Nucleic Acids Research, Vol. 20, No. 20 5389–5396

Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon *Desulfurolobus ambivalens* shows close phylogenetic relationship to eukaryotic ligases

Arnulf Kletzin
Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, W-8033 Martinsried, Germany

Received July 14, 1992; Revised and Accepted September 25, 1992 EMBL accession no. X63438

ABSTRACT

A 3382 bp fragment containing a gene for a DNA ligase from the extremely thermophilic, acidophilic, and facultatively anaerobic archaeon (archaebacterium) *Desulfurolobus ambivalens* was cloned and sequenced. The deduced amino acid sequence (600 amino acids, 67619 molecular weight) showed 30–34% sequence identity with the ATP-dependent eucaryal (eukaryotic) DNA ligases of *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, the human DNA ligase I, and with the Vaccinia DNA ligase. Distant similarity to the DNA ligases from the bacteriophages T3, T4, T6, T7 and the African swine fever virus was found, whereas no similarities were detectable to the NAD-dependent DNA ligases from the bacteria (eubacteria) *Escherichia coli* and *Thermus thermophilus*, to the ATP-dependent RNA-ligase of bacteriophage T4, and to the tRNA-Ligase from *S.cerevisiae*. A detailed comparison of the phylogenetic relationship of the amino acid sequences of all known DNA and RNA ligases is presented including a complete alignment of the ATP-dependent DNA ligases. The *in vivo*-transcription initiation and termination sites of the *D.ambivalens* gene were mapped. The calculated transcript length was 1904–1911 nt.

INTRODUCTION

The known DNA ligases from *Bacteria* (= eubacteria) and *Eucarya* (= eukaryotes; for the definition of the domains see Ref. 1) represent two different, phylogenetically unrelated types. Eucaryal DNA ligases as well as DNA ligases from the bacteriophages of the T-series are ATP-dependent enzymes (2–3) in contrast to the NAD-dependent enzymes of the *Bacteria* (4). Although they vary considerably in size, all known ATP-dependent DNA ligases share a conserved ATP binding site (5).

So far nothing was known about the DNA ligases or their genes from the third domain, the *Archaea*, although it contains many organisms with unusual physiological properties (e.g. extreme thermophiles or halophiles, Refs. 6–7). This work describes for the first time the cloning and sequencing of an archaeal DNA ligase gene from *Desulfurolobus ambivalens*, an extremely thermophilic, acidophilic, chemolithoautotrophic, and facultatively anaerobic archaeon, growing optimally at 80°C (range: 55–85°C) and at pH 2–3 (range 0.8–4) on elemental sulfur, CO_2, and either H_2 or O_2 (8). Prelimnary investigation of the protein expressed in *E.coli* gave the result that the enzyme is ATP-dependent (R.Rüger, Boehringer Mannheim, pers. communication). These studies were carried out separately and will be described elsewhere.

The similarity of the *D.ambivalens* DNA ligase gene to the ATP-dependent eucaryal DNA ligases was high while no relationship to the known bacterial cellular enzymes was found. The phylogenetic analysis of the previously known DNA ligase sequences was fragmentary, dealing only with parts of the entire protein sequence (9–10) or comparing structural features omitting the sequence (2,5). The first complete amino acid alignment of sequences of the ATP-dependent DNA ligases is presented here, and the phylogenetic relationship of all DNA and RNA ligases is analyzed.

MATERIALS AND METHODS

Materials, cells, DNA and RNA purification procedure

All materials were from the highest quality commercially available. Molecular biology enzymes were obtained from Boehringer Mannheim, Germany. *Desulfurolobus ambivalens* (DSM 3772) was grown aerobically and anaerobically according to published procedures (8). DNA was prepared from anaerobically grown cells by standard procedures (11). RNA was prepared from aerobically and anaerobically grown log phase and stationary phase cells by the acidic guanidinium thiocyanate method (12).

Oligodeoxynucleotide probes

The oligonucleotide probes L1, L2, LigT, Lig2T, Lig3T (derived from the nucleotide sequence; Fig. 1), T3, T7 (Stratagene) and SoxI1 were synthesized on a DNA synthesizer 381A according to the manufacturers instructions (Applied Biosystems). Labeling with T4 polynucleotide kinase was done according to standard procedures (11) except that the reaction was incubated for 30 min at room temperature.

Isolation of clones and construction of pDam-L3

An 857 bp HindIII/EcoRI fragment from within the DNA ligase gene was cloned unintentionally into pUC18 (host strain *E.coli* JM83, plasmid designation pL8/2) following endonucleolytic digestion of genomic DNA and Southern hybridization according to standard procedures (11) and identification of the fragment with the ^{32}P-labeled oligonucleotide SoxiI designed for an entirely different gene (13). The oligonucleotides L1 and L2 derived from the nucleotide sequence of pL8/2 were used to identify and clone two different PstI/XbaI-fragments containing the N-terminal (pL1/2) and C-terminal (pL2/4) parts of the DNA ligase gene (Fig. 1). The latter also contained an internal XbaI-site. Subsequently, pDam-L3 was assembled from these three overlapping clones by ligation of the EcoRI/XbaI-fragment of pL2/4 into pBluescript IIKS$^-$ (Stratagene) followed by the EcoRI/HindIII fragment of pL8/2 and by the HindIII fragment of pL1/2. Double strand DNA sequencing, directed by the oligonucleotides T3, T7, L1, and L2, was applied to confirm the results of this procedure (Fig. 1).

