The Journal of Eukaryotic Microbiology

ISOP Published by the International Society of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

ORIGINAL ARTICLE

Molecular Structures and Coding Genes of the Water-Borne Protein Pheromones of *Euplotes petzi*, an Early Diverging Polar Species of *Euplotes*

Bill Pedrini[a], Thea Suter-Stahel[b], Adriana Vallesi[c], Claudio Alimenti[c] & Pierangelo Luporini[c]

a Paul Scherrer Institute, Villigen 5232, Switzerland
b Institute of Molecular Biology and Biophysics, ETH Zürich, Zürich 8093, Switzerland
c Laboratory of Eukaryotic Microbiology and Animal Biology, School of Biosciences and Veterinary Medicine, University of Camerino, Camerino (MC) 62032, Italy

Keywords
Chemical signals.
ciliate pheromones; cysteine-rich proteins; macronuclear genes; protein structure.

Correspondence
P. Luporini, School of Biosciences and Veterinary Medicine, via Gentile III da Varano, 62032 Camerino (MC), Italy
Telephone number: +39-0737403229;
FAX number: +39-0737403290;
e-mail: piero.luporini@unicam.it

Received: 20 May 2016; revised 27 June 2016; accepted July 19, 2016.

doi:10.1111/jeu.12348

ABSTRACT

Euplotes is diversified into dozens of widely distributed species that produce structurally homologous families of water-borne protein pheromones governing self-/nonself-recognition phenomena. Structures of pheromones and pheromone coding genes have so far been studied from species lying in different positions of the *Euplotes* phylogenetic tree. We have now cloned the coding genes and determined the NMR molecular structure of four pheromones isolated from *Euplotes petzi*, a polar species which is phylogenetically distant from previously studied species and forms the deepest branching clade in the tree. The *E. petzi* pheromone genes have significantly shorter sequences than in other congeners, lack introns, and encode products of only 32 amino acids. Likewise, the three-dimensional structure of the *E. petzi* pheromones is markedly simpler than the three-helix up-down-up architecture previously determined in another polar species, *Euplotes nobilii*, and in a temperate-water species, *Euplotes raikovi*. Although sharing the same up-down-up architecture, it includes only two short α-helices that find their topological counterparts with the second and third helices of the *E. raikovi* and *E. nobilii* pheromones. The overall picture that emerges is that the evolution of *Euplotes* pheromones involves progressive increases in the gene sequence length and in the complexity of the three-dimensional molecular structure.

AMONG ciliates, *Euplotes* is probably the most speciose taxon with nearly 100 nominal species described from practically every aquatic habitat of the globe (Borror and Hill 1995; Curds 1975). The ease with which many of these species can be collected and cultivated in laboratory has made *Euplotes* an attractive experimental system to investigate many aspects of ciliate biology, in particular the genetic and molecular basis underlying the ciliate ability to discriminate between self- and nonself (Luporini et al. 2016a,b; as reviews). This ability becomes manifest by observing how cells change their behavior depending on whether they interact with sibs or nonsibs. In the former case, the cells reproduce undergoing binary fission, while in the latter they stop reproducing to socialize, uniting two by two in mating pairs to perform a mutual exchange of "male" gamete-nuclei and cross-fertilization.

As is common among spirotrichous ciliates (Dini and Nyberg 1993), *Euplotes* controls the specificity of its social interactions at a single *mat* locus of the transcriptionally silent germ-line micronucleus, and expresses the multiple alleles of this locus in the transcriptionally active somatic macronucleus with the synthesis of the "mating-type factors" that are responsible for the chemical specificity to each cell (mating) type. As is the case in other ciliates, *Blepharisma* (Miyake 1981) and *Dileptus* (Uspenskaya and Yudin 2016) in particular, these factors are constitutively secreted by *Euplotes* as water-borne signaling pheromones into the extracellular environment (Miceli et al. 1983), which has greatly facilitated their isolation and chemical analysis. A structural characterization of *Euplotes* pheromones has so far been achieved, along with a characterization of the macronuclear pheromone coding genes, in *Euplotes raikovi*,

Euplotes octocarinatus, *Euplotes nobilii*, and *Euplotes crassus* (Luporini et al. 2016a,b; as reviews), which are species clearly separated in the *Euplotes* phylogenetic tree. *Euplotes raikovi* and *E. nobilii* branch off much earlier than *E. octocarinatus* and *E. crassus* which in turn split into distinct clades (Di Giuseppe et al. 2014, 2015; Fotedar et al. 2016).

In all four species, pheromones form species-specific families of structurally homologous, highly stable and cysteine-rich proteins. However, their amino acid sequences markedly vary in length between the species. In *E. raikovi* and *E. nobilii*, they include only 37—40 (51 in a single exceptional case) and 52—63 amino acids, respectively. In *E. octocarinatus*, they include 85—108 amino acids and in *E. crassus* the pheromone sequences split into two distinct sub-families of 45 and 56 amino acids, respectively, due to a phenomenon of gene duplication (Luporini et al. 2005; Vallesi et al. 2014).

These marked interspecific differences in the amino acid sequence length have suggested a possible correlation between the evolutionary diversification of *Euplotes* species and an increase in the pheromone structural complexity. However, this correlation has so far remained quite elusive because three-dimensional pheromone structures could be determined exclusively within the pheromone families of the two closely allied species, *E. raikovi* and *E. nobilii*. The correlation now receives substantial support from the results reported here of genetic analyses and NMR structure determination of the pheromone family of *Euplotes petzi*, a species isolated from polar sea waters that, together with *E. sinicus*, forms the earliest branch of the *Euplotes* phylogenetic tree (Di Giuseppe et al. 2014). In comparison to the *E. raikovi* and *E. nobilii* pheromones, the *E. petzi* pheromones exhibit significantly shorter amino acid sequences, a much simpler three-dimensional structure and, in parallel, shorter nucleotide sequences and a simpler organization of the macronuclear pheromone coding genes.

MATERIALS AND METHODS

Cell cultures

The two Antarctic strains Ad-Cov 2 and Ad-Cov3 used in this study are part of a group of wild-type strains of *E. petzi* collected from polar coastal seawaters (Di Giuseppe et al. 2014, 2015). They were chosen for their strong cross-mating reactivity and their persistent production of pheromone-rich culture filtrates, as revealed by assaying filtrates from one strain for the induction of homotypic (intraclonal) mating pair formation in cultures of the other. The strain cultivation was carried out in a cold room at 4 °C, under a cycle of 12 h of very low light and 12 h of dark, and using the green alga *Dunaliella tertiolecta* as usual food source.

