Volume 327, number 3, 251–255 FEBS 12758 August 1993
© 1993 Federation of European Biochemical Societies 00145793/93/$6.00

Molecular cloning and sequencing of the cDNA coding for a calcium-binding protein regucalcin from rat liver*

Noriaki Shimokawa and Masayoshi Yamaguchi

Laboratory of Metabolism and Endocrinology, Graduate School of Nutritional Sciences, University of Shizuoka, 52-1 Yada, Shizuoka City 422, Japan

Received 5 May 1993; revised version received 11 June 1993

The cDNA of a Ca^{2+}-binding protein regucalcin was cloned from a rat liver cDNA library which was constructed in λZAPII by immunoscreening. Positive clones were obtained from which spanned the region of interest, and they gave a sequence of 1.7 kb by sequencing with the dideoxynucleotide method. Analysis of the sequence of the cloned cDNA showed that the cDNA encoded the complete amino acid sequence of the regucalcin molecule. Regucalcin was composed of 299 amino acid residues and its molecular weight was estimated to be 33,388 Da. The hydropathy profile of regucalcin showed a highly hydrophilic character. The nucleotide and amino acid sequences of regucalcin did not have statistically significant homology, as compared with the registered sequences which are found in the EMBL and GenBank databases containing several other Ca^{2+}-binding proteins (calmodulin, calbindin-D28k and S-100β). The regucalcin molecule did not contain the EF-hand motif as a Ca^{2+}-binding domain. The present study demonstrates that regucalcin is a unique Ca^{2+}-binding protein in the liver of rats.

Regucalcin; cDNA cloning; Primary structure; Rat liver

1. INTRODUCTION

Calcium ion (Ca^{2+}) plays a role as an important second messenger signal in a variety of pathways to produce a Ca^{2+}-mediated physiological responses in many cells. The Ca^{2+} signal is transmitted to an intracellular response partly via a family of Ca^{2+}-binding proteins [1]. Recently, it has been reported that a novel Ca^{2+}-binding protein regucalcin, which differs from calmodulin, is distributed in the hepatic cytosol of rats [2,3]. The isoelectric point of regucalcin was estimated as 5.20, and the Ca^{2+}-binding constant was found to be 4.19×10^5 M^{-1} by equilibrium dialysis [4]. This protein has a reversible effect on the activation and inhibition of various enzymes by Ca^{2+} in liver cells [4–6]. Regucalcin may play a cell physiological role in the regulation of liver cell function related to Ca^{2+}.

On the other hand, the nucleotide and amino acid sequences of regucalcin have not been reported thus far. In the present study, therefore, we have attempted the cloning of a complementary deoxyribonucleic acid (cDNA) and the determination of a complete amino acid sequence for regucalcin in rat liver. It was found that the regucalcin molecule consists of 299 amino acid residues and its sequence does not have significant homology compared with the sequences of other proteins.

Correspondence address: M. Yamaguchi, Laboratory of Metabolism and Endocrinology, Graduate School of Nutritional Sciences, University of Shizuoka, 52-1 Yada, Shizuoka City 422, Japan. Fax: (81) (54) 264 5580.

*The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank Nucleotide Sequence Databases with the following accession number D14327.

2. MATERIALS AND METHODS

2.1. *Chemicals*

Deoxycytidine 5′-[α-^{35}S]thiotriphosphate ([^{35}S]dCTP; 37 TBq/mmol) for DNA sequencing and nitrocellulose membrane (Hybond C) for immunoscreening were obtained from Amersham (Buckinghamshire, UK). Host bacterial cells (SURE Strain) and phages (native and R408 helper phage) were supplied by Stratagene (La Jolla, CA). All media used for the growth of bacteria and phage were purchased from Gibco Laboratories (Grand Island, NY). Reagents (analytical grade) and enzymes were obtained from Stratagene and Sigma Chemical Co. (St. Louis, MO). Any water and solutions used for the preparation of ribonucleic acid (RNA) were treated with chemical diethylpyrocarbonate (DEPC, Sigma) to inhibit RNase activity.