Subcloning and sequencing

The nucleotide sequence of pDam-L3 was determined throughout on both strands by subcloning appropriate and overlapping restriction fragments into the vectors M13mp10, M13tg131, M13mp18, M13mp19 (host strain *E.coli* DH5α), and the four pBluescript II vectors (host strain *E.coli* XL1-Blue). All sequencing was carried out by the dideoxy chain termination method using the Sequenase kit (United States Biochemical Corporation) according to the manufacturers instructions. Sequences from clones in M13 vectors were primed with the universal primer; sequences in pBluescript vectors were primed with either T3 or T7 oligonucleotides.

Cross hybridisation

Genomic DNA of the following organisms was each double digested with BglII/NcoI and with the enzyme(s) given in brackets: *Archaea: Sulfolobus acidocaldarius* (EcoRI), *S.shibatae* (EcoRI), *Desulfurococcus mucosus* (Eco RI), *Thermoproteus tenax* (PstI), *Thermococcus celer* (PstI), *Thermoplasma acidophilum* (EcoRI), *Methanothermus fervidus* (EcoRI), *Methanococcus vannielii* (EcoRI), and *Haloferax volcanei* (BamHI/PstI); *Bacteria: Escherichia coli* (HindIII), and *Bacillus stearothermophilus* (EcoRI); *Eucarya: Bos primigenius taurus* (cattle, EcoRI), and *Saccharomyces cerevisiae* (bakers yeast, EcoRI). The separation and Southern blotting was followed by hybridisation with the α-^{32}P-dATP (Amersham) labeled BglII/NcoI fragment (bp 988−2457, Fig. 1) of pDam-L3 (16). The hybridisation conditions were: 6×SSC, 50% formamid, cooled down from 45°C to room temperature overnight; the washing conditions were: 2×SSC at room temperature. Genomic DNA of *D.ambivalens*, cut with EcoRI and BglII/NcoI, and the BglII/NcoI digested plasmid pDam-L3 were included for control.

Primer extension, nuclease S1-mapping

The *in-vivo* transcription initiation of the DNA ligase gene was mapped by primer extension analysis according to published procedures (14) using the oligonucleotide LigT as primer. The *in-vivo* transcription termination was mapped by S1-analysis (14) using an α-^{32}P-ATP (Amersham) endlabeled NcoI/SnaBI fragment of pL2/4 (Fig. 1, Ref. 11). The oligonucleotides Lig2T and Lig3T were used to investigate the transcription initiation of the two reading frames pDam-L3-ORF2 and 3. Northern hybridisation with randomly primed fragments was carried out by standard procedures after separation of total RNA in glyoxal gels (11).

Sequence comparison and phylogenetic analysis

The searches in the Mipsx protein data bank (Martinsried, Germany) were carried out with the program FIRFASTA (Protein Identification Resource, National Biomedical Research Foundation, 15). The available amino acid sequences of the DNA ligases of *Schizosaccharomyces pombe* (16), *Saccharomyces cerevisiae* (3), *Escherichia coli* (17), *Thermus thermophilus* (18), the human DNA ligase I (9), the Vaccinia virus (10), African swine fever virus DNA ligases (19), and the DNA ligases of the bacteriophages T3 (20), T4 (21), T6 (22) and T7 (23), of the RNA ligase of Bacteriophage T4 (24) and of tRNA ligase of *S.cerevisiae* (25) were subjected to sequence comparison using the programs GAP (University of Wisconsin Genetics Computer Group, Ref. 26) for pairwise alignments and determination of pairwise identities, and to RDF2 analysis for significance evaluation of the results (Protein Identification Resource, Ref. 15, Tab. 1). Because of higher conservation of the DNA ligase genes in the C-terminal domain, long sequences were shortened at the N-terminus to the approximate length of the second sequence prior to RDF2 analysis. The results (Tab. 1) are averages of tests in both directions, showing standard deviation units of the alignment score above the mean random score of the shuffled sequence. The multiple alignment was done with the program PILEUP (UWGCG Vs. 7.0) with manual corrections.

The phylogenetic distances and the branching order were determined with the help of the programs FITCHPRO, PROTPARSPRO, DNAML, and DNA BOOT contained within the PHYLIP 3.4 program package (27) with omission of the third base of each codon to minimize the effects of G+C content. The phylogenetic dendrogram (Fig. 3) was constructed from the amino acid alignment by the distance matrix method (28) using the mutational matrix (29). The distance matrix was calculated from the homology data (30). From the entire sequences only reasonably well aligned parts were used for the calculations (Fig. 2).

RESULTS AND DISCUSSION

A DNA ligase gene from *Desulfurolobus ambivalens* genomic DNA fragment was cloned on three overlapping fragments of genomic DNA. The entire gene was assembled subsequently as stated in the Materials and Methods section yielding the plasmid pDam-L3 (Fig. 1). Both strands of the 3382 bp insert were sequenced after subcloning of appropriate restriction fragments. The nucleotide sequence was submitted to EMBL Data Library and assigned the accession number X63438. The numbering of the nucleotides begins at the HindIII-site upstream of the DNA ligase gene (Fig. 1).

Three possible open reading frames were identified within the sequence: A long ORF extending from bp 908−2710, encoding a 67617 Da protein composed of 600 amino acids, an ORF starting at bp 2961 extending beyond the sequenced region and a third ORF in the opposite strand extending from bp 771 to 250 (Fig. 1). Protein data bank searches (Mipsx, Martinsried, Germany) resulted in high scores for the first ORF to the DNA ligase amino acid sequences of *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, the human DNA ligase I, and the Vaccinia virus DNA ligase (see below). The latter two ORFs showed no significant similarity to any protein in the data bank.

