

## HALOBACTERIUM HALOBIUM Mn-SOD GENE: ARCHAEBACTERIAL AND EUBACTERIAL FEATURES

MARVIN L. SALIN\*, MARY V. DUKE, DIN-POW MA, and JOHN A.  
BOYLE

*Department of Biochemistry and Molecular Biology, P.O. Drawer BB, Mississippi  
State University, Mississippi State, MS 39762, USA*

A 1.8 kb *Pst*I fragment from *Halobacterium halobium* DNA was found to hybridize to synthetic oligonucleotide probes constructed by using the sequence of the N-terminus of a Mn-containing superoxide dismutase purified from *H. halobium*. The entire insert containing a 600-bp coding sequence for Mn-SOD and its 5' and 3' flanking regions was sequenced. The derived amino acid sequence of the structural gene showed a similarity to other manganic and iron-containing superoxide dismutases in normally conserved regions. Primer extension analysis of the *H. halobium* Mn-SOD mRNA showed that gene transcription begins 14 bases upstream of the translational start. A Shine-Dalgarno sequence and archaeobacterial consensus promoter sequences were observed. Several other promoter and terminator nucleotide sequences homologous to prokaryotic and eukaryotic organisms were found.

**KEY WORDS:** Archaeobacteria, halophilic protein, transcript mapping, superoxide dismutase, nucleotide sequence, amino acid sequence.

### INTRODUCTION

Archaeobacteria are microorganisms representing a third and unique line of evolution. The group consists of halophiles, methanogens and thermoacidophiles, which share features with eubacteria as well as eukaryotes.<sup>1</sup> Eubacterial features include prokaryotic cell structure and organization, whereas eukaryotic features include the presence of introns in stable RNA genes, an RNA polymerase that shows similarities to eukaryotic RNA polymerases, and a translation apparatus that is sensitive to antibiotics known to interact with eukaryotic ribosomes.<sup>2</sup>

Halophiles are characterized by growth in salt concentrations in excess of 1M and can usually be found growing in salt brines or salty bodies of water such as the Dead Sea.<sup>3</sup> These organisms generate ATP under anaerobic conditions by using a light-driven proton pump, bacteriorhodopsin, in coordination with an ATP synthase. Under aerobic conditions, they maintain a standard respiratory metabolism.<sup>3</sup> Therefore, halophilic organisms would be exposed transiently to high levels of oxygen and superoxide dismutase should be present in the organism at levels to insure sufficient oxyradical scavenging.

The enzyme from *Halobacterium halobium* is a manganic protein of about 40,000 Mr and consists of two equally sized subunits.<sup>4</sup> The gene has been cloned and the nucleotide sequence determined.<sup>5</sup> Here we report on the identification of the

\* To whom inquiries should be addressed.

transcriptional start site. The 5' -flanking region shows features in common with other archaeobacteria. In addition, some regions have been found which show homology to eubacteria while other regions show homology to eukaryotes.

## MATERIALS AND METHODS

A clone containing the Mn-SOD gene was isolated from a *Halobacterium halobium* cosmid genomic library.<sup>6</sup> A description of the Mn-SOD gene isolation procedure as well as the sequencing strategies and protocols have been published previously.<sup>5</sup> The transcriptional start site was determined by primer extension as described by Golden *et al.*<sup>7</sup> A synthesized 17-mer (TCGTGGTGCCACGTGAC) complementary to the 5' region of SOD mRNA was <sup>32</sup>P-labeled by polynucleotide kinase and used as the primer in the extension reaction.

## RESULTS

The complete nucleotide sequence for the 1.8 kb *Halobacterium halobium* DNA fragment is shown in Figure 1. The open reading frame comprises 600 nucleotides coding for 200 amino acids. The results of the primer extension (Figure 2) indicated that the transcript starts 14 nucleotides upstream from the initiation codon. The first base in the transcript is G. Unfortunately, this region is highly compressed on the gel and the DNA sequence is consequently difficult to unambiguously read. The problem of band compression on gels was also apparent in the original gene sequencing.<sup>5</sup> The final sequence determination was achieved by multiple analyses of the region and its complement.

The amino acid sequence of the open reading frame was compared and aligned to other manganic superoxide dismutases, and the results are shown in Figure 3. Major regions of homology are found interspersed throughout the molecule.

