Structural characterization of a protein pheromone from a cold-adapted (Antarctic) single-cell eukaryote, the ciliate *Euplotes nobilii*

Claudio Alimenti[a], Claudio Ortenzi[a], Vito Carratore[b], Pierangelo Luporini[a,*]

[a] *Dipartimento di Biologia Molecolare Cellulare e Animale, University of Camerino, 62032 Camerino (MC), Italy*
[b] *Istituto di Biochimica delle Proteine ed Enzimologia, CNR/IBPE, Via G. Marconi, 80125 Napoli, Italy*

Received 18 January 2002; accepted 1 February 2002

First published online 15 February 2002

Edited by Hans Eklund

Abstract Free-living species of ciliated Protozoa control their vegetative (mitotic) proliferation and mating (sexual) processes by diffusible, cell type-specific protein signals (pheromones). One of these molecules, designated E*n*-2, was isolated from a species, *Euplotes nobilii*, living in the stably cold marine waters of Antarctica, and its complete amino acid sequence of 60 residues was determined by automated Edman degradation of the whole protein and peptides generated by trypsin digestion. The proposed sequence is: DIEDFYTSETCPYKNDSQLA$_{20}$-WDTCSGGTGNCGTVCCGQCF$_{40}$SFPVSQSCAGMAD-SNDCPNA$_{60}$. The E*n*-2 structure appears to be characterized by an adaptive insertion of a glycine-rich motif potentially capable to confer more flexibility to a functionally critical region of the molecule.

Key words: Cold-adapted protein; Glycine-rich motif; Chemical cell signal; Ciliate mating type; Antarctic biology

1. Introduction

Cell type-specific chemical signals, currently referred to as pheromones, are constitutively secreted into the extra-cellular environment by a number of species of free-living protozoan ciliates [1,2]. For long since the first characterization of two such signals in *Blepharisma japonicum*, one identified with a glycoprotein of about 20 kDa and the other with a tryptophan derivative [3], these signals have been regarded functionally only as 'sex factors' due to their capacity to induce mating pair formation between cells to which they bind in a paracrine-like (heterologous) fashion [2,4]. However, this view was eventually changed by results from more recent studies involving a ciliate, *Euplotes raikovi*, in which pheromones have been shown to be represented by a species-specific family of structurally homologous proteins [5,6], consistently with their multi-allelic determination at a single genetic locus [7,8]. In addition to binding to cells in paracrine-like fashion for mating purposes, *E. raikovi* pheromones act also like prototypic autocrine growth factors, as they can bind to, and promote the vegetative (mitotic) growth (proliferation) of the same cells from which they are constitutively secreted throughout the life cycle [9]. Such a wider spectrum of target cells and activities presumes a coexistence of a multiplicity of specific and shared traits in the structure of these molecules, as well as an evolutionary conservation of some of these traits in other signal molecules. Consistently with this, *E. raikovi* pheromones have been found capable to bind also to subunits of the mammalian cell receptor for interleukin 2 [10], and basic motifs of their three-dimensional organization have been found suitable as templates for fold modelling of other water-borne protein pheromones (denominated 'attractins' to indicate their role in stimulating reproductive adults to approach egg cordons) that have recently been characterized from a number of species of the marine mollusk, *Aplysia* [11].

Here we describe the complete amino acid sequence of a new ciliate pheromone, denominated E*n*-2, that we have isolated from cultures of an Antarctic, marine species of *Euplotes*, *E. nobilii* [12]. This characterization provides new information on eco-physiological properties of organisms adapted to survive in extreme environments, and may facilitate studies on cell signalling at physiologically low temperatures in a simple experimental model.