The G+C content of the coding region of the DNA ligase gene was 36%, slightly above the overall G+C content of *Desulfurolobus ambivalens* DNA (31%, Ref. 8). The isoelectric point of the protein, calculated from the amino acid sequence, was 6.21; the net charge was −6. The content of hydrophobic amino acids (I+L+V+M) was 26.2% which is within the range of the other known DNA ligase genes (20.6−26.8%) showing that this was not responsible for the thermostability required by the growth temperature of the organism (80°C).

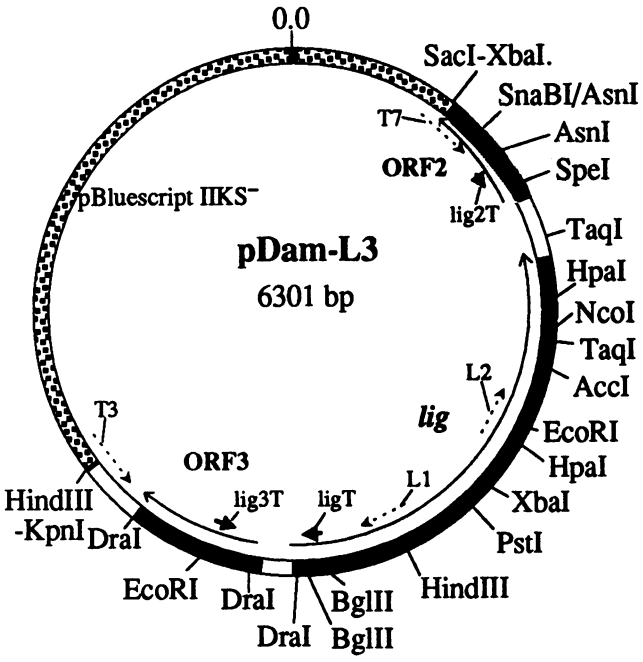

Figure 1. Plasmid map of pDam-L3 with important restriction sites and the orientation of the reading frames. The plasmid was constructed from three independently cloned and overlapping fragments of genomic DNA beginning with the downsteam 1279 bp XbaI/EcoRI fragment followed by the 857 bp EcoRI/HindIII fragment and the 1251 bp HindIII fragment. The results were checked by double strand sequencing using the oligonucleotides L1, L2, T3 and T7 (dashed arrows). The position of the oligonucleotides used for primer extension analysis is marked by short arrows. The zero position is from pBluescript IIKS⁻ (Stratagene); *lig*, DNA ligase gene.

Protein sequence alignments and phylogenetic considerations

The amino acid sequences of the *D.ambivalens* DNA ligase gene and of all other available DNA-ligases were subjected to pairwise alignments for determination of sequence identities and to significance evaluation of similarity using the program RDF2 as stated in the Materials and Methods section (Tab. 1). Additionally, the RNA-ligase from the bacteriophage T4 (24) and the tRNA ligase from *Saccharomyces cerevisiae* (25) were included. The *D.ambivalens* enzyme formed a closely related group with the ATP-dependent eucaryal enzymes. It had 30−34% sequence identity to the known eucaryal enzymes and 48−86 standard deviation units (SD) in the significance evaluation (Tab. 1). The scores of the eucaryal cellular enzymes compared to each other were higher; the scores of the Vaccinia virus enzyme compared to each of the eucaryal and to the *D.ambivalens* ligase were lower. No similarity was found to either of the NAD-dependent, phylogenetically related enzymes of *Escherichia coli* and *Thermus thermophilus* (46,7% identity, Ref. 18) and to the two ATP-dependent RNA-ligases (<20.4% sequence identity, <0.2 SD in the RDF2 analysis, Tab. 1). The T-phage DNA-ligases formed two pairs with high similarity scores (T4 and T6, 98.8% vs. T3 and T7, 72.1%). The RDF2-program did not detect significant similarities between these pairs or of these pairs with most of the eukaryal ligases except with the Vaccinia virus ligase. The pairwise identity scores, however, were higher (20.2−26%). The pairwise similarity scores of the African swine fever virus (ASFV) DNA ligase were similar to the results of the T-phage DNA ligases, the RDF2-results, however, were higher except for the *D.ambivalens* and the T3 enzymes (Tab. 1).

The length of the ATP-dependent DNA ligases differed between 346 (bacteriophage T3) and 919 amino acid residues

Table 1. Results of the pairwise comparison of the known amino acid sequences of DNA ligases including the T4 RNA-ligase and the *S.cerevisiae* tRNA ligase

