Volume 234, number 2, 407–410 FEB 06097 July 1988

The sexual inducer of *Volvox carteri*

Primary structure deduced from cDNA sequence

H.-W. Mages, H. Tschochner and M. Sumper

Lehrstuhl Biochemie I, Universität Regensburg, D-8400 Regensburg, FRG

Received 31 May 1988

The primary structure of the sexual inducer of *Volvox carteri f. nagariensis* has been deduced by cloning and sequence analysis of cDNA. The sexual inducer contains 208 amino acids including a signal sequence. A total of six potential *N*-glycosylation sites are found within the polypeptide chain. At the genomic level, the sexual inducer protein is encoded in five exons.

Sexual inducer; Primary structure; Nucleotide sequence; Gene structure; (*Volvox carteri*)

1. INTRODUCTION

The elegantly simple manner in which developmental programs are presented by the multicellular green flagellate *Volvox* makes this organism well suited for the study of cellular differentiation. During embryogenesis, somatic cells become separated from reproductive cells (gonidia) as clearly as in a diagram [1–4]. The asexual development of *Volvox* is changed to that of sexual reproduction by the action of a sexual inducer [1,5]. The sexual inducer acts on both the male and female strains of *Volvox*, causing development of sperm cells and eggs, respectively.

The sexual inducer is a glycoprotein synthesized and released by sperm cells [2,5,6]. The inducer is one of the most potent biological effector molecules known as it exhibits full effectiveness at about 10^{-16} M [2,7]. Only recently has it become possible to purify this inducer in sufficiently large amounts to obtain amino acid sequence data and to identify the inducer gene [8]. A rather complex exon-intron organization of this gene made it impossible to deduce the complete primary structure of the inducer. Here, we report on the nucleotide sequence of cDNA encoding the sexual inducer and on the principal structural organization of the inducer gene.

Correspondence address: H.-W. Mages, Lehrstuhl Biochemie I, Universität Regensburg, D-8400 Regensburg, FRG

2. MATERIALS AND METHODS

V. carteri f. nagariensis strain 69-1b was grown as in [8]. RNA was prepared as described [9] with some modifications. *Volvox* spheroids were disrupted by forcing them through a 0.4 mm hypodermic needle. Cells were collected by centrifugation and lysed by incubation in 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 5 mM EDTA and 2% SDS for 30 min at 32°C. SDS was then removed by precipitation with 0.132 vol. of 2 M KCl. The supernatant was mixed with 0.33 vol. of 2 M LiCl and incubated overnight at 4°C. Precipitated RNA was collected by centrifugation, washed once with 2 M LiCl and dissolved in 6 M guanidine hydrochloride. Insoluble material was removed by centrifugation. RNA was precipitated by the addition of 0.5 vol. ethanol. Collected RNA was dissolved in water and again precipitated by the addition of 2.5 vols ethanol in the presence of 150 mM Mg acetate. This step was repeated once. Poly(A)$^+$ RNA was prepared as in [9]. A cDNA library was constructed in the vector λgt11, using the Pharmacia cDNA synthesis kit. The library was screened either with a 2.7 kb fragment from the sexual inducer gene [8] or with the 25-mer oligonucleotide 5′-GAGATGTCAGCTGAGTGGCTGCCGA-3′, derived from the genomic sequence of the sexual inducer gene (exon 3). Labelling of the oligonucleotide and screening of the library were performed according to [11]. Recombinant phages were grown on *E. coli* Y 1088. Inserts were subcloned into vector pUC18 and sequenced by the dideoxy chain-termination

