LA-UR -88-2954

LA-UR--88-2954

DE89 000374

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE: ISOLATION AND PARTIAL SEQUENCE OF THE A-PROTEIN GENE OF *Thermus thermophilus* CYTOCHROME c_1aa_3

AUTHOR(S) J. A. Fee, M. W. Mather, P. Springer, S. Hensel, and G. Buse

SUBMITTED TO Annals of the New York Academy of Sciences, 1989

Proc. Conf. on Cytochrome Oxidase: Structure, Function, and Physiopathology, Rome, Italy, June 20 21, 1988.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

MASTER

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy

ISOLATION AND PARTIAL SEQUENCE OF THE A-PROTEIN GENE OF *Thermus thermophilus* CYTOCHROME c_1aa_3

J. A. Fee[§], M. W. Mather[§], P. Springer[§], S. Hensel[¶], and G. Buse[¶]

[§]Isotope and Structural Chemistry Group

Los Alamos National Laboratory

Los Alamos, New Mexico 87545, U. S. A.

and

[¶]Institut für Biochemie

Fachgebiet Molekulare Biologie der Proteine

Rheinisch-Westfälische Technische Hochschule Aachen

D5100 Aachen, German Federal Republic

This work was supported by U.S.P.H.S. grant GM35342 (J. A. F.) and grant Bu 463/1 of the DFG (G. B.).

To appear in the New York Academy of Sciences Annals, 1989

[illegible] 1988.

Thermus thermophilus is a strictly aerobic eubacterium which grows optimally near 70° C. Its respiratory system is very similar to that of eukaryotic mitochondria, and the organism has proven to be a particularly good source of stable, comparatively simple respiratory enzymes [*cf* Refs. 1 and 2]. There are at least two terminal oxidases: The recently discovered cytochrome $\underline{ba}_3$[3] and cytochrome $\underline{c}_1$ $\underline{aa}_3$[2]. Cytochrome $\underline{ba}_3$ is an analog of $\underline{aa}_3$ in which the heme A of cytochrome $\underline{a}$ is replaced with protoporphyrin IX (heme B) while its other redox components appear to be largely identical to those of the now classical mammalian cytochrome $\underline{aa}_3$ [*cf* Ref. 3]; it has only a single 35 kD protein subunit. Cytochrome $\underline{c}_1\underline{aa}_3$ consists of two polypeptides. The ~33 kD C-protein covalently binds one heme C, while the ~55 kD protein is thought to bind the four canonical redox centers of $\underline{aa}_3$, two heme A, and two Cu [*cf* Refs. 1 and 2]. Toward our goal of unequivocally establishing the distribution of the metal centers in cytochrome $\underline{c}_1\underline{aa}_3$, we have isolated the structural gene of the A-protein.

METHODS

Cytochrome $\underline{c}_1\underline{aa}_3$ was prepared by the method of Yoshida et al[2] and shipped on dry ice to the RWTH Aachen for amino acid sequence analysis. The protein was denatured by 5 % SDS, freed from SDS[4], and the large polypeptide subjected to cyanogen bromide cleavage. The peptides were separated on HPLC (Nucleosil 7C4, 4 x 250 mm) in 60 % formic acid (A) with 20 % acetonitrile, 15 % 1 propanol (B) gradient. Four peptides were selected for amino acid sequencing according to Buse et al[4].

DNA cloning and other molecular biology procedures were as described by Maniatis et al[5]. DNA hybridization in dried agarose gels was carried out as described by Mather[6]. The oligodeoxynucleotide probe was synthesized by Synthetic Genetics, San Diego. DNA sequence was obtained by the dideoxy chain termination method of Sanger[7] under the conditions described by L. Johnston Dow et al[8]. Ambiguities resulting from secondary structure in the DNA were resolved by the use of analogs of dGTP, by carrying out the sequencing reactions at elevated temperature, and by inclusion of single strand binding protein in the sequencing reactions.