2.2. *Isolation of RNA*

Total RNAs were prepared as described [7]. Liver was quickly removed from Wistar male rats (3-week-old; Japan SLC Inc., Hamamatsu, Japan), rinsed with ice-cold 0.25 M sucrose solution, and homogenized in buffer solution containing 4 M guanidinium thiocyanate, 25 mM sodium citrate (pH 7), 0.5% sarcosyl, 0.1 M 2-mercaptoethanol and 2 M sodium acetate. Total RNAs were extracted by vigorous shaking in a mixture of phenol, chloroform and isoamyl alcohol, and the phase was separated by centrifugation at $10,000 \times g$ for 20 min at 4°C. RNA located in the aqueous phase was precipitated with isopropanol at −20°C. The RNA precipitate was pelleted by centrifugation, and the pellet was redissolved in 0.5% sodium dodecyl sulfate containing DEPC. The poly(A)$^+$ RNA was purified by using oligo (dT)-cellulose column.

2.3. *Construction of rat liver cDNA library*

The first strand of cDNA was synthesized from 5 μg poly(A)$^+$ RNA

using 50 units Moloney-Murine Leukemia Virus Reverse Transcriptase (M-MuLVRT), and oligo $(dT)_{18}$ primer-linker which has *Xho*I restriction enzyme recognition site. Double-stranded cDNA was generated by replacement synthesis with *E. coli* RNase H and DNA polymerase I [8]. After an *Eco*RI adaptor was added, cDNA was ligated with the *Xho*I–*Eco*RI-digested phage expression vector, a λZAPII (Stratagene) [9], and subsequently packaged into native phages using in vitro packaging extracts.

2.4. *Screening of library*

Nitrocellulose membranes (Hybond C) were soaked with 10 mM isopropyl-thio-β-D-galactoside (IPTG) and dried. Approximately 1×10^6 phage of a rat liver cDNA library were plated onto 20 dishes (diameter 140 mm). After 3.5 h preincubation at 42°C, the dishes were overlayed with the pretreated membranes and incubated for an additional 3.5 h at 37°C. The removed membranes were incubated with rabbit-anti-regucalcin antiserum (1 : 200 dilution) [10] for 120 min. Membranes were then incubated with an anti-rabbit second antibody conjugated to alkaline phosphatase, and finally developed using 0.35 mM nitro blue tetrazolium (NBT) and 0.4 mM 5-bromo-4-chloro-3-indolyl phosphate (BCIP).

2.5. *In vivo excision*

As a part of cDNA is located in the plasmid Bluescript within the phage vector, an automatic excision of the Bluescript vector instead of the subcloning was performed according to the method described in an instruction manual of λZAPII [9]. The purified phage lysate combined with *E. coli* SURE cells was mixed with R408 helper phage and incubated. To obtain the excised Bluescript, the excised DNA was infected with SURE cells and plated for selection on LB plates containing ampicillin (50 μg/ml).

2.6. *DNA sequencing*

Nucleotide sequencing was performed by the method of dideoxynucleotide termination [11] using the Sequenase system (US Biochemical Corp. Cleveland, OH) with 3 μg denatured double-stranded DNA as a template [12]. The cDNA clones were sequenced from both directions by the overlapping manner after characterization by restriction enzyme analysis. The sequences of cDNA were assembled and analyzed by the GENETYX program (SDC Inc., Tokyo).

3. RESULTS

Messenger RNA was isolated from the liver of male Wistar rats by using guanidium thiocyanate and oligo (dT) chromatography. The Poly(A)$^+$ RNA (5 μg), containing sizes in the range of 0.3–4.3 kb, was used for the λZAPII cloning system. By the enzymatic reaction with reverse transcriptase, the cDNA library which consists of 3.2×10^7 independent clones was obtained. Using a rabbit-anti-regucalcin antiserum, 14 immunoreactive clones were detected after the first screening for 1×10^6 recombinant plaques. Finally, eight positive clones were obtained by repeated immunoscreening. The inserts of eight clones obtained from the cDNA library were identified, and the insert size of the cDNA was 1.70 kb. This cDNA of 1.70 kb contained a complete open reading frame of 0.897 kb which is the cDNA encoding a molecule of regucalcin, as shown in Fig. 1.