Species	*S.p.*	*S.c.*	*H.s.*	*D.a.*	Vacc.	ASFV	T4-D	T6	T7	T3	*E.c.*	*T.th.*	T4-R	*S.c.* t
Sch. pombe (767)	421.21	46.2	43.6	35.1	30.1	21.5	20.5	20.5	20.3	24.6	18.0	21.4	15.3	16.6
S. cerevisiae (755)	210.57	397.20	42.0	34.3	29.7	21.2	21.3	22.0	23.4	22.1	16.5	20.3	18.2	15.3
H. sapiens (919)	190.52	174.42	393.44	35.5	27.0	19.7	23.4	23.9	22.6	26.0	18.4	20.5	16.3	16.8
D. ambivalens (600)	86.82	84.01	79.28	270.56	32.8	20.3	20.7	20.6	23.7	20.2	20.4	16.8	16.9	16.0
Vaccinia virus (*552*)	52.94	53.34	51.22	48.54	266.08	20.9	22.8	23.7	21.7	22.8	15.3	18.0	17.4	14.4
Afr. swine fever virus (*419*)	7.60	5.05	5.36	0.12	11.78	227.18	22.9	22.2	16.9	22.6	15.9	15.4	14.5	16.0
T4 DNA ligase (*487*)	1.72	0.13	-0.24	-0.43	4.38	3.00	244.25	98.8	21.6	21.5	20.5	20.3	13.7	15.1
T6 (*487*)	1.52	0.11	-0.08	-0.49	4.82	3.12	231.92	239.88	21.3	21.8	20.7	20.0	13.1	15.3
T7 (*359*)	3.90	4.61	0.94	0.43	4.88	2.35	0.50	0.32	212.73	72.1	17.8	17.8	19.3	14.4
T3 (*346*)	1.33	0.16	1.34	0.48	8.32	-0.10	-1.41	-1.42	121.92	195.81	20.8	20.1	20.0	22.3
E. coli (672)	1.05	2.36	2.12	0.19	-0.98	-0.17	-0.36	-0.46	-1.16	-0.32	334.90	46.7	18.2	22.7
Th. thermophilus (678)	-0.61	0.81	0.64	-0.88	0.58	0.08	0.01	-0.04	-1.44	1.90	141.84	268.24	16.2	15.5
T4 RNA ligase (*374*)	0.90	0.65	-0.48	-0.90	1.25	0.63	0.06	0.06	0.56	0.14	1.27	1.27	183.09	17.0
S. cerev. tRNA ligase (*827*)	0.96	1.86	0.54	0.10	-0.76	-0.33	0.46	0.71	-0.86	-1.02	0.95	-0.18	1.62	428.25

In brackets: total length of the amino acid sequence. References see Materials and Methods section. Upper right triangle: pairwise identities (%); lower left triangle: significance evlauation results using the program RDF2 (SD-units). In the middle diagonal the RDF2 scores of the given sequence tested with itself. Results are standard deviation units of the aligment score above the mean random score of 1000 replicas. Results indicating significant homology are boldface (15). Bacteriophage T4 and T6 DNA ligases differed only by 6 amino acid residues (see Fig. 2).

