGT  
G G A G M A P M R S H I F D Q L K R L H S K R K M S F W Y G A R S K R E M F Y V  
3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080  
AGAAAGCTTCTGATGCTCAAGCTGAGAATGATACTTCTGATGGCACTGTGCACTGTGATGCACTCCAGAGATAAAGTGGGATGGTACATGGCTTTTCAACAGCTTCTTTA  
E D F D M L Q A E N D N F W H C A L S D P L P E D N W D G Y T G F I H N V L Y  
4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200  
CGAGAACTATCTAGCTACGCAAGCGCGAGAAGCTGTGAGTACTAGTGTGGTCAACCTATGATGAATGGCGGTGTTATCGGCTGCTAAAGATCTAGGTGTAGAGATGAAAA  
E N Y L R D H E A P E D C E Y Y M C G P P M M N A A V I G M L K D L G V E D E N  
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320  
CATCTCTAGTAGTACTTGGTGGTAACTCCCTCAAGCTGATTGAATCTATGGCTAGCTCTCAAGAGCTAGGCATTTTGTGAGAGTAAAAATATTAATCAACAGTTGATATAGG  
I L L D D F G G \*

4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440  
TAAATGGATTAAAGTAACCGTCCCATTCATTAATCAACCAATTTGGAGTAAACAGTGAAGCTGGTGTGCTGCTCTCTATTGATTTGGTGGCTGTGAGCAG  
4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560  
CCAGCAGATCAAAATCATTAAGTGGCCCTACCATGGGACTCTTATAACATCAAAATATCGAGCAAGTGGTATCTACCGCAAAAGCGATGACAGCGAAATGATCGCTTACTG  
4570 4580  
GAAGAGGTAAATGACCAATGTGAC