The context surrounding the ATG codon signaling the start of translation, was as defined by Kozak [GCCA(or G)CCATGG] [13]. The translation start site (GCGACCATGT) in the 5′-noncoding region of the regucalcin nucleotide was fitted to the consensus sequence for translation start as defined by Kozak. Thus, we chose methionine as a position 1 of the amino acid sequence of regucalcin (Fig. 1). The number of amino acids in regucalcin as predicted from the nucleotide was 299. The molecular mass of the predicted amino acid sequence was calculated to be 33,388 Da. The amino acid composition of regucalcin is summarized in Table I, and shows that regucalcin has a relatively higher content of glycine, valine, asparatic acid and serine.

The hydropathy profile of regucalcin was calculated according to the method of Kyte and Doolittle [14]. This result is shown in Fig. 2. There was a hydrophobic sequence in both N-terminal and C-terminal regions of the regucalcin molecule. However, regucalcin showed a hydrophilic character as a molecule.

4. DISCUSSION

The present study was undertaken to characterize the molecular structure of regucalcin. From rat liver regucalcin cDNA cloning, a cDNA clone with an insert of 1.7 kb was obtained. Northern hybridization of rat liver mRNA with the obtained regucalcin cDNA probe gave a definite band of about 1.7 kb confirming the full-length size of the regucalcin cDNA clone, as reported previously [15,16]. The amino acid sequence deduced from the regucalcin cDNA coincided completely with the amino acid sequence of regucalcin as analyzed by digestion with proteases. Moreover, the molecular mass of regucalcin consisted of 299 amino acid residues and was calculated to be 33,388 Da. This value was fitted in the result estimated by SDS-polyacrylamide gel electrophoresis for regucalcin (data not shown). Also, it is known that the isoelectric point (pI) of regucalcin is 5.20 [4]. This coincided with the pI of the deduced amino acid sequence. Thus, regucalcin was coded from the obtained cDNA clone.

The nucleotide sequence from cDNA cloning for regucalcin was compared to that of other proteins. A computer-assisted search (FASTA) [17] of the nucleotide sequence of rat liver regucalcin showed that the nucleotide sequence did not have statistically significant homology with the registered sequences which are found in the EMBL and GenBank databases. Moreover, the predicted amino acid sequence of regucalcin did not have significant homology with several other Ca^{2+}-binding proteins; regucalcin had low homology for calmodulin [18], calbindin-D28k [19], and S-100β [20] with homologies of 13.3, 16.3 and 11.0%, respectively. These values from the best homology matching for amino acid sequences were obtained by the introduction of gaps to maximize homology. In fact, the homology was only slight. Thus, regucalcin is a novel protein, and it entirely differs from other Ca^{2+}-binding proteins.