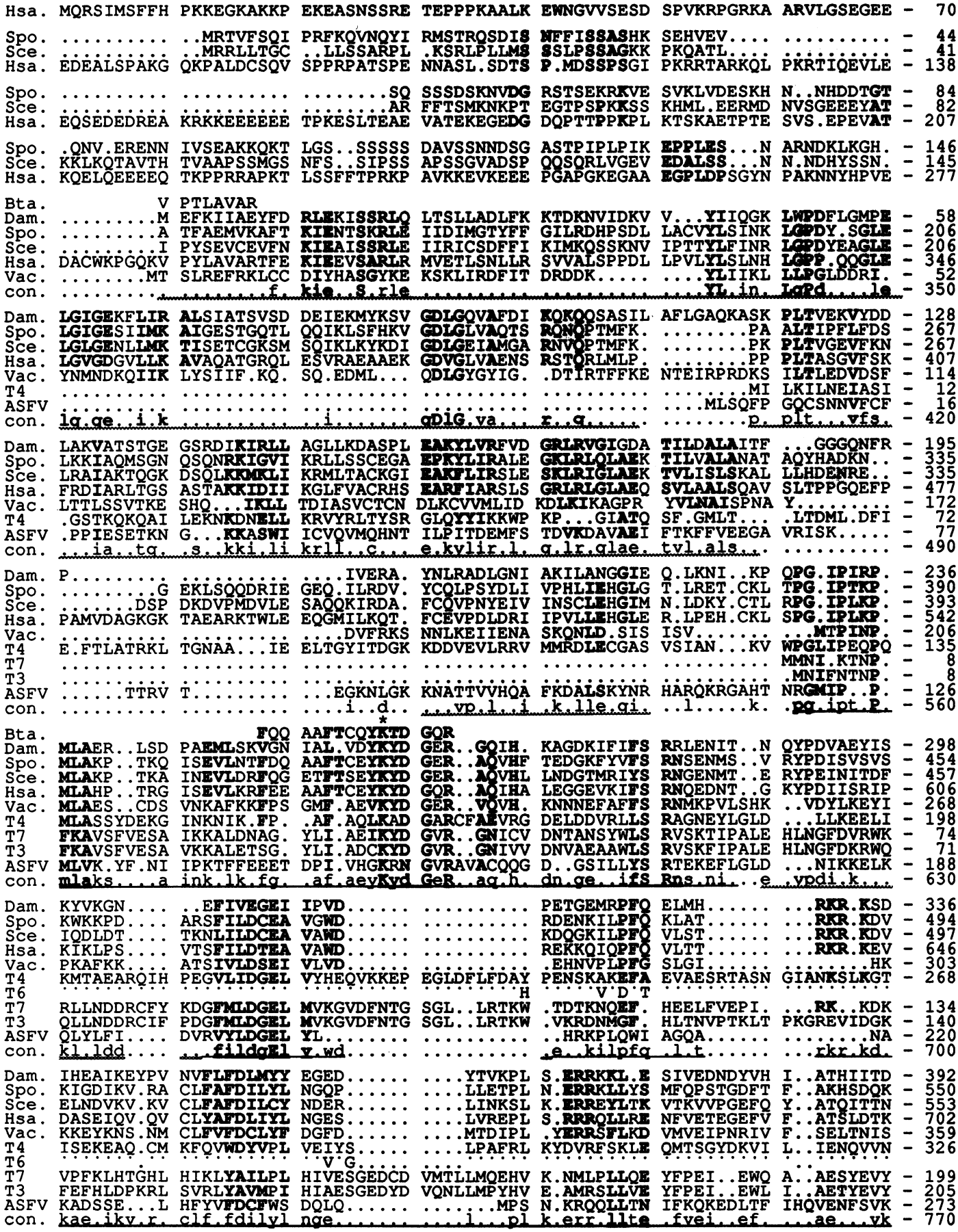

Figure 2. Amino acid alignment of the available DNA-ligases. Abbreviations and references: *Dam. Desulfurolobus ambivalens* (this work); *Sce. Saccharomyces cerevisiae* (3); *Spo. Schizosaccharomyces pombe* (16); *Hsa. Homo sapiens* DNA ligase I (9); Vac. Vaccinia virus (10); *Bta. Bos primigenius taurus* fragments (cattle, Ref. 5); T3, T4, T6, T7: Bacteriophage T3 (20); T4 (21); T6 (22); and T7 (23) DNA ligases; ASFV African swine fever virus (19); con. consensus, an entry was made in the case of at least three identical positions, captial letters indicate identities throughout the sequences. Straight underlined: Parts of the aligned sequences that were used for the calculation of phylgenetic distances (Fig. 3), dotted underline: additionally used for the calculation of distances between the eukaryal, the *D.ambivalens* and the Vaccinia ligases. Numbers: amino acid numbering of the individual sequence, in the consensus line of the total alignment. Bold: some conserved parts, the star shows the ATP-binding lysine residue (5). The amino acid sequence of bacteriophage T6 ligase differed only by six residues from the T4 DNA ligase and was left from the alignment except for those positions (Pos. 669−680 and Pos. 723−725, indicates identities).

(human DNA ligase I). In spite of the differences in size, it was possible to include the ten amino acid sequences of DNA-ligases into a multiple alignment due to highly conserved motifs in the sequences and due to the fact that the longer molecules were conserved in their C-terminal part (Fig. 2). The alignment showed many conserved parts separated by regions of low similarity as previously observed particularly in the N-terminal regions of the eucaryal DNA ligases (9). The first 290 amino acid residues of the human DNA ligase and the first 140 residues of the Yeast ligases preceeding the start of the *D.ambivalens* and Vaccinia ligases in the alignment are highly hydrophilic in contrast to the rest of the molecules and to the other DNA ligases where hydrophilic and hydrophobic domains are found alternating (data not shown).

With one exception neither of the conserved motifs were found in the NAD-dependent enzymes of the *Bacteria* and in the RNA-ligases. The only feature shared by almost all ligases was the motif *K*-x-D(N)-G* around the putative AMP-binding lysine residue (K*) which has been determined for the bovine DNA ligase I (5), the T4 RNA ligase (31) and the tRNA ligase from *S.cerevisiae* (32). The latter showed considarable similarity only in the 30 amino acid residues preceeding the AMP-binding lysine residue to the T4 RNA ligase (32). The sharing of the *K-x-D(n)-G*-motif must be regarded as convergance rather than phylogentic homology, especially since the binding sites in the bacterial enzymes were derived from sequence similarity and had not been determined experimentally (see below and Ref. 5). The consensus sequence for the AMP-binding domain of the ATP-dependend DNA ligases was in most cases *K*-Y-D-G-x-R* (T4 and T6: *K*ADGAR*; ASFV: *KRNGVR*), while the corresponding sequences from the T4 RNA-ligase were *K*EDGSL*, from *E.coli K*LDGLA* (5), and from *T.thermophilus K*VDGLS* (9). The motifs *PMLA* (T4 and T6: *PQMLA*; T3 and T7: *PFKA*, ASFV: *PMLV*) 20−22 residues upstream the ATP-binding lysine residue and the triplett *(f,l,y)SR* 18−21 residues downstream were also strictly conserved suggesting a participation in the reaction (Fig. 2, Pos. 559−611). The region around the motif *EG(l/m/v)(m/i)(v/l)K(5−8)Y(8−10)KxK* was remarkable for three lysine residues at almost identical distances (Pos. 787−818, Fig. 2, Ref. 3). The homologous regions to the C-terminal conserved eukaryal *SLRFPRFIRIR*-motif (Pos. 851−867) could be found in all of the viral enzymes in contrast to earlier findings (2, 10, 19). Neither of these motifs could be found in the NAD-dependent or the RNA-ligases. Several other conserved signatures

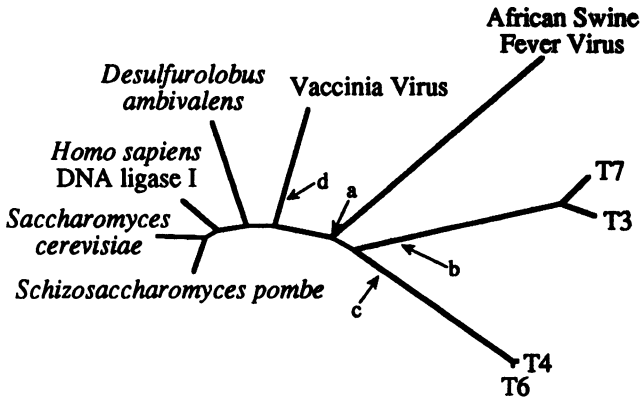

Figure 3. Phylogenetic dendrogram derived from the amino acid alignment shown in Fig. 2 calculated with the distance matrix method (28). The length of the branches was corrected for multiple mutations (30). a–d: Possible branching points of the ASFV DNA ligase. Points a–c were statistically identical, d was significantly worse (27).

allowed a full alignment of the sequences including the ASFV-DNA ligase which differed more than the T-phage ligases did. For these reasons all ATP-dependet DNA ligases were considered to be phylogenetically related in spite of the low overall similarity and the large differences in size (Tab. 1). A significant alignment of the other sequences to any of the ATP-dependent DNA ligases was not possible.