Based upon X-ray structural analysis and sequence alignment, His<sup>26</sup>, His<sup>81</sup>, Asp<sup>175</sup>, and His<sup>179</sup> have been tentatively assigned as ligands to the manganese.<sup>8,9</sup> These four amino acids have been aligned and labeled with asterisks. As can be seen in Figure 3, the positions of these critical amino acids, with the possible exception of His<sup>81</sup> in *E. coli*, have been conserved.

## DISCUSSION

The expression of archaeobacterial genes is less than completely understood. Fewer than 30 transcriptional start sites have been mapped for genes in this entire kingdom.<sup>10</sup> Of this number, greater than half represent either genes coding for stable RNA transcripts or genes found in viral-like particles. Therefore, relatively few genes for archaeobacterial proteins have been studied in detail. A close examination of the sequence depicted in Figure 1 and Figure 3 shows that the gene for Mn-SOD exhibits both eubacterial as well as eukaryotic features.

Previous work on archaeobacterial promoters has suggested the presence of conserved sequences in an A and B box arrangement.<sup>10,11</sup> The B box is a weakly conserved region located in the immediate vicinity of the transcriptional start site. The consensus

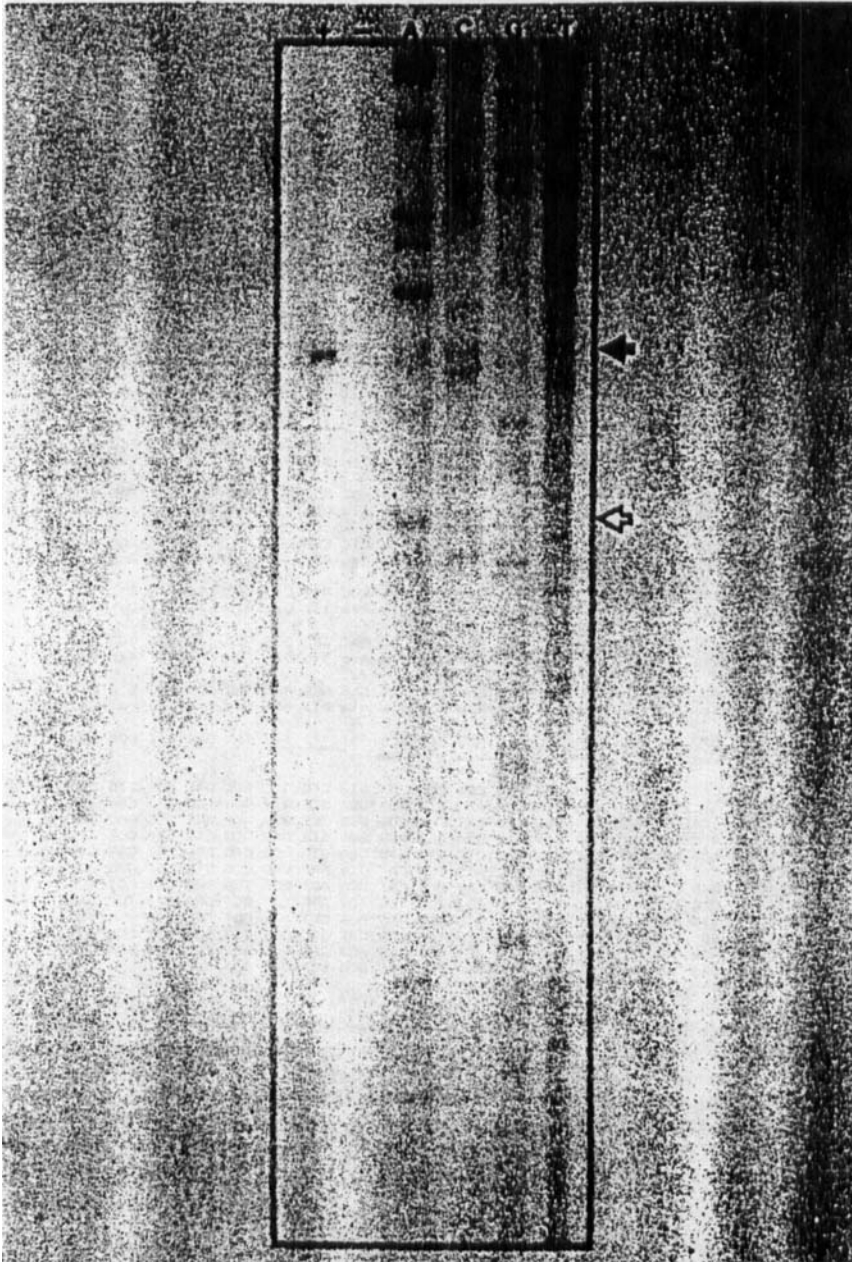
```

CT
-346 GCA GGC CTG GAA GGA CGG CGA AAC CGG CTA TCC GAT CGT GGA CGC CGG CAT GCG
-492 CCA GCT CCG CGC GGA AGC GTA CAT GCA CAA CCG CGT GCG GAT GAT CGT CGC CGC
-438 CTT CCT CAC GAA AGA CCT CCT GGT GGA CTG GCG GGC GGG CTA CGA CTG GTT CCG
-384 GGA GAA GCT CGC GGA CCA CGA CAC CGC CAA CGA CAA CGG CGG CTG GCA GTG GGC
-330 CGC CTC CAC GGG CAC CGA CGC CCA GCC GTA CTT CCG GGT GTT CAA CCC GAT GAC
-276 CCA GGG GGA GCG CTA CGA CCC CGA CGC CGA CTA CAT CAC CGA GTT CGT CCC CGA
-222 ACT CCG GGA CGT GCC CGC GGA CGA TCC ACA GCT GGC ACG AGC TGT CGC TGT CCG
-168 AGC GCC GCC GAC ACG CCC CGG AGT ATC CGG ACC CCA TCG TGG ACC ACA GCC AGC
-114 GCC GCG AGG ACG CGA TCG CGA TGT TCG AGC GTG GCC GCG GCB ACG AGT GAG CCC
-60 ACG ATC CTC ACG AAC ATT TAA CAT GAC GCC GCG TGA TCA CTG ATC CGG TGG ATT
-6 CCA CCG ATG AGC CAG CAC GAA CTC CCA TCG CTG CCG TAC GAC TAC GAC GCA CTC
Met Ser Gln His Glu Leu Pro Ser Leu Pro Tyr
49 GAA CCA CAC ATC AGT GAG CAG GTG GTC ACG TGG CAC CAC GAC ACC CAC CAC CAG
Glu Pro His Ile Ser Glu Gln Val Val Thr Trp His His Asp Thr His His Gln