Furthermore, regucalcin showed a hydrophilic character. In particular, regucalcin had comparatively higher hydrophilic regions at positions between 100 and

```
TGGATGCTGGAGTGTTTCCTTTGTCTTCTATTTTAAAGATATCTTGAAAAAAACCTGTCACTGTCCTTTTCCTGCGACC

ATG.TCT.TCC.ATC.AAG.ATT.GAA.TGT.GTT.TTA.AGG.GAG.AAC.TAC.AGG.TGT.GGG.GAG.TCC.CCT.
Met-Ser-Ser-Ile-Lys-Ile-Glu-Cys-Val-Leu-Arg-Glu-Asn-Tyr-Arg-Cys-Gly-Glu-Ser-Pro-
                                        10                                      20

GTG.TGG.GAG.GAG.GCA.TCA.AAG.TGT.CTG.CTG.TTT.GTA.GAC.ATC.CCT.TCA.AAG.ACT.GTC.TGC.
Val-Trp-Glu-Glu-Ala-Ser-Lys-Cys-Leu-Leu-Phe-Val-Asp-Ile-Pro-Ser-Lys-Thr-Val-Cys-
                                        30                                      40

CGA.TGG.GAT.TCG.ATC.AGC.AAT.CGA.GTG.CAG.CGA.GTT.GGT.GTA.GAT.GCC.CCA.GTC.AGT.TCA.
Arg-Trp-Asp-Ser-Ile-Ser-Asn-Arg-Val-Gln-Arg-Val-Gly-Val-Asp-Ala-Pro-Val-Ser-Ser-
                                        50                                      60

GTG.GCA.CTT.CGA.CAG.TCA.GGA.GGC.TAT.GTT.GCC.ACC.ATT.GGA.ACC.AAG.TTC.TGT.GCT.TTG.
Val-Ala-Leu-Arg-Gln-Ser-Gly-Gly-Tyl-Val-Ala-Thr-Ile-Gly-Thr-Lys-Phe-Cys-Ala-Leu-
                                        70                                      80

AAC.TGG.GAA.GAT.CAA.TCA.GTA.TTT.ATC.CTA.GCC.ATG.GTG.GAT.GAA.GAT.AAG.AAA.AAC.AAT.
Asn-Trp-Glu-Asp-Gln-Ser-Val-Phe-Ile-Leu-Ala-Met-Val-Asp-Glu-Asp-Lys-Lys-Asn-Asn-
                                        90                                     100

CGA.TTC.AAT.GAT.GGG.AAG.GTG.GAT.CCT.GCT.GGG.AGA.TAC.TTT.GCT.GGT.ACC.ATG.GCT.GAG.
Arg-Phe-Asn-Asp-Gly-Lys-Val-Asp-Pro-Ala-Gly-Arg-Tyr-Phe-Ala-Gly-Thr-Met-Ala-Glu-
                                       110                                     120

GAA.ACC.GCC.CCA.GCT.GTT.CTG.GAG.CGG.CAC.CAA.GGG.TCC.TTG.TAC.TCC.CTT.TTT.CCT.GAT.
Glu-Thr-Ala-Pro-Ala-Val-Leu-Glu-Arg-His-Gln-Gly-Ser-Leu-Tyr-Ser-Leu-Phe-Pro-Asp-
                                       130                                     140

CAC.AGT.GTG.AAG.AAA.TAC.TTT.AAC.CAA.GTG.GAT.ATC.TCC.AAT.GGT.TTG.GAT.TGG.TCC.CTG.
His-Ser-Val-Lys-Lys-Tyr-Phe-Asn-Gln-Val-Asp-Ile-Ser-Asn-Gly-Leu-Asp-Trp-Ser-Leu-
                                       150                                     160

GAC.CAT.AAA.ATC.TTC.TAC.TAC.ATT.GAC.AGC.CTG.TCC.TAC.ACT.GTG.GAT.GCC.TTT.GAC.TAT.
Asp-His-Lys-Ile-Phe-Tyr-Tyr-Ile-Asp-Ser-Leu-Ser-Tyr-Thr-Val-Asp-Ala-Phe-Asp-Tyr-
                                       170                                     180

GAC.CTG.CCA.ACA.GGA.CAG.ATT.TCC.AAC.CGC.AGG.ACT.GTT.TAC.AAG.ATG.GAA.AAA.GAT.GAA.
Asp-Leu-Pro-Thr-Gly-Gln-Ile-Ser-Asn-Arg-Arg-Thr-Val-Tyr-Lys-Met-Glu-Lys-Asp-Glu-
                                       190                                     200

CAA.ATC.CCA.GAT.GGA.ATG.TGC.ATT.GAT.GTT.GAG.GGG.AAG.CTT.TGG.GTG.GCC.TGT.TAC.AAT.
Gln-Ile-Pro-Asp-Gly-Met-Cys-Ile-Asp-Val-Glu-Gly-Lys-Leu-Trp-Val-Ala-Cys-Tyr-Asn-
                                       210                                     220

GGA.GGA.AGA.GTA.ATT.CGC.CTA.GAT.CCT.GAG.ACA.GGG.AAA.AGA.CTG.CAA.ACT.GTG.AAG.TTG.
Gly-Gly-Arg-Val-Ile-Arg-Leu-Asp-Pro-Glu-Thr-Gly-Lys-Arg-Leu-Gln-Thr-Val-Lys-Leu-
                                       230                                     240

CCT.GTT.GAT.AAA.ACA.ACT.TCA.TGC.TGC.TTT.GGA.GGG.AAG.GAT.TAC.TCT.GAA.ATG.TAC.GTG.
Pro-Val-Asp-Lys-Thr-Thr-Ser-Cys-Cys-Phe-Gly-Gly-Lys-Asp-Tyr-Ser-Glu-Met-Tyr-Val-
                                       250                                     260

ACA.TGT.GCC.AGG.GAT.GGG.ATG.AGC.GCC.GAA.GGT.CTT.TTG.AGG.CAG.CCT.GAT.GCT.GGT.AAC.
Thr-Cys-Ala-Arg-Asp-Gly-Met-Ser-Ala-Glu-Gly-Leu-Leu-Arg-Gln-Pro-Asp-Ala-Gly-Asn-
                                       270                                     280

ATT.TTC.AAG.ATA.ACA.GGT.CTT.GGG.GTC.AAA.GGA.ATT.GCT.CCA.TAT.TCC.TAT.GCA.GGG.TAA
Ile-Phe-Lys-Ile-Thr-Gly-Leu-Gly-Val-Lys-Gly-Ile-Ala-Pro-Tyr-Ser-Tyr-Ala-Gly-***
                                       290                                     300

ACTGCAGCTCTTCCTTGCTGTCAGAAGAAAAAGCTTGAAGACAACTGAGAATTAAACTGCAGCTCTTCCTTGCTGTCAG
AAGAAAAAGCTTGAAGACAACTGAGAATTAAGGGAGAGAAATCAATGAACTTTCATATTGTTTTTTTAATGAGGCAGTG
ATATTGACATGGTTAAACTGCTTTAATTTACACTTTTGATTGGGTGCTGGGGAATAAACCTAAAGCCATGGCATATTAA
```