From the multiple alignment a phylogenetic dendrogram was calculated using the distance matrix method (Fig. 3, Ref. 28). The *D.ambivalens* DNA ligase was more closely related to the eucaryal cellular enzymes than to the Vaccinia ligase (Figs. 2+3). This was confirmed by the results of the pairwise comparisons (Tab. 1) and of the multiple alignment (Fig. 2, e.g. pos. 301–309 and 349–356) where in several cases the *D.ambivalens* enzyme shared more identities with the eukaryal cellular ligases than the Vaccinia enzyme did. The topology of the dendrogram was confirmed by the maximum likelyhood, by the protein parsimony and by the bootstrap analysis (27) with respect to most branching points, especially confirming the position of the *D.ambivalens*

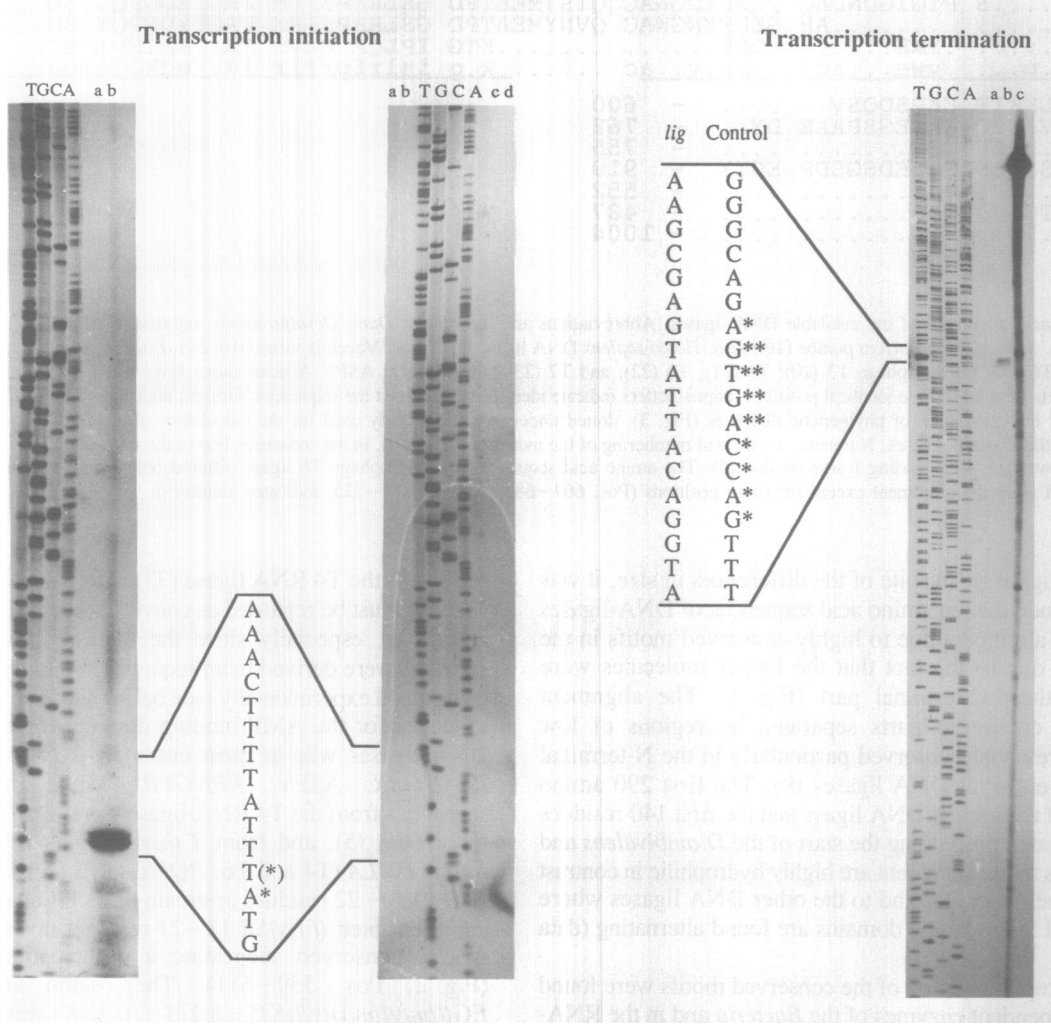

Figure 4. Results of the transcriptional mapping of the DNA ligase gene. Left and middle: Primer extension and nuclease S1 signals obtained with oligonucleotide LigT with RNA of log phase (left) and of statonary phase cells (middle); astrerisks: the initiation site. Left, lanes: T,G,C,A: the sequencing reaction primed with the oligonucleotide LigT (anti-sense strand), a,b: primer extension signal with RNA from aerobically and anaerobically grown cells. Middle, lanes: a,b: Nuclease S1- and primer extension signals of RNA from aerobically grown cells, c,d: of RNA from anaerobically grown cells; sequence as left panel. Right: Nuclease S1 analysis of transcription termination. The control DNA was from the Sequenase kit primed with the −40 oligonucleotide included. Asterisks indicate termination signal. Lanes, a: probe subjected to the S1 mapping procedure without RNA; b: Nuclease S1 transcription termination signal with stationary phase RNA from anaerobically grown cells; c: undigested, labeled probe (NcoI/SnaBI fragment, 764 bp, Fig. 1).

DNA ligase in relation to the Vaccinia and eucaryal cellular enzymes. A separate calculation only with the five sequences from *D.ambivalens*, Vaccinia virus and the eukaryal cellular enzymes as shown in Fig. 2 gave the same result.