103 AGC TAC GTG GAC GGC CTC AAC AGC GCC GAG GAG ACG CTG GCG GAG AAC CGT GAG
Ser Tyr Val Asp Gly Leu Asn Ser Ala Glu Glu Thr Leu Ala Asn Arg Glu
157 ACC GGC GAC CAC GCT TCG ACT GCC GGC GCG CTC GGG GAC GTC ACG CAC AAC GGC
Thr Gly Asp His Ala Ser Thr Ala Gly Ala Leu Gly Asp Val Thr His Asn Gly
211 TGT GGG CAC TAC CTC CAC ACG ATG TTC TGG GAG CAC ATG AGT CCC GAC GGG GGC
Cys Gly His Tyr Leu His Thr Met Phe Trp Glu His Met Ser Pro Asp Gly Gly
265 GGC GAG CCG TCC GGG GCG CTC GCC GAC CCG ATC GCG GCG GAC TTC GGC TCC TAC
Gly Glu Pro Ser Tyr Asp Ala Leu Ala Asp Arg Ile Ala Ala Asp Phe Gly Ser Tyr
319 GAG AAC TGG CCG GCT GAA TTC GAG GTG GCG GCC GGC GCG GCC AGC GGC TGG GCG
Glu Asn Trp Arg Ala Glu Phe Glu Val Ala Ala Gly Ala Ala Ser Gly Trp Ala
373 CTG CTC GTC TAC GAT CCG GTC GCC AAG CAG CTC CCG AAC GTG GCC ATC GTC GAC AAC
Leu Leu Val Tyr Asp Pro Val Ala Lys Gln Leu Arg Asn Val Ala Val Asp Asn
427 CAC GAC GAG GGC GCG CTC TGG GGC AGC CAC CCC ATC CTC GCC CTC GAC GTC TGG
His Asp Glu Gly Ala Leu Trp Gly Ser His Pro Ile Leu Ala Leu Asp Val Trp
481 GAG CAC TCC TAC TAC TAC GAC TAC GGC CCC GAC GCG GGC ACG TTC GTC GAC GCC
Gly His Ser Tyr Tyr Tyr Asp Tyr Gly Pro Asp Arg Gly Thr Phe Val Asp Ala
535 TTC TTC GAG GTG ATC GAC TGG GAC CCC ATC GCG GCG AAC TAC GAC GAC GTG GTG
Phe Phe Glu Val Ile Asp Trp Asp Pro Ile Ala Ala Asn Tyr Asp Asp Val Val
589 TCG CTG TTC GAG TGA CCG AAC ACG CTC CCG TGT TTT TTT CGC GTC ATG GCG GCT
Ser Leu Phe Glu ---
643 CAT CAG TGG CTG GCG TGA GTG CCG CGA CGT GTG CTC GTC ACC GAC CGC GTG CAC
697 GTG GTC GGT GAG CGC GTC GAA ATC GTA GCA GGT GCC GTC CCG TCG AGG CCA TGA
731 CGA CCA GCG AAG TGT CCA TCG CAT CAA GGA ACG TCC GGA ACA GCT CCC AGG TGT
805 CGA ACC CAC GGT GCG TGT ACC GCA GGG ACG GGA GTG GAG GCG GTT GTA GGT GGC
859 GCG AGT GAT GAA CCC GTA GCG CTC GTT GTC GAG GTA CCG CTG GGC GTC CAC GGT
913 CGT CGT GCG GGC TGC AAC GTC ATC GCG CAT GTA CAG CAG CCG CTG GCG GTG CCG
967 GTC AAG CAC CCA GAT GTC CCG GAG GCT CGT CCG ACT CGA TCA TGA CCG CGT CCA
1021 CGA GCC TGC ATG GCG TCT GTG CGT TGG TCG TCC AAC TGC GGT GAA CTA TCT GAC
1075 ATG TTG GGC GGA TAC ATG GAC TTC GGC GAG ACA GAT AAC CGT TGG GTC AAC TCG
1129 CTG GGT TTT CGT CCG GCG GCG CTG TCA GCA CAT GCC GGT GGG GCA AAT CCG AGT
1183 TCG ACG ACC TGT ACC CGT GTG ACT TCT ACG ACG CAG CCG CAT TCG ACG ACG
1237 ACC GCA TGT ACA CCG TCT ACG AGA TCG GCG GCG CCT GCT GCA G