2. Materials and methods

2.1. Cells and protein purification

Cells of strain AC-1 represented the E*n*-2 source. They were grown at 4°C on the green alga *Dunaliella tertiolecta*, and their cultures were expanded and prepared to purify E*n*-2 as previously reported [12]. The purification procedure was carried out according to a standard protocol originally devised for *E. raikovi* pheromones [8]. It essentially involved pheromone adsorption from cell-free supernatant onto Sep-Pak C_{18} cartridges (Water Corporation, Milford, MA, USA) and subsequent chromatographic fractionations of the cartridge eluate on Superose-12 and Mono-Q columns (Amersham Pharmacia Biotech, Piscataway, NJ, USA). Typically, 5 μg of homogeneous protein was prepared from 1 l of culture supernatant and each preparation was found to be active, on the basis of mating-inducing assays [12], at concentrations from 10^{-8} to 10^{-9} M. Purified material, unless used within a few days, was stored at −20°C (without losing activity) after lyophilization.

2.2. Protein modification and cleavage

To carry out E*n*-2 trypsin digestion, a purified preparation was (1) re-suspended at a final concentration of 0.5 μg/μl in 0.1 M Tris, pH 8.1, containing 8 M urea (Mallinckrodt Baker, Phillipsburg, NJ, USA), (2) reduced with 100 mM dithiothreitol (DTT; Sigma-Aldrich, St. Louis, MO, USA) for 2 h at 50°C, (3) *S*-pyridyl-ethylated with 1% (v/v) 4-vinylpyridine (Sigma-Aldrich) for 2 h at room temperature, (4) immediately desalted by gel-filtration chromatography on a Superdex Peptide HR 10/30 column (Amersham Pharmacia Biotech) equilibrated with 30% acetonitrile (Mallinckrodt Baker), (5) lyophilized, and (6) incubated at a final concentration of 0.2 μg/μl, for 15 h at room

*Corresponding author. Fax: (39)-737-636216.
E-mail address: piero.luporini@unicam.it (P. Luporini).

Abbreviations: DTT, dithiothreitol; RP-HPLC, reverse-phase high performance liquid chromatography

0014-5793/02/$22.00
PII: S0014-5793(02)02393-1

temperature, in 0.1 M Tris, pH 8.1, with 2 M urea, 2 mM $CaCl_2$, and trypsin (Sigma-Aldrich) (enzyme/substrate ratio, 1/50 w/w).

2.3. Peptide separation

The E*n*-2 enzymatic digest was fractionated by reverse-phase high performance liquid chromatography (RP-HPLC) on a 2.1×25-cm C_{18} column (Supelco, Bellefonte, PA, USA) that was eluted with a linear gradient of acetonitrile (2–37.5% in 60 min) in the presence of 0.06% (v/v) trifluoroacetic acid, at a flow rate of 200 μl/min. The separated E*n*-2 peptides were lyophilized and stored at −20°C prior to use.

2.4. Mass spectrometry and amino acid sequence analysis

Molecular mass determinations were carried out on a matrix-assisted laser desorption/ionization time of flight Voyager DE-PRO apparatus (Perseptive Biosystems, Framingham, MA, USA) equipped with a nitrogen laser (operated at 337 nm). Automated Edman degradation was carried out in a Procise 492 protein sequencer (Applied Biosystem, Perkin Elmer Division, Foster City, CA, USA).

3. Results

Purified preparations of the pheromone E*n*-2 were used for the amino acid sequence determination after having been reduced with DTT and alkylated by incubation with 4-vinylpyridine, to allow identification of cysteine residues modified by pyridylethylation. Eight cysteine residues had previously been suggested to be present in E*n*-2 by data (not shown) of amino acid composition. This number was subsequently confirmed by mass spectrometric comparison between native and alkylated E*n*-2 samples (Fig. 1). In fact, while the mass of a native sample was measured to be 6298.92 Da, that of an alkylated sample was 850.46 Da greater, as was to be expected from the conversion of eight cysteines into 4-pyridylethylcysteine residues (the mass of each modified cysteine involving an increase of 105 Da).