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

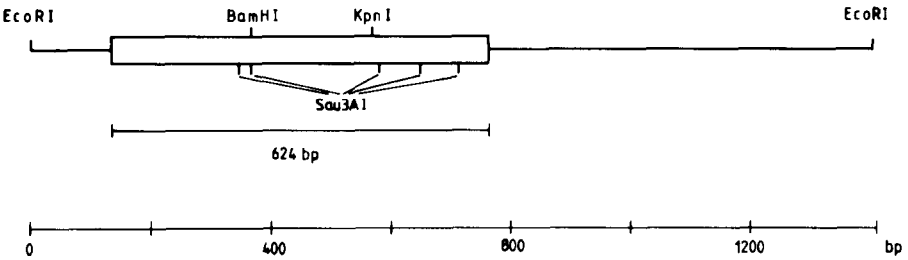

Fig.1. Restriction endonuclease map of the cDNA insert in λgt11 encoding the sexual inducer glycoprotein.

method [12], using synthetic oligodeoxynucleotides as primers. Oligodeoxynucleotides were synthesized on an Applied Biosystems DNA synthesizer.

3. RESULTS AND DISCUSSION

A cDNA library was constructed using poly(A)$^+$ RNA from sexually induced *Volvox* spheroids of the male strain 69-1b [5]. Since the exact period of transcription of the sexual inducer gene is not yet known, the following stages of developing sperm cells were mixed together and used for RNA extraction: Spheroids containing androgonidia, spheroids containing sperm packets, free-swimming sperm packets and individual sperm cells (after dissociation of the sperm packets).

```
                                                                  -135  AACACACCGCTGCCTTGTTCAGTCATCCCAGCAGCT  -100

ACGTACCTCAGCCAGCTACGTTTGTAGCGGCGCTCCCCTCACGCAGGCCCGCGCTCGGCATCCCGTTCGACACGAACAAGGAGATACCACTTCTACAGA    -1

ATG GCA GTA GTG GTC GTC AAT TCT GCA ACC GCC TCA CTT TTG GCG GTG TGT TTG GTC TTC ATG GCT GTT GGG CTC     75
Met Ala Val Val Val Val Asn Ser Ala Thr Ala Ser Leu Leu Ala Val Cys Leu Val Phe Met Ala Val Gly Leu

TGC ACG GGC CAA ATT GTC GAT GTC AAA TTT CCC AGC TGC AGG TGC GAG CGA GAG CTT ACA CCC TTC GCC ATC AAG    150
Cys Thr Gly Gln Ile Val Asp Val Lys Phe Pro Ser Cys Arg Cys Glu Arg Glu Leu Thr Pro Phe Ala Ile Lys

TCG GCA GCC ACT CAG CTG ACA TCT CGC AAT CCC GGC GTG GTC AAC TTG TAC TGC TTT GAG ATC GGC ATC GTG AAC    225
Ser Ala Ala Thr Gln Leu Thr Ser Arg Asn Pro Gly Val Val Asn Leu Tyr Cys Phe Glu Ile Gly Ile Val Asn

TCT GGA TCC GGC GCA TGC TAC ACA GAA CCC GCT TCC CAG AAC TTA TCC AAG GTA TCC GTC TAC GCT CAG GCC GCC    300
Ser Gly Ser Gly Ala Cys Tyr Thr Glu Pro Ala Ser Gln Asn Leu Ser Lys Val Ser Val Tyr Ala Gln Ala Ala

CAG CGC GAC CGT CTG TCG GCC TTT GGG GTC CTC CTG GCT GGT GCG CCT GTC AGC AAC ATG ACC TAT CTC ACT CCC    375
Gln Arg Asp Arg Leu Ser Ala Phe Gly Val Leu Leu Ala Gly Ala Pro Val Ser Asn Met Thr Tyr Leu Thr Pro

AGA TGG GAC TCA CTG AAC ATG ACC ACC ATC AGC AAC CTT AAC TTC AGC AAG ACG CAG GCG AAT GGT ACC AGG ATC    450
Arg Trp Asp Ser Leu Asn Met Thr Thr Ile Ser Asn Leu Asn Phe Ser Lys Thr Gln Ala Asn Gly Thr Arg Ile

TGT CTT GAG CTG TTC AAG CCT ACC ACC ATC AAC GAG TTT TGC GAA CGA GAG GGC GCG TCA GGA TCG TTC TGC TGG    525
Cys Leu Glu Leu Phe Lys Pro Thr Thr Ile Asn Glu Phe Cys Glu Arg Glu Gly Ala Ser Gly Ser Phe Cys Trp

GTT GCC CTG TTT AAT GAC AAC AAC TGT GTA CCG CCT AAC TCA ACT GTG GTG ATC TCT AAG CGC CTG TGC TGC CCA    600
Val Ala Leu Phe Asn Asp Asn Asn Cys Val Pro Pro Asn Ser Thr Val Val Ile Ser Lys Arg Leu Cys Cys Pro

AGA TTT CAG TCT TTC CTG TCC CCG TGA TGATATTATTGTCATGATGGATGATGTGGAGGCTGCCTGACGTTAAAACCCTAAGTAATAATA    690
Arg Phe Gln Ser Phe Leu Ser Pro ***

ATGATGTCATGACGTGGCCTTTATTTTCTGCATATGCAGGGAGGAGTGGTCATTGATGGAACGCAAAAAATATTAATTCGTACTATCCTCTTATTATTC    789

ATTCATCATTATCATTTATATAACATGTGTTGCTTAGATATTATTCATTATTTCTAATCAGGGTATTGTGTATGGTCAGGTTTTTGCGGGCGCCTGTGA    888

ATTTTATACCTTTCATTCGCTTTAATTTTATACCTTTCATTCATCACTGATGTGGCTGCTCCGATATATACTAGATTCGATGAATGGAATGGTATGTAT    987

GTGTATGTCTGGTGCTGTATATGAGCCCCTAATTAGGCGAATACATATTCCAACAATCTCCTTTAATGATTGCTAATTAATTAATTCTTCTTCTTCTGA   1086

TTCTTCTGTAATCGTTTCTGTATGGGATAAAACCCCATACGTGCATAAATTTCGTTAATTTAATCGTTATTCAATTCATTCATTCTTCTTCTTCCTTTA   1185

ATTCCTCTTATTTTTTCTTTACATACGTGTGAGTATATTTACCTGAAACTATTAATAATTTCAGAATGGATGTGAAA                          1262
```