```

FIGURE 1 Nucleotide sequence of *Halobacterium halobium* Mn-SOD gene and its flanking region. Presumed consensus archaeobacterial transcriptional promoters at the 5' end are noted by underlines containing dots. The arrowhead marks the transcriptional start site. Dashes above the nucleotides immediately preceding the open reading frame denote a putative modulator of eukaryotic translation. A double line at the 3' end denotes the poly T transcriptional termination signal; the arrows represent an inverted repeat required for formation of a loop structure. A 17-mer complementary to the overlined sequence was used in primer extension analysis.

sequence is (A/T)TG(A/C). There are two regions in the gene transcript that could be this box (Figure 1). The sequence CTGA ends three nucleotides upstream from the start site (arrow in Figure 1). The sequence GTGG is located immediately after the start site. These sequences are underlined with dots in Figure 1. The A box is a more highly conserved sequence ordinarily centered about 25 nucleotides upstream from box B and from the start of transcription. Its consensus sequence is usually rendered as TTA(T/A)ATA. The sequence TTAAACAT is found at positions -23 to -30. This appears to be an excellent candidate for the A box both in sequence and location.



**FIGURE 2** Determination of the transcriptional start site by the primer extension method. The plus (+) and minus (–) symbols are the extension reaction with and without mRNAs, respectively. Four lanes A,C,G,T represent dideoxy sequence reactions of the M13 single-stranded DNA template using a 17-mer oligonucleotide as a sequencing primer. The extended product is marked by a closed arrowhead and the translational initiation site is denoted by an open arrowhead.