RESULTS

The availability of fragmentary amino acid sequence information can make possible the detection of specific DNA sequence and thereby aid in the identification of structural genes. Our approach to isolating the A-protein gene involves the use of most probable codons, as discussed by Lathe[9], to construct a single 38 base oligodeoxynucleotide probe as shown in FIGURE 1. When the probe was labeled at its 5' terminus with ^{32}P-phosphate and hybridized to restricted chromosomal DNA in dried agarose gels, specific binding to ~2.7 kbp (Hind III) and ~2.8 kbp (Pst I) DNA fragments was observed (photos not shown). These fragments were purified from the gel, ligated into plasmid pUC8, and the recombinant plasmids were transformed into *E.coli*; colonies bearing the HindIII and PstI restriction fragments were identified using colony hybridization. When the probe was hybridized to plasmid DNA from these strains, which had been cut with both HindIII and PstI, it associated, in both cases, with a common fragment of approximately 600 bp. This along with other HindIII/PstI fragments, as shown in FIGURE 2, were cloned into bacteriophage M13 vectors for sequencing. The results of our sequencing effort up to the time of this writing are shown in FIGURE 3. Preliminary amino acid sequence information available for *Thermus* is compared in FIGURE 4 with relevant sequence information from bovine and yeast subunits I and the large subunits of *Paracoccus denitrificans* and *Bacillus* PS3; the homology among these proteins is extensive. Regions of particularly pronounced homology are evident near positions 95, 240, 290, 370, and 435 (bovine numbering). These regions include the conserved histidines which have been predicted to form ligands to the two hemes and Cu_B of cytochrome aa_3[18]

DISCUSSION

An outstanding question in bioenergetics is the distribution of the cannonical redox metals among the various subunits of cytochrome c oxidase. Wikstrom et al[19] have concluded from a study of the extant literature that only Subunits I and II are reasonable candidates for metalloproteins, and Holm et al[18] devised a structural model in which cytochrome a and the cytochrome a_3/Cu_B pair are bound to Subunit I while Cu_A is bound to Subunit II. Purification of a two subunit oxidase from *Paracoccus denitrificans* having functional and spectral identity as well as one-dimensional structures homologous to Subunits I and II of eukaryotic enzyme[20], further supports the view that the metals are localized in the two large subunits.

The enzymes from *Thermus* and *Bacillus* PS3 seem to have stood outside this derivation. They are isolated as non-covalent complexes of a large subunit and a smaller subunit which binds the heme C. Based on the inability to resolve cytochrome $\underline{c}_1\underline{aa}_3$ into more than two subunits and laboratory "incidents" in which either the cytochrome $\underline{c}_1$ or $\underline{aa}_3$ was obtained separately in pure form (unpublished results), it has been argued that the redox metals of $\underline{aa}_3$ are associated exclusively with the single large subunit[12]. In these accidental situations, the cytochrome $\underline{c}_1$ has native optical properties and lacks the typical Cu_A EPR spectrum while the cytochrome $\underline{aa}_3$ has native optical properties and possesses cytochrome $\underline{c}$ oxidase activity; unfortunately, it has not been possible to reproducibly separate the two proteins of $\underline{c}_1\underline{aa}_3$ without denaturation, and their individual characteristics have not been fully explored. Nevertheless, the above observations lend merit to the argument that Subunit I of the *Thermus* enzyme binds the cannonical metals.

The data presented in FIGURE 4 show that Subunits I from the represented phyla are probably the same protein and thus bring the *Thermus* and PS3 enzymes into the fold of all other $\underline{aa}_3$ type oxidases. Indeed, we will hereafter refer to the A protein as Subunit I of *Thermus*. In these enzymes then, either all metals are in Subunit I or the C proteins contain sequences that allow formation of the Cu_A binding site. Complete sequence information on the C protein of *Thermus* is not yet available, however, a few peptide sequences from this protein (G. Buse et al. In preparation) and unpublished work from Sone's (N. Sone, Personal communication) laboratory suggest that the C protein may be a gene fusion of a cytochrome $\underline{c}$ and a portion of Subunit II which contains the presumed Cu_A binding region with two histidine and two cysteine residues. A crucial issue then is whether Cu_A requires these cysteines to account for its unusual electronic properties. The recent finding that cytochrome $\underline{ba}_3$ from *Thermus* contains a Cu_A binding site but appears to have only one cysteine residue casts doubt on the two cysteine hypothesis (*cf* Ref. [3] and references therein); almost heretically, one can ask if the Cu_A site contains any sulfur ligands

While the problem of metal distribution remains unresolved, it is clear from the sequence data that if Cu$_A$ is in Subunit I it cannot be coordinated to cysteine as there are no invariant cysteine residues in this subunit.