Fig.2. Nucleotide sequence of cloned cDNA for *Volvox* sexual inducer and deduced amino acid sequence of the protein. Amino acid sequences determined by peptide analysis are underlined. The potential *N*-glycosylation sites are boxed. Exon boundaries are indicated by arrowheads.

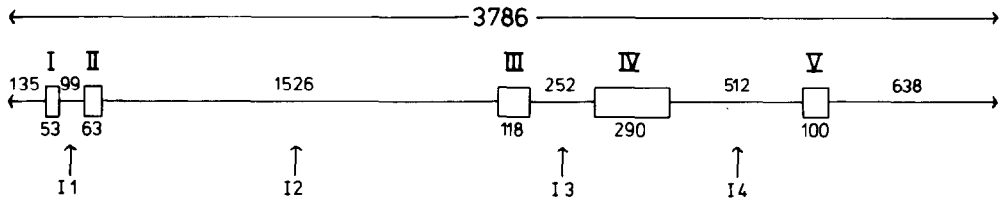

Fig.3. Exon-intron organization of the sexual inducer gene. Boxes represent exons, lines introns. Numbers indicate the length of DNA stretches in bp.

The λgt11 cDNA library was screened with inducer-specific genomic DNA (2.7 kb fragment described in [8]). Five recombinant clones were obtained from 1×10^5 transformants. In order to select a full-length cDNA clone, the positive clones were hybridized with a different probe, the oligodeoxynucleotide 5′-GAGATGTCAGCTGA-GTGGCTGCCGA-3′, corresponding to the peptide SAATQLTS located closest to the N-terminus of the partial amino acid sequence known so far. A single clone (Ind cK2) remained positive and was subsequently subcloned. Determination of its nucleotide sequence confirmed that this clone covered the entire coding region of the sexual inducer cDNA.