Fig. 1. Nucleotide and deduced amino acid sequences of the rat regucalcin cDNA. The predicted amino acid sequence was presented below the nucleotide sequence. The numbers in the sequence refer to the positions of amino acids; position 1 represents methionine (Met) residue. The termination codon (TAA) was indicated by ***. The putative polyadenylation signal AATAAA is underlined.

200 of the amino acid sequence. This hydrophilic region may be a functional domain which is related to the binding of Ca^{2+}. Many Ca^{2+}-binding proteins have the common structural feature of an EF-hand, which binds

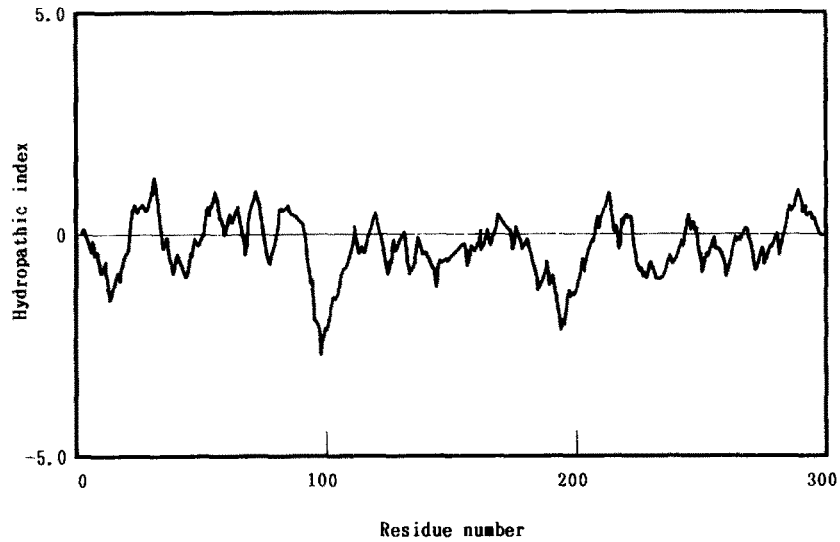

Fig. 2. Hydropathy profiles of rat regucalcin. Hydropathy analysis was performed according to Kyte and Doolittle for the 11 amino acid residues. The N-terminal was on the left and the C-terminal was on the right of the figure. Area of relative hydrophobicity or hydrophilicity was seen the above or below of horizontal line in the figure, respectively.