Pheromone isolation and molecular mass determination

The major pheromone source was provided by cultures of strain Ad-Cov2. They were maintained in the growth phase for 4—5 wks at a concentration of approximately 5,000 cells/ml and a reproduction rate of approximately one cell fission every 3 d, and finally deprived of food for 1 wk to homogeneously arrest cells in the G_1 stage of their reproductive cycle before being used to prepare cell-free filtrates. A routine three-step chromatographic procedure, originally devised for the *E. raikovi* pheromones (Raffioni et al. 1992), was used for the pheromone purification. It basically involves protein adsorption from culture filtrates onto reverse-phase Sep-Pak C_{18} cartridges (Waters, Milford, CA), and separation by gel filtration on a Superdex peptide HR10/30 column (GE Healthcare, Little Chalfont, U.K.) and anion exchange chromatography on a Mono-Q column (Supelco, Bellefonte, PA). MALDI/TOF measurements were performed on a Bruker Daltonics Ultraflex mass spectrometer (Bruker Daltonics, Bremen, Germany), with the acceleration voltage set at 20 kV. Samples were directly applied onto a stainless-steel spectrometer plate as 1-μl droplet of synapinic acid-matrix solution.

Pheromone gene amplification, cloning, and transcript analysis

Aliquots (0.5 μg) of DNA, purified following a standard procedure (La Terza et al. 2009), were used as template in polymerase chain reaction (PCR) amplifications run in the Eppendorf Ep-gradient Mastercycler (Eppendorf AG, Hamburg, Germany). Oligonucleotides used as PCR primers were synthesized by Invitrogen (Life Technologies, Carslbad, CA), and their designations and sequences are reported in Table 1. Amplification products were purified with Quantum prep PCR Kleen spin columns (Bio-Rad, Hercules, CA), and cloned using the TOPO-TA cloning kit (Life Technologies) according to the manufacturer's recommendation. Different clones from each cloning reactions were sequenced at the BMR Genomics of Padua (Italy). Gene transcript analysis was carried out on total RNA samples prepared following the protocol provided by the TRIzol plus RNA isolation kit (Life Technologies). RNA aliquots (1 μg) were converted into single-stranded cDNA (ss-cDNA) using the SuperScript™ II reverse transcriptase and an adapter primer ("oligo(dT)AP" in Table 1), according to the procedure of the 3′ RACE system (Thermo Fisher Scientific, Waltham, MA). The resulting ss-cDNA was amplified using the gene-specific oligonucleotide

Table 1. PCR primer designations and sequences

Name	Nucleotide sequence (5′–3′)
Tel	CCCCAAAACCCCAAAACCCC
Ep-deg-FW	GCTCCAGAACCWGAYTGYTG
Ep-RV	AGCTGATGTTTCTCGACGGA
Ep-5′-FW	GTTTCACTATAAATACAGAGAAATT
Ep-3′-RV	AATTGCATCAATATGTGATC
Ep-FW1	TCTCTCTACTCAATGAACACCA
Oligo(dT_{18})AP	GGCCACGCGTCGACTAGTACT_{18}
UAP	GGCCACGCGTCGACTAGTAC

("Ep-FW1" in Table 1) as forward primer, and an "abridged universal amplification" oligonucleotide (provided with the system, and "UAP" in Table 1) as reverse. Amplification products were cloned and sequenced as described above.

Molecular structure determination

The pheromone molecular structure was determined following a procedure based on homonuclear ^{1}H-NMR in solution (Wüthrich 1986), which has previously been successfully applied to determine the pheromone structure in other *Euplotes* species (Brown et al. 1993; Pedrini et al. 2007). A single 500-µl NMR sample of the E*p*-1 pheromone was prepared by dissolving the lyophilized protein into a 10 mM sodium phosphate buffer at pH 6.0, containing 5% D_20 for the purpose of locking the NMR spectrometer base frequency. The final E*p*-1 concentration in the sample solution was about 0.3 mM. The following two-dimensional NMR spectra (Wüthrich 1986) were collected at 20 °C on Bruker Advance III instruments equipped with cryo-cooled probe heads: double quantum filtered 2D [^{1}H,^{1}H]-COSY at 700 MHz, 2D [^{1}H,^{1}H]-TOCSY with a 60 ms mixing time at 600 MHz, and 2D [^{1}H,^{1}H]-NOESY with 120 ms mixing time at 700 MHz. The ^{1}H resonance frequencies were assigned interactively relying on all three NMR spectra. The assignment was complete with the following few exceptions: Ser_1, HN of Cys_2, HB2 and HB3 of Ser_4, HG2 and HG3 of Glu_5, HD2 and HD3 of Pro_{10}. The input for the automated structure calculation consisted of the E*p*-1 amino acid sequence, the 2D [^{1}H,^{1}H]-NOESY spectrum, and the list of the determined ^{1}H resonance frequencies. The disulfide bridge pattern was derived by analogy from the alignment of the E*p*-1 sequence with those of the *E. nobilii* pheromones. Each bridge was accounted for by additional distance restraints (Fadel et al. 2005). The calculation consisted in a standard iterative, 7-cycle protocol (Herrmann et al. 2002a). In each cycle, the NOE peak identification and assignment were performed with the ATNOS/CANDID software suite (Herrmann et al. 2002a,b), while the representative conformer bundle was obtained with the torsion angle dynamics program CYANA (Güntert et al. 1997) on the basis of the derived upper distance and torsion angle limits (Table 2). The 20 conformers with lowest CYANA target function obtained after the seventh cycle were refined inside a water molecule shell with the CNS program (Brünger et al. 1998), and represent the final E*p*-1 structure.

Data availability

The ^{1}H chemical shift list used for the E*p*-1 structure calculation is deposited at the Biological Magnetic Resonance Data Bank (BMRB) under the accession code 25610, and the final E*p*-1 structure is deposited at the Protein Data Bank (PDB) under the accession code 2N2S. The pheromone gene sequences are deposited at the GenBank database under the accession codes: *mac-ep-1*, KX279861; *mac-ep-2*, KX279862; *mac-ep-3*, KX279863; *mac-ep-4*, KX279864.