Fig.1 shows the restriction map of Ind cK2 insert DNA. Sequence analysis of Ind cK2 cDNA predicts an open reading frame of 624 nucleotides (fig.2). All the amino acid sequences of tryptic and chymotryptic peptides [8] were found to be encoded in the same reading frame by the cDNA sequence. Triplet ATG at positions 1–3 is the initiation codon of the sexual inducer, since it is the first ATG codon that is located downstream of a nonsense codon (TAG at positions −76 to −73). The 208th codon which specifies proline is followed by two termination codon. Thus, the amino acid sequence derived from the cDNA contains 208 amino acids of molecular mass 22 360 Da. Since the sexual inducer is a secreted glycoprotein, this sequence should include a signal peptide sequence. Amino acid residues 1–11 indeed represent a typical signal sequence with a potential recognition sequence Ala-X-Ala [13] for a signal peptidase at amino acid positions 9–11. Unfortunately, Edman degradation of the intact inducer protein turned out to be impossible. Therefore, the exact N-terminus of the mature sexual inducer remains to be established. As expected for a glycoprotein, consensus sequences for *N*-linked glycosylation (Asn-X-Ser or Thr) were found (a total of six).

Since we had previously isolated and analysed a genomic clone of the inducer gene, it is now possible to reconstruct the exon-intron organization of the inducer gene. The complete amino acid sequence of the sexual inducer is encoded in five exons (fig.3). Exon 1 encodes amino acids 1–18, exon 2 amino acids 19–39, exon 3 amino acids 40–78, exon 4 amino acids 79–175 and finally, exon 5 encodes the remaining amino acids of the C-terminal part of the protein.

The amino acid sequences of the sexual inducer was compared to known protein sequences in the NBRF data base [14]. Surprisingly, the best homology to another protein was found to Coypu insulin (B-chain) [15]. The amino acid sequence encoded by exon 2 exhibits 35% identity in a 20-amino-acid overlap. The biological significance of this homology is not known, but remarkably those residues of the insulin B-chain regarded as being essential for an insulin-like hormone (according to [16] and denoted below by asterisks) are indeed conserved in the exon 2 sequence:

```
          20
inducer   FMAVGLCTGQIVDVKFPSCR
          ...  :::.:.:::. .. ::
insulin   YVSQRLCGSQLVDTLYSVCR
              *    *   *  *
```

However, preliminary experiments indicate that bovine insulin is not able to mimic the biological action of the sexual inducer.

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft (SFB 43). We wish to thank E. Berg and S. Stammler for excellent technical assistance.

REFERENCES

[1] Starr, R.C. (1969) Arch. Protistenk. 111, 204–222.
[2] Starr, R.C. (1970) Dev. Biol. Suppl. 4, 59–100.
[3] Green, K.J. and Kirk, D.L. (1981) J. Cell Biol. 91,743–752.
[4] Green, K.J. and Kirk, D.L. (1982) J. Cell Biol. 94, 741–742.
[5] Starr, R.C. and Jaenicke, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1050–1054.
[6] Gilles, R., Bittner, C. and Jaenicke, L. (1981) FEBS Lett. 124, 57–61.
[7] Gilles, R., Gilles, C. and Jaenicke, L. (1984) Z. Naturforsch. 39c, 584–592.
[8] Tschochner, H., Lottspeich, F. and Sumper, M. (1987) EMBO J. 6, 2203–2207.
[9] Kirk, M.M. and Kirk, D.L. (1985) Cell 41, 419–428.
[10] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[11] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[12] Sanger, F., Niclen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Perlman, D. and Halvarson, H.O. (1983) J. Mol. Biol. 167, 391–409.
[14] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
[15] Smith, L.F. (1972) Diabetes 21 (suppl.2), 457–460.
[16] Thamm, P.M. (1980) Umsch. Wiss. Techn. 80, 592–596.