



University of Groningen

Cloning and characterization of an aromatic amino acid and leucine permease of Penicillium chrysogenum

Trip, Hein; Evers, Melchior E.; Konings, Wil N.; Driessen, Arnold J.M.

Published in: Biochimica et Biophysica Acta-Biomembranes

DOI: 10.1016/S0005-2736(02)00510-2

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version Publisher's PDF, also known as Version of record

Publication date: 2002

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA): Trip, H., Evers, M. E., Konings, W. N., & Driessen, A. J. M. (2002). Cloning and characterization of an aromatic amino acid and leucine permease of Penicillium chrysogenum. *Biochimica et Biophysica Acta-Biomembranes*, 1565(1), 73 - 80. https://doi.org/10.1016/S0005-2736(02)00510-2

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Biochimica et Biophysica Acta 1565 (2002) 73-80



Cloning and characterization of an aromatic amino acid and leucine permease of *Penicillium chrysogenum*

Hein Trip, Melchior E. Evers, Wil N. Konings, Arnold J.M. Driessen*

Department of Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

Received 26 March 2002; accepted 10 July 2002

Abstract

The gene encoding the amino acid permease ArlP (*Aromatic and leucine Permease*) was isolated from the filamentous fungus *Penicillium chrysogenum* after PCR using degenerated oligonucleotides based on conserved regions of fungal amino acid permeases. The cDNA clone was used for expression of the permease in *Saccharomyces cerevisiae* M4054, which is defective in the general amino acid permease Gap1. Upon overexpression, an increase in the uptake of L-tyrosine, L-phenylalanine, L-tryptophan and L-leucine was observed. Further competition experiments indicate that ArlP recognizes neutral and aromatic amino acids with an unbranched β -carbon atom. \bigcirc 2002 Elsevier Science B.V. All rights reserved.

Keywords: Amino acid permease; Penicillium chrysogenum; Gene expression; Saccharomyces cerevisiae; Amino acid uptake

1. Introduction

Fungi can use amino acids either as nitrogen and/or carbon source or as building blocks for protein synthesis. The uptake of amino acids is mediated by a multiplicity of carriers with different, overlapping degree of specificity. In addition, cells harbor one or more broadly specific, large capacity, general amino acid permeases [1,2]. Most of the sequenced and characterized fungal amino acid permeases belong to one related group, referred to as the AAP family [3]. This family also includes some bacterial amino acid permeases. These transport systems have a common structural organization with 12 putative α -helical transmembrane segments and cytoplasmically located N- and C-terminal hydrophilic regions [2]. Transport occurs by proton symport with the pmf as driving force [4,5].

In *Neurospora crassa*, an amino acid permease specific for neutral aliphatic and aromatic amino acids was identified, which is not related to the AAP family [6-8]. It is a member of the amino acid/auxin permease (AAAP) family, which includes auxin and amino acid permeases from plants and animals [9].

In the yeast Saccharomyces cerevisiae, 24 members of the AAP family have been found [3,10,11] of which most have been functionally characterized [12]. Based on their regulation at the transcriptional level, the permeases in S. cerevisiae are divided in two classes. The general amino acid permease Gap1p, which transports all D- and L-amino acids with high capacity [13], the proline permease Put4p [14] and the acidic amino acid permease Dip5p [15] are subject to nitrogen catabolite repression. The transcription of their genes is repressed when a rich nitrogen source like ammonia or arginine is available [16]. The expression of most low-capacity, highly specific and high-affinity permeases requires induction by sensing of an extracellular amino acid [17]. Examples are the branched-chain amino acid permeases, Bap2p [18] and Bap3p, the tyrosine and tryptophan permease Tat1p [19] and the histidine permease, Hip1p [20].

In filamentous fungi, amino acid uptake has been studied mainly by transport assays and characterization of mutants defective in uptake of one or more amino acids. In *N. crassa*, five distinct transport systems (systems I–V) have been identified. System III corresponds to the general amino acid permease. The other systems are specific for a group of related amino acids. A similar situation seems to hold for *Aspergillus nidulans* (Ref. [1] and references therein). The L-proline specific transport system, encoded by *prnB*, was

^{*} Corresponding author. Tel.: +31-50-3632164; fax: +31-50-3632154. *E-mail address:* a.j.m.driessen@biol.rug.nl (A.J.M. Driessen).