REFERENCES

1. FEE, J. A., D. KUILA, M.W. MATHER & T. YOSHIDA. 1986. Respiratory proteins from extremely thermophilic, aerobic bacteria. Biochim. Biophys. Acta **853**: 153-185.
2. YOSHIDA, T., R.M. LORENCE, M. G. CHOC, G. E. TARR, K. L. FINDLING & J. A. FEE. 1984. Respiratory proteins from the extremely thermophilic aerobic bacterium, *Thermus thermophilus*. J. Biol. Chem. **259**: 112-123.
3. ZIMMERMANN, B.H., C. I. NITSCHE, J.A. FEE, F. RUSNAK & E. MÜNCK. 1988. Properties of a copper-containing cytochrome ba$_3$: A second terminal oxidase from the extreme thermophile, *Thermus thermophilus*. Proc. Natl. Acad. Sci. USA *in press*.
4. BUSE, G., STEFFENS, G. J., STEFFENS, G. C. M., MEINECKE, L., HENSEL, S., & REUMKENS, J. 1986. Sequence analysis of complex membrane proteins (cytochrome c oxidase). *In* Advanced Methods in Protein Microsequence Analysis. B. Wittmann Liebold et al, Eds. Springer Verlag. Berlin.
5. MANIATIS, T., E.F. FRITSCH & J. SAMBROOK. 1982. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.
6. MATHER, M. W. Base composition independent hybridization in dried agarose gels. Screening and recovery for cloning of genomic DNA fragments. 1988. BioTechniques **6**: 444-447.
7. SANGER F., S. NICKLEN & A.R. COULSON. 1980. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 6363-6468.
8. JOHNSTON-DOW, L., E. MARDIS, C. HEINER & B. A. ROE. 1987. Optimized methods for flourescent and radio labeled DNA sequencing. BioTechniques 5: 754-765.

9. LATHE, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data. Theoretical and practical considerations. J. Mol. Biol. **183**: 1-12.

10. KAGAWA, Y., H. NOJIMA, N. NUKIWA, M. ISHIZUKA, T. NAKAJIMA, T. YASUHARA, T. TANAKA & T. OSHIMA. 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. J. Biol. Chem. **259**: 2956-2960.

11. KUNAI, K., M. MACHIDA, H. MATSUZAWA & T. OHTA. 1986. Nucleotide sequence and characteristics of the gene for L-lactate dehydrogenase of *Thermus caldophilus* GK24 and the deduced amino-acid sequence of the enzyme. Eur. J. Biochem. **160**: 433-440.

12. NISHIYAMA, M., N. MATSUBARA, K. YAMAMOTO, S. IIJIMA, T. UOZUMI & T. BEPPU. 1986. Nucleotide sequence of the malate dehydrogenase gene of *Thermus flavus* and its mutation directing an increase in enzyme activity. J. Biol. Chem. **261**: 14178-14183.

13. ANDERSON, S., M. H. L. DE BRUIJN, A. R. COULSON, I. C. EPERON, F. SANGER & I. G. YOUNG. 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156**: 683-717.

14. BONITZ, S. G., G. CORUZZI, B. E. THALENFELD & A. TZAGOLOFF. 1980. Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit I of yeast cytochrome oxidase. J. Biol. Chem. **255**: 11927-11941.

15. RAITIO, M., T. JALLI & M. SARASTE. 1987. Isolation and analysis of the genes for cytochrome c oxidase in *Paracoccus denitrificans*. EMBO J. **6**: 2825-2833.