It was first possible to identify unequivocally 30 residues at

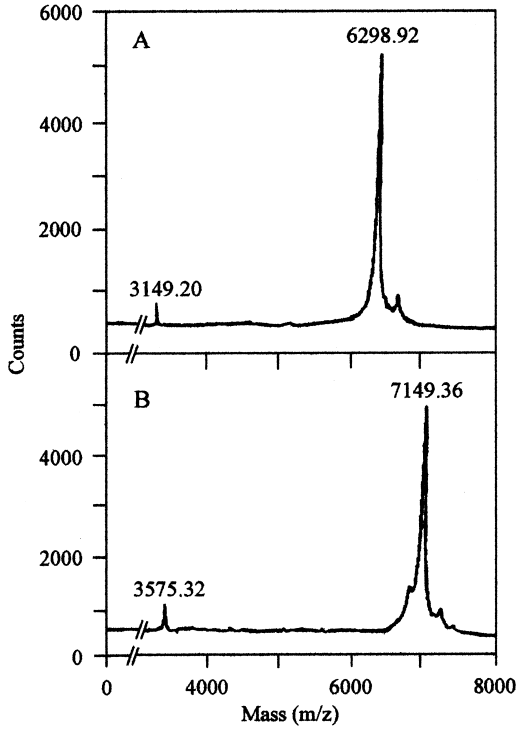

Fig. 1. Mass spectrometric analysis of native and alkylated E*n*-2. Only the regions of the spectra up to 8000 are shown. A: Native E*n*-2; B: alkylated E*n*-2.

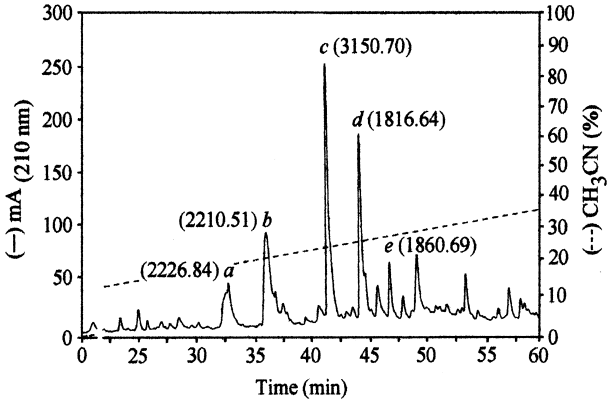

Fig. 2. Tryptic map of E*n*-2. A purified E*n*-2 preparation was reduced with DTT and alkylated with 4-vinylpyridine before being exposed to digestion with trypsin. The chromatography of the E*n*-2 digest was eluted with an acetonitrile linear gradient and the E*n*-2 fragments detected by their absorbance at 210 nm. The fragments relative to peaks *a* to *e* were analyzed by mass spectrometry to determine their molecular mass (reported in parentheses) and automated Edman degradation to determine their sequence.

the amino-terminus of the E*n*-2 sequence by automated Edman degradation of an intact sample of alkylated E*n*-2. The other residues necessary for completing this sequence were then identified by analysis of peptides generated by trypsin digestion of other alkylated samples of E*n*-2. Numerous peaks were produced by the E*n*-2 digest in RP-HPLC (Fig. 2). However, only the peptides of those five peaks (designated *a* to *e* in Fig. 2) which had no counterpart in the chromatographic profile of peptides resulting from trypsin autolysis were further analyzed. Of each peptide, hereafter designated with the same letter as that of its relative peak, first the molecular mass was determined and, then, the amino acid sequence.

Only the sequence of peptide *c*, showing a mass of 3150.7 Da, was found to be unique. Instead, the 2226.84-Da sequence of peptide *a* turned out to be the same as the 2210.51-Da sequence of peptide *b*, and the 1816.64-Da sequence of peptide *d* to be the same as the 1860.69-Da sequence of peptide *e*. The mass variation of 16 Da between peptides *a* and *b* was assumed to arise from oxidation of the single methionine residue of peptide *a* to the corresponding sulfoxide derivative. Similarly, the mass variation of 44 Da between peptides *e* and *d* was assumed to be generated by carboxylation of one residue of glutamic, or of aspartic, acid in peptide *e*. In effect, both these modifications can occur readily in protein analysis [13].