Table 2. Input for the structure determination, and characterization of the 20 conformer bundle representing the E*p*-1 structure

Quantity	Value
NOE upper distance limits[a]	372
Intraresidual[a]	71
Short range[a]	110
Medium range[a]	130
Long range[a]	61
Dihedral angle constraints	107
Residual target function value[a] [Å^2]	0.80 ± 0.07
Residual NOE violations[a]	
Number ≥ 0.1 Å	5 ± 2
Maximum [Å]	0.20 ± 0.02
Residual dihedral angle violations[a]	
Number ≥ 2.5°	2 ± 0
Maximum [°]	4.36 ± 1.02
CNS energies [kcal/mol]	
Total	−553 ± 34
van der Waals	−104 ± 17
Electrostatic	−924 ± 26
RMSD from ideal geometry	
Bond length [Å]	0.0133 ± 0.0004
Bond angles [°]	1.08 ± 0.04
RMSD from mean coordinates[b] [Å]	
Backbone (3–28)	0.63 ± 0.19
All heavy atoms (3–28)	1.03 ± 0.19
Ramachandran plot statistics[c]	
Most favored regions [%]	71.1
Additional allowed regions [%]	26.4
Generously allowed regions [%]	1.8
Disallowed regions [%]	0.7
Structure quality scores	
PROCHECK global quality score (*Z* score)	−3.78
Verify3D (raw score)	0.46

The given standard deviations are over the conformer bundle
[a]Before CNS energy minimization.
[b]The numbers in parentheses indicate the residues for which the RMSD was calculated.
[c]As determined by PROCHECK (Laskowski et al. 1993).

RESULTS

Pheromone and pheromone gene sequence determination

A combined chemical and genetic approach was necessary to determine the sequences of the pheromones and pheromone genes of the Ad-Cov2 and Ad-Cov3 cells, known from previous Mendelian analyses of mating-type inheritance (P. Luporini, unpubl. data) to carry distinct heterozygous combinations of co-dominantly expressed alleles at their *mat* locus resulting in the co-secretion of two distinct pheromones. As initial chemical step, the two pheromones, designated E*p*-1 and E*p*-2, of the Ad-Cov2 cells were purified in amounts of 10—15 µg/liter of culture filtrates, assessed to be active in inducing mating pair formation between Ad-Cov3 cells at micro-molar concentrations, and measured to have molecular masses of 3180.6 and 3282.7 Da, respectively (Fig. 1). A purified sample of the E*p*-1 pheromone

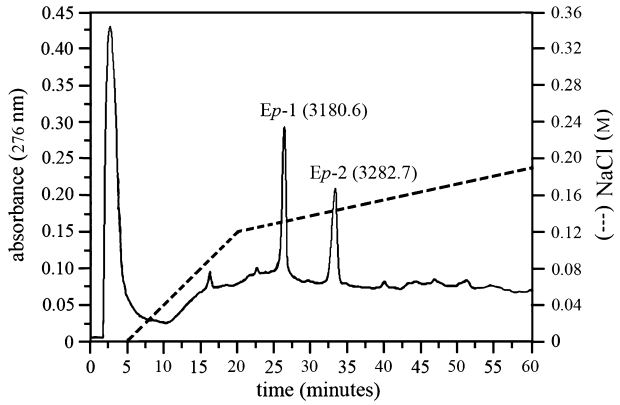

Figure 1 Chromatographic fractionation of pheromones E*p*-1 and E*p*-2 secreted by Ad-Cov2 cells of *Euplotes petzi* on Mono-Q column. The elution peaks are indicated together with the molecular masses of the associated proteins, as determined by MALDI-TOF measurements. The dashed line indicates the applied NaCl elution gradient.

was then alkylated with 4-vynilpyridine to detect the Cys residues and subjected to Edman degradation to determine a segment of the amino acid sequence at the amino-terminus. The 14-residue amino-terminal sequence Ser_1-Cys-Gly-Ser-Glu-Cys-Ala-Pro-Glu-Pro-Asp-Cys-Trp-Gly_{14} was obtained. The knowledge of this sequence paved the way to the second genetic step directed to determine the complete pheromone amino acid sequences via PCR amplification and analysis of the pheromone coding genes.

A degenerate oligonucleotide ("Ep-deg-FW", Table 1) was synthesized on the basis of the Ala_7—Trp_{13} sequence stretch and used as PCR primer to amplify DNA preparations from Ad-Cov2 cells in combination with a primer ("Tel", Table 1) specific to the telomeric $(C_4A_4)_2CCCC$ repeats distinctive of every *Euplotes* macronuclear gene (Jahn and Klobutcher 2002). The 450-bp products of this amplification were resolved into nucleotide sequences covering the complete 3′ trailer region and portions of the coding region of two distinct macronuclear coding genes. The complete coding regions were next obtained, together with the 5′ leader regions, as 380-bp products of a second PCR amplification run with the telomeric ("Tel") primer and an oligonucleotide ("Ep-RV", Table 1) specific to a conserved stretch of the 3′ trailer region. Using two oligonucleotides ("Ep-5′-FW" and "Ep-3′-RV", Table 1) specific to sequences immediately adjacent to the 5′ and 3′ telomeric ends, it was finally possible to clone 650-bp amplification products containing the full-length sequences of the two macronuclear pheromone genes encoding the E*p*-1 and E*p*-2 pheromones of the Ad-Cov2 cells. In the same way, we then cloned 650-bp amplification products containing the full-length sequences of the two macronuclear pheromone genes encoding the two pheromones, designated as E*p*-3 and E*p*-4, of the Ad-Cov3 cells.

Gene sequences

The four pheromone gene sequences encoding E*p*-1, E*p*-2, E*p*-3, and E*p*-4, and accordingly designated as *mac-ep-1*, *mac-ep-2*, *mac-ep-3*, and *mac-ep-4* (with the abbreviation *mac* indicating their origin in the transcriptionally active macronucleus of the cell), extend for 715 bp and differ at only 10 positions, seven of which lie in the open reading frame (ORF) and determine five amino acid substitutions (Fig. 2). In all sequences, the ORF extends for 219 bp from the first ATG codon in position 101 to the TAA codon in position 317. It encodes the pheromones in the form of 72-amino acid immature cytoplasmic precursors. The first two 20-amino acid sequence stretches are identical among the four sequences and represent the signal (or pre) and pro peptides, that are destined to be removed by proteolytic cleavages (Nielsen et al. 1997). The third 32-amino acid stretch, which contains all five amino acid substitutions, represents the mature, biologically active form that is secreted into the extracellular environment.