16. SONE, N., F. YOKOI, T. FU, S. OHTA, T. METSO, M. RAITIO & M. SARASTE. 1988. Nucleotide sequence of the gene coding for cytochrome oxidase subunit I from the thermophilic bacterium PS3. J. Biochem. (Tokyo) **103**: 606-610.

17. DEVEREUX, J., P. HAEBERLI & O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**: 387-395.

18. HOLM, L., M. SARASTE & M. WIKSTRÖM. 1987. Structural models of the redox centers in cytochrome oxidase. EMBO J. **6**: 2819-2823.

19. WIKSTRÖM, M., M. SARASTE & T. PENTTILA. 1985. Relationship between structure and function in cytochrome oxidase. *In* The Enzymes of Biological Membranes. Vol. 4. A. N. Martonosi Ed. Plenum Press. New York.

20. STEFFENS, G. C. M., BUSE, G., OPPLIGER, W., AND LUDWIG, B. 1983. Sequence homology of bacterial and mitochronrial cytochrome c oxidases. Biochem. Biophys. Res. Commun. **116**. 335 - 340.

FIGURE LEGENDS

FIGURE 1. Oligodeoxynucleotide probe construction using most probable codons. Peptide sequence from cyanogen bromide peptide 4 (see Fig. 3) was back translated to nucleotide sequence using only the most probable codon for each amino acid. Codon frequencies for *Thermus* proteins were determined by combining the codon usage data from References 10 - 12. The estimated frequencies of the most probable codons used in the probe are shown in the last line of the figure.

FIGURE 2. Schematic of the chromosomal region containing the A peptide gene of *Thermus thermophilus* cytochrome oxidase. The position of the gene is fixed by partial DNA sequence, and the positions of the ends of the genes were estimated using the known size of the mature polypeptide and its homology to other subunits I (see Fig. 4). H indicates a HindIII restriction site and P indicates a PstI site. The positions of the cyanogen bromide peptide fragments for which amino acid sequence was obtained are indicated below the line for the gene.

FIGURE 3. DNA sequence of a portion of the cytochrome oxidase subunit I gene of *Thermus thermophilus*. The region sequenced is in the C terminal half of the gene. The corresponding amino acid sequence is shown below the DNA sequence. Cyanogen bromide fragments 2,3, and 4 are encompassed by the DNA sequence and are shown in italics and underlined. The position of

hybridization to the DNA probe is indicated by underlining the DNA sequence. Comparison with FIGURE 1 will show that there is only one mismatch between the probe and the actual sequence.

FIGURE 4. Alignment of Subunit I sequences from several different species. Amino acid sequences for bovine[13], yeast[14], *Paracoccus denitrificans* (Pd)[15], *Bacillus* PS3 (PS3)[16], and partial sequences for *Thermus thermophilus* (Tt) are shown. Amino acid residues conserved in all subunit I sequences reported to date are indicated by an asterisk above the sequence. Positions at which the *Thermus thermophilus* subunit matches all four other subunits presented here are indicated by an asterisk below the sequence. The sequences were aligned pairwise using the program GAP from the University of Wisconsin Genetics Computer Group[17]. The alignments were then assembled into the composite figure, with minor discrepancies resolved in favor of preserving the alignment of residues conserved throughout all the above sequences.

Peptide Sequence	**Met**	**Pro**	**Arg**	**Arg**	**Tyr**	**Tyr**	**Thr**	**Tyr**	**Asn**	**Ala**	**Asp**	**Ile**	**Ala**
Codons possible	AUG	CCA CCU CCC CCG	CGA CGU CGC CGG AGA AGG	CGA CGU CGC CGG AGA AGG	UAU UAC	UAU UAC	ACA ACU ACC ACG	UAU UAC	AUU AAC	GCA GCU GCC GCG	GAU GAC	AUA AUU AUC	GC-
Codon chosen	AUG	CCC	CGG	CGG	UAC	UAC	ACC	UAC	AAC	GCC	GAC	AUC	GC
% use	100	63	51	51	95	95	63	95	99	74	98	89	(100)