The complete E*n*-2 sequence of 60 residues was reconstructed by aligning the sequences of peptides *d* (=*e*), *c*, and *a*(=*b*) in that order (Fig. 3). This alignment was decided on the basis of the knowledge of the 30 residues of the amino-terminal half of the E*n*-2 sequence. It was in fact found that: (i) the sequence of peptide *d* (ending with a lysine specific for the trypsin activity) overlaps the first 14 of these residues, and (ii) the first half of the sequence of peptide *c* (ending with a phenylalanine that is likely to have represented a site of chemotrypsin activity present in the trypsin preparation used) overlaps the other 16 residues. Thus, the second half of the sequence of peptide *c* and the sequence of peptide *a* necessarily completed the E*n*-2 sequence. Two considerations relative to the peptide-*a* sequence further supported this conclusion: (i) it contains two cysteines that, added to the six of peptides *d*

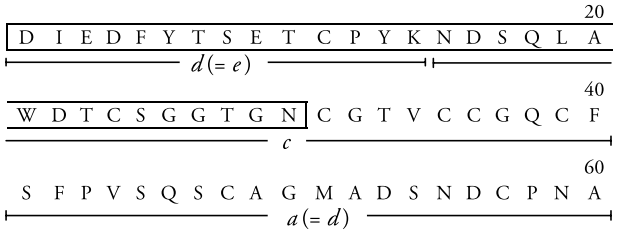

Fig. 3. Proposed amino acid sequence of E*n*-2. The sequence segment determined by analysis of intact E*n*-2 is enclosed in a box. Sequence segments determined by analysis of E*n*-2 tryptic fragments are underlined, whose designations *d* (= *e*), *c*, and *a* (= *d*) reflect the peak designations in the chromatographic profile of Fig. 2. The sequence is available from the Swiss-Prot database under the accession number P83235.

and *c*, bring the total of the E*n*-2 cysteines to the eight initially determined; (ii) the addition of its mass to those of peptides *c* and *d* results in the overall mass of an alkylated E*n*-2 molecule. Lastly, the calculated molecular weight of 6303.8 for the proposed E*n*-2 sequence is in good agreement with the value of 6298.92 Da determined for one molecule of native E*n*-2 by mass spectrometry analysis.

4. Discussion

Pheromone E*n*-2 of *E. nobilii* is the first diffusible protein hormone to be characterized at level of complete amino acid sequence from a ciliate (and, more in general, a protist) living in permanently cold waters (from −1.8°C in winter to about 3–4°C in summer), such as those surrounding the Antarctic coasts. Other ciliate pheromone structures are known only from two other *Euplotes* species of temperate waters, *E. raikovi* [6,14–17] and *E. octocarinatus* [18–20].

Structural similarities link E*n*-2 closer to the group of *E. raikovi* pheromones with sequences of 37–40 residues and six cysteines [6,14–16], than to *E. octocarinatus* pheromones whose sequences comprise 85–109 residues and usually 10 cysteines [18–20]. Similarly to *E. raikovi* pheromones, E*n*-2 has a relatively small molecular mass (6300 Da), an aspartic acid at the N-terminus, a high cysteine component (eight residues), a rather acidic p*I* (3.5), and a strong resistance to degradation (consistently with its nature of environmental chemical signal). Also, the E*n*-2 configuration might be essentially α-helical as in *E. raikovi* pheromones, notwithstanding the inconsistency of this assumption with most of the indications provided by predictive methods of protein secondary structures (Fig. 4).