As deduced from an analysis of the expression of the *mac-ep-1* and *mac-ep-2* genes by RT-PCR (see Materials and Methods), no intron is present within the ORF. In both cases, amplification products of 425 bp, excluding the polyA tail, were obtained, and the sequences of these products included, in addition to the 12-bp of the primer sequence, the uninterrupted ORF of 219 bp, and the 3′ untranslated region of 194 bp (data not shown).

In all four *mac-ep* genes, the 5′ leader region is only 72-bp long, fully conserved and particularly rich in A and T bases (82% of the total nucleotide number) most of which are arranged to form adjacent TAA repetitions as in other macronuclear genes of *Euplotes* and spirotrichous ciliates (Cavalcanti et al. 2004; Ghosh et al. 1994; Lescasse et al. 2005; Vinogradov et al. 2012). Lacking conventional TATA and CAAT boxes for the transcription regulation, these repetitions qualify as candidates for this regulation acting either independently, or in cooperation with the GAAAT motif located 21-bp downstream the telomeric end (Ghosh et al. 1994).

The 3′ trailer region is much more extended (396 bp) than the 5′ leader region, but equally conserved except for three nucleotide substitutions. In addition to containing a polyadenylation signal represented by an AATAAT motif lying 16 nucleotides upstream the poly(A) tail of the transcript, it is particularly rich in repetitions of 3–5 identical nucleotides and short inverted and direct repeats whose function, if any, remains to be determined.

Amino acid sequences and molecular structure

The 32-amino acid sequences of all four *E. petzi* pheromones contain eight Cys residues designated Cys-I to Cys-VIII and located in fully conserved positions. These sequences start with a Ser residue, which is unusual compared to the pheromone sequences of other *Euplotes* species, most of which carry an Asp residue at their amino-terminus or, alternatively, Gly/Asn/Tyr/Leu residues (Luporini et al. 2005; Vallesi et al. 2014) (Fig. 3). Among the five nonconserved amino acid positions, two lie in the 11-amino acid sequence stretch at the amino-terminus and three in the 12-amino acid stretch at the carboxyl

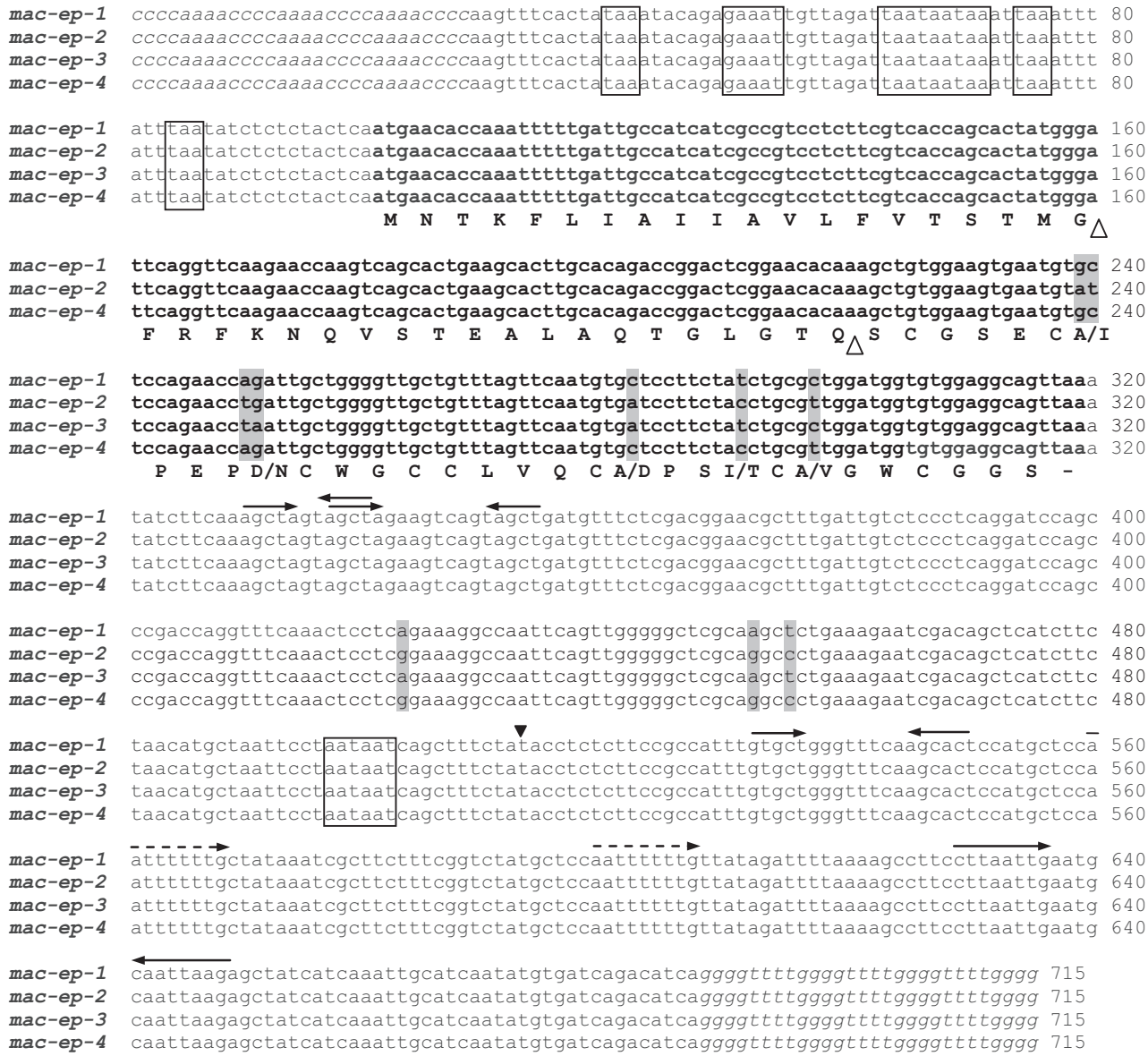

Figure 2 Nucleotide sequence alignment of *Euplotes petzi* macronuclear pheromone genes and deduced amino acid sequences. Nucleotide sequences, lower case letters; extensions indicated by progressive numbers on the right; telomeres, italics; coding regions specifying pre-pro-pheromone, bold; nucleotide variations, highlighted in gray; putative signals for transcription regulation and polyadenylation, boxed; short inverted and direct repeats, solid and dotted arrows, respectively; last nucleotide in the gene transcripts, filled arrowhead; amino acid sequence, capital letters and single-letter code; amino acid variations separated by a slash; putative cleavage sites of the pre- and pro segments, light arrowheads.

terminus. The central nine-amino acid stretch, delimited by CysIII and CysVI and including the CysIV/CysV doublet, is devoid of any amino acid substitution and thus appears to be the most conserved sequence domain of the *E. petzi* pheromones.

