Plant Molecular Biology **18**: 399–402, 1992.

Update section

Sequence

Nucleotide sequence of a high-pI rice (*Oryza sativa*) -amylase gene

Ju-Kon Kim [1] and Ray Wu [1,2]*

[1] *Field of Botany and* [2] *Section of Biochemistry, Molecular and Cell Biology, Division of Biological Sciences, Cornell University, Ithaca, NY 14853, USA (*author for correspondence)*

Received 19 August 1991; accepted 28 August 1991

α-Amylase (EC 3.2.1.1) is an enzyme which catalyzes an α (1–4) endoglycolytic cleavage of amylose and amylopectin. During cereal seed germination, a number of hydrolytic enzymes including α-amylases and proteases are produced to hydrolyze starch and proteins to provide energy and building blocks for seedling growth. The α-amylase has been of considerable interest to plant biologists primarily because exogenously added gibberellic acid (GA_3) promotes the synthesis and secretion of large amounts of the enzyme by aleurone or scutellar tissue in several important cereals such as barley, wheat and rice [1, 7]. α-Amylase in cereal seeds is known to consist of several isozymes encoded by a multigene family [4, 5, 6, 7, 8, 11]. Many of those α-amylase genes have been cloned; their nucleotide and amino acid sequence determined. In rice, DNA slot-blot hybridization reveals at least five hybridization groups containing a total of ten α-amylase genes [4]. To date, two cDNA and three genomic rice α-amylase genes have been sequenced [4, 5, 11].

We have previously isolated three different α-amylase clones, *OSamy-a*, *OSamy-b*, and *OSamy-c*, from a rice (*Oryza sativa* cv. IR26) genomic library [12]. Here we report the nucleotide and the deduced amino acid sequence of the *OSamy-c* clone. The complete nucleotide sequence of the clone is shown in Fig. 1. Sequence analysis revealed the presence of three putative introns shown in Fig. 1, which interrupt the coding region in this clone, bounded by intron/exon splice junctions consistent with the consensus sequence for such junctions [2]. Sequence comparison was done on a VAX computer using the GAP program of the University of Wisconsin Genetics Computer Group package (GCG) [3] in conjunction with PUBLISH to align the nucleotide and amino acid sequences. The nucleotide sequence of *OSamy-c* was compared to several of the α-amylase genomic clones including *RAmy1A* and *RAmy1B* isolated and sequenced by Huang *et al.* [4, 5] from a different rice cultivar, M202. Huang *et al.* [5] reported a restriction map and part of DNA sequence information of *RAmy1B* including 5′ flanking and part of the coding regions [5]. DNA sequence of *OSamy-c* is almost identical to that of *RAmy1B* in corresponding regions. The sequences of *OSamy-c* and *RAmy1A* show 83% DNA sequence identity in the regions of the first three exons and introns, but only 68% identity in the fourth exon and no homology in 3′ untranslated regions. These results suggest that the two genes are different, but belong to the same α-amylase gene subfamily in rice. *OSamy-c* also shows 90% DNA sequence identity to *RAmy1A* in the 5′ upstream region up to −270 from the transcription start site whereas it does not have sequence identity with that of the *OSamy-a* and *OSamy-b* clones [12] in the same region except