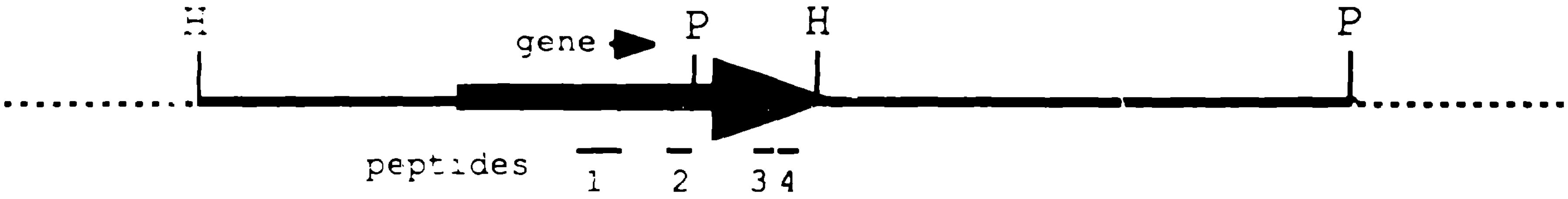

Fig 2

```
GGGACCATGGTCTGGGCCCACCACATGTTCACCGTGGGCGAGTCCACCCTCTTCCAGATCGCCTTCGCCTTCTTCACCGCCCTCATCGCC
---------+---------+---------+---------+---------+---------+---------+---------+---------+ 90
GlyThrMetValTrpAlaHisHisMetPheThrValGlyGluSerThrLeuPheGlnIleAlaPheAlaPhePheThrAlaLeuIleAla -

GTGCCCACGGGGGTCAAGCTCTTCAACATCATCGGCACCCTCTGGGGCGGGAAGCTGCAGATGAAGACCCCCCTCTACTGGGTTTTGGGC
---------+---------+---------+---------+---------+---------+---------+---------+---------+ 180
ValProThrGlyValLysLeuPheAsnIleIleGlyThrLeuTrpGlyGlyLysLeuGlnMetLysThrProLeuTyrTrpValLeuGly -

TTCATCTTCAACTTCCTCCTCGGGGGGATCACCGGGGTCATGCTCTCCATGACCCCCCTGGACTACCAGTTCCACGACTCCTACTTCGTG
---------+---------+---------+---------+---------+---------+---------+---------+---------+ 270
PheIlePheAsnPheLeuLeuGlyGlyIleThrGlyValMetLeuSerMetThrProLeuAspTyrGlnPheHisAspSerTyrPheVal -

GTGGCCCACTTCCACAACGTCCTCATGGCGGGCTCCGGCTTCGGGGCCTTCGCCGGCCTTTACTACTGGTGGCCCAAGATGACGGGCCGG
---------+---------+---------+---------+---------+---------+---------+---------+---------+ 360
ValAlaHisPheHisAsnValLeuMetAlaGlySerGlyPheGlyAlaPheAlaGlyLeuTyrTyrTrpTrpProLysMetThrGlyArg -

ATGTACGACGAGAGGCTGGGCCGGCTCCACTTCTGGCTCTTCCTCGTGGGCTACCTCCTCACCTTCCTGCCCCAGTACGCCCTGGGCTAC
---------+---------+---------+---------+---------+---------+---------+---------+---------+ 450
MetTyrAspGluArgLeuGlyArgLeuHisPheTrpLeuPheLeuValGlyTyrLeuLeuThrPheLeuProGlnTyrAlaLeuGlyTyr -

CTGGGGATGCCCCGGCGCTACTACACCTACAACGCCGACATCGCCGGCTGGCCCGAGCTCAACCTCCTCTCCACCATCGGCGCCTACATC
---------+---------+---------+---------+---------+---------+---------+---------+---------+ 540
LeuGlyMetProArgArgTyrTyrThrTyrAsnAlaAspIleAlaGlyTrpProGluLeuAsnLeuLeuSerThrIleGlyAlaTyrIle -

CTGGGCCTGGGCGGGCTGGTCTGGATCTACACCATGTGGAAAAGCCTCCGCTCCGGCCCCAAGGCCCCCGACAACCCTTGGGGCGGTTAC
---------+---------+---------+---------+---------+---------+---------+---------+---------+ 630
LeuGlyLeuGlyGlyLeuValTrpIleTyrThrMetTrpLysSerLeuArgSerGlyProLysAlaProAspAsnProTrpGlyGlyTyr -

ACCCTGGAGTGGCTCACCGCCTCGCCTCCCAAGGCCCACAACTTTGACGTGAAGCTT
---------+---------+---------+---------+---------+------- 687
ThrLeuGluTrpLeuThrAlaSerProProLysAlaHisAsnPheAspValLysLeu -
```