The E*p*-1 pheromone was chosen for the determination of the three-dimensional structure by NMR spectroscopy because it could be purified in amounts significantly larger than the other *E. petzi* pheromones. For the structure calculations, it was assumed that the disulfide bond pattern of the eight E*p*-1 Cys residues is the same as the pattern previously established for the Cys residues of the *E. nobilii* pheromones (Di Giuseppe et al. 2011; Pedrini et al. 2007; Stewart et al. 1992). This assumption was supported by the full alignment of the Cys residues (Fig. 3).

The bundle of the 20 lowest target function conformers and the resulting secondary structural elements of the conformer closest to the mean are shown in Fig. 4, 5. The E*p*-1 three-dimensional structure consists of three segments arranged in a right-handed up-down-up motif. The first up-segment (Ser_1—Pro_8) is devoid of secondary structure. The residues Ser_4—Ala_7 take an arrangement close

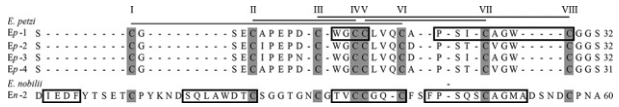

Figure 3 Amino acid sequence alignment of the four *Euplotes petzi* pheromones with pheromone E*n*-2 taken as representative of the *Euplotes nobilii* pheromone family. The alignment is optimized by gap insertions. The cysteine residues are highlighted in gray, numbered progressively from the amino to the carboxyl terminus, and connected according to their disulfide pairings. Boxes include the residues arranged in secondary α-helical structures.

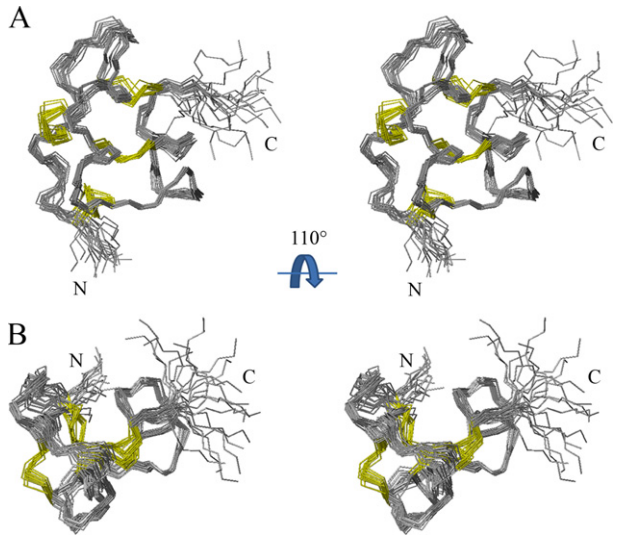

Figure 4 NMR solution structure bundle of the pheromone E*p*-1. Lateral (**A**) and top (**B**) stereo view of the bundle of 20 energy-minimized DYANA conformers, superimposed for minimum RMSD value to the lowest target function conformer. The polypeptide backbone is in gray and the disulfide bridges are in yellow. The amino (N) and carboxyl (C) termini are indicated.

to a 3_{10}-helix, which however does not form completely. The second down-segment (Asp_{11}—Gln_{19}) includes a short four-residue (Trp13—Cys16) alpha-helix (α_1). The third up-segment builds a second, longer alpha-helix (α_2) with eight residues (Pro_{22}—Cys_{29}). The overall E*p*-1 structure is stabilized by the arrangement of the four disulfide bridges, two of which (Cys_{12}-III/Cys_{29}-VIII and Cys_{16}-V/Cys_{25}-VII) hold the two helices together to form a structured scaffold, and the other two (Cys_2-I/Cys_{20}-VI and Cys_6-II/Cys_{15}-IV) anchor the first unstructured segment to the scaffold. The shortness of the stretches Glu_9—Pro_{10} and Cys_{20}—Ala_{21} connecting the three segments contributes to making the E*p*-1 molecule very compact. The sole region of the molecule that remains disordered is the three-residue (Gly_{30}—Ser_{32}) carboxyl terminus.

The compactness of the E*p*-1 molecular structure also appears to be reflected in preliminary analyses of thermostability carried out by circular dichroism spectrometry showing that the E*p*-1 pheromone fold is stable upon heating up to 95 °C (Russo R., Verde C., unpubl. data), like

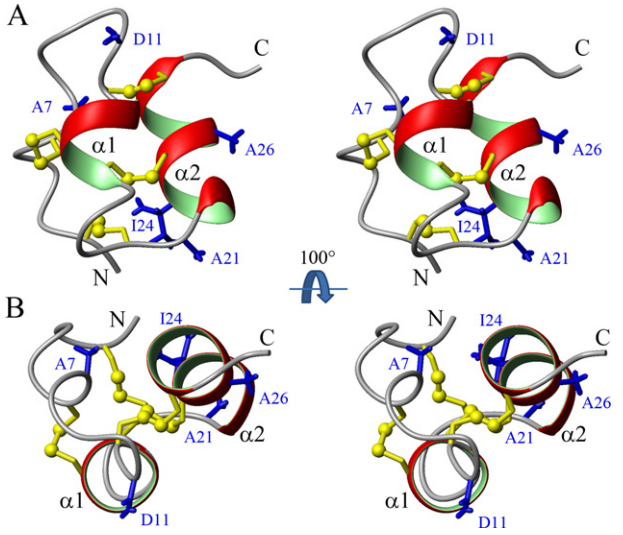

Figure 5 Representative conformer for the pheromone E*p*-1 molecular structure. Lateral (**A**) and top (**B**) stereo view of the E*p*-1 conformer with the lowest backbone RMSD to the mean atom coordinates. The helices α1 and α2 are represented as red-green ribbons, and the disulfide bonds as yellow sticks and balls. The side chains of the amino acids not conserved in the other E*p*-2, E*p*-3, and E*p*-4 pheromones are in blue and labeled with their single-letter code and sequence position. The amino (N) and carboxyl (C) termini are indicated.

the temperate sea-water pheromones of *E. raikovi* and unlike *E. nobilii* pheromones which unfold in the temperature range 55–70 °C (Geralt et al. 2013) and from which more similar properties were expected because of the same polar origin.