Differences were obtained concerning the branching topology of the human and yeast ligases showing the *Sch. pombe* and the human ligases to be closer related than the two yeast ligases. The protein parsimony analysis did not give significant differences between the two branching topologies. However, the dendrogram shown in Fig. 3 was believed to be correct because of the almost identical total length of both yeast DNA ligase enzymes and of the score of pairwise identities (Tab. 1). The artefactual results of the bootstrap and the maximum likelyhood methods at this point were probably due to the comparatively short stretches of the aligned sequences which could be used for the calculation of the dendrogram. The branching point of the ASFV sequence could not be determined precisely. With no method significant differences between the branching points a, b, c were found (Fig. 3). Branching point d, closest to the Vaccinia virus enzyme could be excluded by comparison of user-defined tree in the protein parsimony method (27). The dendrogram shown in Fig. 3 was believed to be correct because of the results of the pairwise comparisons (Tab. 1). The large differences to the other sequences (Fig. 2) explain the length of the branch. This and the large distances of the T-phage enzymes may reflect rapid evolution of these viruses.

Cross hybridisation

A randomly primed labeled BglII/NcoI fragment covering 82% of the DNA ligase gene (Fig. 2) was used to hybridize Southern blots of genomic DNA of various organisms listed in the Materials and Methods section, each of them cut with two sets of restriction enzymes. Under conditions of low stringency in the hybridisation and washing steps defined bands were only obtained with DNA from *Sulfolobus acidocaldarius* and from *S.shibatae*. Both of them are relatives of *D.ambivalens* belonging to the same order *Sulfolobales* (33). No signals were obtained with DNA from other archaeal, bacterial and eucaryal sources including *Thermoproteus tenax* and *Desulfurococcus mucosus* showing that the gene is not highly conserved on the DNA level (not shown). Hybridisation with genomic DNA from *D.ambivalens*, digested with NcoI/BglII, gave a single band. Digestion with EcoRI, cutting within the gene, gave double bands as expected. Both results showed that a single copy of the DNA ligase gene is present in the genome (data not shown).

Transcriptional analysis

The *in vivo* transcriptional start and termination sites of the DNA ligase gene were mapped. RNA isolated from log-phase and from stationary phase cells produced two different primer extension signals for transcription initiation (Fig. 4). Weak signals were obtained when total RNA preparations were from log phase cells. A strong induction was observed in RNA-preparations from stationary phase cells grown for 10 days. The signal splitting seen in this case might be due to incomplete extension with the reverse transcriptase (Fig. 4). When comparing the signal strength of RNA from aerobically and anaerobically grown cells no difference was observed, showing that the gene expression was not influenced by the mode of growth of the organism. The transcription was initated at the ATG start codon of the reading frame. The derived promoter sequence could easily be aligned to known archaeal promoters (Fig. 5, Ref. 34). No transcription signals were obtained from the other two open reading frames.

The analysis of the transcription termination site by Nuclease S1 mapping is shown in Fig. 4. The derived terminator sequence is given in Fig. 5. The transcription termination signal did not match known termination sequences (Fig. 5, Ref. 35). Two stem and loop structures could be derived from the nucleotide sequence, although they were not immediately in front of the termination site. The function of a 10 nt palindromic sequence immediately downstream of the stop codon (Fig. 5) remains uncertain. Transcript processing was not investigated in detail. However, the oligonucleotide Lig2T, designed for primer extension analysis of the transcription of pDam-L3-ORF2 (Fig. 1), did not give any signal showing that readthrough and subsequent transcript processing did not occur (data not shown). The total length of the transcript derived from the data was 1904−1911 nt. No corresponding RNA species was detected by Northern analysis. This might be due to low level of expression as shown in Fig. 4.

ACKNOWLEDGEMENTS

Special thanks are due to Prof. W.Zillig for the continuing support during the entire work to Manfred Reitmeier and Wolfgang Melchinger for assistence in the sequencing of the gene.

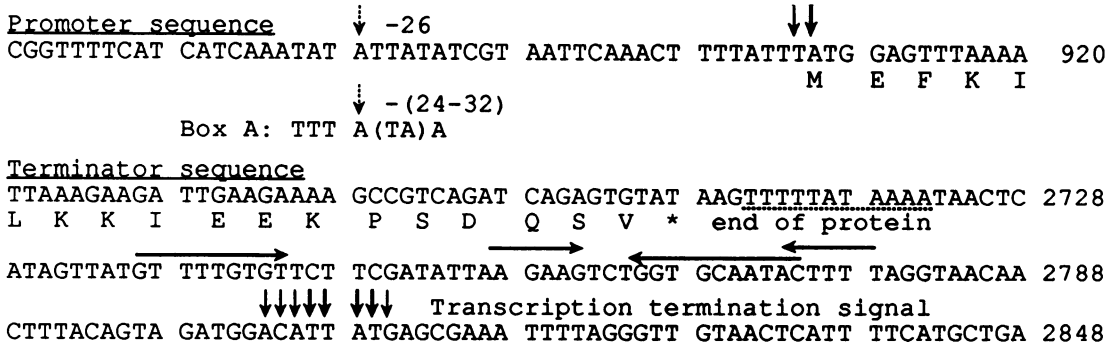

Figure 5. Promoter and terminator sequences of the *D.ambivalens* DNA ligase gene (Fig. 1). Promoter, vertical arrows: transcriptional start sites; figures, distance of the start site to the TA position of box A (bold); lower line, Box A consensus sequence (34). Terminator: vertical arrows, position of the termination signal (weak and strong, Fig. 4); horizontal arrows, possible stem and loop structures; dotted underline, palindromic sequence of unknown function.