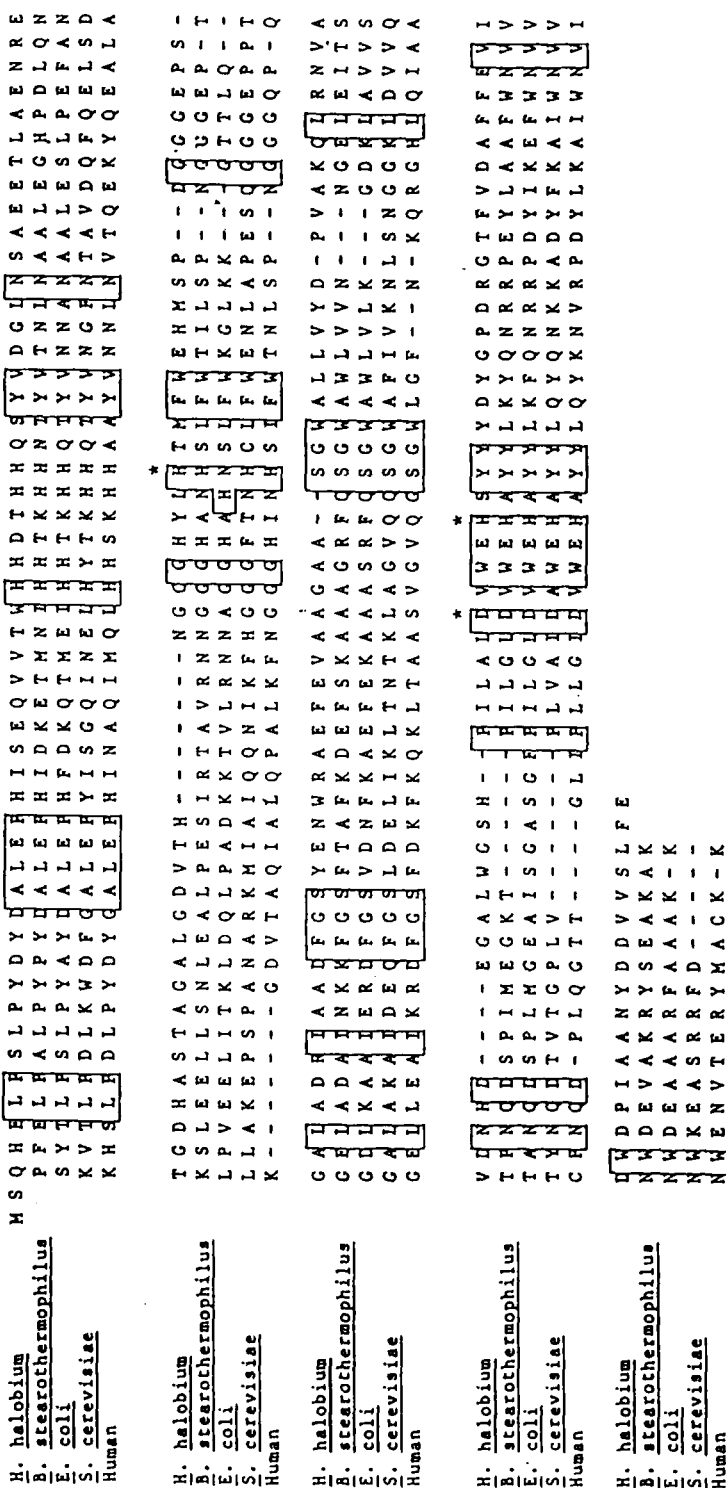


FIGURE 3 Amino acid sequence of *Halobacterium halobium* Mn-SOD in comparison to other sequenced Mn-SODs. Data from *Bacillus stearothermophilus*, *Escherichia coli*, *Saccharomyces cerevisiae*, and human liver Mn-SODs were obtained from a compilation in Bannister *et al.*\* Amino acids with greatest homology are in boxes. Asterisks are placed over amino acids postulated to be critical in metal liganding. Gaps have been introduced to achieve greatest sequence homology.