Fig 3

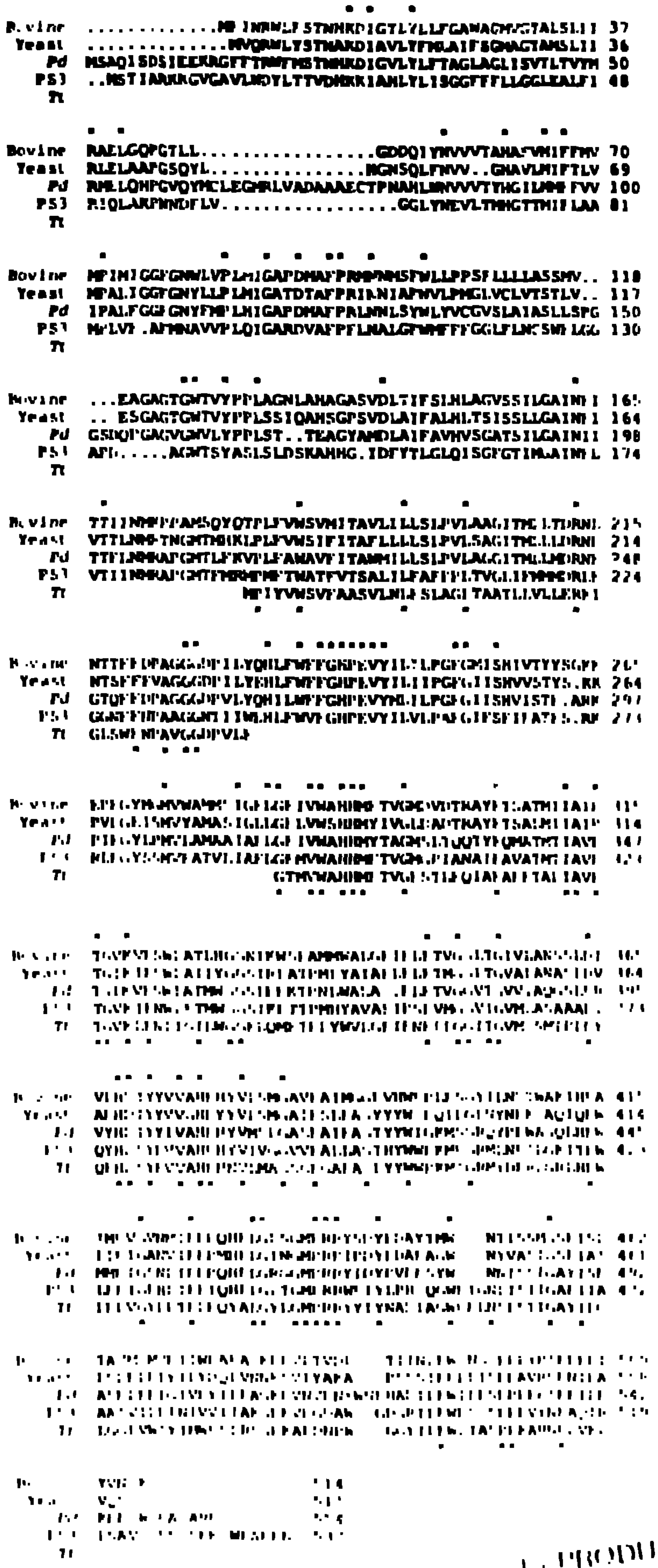

Fig 4