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X52240.

```
-618  GTCGACAAGTTACACCATGATGAAAGAAATCCATGTCCGTACCACAACTCGCAACATGTA

-558  CTCCATACGTCCCATATTAAAAGATATTTTAGGTGAATATGACACATTCTATTACTACAG

-498  ACAGTTAAATATATGTGAACAATGTTATACATGTTTCTTTTTAAATGCTTTGTCATGCCT

-438  AATAAGAGTGCTTGTATTATCTTTATAGTGTTTTAAGCCTTTCAATATAAAAAAAAGGCC

-378  TTTCAACATATAGATAGACTTTTACAGACTTTGTCAAAATTTGTATGCATCTAAGATCTC

-318  TTATAACATGGAGAGGGAGTAATTTACAAGGTCGCTCTTCAGCAAAAAAATTCATGTTCA

-258  GCTCGCAGAAGCACAAGCAAGAAAGATGGAATTCCTACTGACTTGTGCCTTTTGAGTGCT

-198  CCATCTCTCAAGGCCATTAAATTGCCTATGGGCTCACCGGCCAATAACAAACTCCTACTG

-138  ACCAGTGTTATCCATCCAATCCAGTGTCCCAAAGCAACATTCAAGCCCAGCCAGGCCTCC

 -78  AAAAGTTGCAAGTTGAGCATGGCAAAATCCCCGGCAATTCTCGACTATAAATACCTCAGC
                    +1
 -18  AGACACACACCCAGCTTCATCAATCCACCCACCTCCGAAGTGTATCTGCAGCAGTGTCAG

 +39  CCGATCGAACACAATGGTGGCGAACAAACACTTCTTGTCCCTTTCCGTGCTCATCGTCCT
                M  V  A  N  K  H  F  L  S  L  S  V  L  I  V  L

 +99  CCTCGGCCTCTCCTCCAACTTGGCAGCCGGGCAAGTCCTGTTTCAGgtaagagatcgaca
       L  G  L  S  S  N  L  A  A  G  Q  V  L  F  Q

+159  tgagttgggttggctgcatcgaactgatcgtgtttttgtactactgagcctgagagaatg

+219  atgctcttgttgtttgatgatgatgcagGGTTTCAACTGGGAGTCGTGGAAGGAGAATGG
                                  G  F  N  W  E  S  W  K  E  N  G

+279  CGGGTGGTACAACCTGCTTATGGGCAAGCTAGGTGGACGACATCGCGGCCGGCATCACCC
       G  W  Y  N  L  L  M  G  K  L  G  G  R  H  R  G  R  H  H  P

+339  ACGTCTGGCTCCCTCGCGTCCCACTCTGTCGGCGAGCAAGgtcggtgctctctcgatccc
       R  L  A  P  S  R  P  T  L  S  A  S  K

+399  tccttcgtcgcaccattgccggcaaaatacatgcataggtcgttgaattgcttgaatgct

+459  gatctgcagGCTACATCGGGGCGGCTGTACGATCTGACGCGTCCAATCATCGAGGCGTTC
             A  T  S  G  R  L  Y  D  L  T  R  P  I  I  E  A  F

+519  CATGGCAAGGGCGTCCAGCGTGATCGCCGACAAGTCGTCATCAAGGCACACGGCGAGCAC
      H  G  K  G  V  Q  R  D  R  R  Q  V  V  I  K  A  H  G  E  H

+579  AAGGACGGCCGGGCATCTCTGCGACTCCCGCCTCGACTGGGGCCCGCACATGATCTGCCG
      K  D  G  R  A  S  L  R  L  P  P  R  L  G  P  A  H  D  L  P

+639  CGACGACCCCTACGGCGACGGCACCGGAAACCCGACACCGCGGCACATCGACCACCTCAA
      R  R  P  L  R  R  R  H  R  K  P  D  T  A  A  H  R  P  P  Q

+699  CAAGCGCGTCCACGGAGCTCATCGGCTGGCTCGACTGGCTCAAGATGGACATCGGCTTCG
      Q  A  R  P  R  S  S  S  A  G  S  T  G  S  R  W  T  S  A  S