The 3' end of the 16S rRNA of *Halobacterium halobium* is 3' UCCUCCACU.<sup>12</sup> This sequence is remarkably similar to the sequences seen in eubacterial organisms.<sup>12</sup> The 5' end of the gene transcript, GGUGG, is a good Shine-Dalgarno sequence which is capable of binding to the 16S rRNA. Several halobacterial and other archaeobacterial gene transcripts for proteins have start sites within one, two, or three nucleotides of the translation initiation codon.<sup>10,11</sup> These genes appear to have sequences complementary to the 3' end or the 16S RNA within the gene rather than in the untranslated leader. Therefore, there seems to be no obvious positional rules directing translational initiation beyond the necessity for rRNA-mRNA complementary sequences. However, the fact that the archaeobacterial sequences are essentially the same as the eubacterial sequences may indicate that this method of assisting translation initiation predates the split between the two kingdoms. Eukaryotes do not use such complementary sequences and therefore have either lost the necessity or developed independently of the other two kingdoms prior to the appearance of Shine-Dalgarno-type interactions. The sequence, CCACCT, (shown by dots over the nucleotides) directly adjacent to the start codon is analogous to the consensus sequence used in eukaryotic translation.<sup>13</sup>

At the 3' end of the gene a series of T's are observed (double underline). This string of seven T's is preceded by an inverted repeat of 3 nucleotides. If one counts two of the seven T's as part of the inverted repeat, then the repeat becomes 5 nucleotides long and surrounds a potential loop of 4 bases. Potential stem and loop regions preceding a string of T's are characteristic of rho independent transcriptional termination sequences associated with eubacteria.

In conclusion, it would appear therefore, that this *H. halobium* gene possesses certain eukaryotic as well as eubacterial features in its promoter, translation initiation and terminator sequences.

### Acknowledgements

This work was supported by funds made available from the Mississippi Agricultural and Forestry Experiment Station.

### References

1. C.R. Woese, L.J. Magrum, and G.E. Fox, (1978) Archaeobacteria. *Journal Molecular Evolution*, 11; 245-252.
2. P.P. Dennis, (1986) Molecular biology of archaeobacteria. *Journal of Bacteriology*; 168, 471-478.
3. D.J. Kushner, (1985) The halobacteriaceae. In Woese, C.R., and Wolfe, R.S. (eds). *The Bacteria- A Treatise on Structure and Function*. Academic Press, NY, pp. 171-214.
4. M.L. Salin, and D. Oesterhelt, (1988) Purification of a manganese-containing superoxide dismutase from *Halobacterium halobium*. *Archives in Biochemistry and Biophysics*, 260; 806-810.
5. M.L. Salin, M.V. Duke, D. Oesterhelt and D-P. Ma, (1988) Cloning and determination of the nucleotide sequence of the Mn-containing superoxide dismutase gene from *Halobacterium halobium*. *Gene*, 70; 153-159.
6. P. Hegemann, A. Blanck, H. Vogelsang-Wenke, F. Lottspeich and D. Oesterhelt, (1987) The halopsin gene. I. Identification and isolation. *EMBO Journal*, 6; 259-264.
7. S.S. Golden, J. Brusslan and R. Haselkorn, (1986) Expression of a family of psbA genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans*R2. *EMBO Journal*, 5; 2789-2798.
8. W.C. Stallings, K.A. Patridge, R.K. Strong and M.L. Ludwig, (1985) The structure of manganese superoxide dismutase from *Thermus thermophilus* HB8 at 2.4Å resolution. *Journal of Biological Chemistry*, 260; 16424-16432.

9. J.V. Bannister, W.H. Bannister and G. Rotilio, (1987) Aspects of the structure, function, and applications of superoxide dismutase. *CRC Critical Reviews in Biochemistry*. CRC Press, Boca Raton, FL. pp. 111–180.
10. W. Zilling, P. Palm, W-D. Reiter, F. Gropp, G. Pühler and H-P. Klenk, (1988) Comparative evaluation of gene expression in archaeobacteria. *European Journal of Biochemistry*, 173; 474–482.
11. J.W. Brown, C.J. Daniels and J.N. Reeve (1989) Gene structure, organization, and expression in archaeobacteria. *CRC Critical Reviews in Microbiology* 16; 287–338.
12. Shine, J. and L. Dalgarno. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences USA*, 71; 1342–1346.
13. M. Kozak (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44; 283–292.

Accepted by Prof. G. Czapski