Ca^{2+} [21]. The S-100 protein has two EF-hands in the region with a relatively high hydrophilicity [22]. The most common EF-hand is composed of the helix-loop-helix-domain. The prototype loop consists of 12 amino acids, of which five have a carboxyl (or a hydroxyl group) in their side chain, precisely spaced so as to coordinate the Ca^{2+} [23]. Analysis of the structure of the EF-hand from the regucalcin sequence did not give the expected pattern of amino acids conforming to the typical EF-hand structure of a Ca^{2+}-binding site. Regucalcin may differ from a Ca^{2+}-binding protein of the EF-hand type, although further investigation is required to identify the sites of Ca^{2+}-binding. Further analysis suggests that the regucalcin sequence does not contain a potential *N*-linked glycosylation site (Asn-X-Ser (or Thr)] [24].

In conclusion, we have identified a regucalcin cDNA encoding a complete regucalcin molecule, and have determined the primary structure of regucalcin. Regucalcin may be a unique Ca^{2+}-binding protein.

Table I

Amino acid composition of regucalcin from rat liver

Amino acid	Mol residue/mol protein	Ratio (%)
Alanine	18	6.02
Arginine	16	5.35
Asparagine	11	3.68
Aspartic acid	24	8.03
Cysteine	10	3.34
Glutamine	9	3.01
Glutamic acid	16	5.35
Glycine	25	8.36
Histidine	3	1.00
Isoleucine	16	5.35
Leucine	20	6.69
Lysine	19	6.35
Methionine	7	2.34
Phenylalanine	11	3.68
Proline	12	4.01
Serine	23	7.69
Threonine	14	4.68
Tryptophan	5	1.67
Tyrosine	15	5.02
Valine	25	8.36

Acknowledgement: We would like to thank Dr. Yasuhiro Tomooka (Institute of Physical and Chemical Research, Japan) for helping us with the homology search of regucalcin cDNA and Dr. Mituo Hayashi (Asahi Chemical Industry Co., Ltd., Japan) for excellent help in amino acid analysis of regucalcin.

REFERENCES

[1] Wasserman, R.H. (1991) in: Novel Calcium-Binding Proteins (C.W. Heizmann, Ed.), pp. 7–12, Springer, Berlin.
[2] Yamaguchi, M. and Yamamoto, T. (1978) Chem. Pharm. Bull. 26, 1915–1918.
[3] Yamaguchi, M. and Sugii, K. (1981) Chem. Pharm. Bull. 29, 567–570.
[4] Yamaguchi, M. (1992) in: Calcium Inhibition (K. Kohama, Ed.), pp. 19–41, Japan Sci. Soc. Press, Tokyo and CRC Press, Boca Raton.
[5] Yamaguchi, M. and Sakurai, T. (1991) FEBS Lett. 279, 281–284.
[6] Yamaguchi, M. and Tai, H. (1991) Mol. Cell Biochem. 106, 25–30.
[7] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[8] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
[9] Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) Nucleic Acids Res. 16, 7583–7600.
[10] Yamaguchi, M. and Isogai, M. (1993) Mol. Cell Biochem. (in press).
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. 152, 232–238.

[13] Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8132.
[14] Kyte, J. and Doolittle, R. (1982) J. Mol. Biol. 157, 105–132.
[15] Shimokawa, N. and Yamaguchi, M. (1992) FEBS Lett. 305, 151–154.
[16] Shimokawa, N. and Yamaguchi, M. (1993) FEBS Lett. 316, 79–84.
[17] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
[18] Nojima, H. and Sokabe, H. (1987) J. Mol. Biol. 193, 439–445.
[19] Yamakuni, T., Kuwano, R., Odani, S., Miki, N., Yamaguchi, Y. and Takahashi, Y. (1986) Nucleic Acids Res. 14, 6768.
[20] Isobe, T. and Okuyama, T. (1978) Eur. J. Biochem. 89, 379–388.
[21] Tufty, R.M. and Kretsinger, R.H. (1975) Science 187, 167–169.
[22] Hilt, D.C. and Kligman, D. (1991) in: Novel Calcium-Binding Proteins (C.W. Heizmann, Ed.), pp. 65–103, Springer, Berlin.
[23] Heizmann, C.W. and Hunziker, W. (1991) Trends Biochem. Sci. 16, 98–103.
[24] Marshall, R.D. (1972) Annu. Rev. Biochem. 41, 673–702.