+759  ACGCGTGGCCTCGACTTCGCCAAGGCTACTCCGCCGACATGCAAAGATCTACATCGATGC
      T  R  G  L  D  F  A  K  A  T  P  P  T  C  K  D  L  H  R  C

+819  CACCGAGCCGAGCTTCGCCGTGGCCGAGATATGGACGTCGATGGCGAACGCGGGGACGGC
      H  R  A  E  L  R  R  G  R  D  M  D  V  D  G  E  R  G  D  G

+879  AAGCGCAACTACGACCAGAACGCGCACCGGCAGGAGCTGGTCAACTGGGTCAATCGTGTC
      K  R  N  Y  D  Q  N  A  H  R  Q  E  L  V  N  W  V  N  R  V

+939  GGCGGCCAACAGCAATGCCACGGCGTTCGACTTCACCACCAAGGGCATCCTCAACGTGCC
      G  G  Q  Q  Q  C  H  G  V  R  L  H  H  Q  G  H  P  Q  R  A

+999  GTGGAGGGCGATCGAGCTGTGGCTCCGCGGCGAGGACGCAAGGCGCCCGGCATGATCGGG
      V  E  G  D  R  A  V  A  P  R  R  G  R  K  A  P  G  M  I  G

+1059 TGGTGGCCGGCCAAGGCGGACCTTCGTCGACAACCACGACACCGGCTCGACGCAGCACCT
      W  W  P  A  K  A  D  L  R  R  Q  P  R  H  R  L  D  A  A  P

+1119 GTGGCCGTTTCCCTCCGACAAGGATATGCCTACATCCTCACCCACCCCGGCAACCCATGC
      V  A  V  S  L  R  Q  G  Y  A  Y  I  L  T  H  P  G  N  P  C
```

```
+1179   ATCgtgagtagccaactcgattagaaattctgaatcatcctgctgcaaactgatcgatga
        I

+1239   actgaagataaattctgcaaaactctttcagTTCTACGACCATTTCTTCGACTGGGGTCT
                                        F  Y  D  H  F  F  D  W  G  L

+1299   CAAGGATGAGATCGAGTGCCTCGTGTCAATCAGAAACCGGCAGGGGATCCACCCGGCGCA
         K  D  E  I  E  C  L  V  S  I  R  N  R  Q  G  I  H  P  A  Q

+1359   GGGGCCTGGTTGCTGGCTCTTGCGAGTTGAAAAGGGCAGATGCTCGCAGAGATGGAAGCC
         G  P  G  C  W  L  L  R  V  E  K  G  R  C  S  Q  R  W  K  P

+1419   ACGAGAGGGAAATTTACGAATGTGCCCATGCGCCGGCGTATTTAGGGCAGCTTGTTTACT
         R  E  G  N  L  R  M  C  P  C  A  G  V  F  R  A  A  C  L  L

+1479   CTAGTAAAAGAAAAGTATCACCGATCACGACATATCGTTTAGAAAGTGTGTGAGTTAGGC

+1539   ATGGGCGAACCACATCCTTTACGACGAGGCCTCTCGATCTAGGTCTCTCGATCTGCTAGA

+1599   CGAGAGATCAACCAACAAAAGCTACTATACATGATATACACCGTGGCATGATACAAATCC

+1659   TTACAAATAGCAACAATTATTGTTCAATAGCCATATCAGTTGTCTAGGCTTACCAATATA

+1719   CTCCCTCTGTTCTTATATATCGTTTGATTTTTTTTCTTAGTTTAAAGTTATTTGACTAAC

+1779   TTATAGAAAAAATAGCAATATTTATGACCGAAATTAGTTTTATTCAATTTAAAGTTGAAT

+1839   ATATTTTGATAAGATGTTTGCTTTGTATTGAAAATATTACTACATTTTACTATAAACTTG

+1899   GTCAAATTTAAGAAGTTTTAACTAAGAAAAAAGTCAAACGACTTATAAAACTAAAACGGA

+1959   GGGACTGTCCTAAAAGGTGGAATTTCATATTGCTAGTATGAGATCAATGCTAGTAACTCT

+2019   TTGCCATAATGAGCTTTGTCTAGGCTATAAAAAATATATGGATCC
```