REFERENCES

1. Woese,C.R., Kandler,O. and Wheelis,M.L. (1990) *Proc. Acad. Natl. Sci. USA*, **87**, 4576
2. Lasko,D.D., Tomkinson,A.E. and Lindahl,T. (1990) *Mutation Res.*, **236**, 277–287.
3. Barker, D G., Johnson, A L. and Johnston, L H. (1985) *Mol. Gen. Genet.*, **200**, 458–462.
4. Ishino,Y., Shinagawa,H., Makino,K., Tsunasawa,S., Sakiyama,F. and Nakata,A. (1986) *Mol. Gen. Genet.*, **204**, 1–7.
5. Tomkinson,A.E., Totty,N.F., Ginsburg,M. and Lindahl,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 400–404.
6. Kushner,D.J. (1985) In Woese,C.R. and Wolfe,R.S. (ed.) The Bacteria. A treatise in structure and function. Academic Press, Orlando, Vol.VIII, pp. 171–214.
7. Stetter,K.O., Fiala,G., Huber,G., Huber,R. and Segerer,A. (1990) *FEMS Microbiol. Rev.*, **75**, 117–124.
8. Zillig,W., Yeats,S., Holz,I., Böck,A., Rettenberger,M., Gropp,F. and Simon,G. (1986) *System. Appl. Microbiol.*, **8**, 197–203.
9. Barnes,D.E., Johnston,L.H., Kodama,K., Tomkinson,A.E., Lasko,D.D. and Lindahl,T. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6679–6683.
10. Smith,G.L., Chan,Y.S. and Kerr,S.M. (1989) *Nucl. Acids Res.*, **17**, 9051–9062.
11. Sambrock,J., Fritsch,E.F., and Maniatis,T. (1989) Molecular cloning. A laboratory manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour.
12. Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156–159.
13. Kletzin,A. (1992) *J. Bacteriol.*, **174**, *in press*, X56616
14. Hüdepohl,U., Reiter,W.-D. and Zillig,W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5851–5855.
15. Pearson,W.R. (1990) *Meth. Enzymol.*, **183**, 63–99.
16. Barker,D.G., White,J.H.M. and Johnston,L.H. (1987) *Eur. J. Biochem.*, **162**, 659–667.
17. Ishino,Y., Shinagawa,H., Makino,K., Tsunasawa,S., Sakiyama,F. and Nakata,A. (1986) *Mol. Gen. Genet.*, **204**, 1–7.
18. Lauer,G., Rudd,E.A., Mckay,D.L., Ally,A., Ally,D. and Backman,K.C. (1991) *J. Bacteriol.*, **173**, 5047–5053.26. Schmitt,M.P. Beck,P.J., Kearney,C.A., Spence,J.L., DiGiovanni,D., Condreay,J.P. and Molineux,I.J. (1987) *J. Mol. Biol.*, **193**, 479–495.
19. Hammond,J.M., Kerr,S.M., Smith,G. and Dixon,L.K. (1992) *Nucl. Acids Res.*, **20**, 2667–2671
20. Schmitt,M.P., Beck,P.J., Kearney,C.A., Spence,J.L., DiGiovanni,D., Condreay,J.P., and Molineux,I.J. (1987) *J. Mol. Biol.* **193**, 479–495.
21. Armstrong,J., Brown,R.S. and Tsugita,A. (1983) *Nucl. Acids Res.*, **11**, 7145–7156.
22. Kaliman,A.V., Zimin,A.A., Nazipova,N.N., Kryukov,V.M., Tanyashin,V.I., Kraev,A.S., Mironova,M.V., Skyrabin,K.G. and Baev,A.A. (1988) *Dokl. Akad. Nauk SSSR*, **299**, 737–742.
23. Dunn,J.J. and Studier,F.W. (1981) *J. Mol. Biol.*, **148**, 303–330.
24. Rand,K.N. and Gait,M.J. (1988) *EMBO J.*, **3**, 397–402.
25. Westaway,S.K., Phizicky,E.M. and Abelson,J. (1988) *J. Biol. Chem.*, **263**, 3171–3176.
26. Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucl. Acids Res.*, **12**, 387–395.
27. Felsenstein,J. (1982) *Quart. Rev. Biology*, **57**, 379–404.
28. Fitch,W.M. and Margoliash,E. (1967) *Science*, **155**, 279–284.
29. Dayhoff,M.O., Schwarz,R.M. and Orcutt,B.C. (1979) In Dayhoff,M.O. (ed.), Atlas of protein sequence and structure. National Biomedical Research. Foundation, Washington D C, Vol. 5, Suppl. 3, p. 345.
30. Feng,D.F., Johnson,M.S. and Doolittle,R.F. (1985) *J. Mol. Evol.*, **21**, 112–125.
31. Thøgersen,H.C., Morris,H.R., Rand,K.N. and Gait,M.J., (1985) *Eur. J. Biochem.*, **147**, 325–329.
32. Xu,Q., Teplow,D., Lee,T.D. and Abelson,J. (1990) *Biochmistry*, **29**, 6132–6138.
33. Klenk,H.-P., Haas,B., Schwass,V. and Zillig,W. (1986) *J Mol Evol.*, **24**, 167–173.
34. Reiter,W.-D.; Palm,P.; Zillig,W. (1988) *Nucl. Acids Res.*, **16**, 1–19.
35. Brown,J.W., Daniels,C.J. and Reeve,J.N. (1989) *CRC Crit. Rev. Microbiol.*, **16**, 287–338.