DISCUSSION

We used *E. petzi* for its basal position in the *Euplotes* phylogenetic tree to obtain insights on the first steps of the structural evolution of *Euplotes* pheromones and pheromone coding genes. With regard to the pheromone genes, Fig. 6 compares their basic structure in *E. petzi* with those known from *E. octocarinatus*, *E. nobilii*, *E. raikovi*, and *E. crassus* (for references, see legend of Fig. 6). It appears that *E. petzi* pheromone genes are unique in exhibiting by far the smallest size, resulting both from a 5′ leader region that is from 2- to 10-fold shorter than in the other *Euplotes* species, and from a lack of intron sequences which are instead a common trait of the 5′-leader and coding gene regions in other *Euplotes* species. In *E. octocarinatus*, the extension of the intron sequences is maximal including from 366 to 515 nucleotides (Brünen-Nieweler et al. 1998; Möllenbeck and Heckmann 1999), and in *E. raikovi* their removal has been shown to involve splicing mechanisms resulting in the synthesis of additional, longer pheromone isoforms that interact with the secreted form in the cell–cell recognition phenomena that are intrinsic to the mating-type mechanism (Di Giuseppe et al. 2002; Miceli et al. 1992; Ortenzi

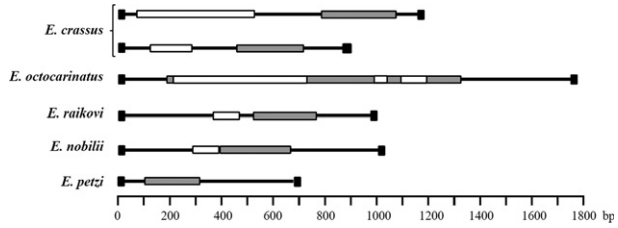

Figure 6 Comparison of the pheromone genes from *Euplotes petzi* and other *Euplotes* species. ORF encoding pre-propheromone, gray boxes; 5′ leader and 3′ trailer regions, solid lines; telomeres, black boxes; intron sequences, light boxes. In *Euplotes crassus*, two pheromone gene families are present due to an event of gene duplication. The data for *Euplotes nobilii* are from Vallesi et al. (2012) and A. Vallesi (unpubl. data), *Euplotes raikovi* from Miceli et al. (1992) and Di Giuseppe et al. (2002), *Euplotes octocarinatus* from Brünen-Nieweler et al. (1998) and Möllenbeck and Heckmann (1999), *E. crassus* from Vallesi et al. (2014) and A. Vallesi (unpubl. data).

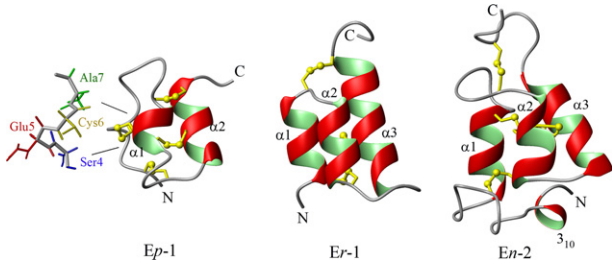

Figure 7 Comparison of the NMR structure of the *Euplotes petzi* pheromone E*p*-1 with the structures of E*r*-1 (Mronga et al. 1994; PDB, 1ERC) and E*n*-2 (Placzek et al. 2007; PDB, 2NSW), taken as representatives of the *Euplotes raikovi* and *Euplotes nobilii* pheromone families, respectively. The layout is the same as in Fig. 4. The regular α helices (α) are numbered progressively from the amino (N) to the carboxyl (C) terminus, along with the single-turn 3_{10} helix specific to pheromone E*n*-2. For *Euplotes petzi*, the four-residue stretch Ser_4–Ala_7 is magnified and shown in stick representation with the side chains.

et al. 2000; Vallesi et al. 2005). This lack of intron sequences would thus imply that the *E. petzi* pheromone genes are expressed without splicing phenomena, and motivates a specific investigation on whether and how *E. petzi* is able to generate other pheromone isoforms in addition to the secreted one.

Like their coding genes, also the *E. petzi* pheromone molecules have quite short sequences, of only 32 amino acids in the four analyzed pheromones, and take a comparatively simpler two α-helix molecular structure. This structure has been determined on native protein preparations only for the E*p*-1 pheromone. Nevertheless, it can safely be credited to be common to the whole *E. petzi* pheromone family in consideration of the close amino acid sequence similarity. Furthermore, the structures of the E*p*-2, E*p*-3, and E*p*-4 pheromones obtained applying common molecular modeling programs, such as the "Swiss-Pdb Wiewer" (swissmodel.expasy.org), are practically identical to the E*p*-1 structure used as template (data not shown). In Fig. 7, the two α-helix fold of the *E. petzi* pheromones is illustrated in comparison with the *E. raikovi* and *E. nobilii* pheromones folds, which are characterized by a common right-handed three-helix up-down-up architecture, with the sole known deviation of the *E. raikovi* E*r*-23 pheromone fold that includes two very short additional helices (Zahn et al. 2001). Yet structurally unique for including only two α-helices and lacking extended peripheral coil regions specific to the cold-adapted *E. nobilii* pheromones (Alimenti et al. 2009), the molecular structure of the *E. petzi* pheromones clearly looks as a precursor of the more engineered structures of the *E. raikovi* and *E. nobilii* pheromones. The second and third of the three helices of these pheromones find their topological equivalent in the first and the second helices, respectively, of the *E. petzi* pheromones, and the first one finds its counterpart in the four-residue segment, Ser_4–Gln-Cys-Ala/Ile_7, whose spatial arrangement in the amino-terminal segment of the *E. petzi* pheromones presages the formation of a helical motif.

At functional level, it is particularly significant that among the five amino acid substitutions distinctive of the *E. petzi* pheromones, the three Ala21/Asp, Ile24/Thr, and Ala26/Val substitutions concentrate in positions included in, or in close proximity to the second helix which thus candidates to be the primary site for receptor recognition and binding. The topological equivalence of this helix with the third one of *E. raikovi* and *E. nobilii* pheromones strongly supports this hypothesis. Indeed, the primary role of this third helix in receptor binding is known from analyses of the crystallographic structure of the *E. raikovi* pheromone E*r*-1 (Weiss et al. 1995), as well as by observations that even the oxidation of a single Met residue exposed on its surface is sufficient to cause an autocrine-to-paracrine shift in the pheromone biological activity (Alimenti et al. 2012).