According to these indications, non-regular and/or β-sheet elements would predominate in the E*n*-2 structure. However, similar predictions for *E. raikovi* pheromones have been disproved by results from crystallographic and nuclear magnetic resonance analyses [15,17]. Not only have these analyses established that the structure of these molecules exclusively consists of a bundle of three α-helices tightly associated by three disulfide bonds [6,14–16], or (limitedly to the case of pheromone E*r*-23) of a bundle of five helices interconnected by five disulfide bonds [17]; but they have also provided evidence that a key role in enforcing and stabilizing these configurations is played by the high density and strategic disposition of the disulfide bonds [17]. The fact that six of the eight E*n*-2 cysteines find their precise counterparts in the 'standard' six cysteines of *E. raikovi* pheromones (Fig. 4) implies the formation of an analogous complex of disulfide bonds and a similar function of this complex (further reinforced by a putative fourth bond between the two additional E*n*-2 cysteines) in the stabilization of an essentially α-helical organization of the E*n*-2 molecule.

A prominent feature that, instead, distinguishes E*n*-2 from

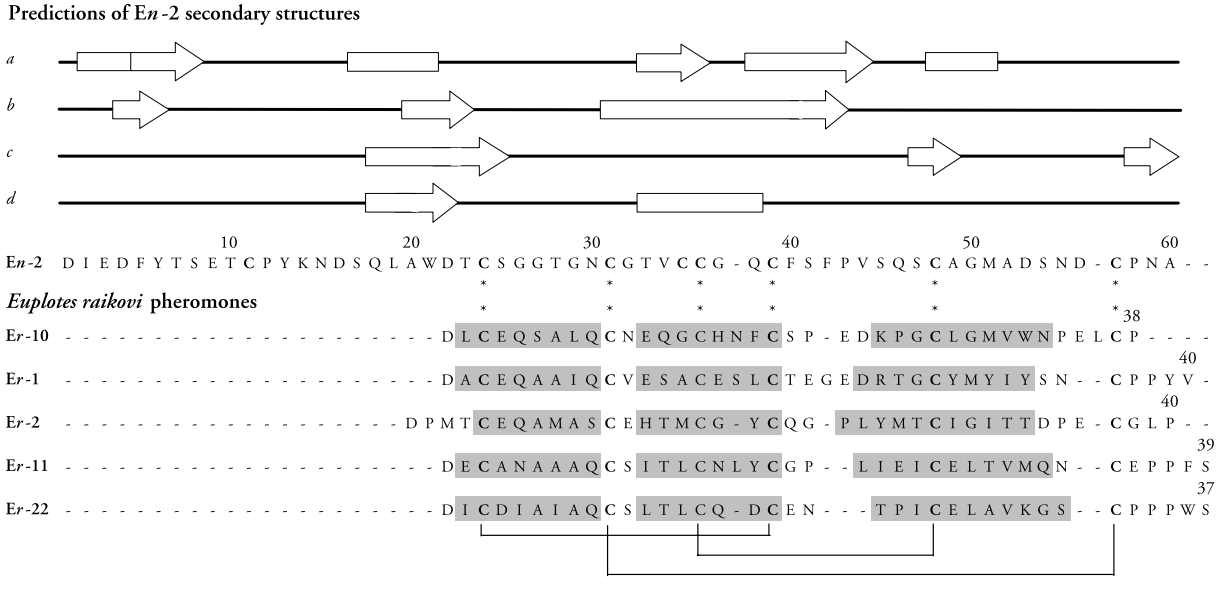

Fig. 4. Predictions of E*n*-2 secondary structures and E*n*-2 sequence alignment with *E. raikovi* pheromones. Structures were predicted by DPM [21], PHD [22], GOR IV [23], and PSIPRED [24] methods (*a–d*, respectively). Coil, β-sheet, and α-helix regions are represented by lines, arrows, and cylinders, respectively. Multiple sequence alignment between E*n*-2 and the *E. raikovi* pheromones was based on the CLUSTAL algorithm [25]. Only *E. raikovi* pheromones with six cysteines and a known three-dimensional structure have been considered [6,14–16]. Cysteine residues (printed in bold) located in conserved positions between E*n*-2 and *E. raikovi* pheromones are marked by asterisks and paired by lines to denote the disulfide arrangement originally determined in *E. raikovi* pheromones E*r*-1 and E*r*-2 [26]. In *E. raikovi* pheromones, sequence segments involved in the formation of α-helix structures are shaded.