Fig. 1. Nucleotide sequence and predicted amino acid sequence of *OSamy-c*. The translation initiation codon ATG, the putative TATAAATA box, the termination codon TAG, and the polyadenylation signal AATATA are underlined. Introns are shown in lower case. The nucleotide sequence is numbered from the putative transcription start at position +1 which was chosen by comparing with other known rice α-amylases genes [4, 11]. The single-letter amino acid code is used. The amino acid sequence is positioned under the first nucleotide of each codon.

for the pyrimidine box which is conserved in the upstream sequence of GA-inducible genes [4, 12]. During rice seed germination the accumulation of the mRNA corresponding to this high-pI subfamily was stimulated by exogenous GA_3 (Kim, Cao, and Wu, manuscript in preparation).

The amino acid sequence of the *OSamy-c* gene product was derived from DNA sequence data. The largest open reading frame generated after intron-splicing of the *OSamy-c* clone can code for an α-amylase containing 383 amino acids (molecular mass 45 kDa), while *RAmy1A* can code for a protein with 428 amino acids. Aligning the predicted protein sequence of *OSamy-c* with that of *RAmy1A* using the GAP program, indicates that there is only 63% identity and 82% similarity between the amino acid sequences. These differences are scattered over the length of the protein except for the first 51 amino acids where two genes are highly conserved with regard to both nucleotide and peptide sequence. Thus, the two genes are less conserved at the level of peptide sequence than that of nucleotide sequence.

It is interesting to note that the α-amylase encoded by the *OSamy-c* clone appears to be a basic protein. It has many more positively charged amino acids than negatively charged ones (52 Arg, 16 Lys; 12 Glu, 18 Asp), resulting in a very high calculated pI value (pI 11.7). Even if the tentative signal peptide (first 31 amino acids) is removed, the pI value still remains to be very high (pI 11.5). This was calculated by using the ISOELECTRIC program of GCG and it was assumed that there was no post-translational modification such as phosphorylation of the protein. Deduced protein sequences for several cereal α-amylase isoen-

zymes are available [11]. No α-amylase isozyme, however, has been reported to have such a high calculated pI value. Two deduced polypeptides from pOS103 and pOS137 have predicted isoelectric points of approximately 6.0 [11]. It has also been demonstrated by isoelectric focusing that germinating rice seed produces three α-amylase isozymes whose pI values seem to be lower than 6.0 [9]. Therefore, the exceptionally high calculated pI value of *OSamy-c* gene product is very unusual although we do not rule out of the possibility that *in vivo* pI value is different due to posttranslational modification.

Acknowledgements

We are grateful to Tsai-Mei Ou-Lee for her gift of the *OSamy-c* genomic clone. J.K. was supported by a predoctoral fellowship from the Rockefeller Foundation and the Rural Development Administration of Korea. This work was supported by research grant RF86058, No. 60, from the Rockefeller Foundation.

References

1. Akazawa T, Hara-Nishimura I: Topographic aspects of biosynthesis, extracellular secretion, and intracellular storage of proteins in plant cells. Annu Rev Physiol 36: 441–472 (1985).
2. Brown JWS: A catalogue of splice junctions and putative branch sequences from plant introns. Nucl Acids Res 14: 9549–9559 (1986).
3. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
4. Huang N, Sutliff TD, Litts JC, Rodriguez RL: Classification and characterization of the rice α-amylase multigene family. Plant Mol Biol 14: 655–668 (1990).
5. Huang N, Koizumi N, Reinl S, Rodriguez RL: Structural organization and differential expression of rice α-amylase genes. Nucl Acids Res 18: 7007–7014 (1990).
6. Huttly AK, Martienssen RA, Baulcombe DC: Sequence heterogeneity and differential expression of the α-Amy2 gene family in wheat. Mol Gen Genet 214: 232–240 (1988).
7. Jacobsen JV, Chandler PM: In: Davies PJ (ed) Plant Hormones and their Role in Plant Growth and Development, pp. 164–193. Martinus Nijhoff, Dordrecht, The Netherlands (1987).
8. Khursheed B, Rogers JC: Barley α-amylase genes. J Biol Chem 263: 18953–18960 (1988).
9. Knox CAP, Sonthayanon B, Chandra GR, Muthukrishnan S: Structure and organization of two divergent α-amylase genes from barley. Plant Mol Biol 9: 3–17 (1987).
10. Miyata S, Akazawa T: Enzymic mechanism of starch breakdown in germinating rice seeds 12. Biosynthesis of α-amylase in relation to protein glycosylation. Plant Physiol 70: 147–153 (1982).
11. O'Neill SD, Kumagai MH, Majumdar A, Huang N, Sutliff TD, Rodriguez RL: The α-amylase genes in *Oryza sativa*: Characterization of cDNA clones and mRNA expression during seed germination. Mol Gen Genet 221: 235–244 (1990).
12. Ou-Lee TM, Turgeon R, Wu R: Interaction of a gibberellin-induced factor with the upstream region of an α-amylase gene in rice aleurone tissue. Proc Natl Acad Sci USA 85: 6366–6370 (1988).