ACKNOWLEDGMENTS

This article is dedicated to Kurt Wüthrich. Without his pioneering vision during the early days of NMR protein spectroscopy and his support and encouragement over the past two decades, the present extensive knowledge about *Euplotes* pheromone structures—including the results presented in this article—would not have been achieved. This work was financially supported by the Italian Programma Nazionale di Ricerca in Antartide (PNRA) and Progetti di Ricerca di Interesse Nazionale (PRIN). The NMR measurements were performed exploiting the infrastructures provided by the Biomolecular NMR Spectroscopy Platform (BNSP) of the Department of Biology at the ETH, Zürich.

COMPETING INTERESTS

The authors declare that they have no competing interests.

LITERATURE CITED

Alimenti, C., Vallesi, A., Luporini, P., Buonanno, F. & Ortenzi, C. 2012. Cell aging-induced methionine oxidation causes an

autocrine to paracrine shift of the pheromone activity in the protozoan ciliate, *Euplotes raikovi*. *Exp. Cell Res.*, 318:144–151.

Alimenti, C., Vallesi, A., Pedrini, B., Wüthrich, K. & Luporini, P. 2009. Molecular cold-adaptation: comparative analysis of two homologous families of psychrophilic and mesophilic signal proteins of the protozoan ciliate, *Euplotes*. *IUBMB Life*, 61:838–845.

Borror, A. C. & Hill, B. F. 1995. The order Euplotida (Ciliophora): taxonomy, with division of *Euplotes* into several genera. *J. Eukaryot. Microbiol.*, 42:457–466.

Brown, L. R., Mronga, S., Bradshaw, R. A., Ortenzi, C., Luporini, P. & Wüthrich, K. 1993. Nuclear magnetic resonance solution structure of the pheromone E*r*-10 from the ciliated protozoan *Euplotes raikovi*. *J. Mol. Biol.*, 231:800–816.

Brünen-Nieweler, C., Weiligmann, J. C., Hansen, B., Kuhlmann, H. W., Möllenbeck, M. & Heckmann, K. 1998. The pheromones and pheromone genes of new stocks of the *Euplotes octocarinatus* species complex. *Eur. J. Protistol.*, 34:124–132.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, N. J., Rice, L. M., Simonson, T., Gregory, L. & Warren, G. L. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, 54:905–921.

Cavalcanti, A. R. O., Dunn, D. M., Weiss, R., Herrick, G., Landweber, L. F. & Doak, T. G. 2004. Sequence features of *Oxytricha trifallax* (Class Spirotrichea) macronuclear telomeric and subtelomeric sequences. *Protist*, 155:311–322.

Curds, C. R. 1975. A guide to the species of genus *Euplotes* (Hypotrichida, Ciliatea). *Bull. Br. Mus. Nat. Hist.*, 28:1–61.

Di Giuseppe, G., Dini, F., Vallesi, A. & Luporini, P. 2015. Genetic relationships in bipolar species of the protist ciliate, *Euplotes*. *Hydrobiologia*, 761:71–83.

Di Giuseppe, G., Erra, F., Dini, F., Alimenti, C., Vallesi, A., Pedrini, B., Wüthrich, K. & Luporini, P. 2011. Antarctic and Arctic populations of the ciliate *Euplotes nobilii* show common pheromone-mediated cell-cell signaling and cross-mating. *Proc. Natl Acad. Sci. USA*, 108:3181–3186.

Di Giuseppe, G., Erra, F., Frontini, F. P., Dini, F., Vallesi, A. & Luporini, P. 2014. Improved description of the bipolar ciliate, *Euplotes petzi*, and definition of its basal position in the *Euplotes* phylogenetic tree. *Eur. J. Protistol.*, 50:402–411.

Di Giuseppe, G., Miceli, C., Zahn, R., Damberger, F., Wüthrich, K. & Luporini, P. 2002. A structurally deviant member of the *Euplotes raikovi* pheromone family: E*r*-23. *J. Eukaryot. Microbiol.*, 49:86–92.

Dini, F. & Nyberg, D. 1993. Sex in ciliates. *In*: Jones, J. G. (ed.), Advances in Microbial Ecology. Plenum Press, New York. p. 85–153.

Fadel, V., Bettendorff, P., Herrmann, T., de Azevedo Jr, W. F., Oliveira, E. B., Yamane, T. & Wüthrich, K. 2005. Automated NMR structure determination and disulfide bond identification of the myotoxin crotamine from *Crotalus durissus* terrificus. *Toxicon*, 46:759–767.

Fotedar, R., Stoeck, T., Filker, S., Fell, J. W., Agatha, S., Masoud Al Marri, M. & Jiang, J. 2016. Description of the halophile *Euplotes qatarensis* nov. spec. (Ciliophora, Spirotrichea, Euplotida) isolated from the hypersaline Khor Al-Adaid lagoon in Qatar. *J. Eukaryot. Microbiol.*, doi:10.1111/jeu.12305.

Geralt, M., Alimenti, C., Vallesi, A., Luporini, P. & Wüthrich, K. 2013. Thermodynamic stability of psychrophilic and mesophilic pheromones of the protozoan ciliate *Euplotes*. *Biology*, 2:142–150.

Ghosh, S., Jaraczewski, J. W., Klobutcher, L. A. & Jahn, C. L. 1994. Characterization of transcription initiation, translation initiation, and poly(A) addition sites in the gene-sized macronuclear DNA molecules of *Euplotes*. *Nucleic Acids Res.*, 22:214–221.

Güntert, P., Mumenthaler, C. & Wüthrich, K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, 273:283–298.

Herrmann, T., Güntert, P. & Wüthrich, K. 2002a. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, 24:171–189.

Herrmann, T., Güntert, P. & Wüthrich, K. 2002b. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, 319:209–227.

Jahn, C. L. & Klobutcher, L. A. 2002. Genome remodeling in ciliated protozoa. *Annu. Rev. Microbiol.*, 56:489–520.

La Terza, A., Dobri, N., Alimenti, C., Vallesi, A. & Luporini, P. 2009. The water-borne protein signals (pheromones) of the Antarctic ciliated protozoan *Euplotes nobilii*: structure of the gene coding for the E*n*-6 pheromone. *Can. J. Microbiol.*, 55:57–62.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 26:283–291.