the other *Euplotes* pheromones is its relatively higher numbers of glycine and serine residues, 13 in total, representing nearly one fourth of the 60 residues of the E*n*-2 sequence. Multiple local insertions of these amino acids have been described in various proteins of psychrophilic organisms, and are usually thought to be associated with functional aspects of adaptive strategies of proteins that need to withstand thermal variations [27–32]. Particularly credited is the concept that the introduction of multiple glycine residues into a critical region of the molecular structure represents a common device of cold-adapted proteins for increasing their local flexibility, necessary to compensate for the reduction of chemical reaction rates at low temperatures [29,30,33]. A closer look at the arrangement of the glycines in the E*n*-2 sequence provides further support to this concept. Three of these glycines are arranged in a cluster, $G_{26}GTGNC_{31}$, that is closely reminiscent of the glycine-rich repeats (GGXGXD) described as a specific structural feature of extracellular metalloproteases of the serralysin family characterized from the psychrotrophic bacterium, *Pseudomonas fluorescens* [34]. A specific role played by this cluster in making more flexible a functionally important domain of the E*n*-2 structure is strongly suggested by its localization within a sequence segment whose counterpart in *E. raikovi* pheromones appears to be systematically inhibited for glycine residues (Fig. 4). This segment in fact represents a structural block of these molecules that is conservatively organized into a regular α-helix [6,14,16,20], and is known to be directly involved in processes of cooperative oligomerization in connection with pheromone binding to their membrane pheromone receptors [15,35]. Any modification of its structure thus appears destined to affect the overall biological activity of the molecule.

Acknowledgements: This work is dedicated to the memory of Prof. Renzo Nobili. Financial support was provided by the Italian National Program of Antarctic Research (P.N.R.A.).