Lescasse, R., Yang, T., Grisvard, J., Villalobo, E., Moch, C., Baroin-Tourancheau, A. & Morin, L. 2005. Gene structure of the ciliate *Sterkiella histriomuscorum* based on a combined analysis of DNA and cDNA sequences from 21 macronuclear chromosomes. *Chromosoma*, 114:344–351.

Luporini, P., Alimenti, C., Ortenzi, C. & Vallesi, A. 2005. Ciliate mating types and their specific signal pheromones. *Acta Protozool.*, 44:89–101.

Luporini, P., Alimenti, C., Pedrini, B. & Vallesi, A. 2016a. Ciliate communication via water-borne pheromones. *In*: Witzany, G. & Nowacki, M. (eds.), Biocommunication of Ciliate. Springer, Berlin, p. 159–174.

Luporini, P., Pedrini, B., Alimenti, C. & Vallesi, A. 2016b. Revisiting fifty years of research on pheromone signaling in ciliates. *Eur. J. Protistol*, doi:10.1016/j.ejop.2016.04.006.

Miceli, C., Concetti, A., & Luporini, P. 1983. Isolation of the mating-inducing factor of the ciliate Euplotes. *Exp. Cell Res.*, 149:593–598.

Miceli, C., La Terza, A., Bradshaw, R. A. & Luporini, P. 1992. Identification and structural characterization of a cDNA clone encoding a membrane-bound form of the polypeptide pheromone E*r*-1 in the ciliate protozoan *Euplotes raikovi*. *Proc. Natl Acad. Sci. USA*, 89:1988–1992.

Miyake, A. 1981. Cell interactions by gamones in *Blepharisma*. *In*: O'Day, D. H. & Horgan, P. A. (ed.), Sexual Interactions in Eukaryotic Microbes. Academic Press, New York. p. 95–129.

Möllenbeck, M. & Heckmann, K. 1999. Characterization of two genes encoding a fifth so far unknown pheromone of *Euplotes octocarinatus*. *Eur. J. Protistol.*, 35:225–230.

Mronga, S., Luginbühl, P., Brown, L. R., Ortenzi, C., Luporini, P., Bradshaw, R. A. & Wüthrich, K. 1994. The NMR solution structure of the pheromone E*r*-1 from the ciliated protozoan *Euplotes raikovi*. *Protein Sci.*, 3:1527–1536.

Nielsen, H., Engelbretch, J., Brunak, S. & von Hijene, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.*, 10:1–6.

Ortenzi, C., Alimenti, C., Vallesi, A., Di Pretoro, B., La Terza, A. & Luporini, P. 2000. The autocrine mitogenic loop of the ciliate *Euplotes raikovi*: the pheromone membrane-bound forms are the cell binding sites and potential signalling receptors of soluble pheromones. *Mol. Biol. Cell*, 11:1445–1450.

Pedrini, B., Placzek, W. J., Koculi, E., Alimenti, C., La Terza, A., Luporini, P. & Wüthrich, K. 2007. Cold-adaptation in sea-water-borne signal proteins: sequence and NMR structure of the pheromone E*n*-6 from the Antarctic ciliate *Euplotes nobilii*. *J. Mol. Biol.*, 372:277–286.

Placzek, W. J., Etezady-Esfarjani, T., Hermann, T., Pedrini, B., Peti, W., Alimenti, C., Luporini, P. & Wuthrich, K. 2007. Cold-adapted signal proteins: NMR structures of pheromones from the Antarctic ciliate *Euplotes nobilii*. *IUBMB Life*, 59:578–585.

Raffioni, S., Miceli, C., Vallesi, A., Chowdhury, S. K., Chait, B. T., Luporini, P. & Bradshaw, R. A. 1992. Primary structure of *Euplotes raikovi* pheromones: comparison of five sequences of pheromones from cells with variable mating interactions. *Proc. Natl Acad. Sci. USA*, 89:2071–2075.

Stewart, A. E., Raffioni, S., Chaudhary, T., Chait, B. T., Luporini, P. & Bradshaw, R. A. 1992. The disulphide bond pairing of the pheromones E*r*-1 and E*r*-2 of the ciliated protozoan *Euplotes raikovi*. *Protein Sci.*, 1:777–785.

Uspenskaya, Z. I. & Yudin, A. L. 2016. Fifty years of research on serotypes and mating types in *Dileptus anser*: a review. *Eur. J. Protistol.*, 53:31–44.

Vallesi, A., Alimenti, C., Federici, S., Di Giuseppe, G., Dini, F., Guella, G. & Luporini, P. 2014. Evidence for gene duplication and allelic codominance (not hierarchical dominance) at the mating type locus of the ciliate, *Euplotes crassus*. *J. Eukaryot. Microbiol.*, 61:620–629.

Vallesi, A., Alimenti, C., Pedrini, B., Di Giuseppe, G., Dini, F., Wüthrich, K. & Luporini, P. 2012. Coding genes and molecular structures of the diffusible signalling proteins (pheromones) of the polar ciliate, *Euplotes nobilii*. *Mar. Genomics*, 8:9–13.

Vallesi, A., Ballarini, P., Di Pretoro, B., Alimenti, C., Miceli, C. & Luporini, P. 2005. The autocrine, mitogenic pheromone-receptor loop of the ciliate *Euplotes raikovi*: pheromone-induced receptor internalization. *Eukaryot. Cell*, 4:1221–1227.

Vinogradov, D. V., Tsoi, O. V., Zaika, A. V., Lobanov, A. V., Turanov, A. A., Gladishev, V. N. & Gel'fand, M. S. 2012. Draft macronucleus genome of *Euplotes crassus* ciliate. *Mol. Biol.*, 46:328–333.

Weiss, M. S., Anderson, D. H., Raffioni, S., Bradshaw, R. A., Ortenzi, C., Luporini, P. & Eisenberg, D. 1995. A cooperative model for ligand recognition and cell adhesion: evidence from the molecular packing in the 1.6 Å crystal structure of the pheromone E*r*-1 from the ciliate protozoan *Euplotes raikovi*. *Proc. Natl Acad. Sci. USA*, 92:10172–10176.

Wüthrich, K. 1986. NMR of proteins and nucleic acids. Wiley, New York.

Zahn, R., Damberger, F., Ortenzi, C., Luporini, P. & Wüthrich, K. 2001. NMR structure of the *Euplotes raikovi* pheromone E*r*-23 and identification of its five disulfide bonds. *J. Mol. Biol.*, 313:923–931.