References

[1] Luporini, P., Vallesi, A., Miceli, C. and Bradshaw, R.A. (1995) J. Eukaryot. Microbiol. 42, 208–212.
[2] Miyake, A. (1996) in: Ciliates – Cells as Organisms (Hausmann, K. and Bradbury, P.C., Eds.), pp. 243–290, Gustav Fischer, Stuggart.
[3] Miyake, A. (1981) in: Sexual Interactions in Eukaryotic Microbes (O'Day, D.H. and Horgen, P.A., Eds.), pp. 95–129, Academic Press, New York.
[4] Kuhlmann, H.W. and Heckmann, K. (1989) J. Exp. Zool. 251, 316–328.
[5] Raffioni, S., Miceli, C., Vallesi, A., Chowdhury, S.K., Chait, B.T. and Luporini, P. (1992) Proc. Natl. Acad. Sci. USA 89, 2071–2075.
[6] Luginbühl, P., Ottiger, M., Mronga, S. and Wüthrich, K. (1994) Protein Sci. 3, 1537–1546.
[7] Luporini, P. and Miceli, C. (1986) in: Molecular Biology of Ciliated Protozoa (Gall, J., Ed.), pp. 263–299, Academic Press, New York.
[8] Luporini, P., Raffioni, S., Concetti, A. and Miceli, C. (1986) Proc. Natl. Acad. Sci. USA 83, 2889–2893.
[9] Vallesi, A., Giuli, G., Bradshaw, R.A. and Luporini, P. (1995) Nature 376, 522–524.
[10] Vallesi, A., Giuli, G., Ghiara, P., Scapigliati, G. and Luporini, P. (1998) Exp. Cell Res. 241, 253–259.
[11] Schein, C.H., Nagle, G.T., Page, J.S., Sweedler, J.V., Xu, Y., Painter, S.D. and Braun, W. (2001) Biophys. J. 81, 463–472.
[12] Felici, A., Alimenti, C., Ortenzi, C. and Luporini, P. (1999) Ital. J. Zool. 66, 355–360.
[13] Jardine, I. (1990) Methods Enzymol. 193, 441–445.
[14] Luginbühl, P., Wu, J., Zerbe, O., Ortenzi, C., Luporini, P. and Wüthrich, K. (1966) Protein Sci. 5, 1512–1522.
[15] Weiss, M.S., Anderson, D.H., Raffioni, S., Bradshaw, R.A., Ortenzi, C., Luporini, P. and Eisenberg, D. (1995) Proc. Natl. Acad. Sci. USA 92, 10172–10176.
[16] Liu, A., Luginbühl, P., Zerbe, O., Ortenzi, C., Luporini, P. and Wüthrich, K. (2001) J. Biomol. NMR 19, 75–78.
[17] Zahn, R., Damberger, F., Ortenzi, C., Luporini, P. and Wüthrich, K. (2001) J. Mol. Biol. 313, 923–931.
[18] Schlze-Dieckhoff, H., Freiburg, M. and Heckmann, K. (1987) Eur. J. Biochem. 168, 89–94.
[19] Brünen-Nieweler, C., Weiligmann, J.C., Hansen, B., Kuhlmann, H.W., Möllenbeck, M. and Heckmann, K. (1998) Eur. J. Protistol. 34, 124–132.
[20] Möllenbeck, M. and Heckmann, K. (1999) Eur. J. Protistol. 35, 225–230.
[21] Deléage, G. and Roux, B. (1987) Protein Eng. 1, 289–294.
[22] Rost, B. (1996) Methods Enzymol. 266, 525–539.
[23] Garnier, J., Gibrat, J.-F. and Robson, B. (1996) Methods Enzymol. 266, 540–553.
[24] Jones, D.T. (1999) J. Mol. Biol. 292, 195–202.
[25] Jeammougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and Gibson, T.J. (1998) Trends Biochem. Sci. 23, 403–405.
[26] Stewart, A.E., Raffioni, S., Chaudhary, T., Chait, B.T., Luporini, P. and Bradshaw, R.A. (1992) Protein Sci. 1, 777–785.
[27] Feller, G., Narinx, E., Arpigny, J.L., Aittaleb, M., Baise, E., Genicot, S. and Gerday, C. (1996) FEMS Microb. Rev. 18, 189–202.
[28] Feller, G., Arpigny, J.L., Narinx, E. and Gerday, C. (1997) Comp. Biochem. Physiol. 118, 495–499.
[29] Gerday, C., Aittaleb, M., Arpigny, J.L., Baise, E., Chessa, J.P., Garsoux, G., Petrescu, I. and Feller, G. (1997) Biochem. Biophys. Acta 1342, 119–131.
[30] Marshall, C.J. (1997) Trends Biotechnol. 15, 359–364.
[31] Devos, N., Ingouff, M., Loppes, R. and Matagne, R.F. (1998) J. Phycol. 34, 655–660.
[32] Wintrode, P.L., Miyazaki, K. and Arnold, F.H. (2001) Biochem. Biophys. Acta 1549, 1–8.
[33] Laurell, H., Contreras, J.A., Castan, I., Langin, D. and Holm, C. (2000) Protein Eng. 13, 711–717.
[34] Kumeta, H., Hoshino, T., Goda, T., Okayama, T., Shimada, T., Ohgiya, S., Matsuyama, H. and Ishizaki, K. (1999) Biosci. Biotechnol. Biochem. 63, 1165–1170.
[35] Ortenzi, C., Alimenti, C., Vallesi, A., Di Pretoro, B., La Terza, A. and Luporini, P. (2000) Mol. Biol. Cell 11, 1445–1455.