the first amino acid permease from a filamentous fungus that has been sequenced [21] and characterized with respect to kinetics and regulation [22]. For the penicillin-producing filamentous fungus *Penicillium chrysogenum*, nine different transport systems have been reported: system I for Lmethionine [23]; II for L-cysteine [24]; III for neutral and basic amino acids [25]; IV for acidic amino acids; V for L-proline; VI for L-lysine and L-arginine, VII for Larginine; VII for L-lysine and IX for L-cysteine [26]. The first two systems are expressed under sulfur starvation, while systems III–V are expressed under nitrogen and carbon starvation (NCR and CCR). Systems VI–VIII appear constitutively [1]. Here we describe the cloning of a gene that specifies a novel amino acid permease from *P. chrysogenum* with an unusual specificity.

2. Materials and methods

2.1. Strains and media

The P. chrysogenum strain used in this study is Wisconsin 54-1255, a low-level penicillin-producing strain (kindly provided by DSM). Cultures were started on YPG medium (1% yeast extract, 2% peptone and 2% glucose) and incubated for 16 h and subsequently diluted into minimal medium supplemented with 4% glucose and 0.4% ammonium acetate as described by Ref. [27]. S. cerevisiae M4054 strain (MAT α , ura3, $\Delta gap1$) [18] was used for the functional expression and transport assays. Cells were grown in a buffered minimal medium containing 1% succinate, 0.6% NaOH, 0.16% Yeast Nitrogen Base without ammonium sulfate and without amino acids (Difco), 2% glucose and 0.1% proline as a nitrogen source (MP), or on YPg. Minimal medium containing 0.5% (NH₄)₂SO₄ instead of proline was used for selection of transformants containing the plasmid yCA and yCN (see below). E. coli DH5a was employed for all cloning steps, carried out as described in Ref. [28].

2.2. Amplification of amino acid permease gene fragment by *PCR*

Degenerated oligonucleotides (5'-ATGATSKCCMTSG-GYGSYISYMTIGGIACIGG-3' and 5'-GAAYTC-GRIYTCRSCRWAIGGICG-3') were designed based on conserved regions of fungal amino acid permease genes. PCR was performed on chromosomal DNA of *P. chrysogenum* Wisconsin 54–1255 in a Perkin-Elmer thermocycler for 30 cycles, with each cycle consisting of 94 °C for 45 s, 40 °C for 1 min and 72 °C for 1 min. At the end of 30 cycles, samples were incubated for 5 min at 72 °C. Gel purified products, ranging from approximately 350 to 500 bp were subcloned into the pGEM-T Easy vector (Promega) for DNA sequence determination. Of the 25 clones sequenced, two identical clones showed homology with known amino acid transporter-encoding genes.

2.3. Isolation of permease gene from genomic library

A genomic library of *P. chrysogenum* DNA in phage Lambda ZAP II (Stratagene), was screened by plaque hybridization as described by Ref. [28]. Approximately 30,000 plaques were transferred to Zeta-Probe blotting membranes (BioRad), and probed with the radioactively labeled abovementioned PCR product. One positive plaque was isolated, phages were released and pBluescript was rescued from the Lambda ZAP II vector by in vivo excision, using M13. The 8kb insert carried the complete permease gene.

2.4. Expression in S. cerevisiae M4054

For expression, a construct was created based on the yeast-*E. coli* shuttle vector yEP352 [29]. The copper inducible promoter pCUP1 of *S. cerevisiae* was inserted into yEP352. The cDNA of the amino acid permease gene was amplified by PCR from a cDNA library and cloned behind pCUP1. The terminator of *AatA*, encoding acyl coenzyme A:isopenicillin N acyltransferase of *P. chrysogenum*, was cloned behind the permease gene resulting in the plasmid yCA. For control experiments, the same construct without the permease gene was used, denoted yCN. Plasmids were transformed into *S. cerevisiae* strain M4054. To test transcript levels, total RNA was isolated from cells disrupted by vortexing with glassbeads and extracted with Trizol reagent (Gibco BRL). Specific *arlP* mRNA levels were measured by quantitative RT-PCR.

2.5. Transport studies

Overnight cultures grown in minimal medium containing proline as the sole nitrogen source were diluted with fresh medium, and growth was continued at 30 °C until an OD₆₀₀ of 0.4-0.8. Cells from 50 ml of culture were harvested, washed in minimal medium without any nitrogen source and resuspended to OD_{600} of 10 in the same medium. After 10 min of preincubation at 30 °C, 100 µl of cells were added to 150 μ l of the same medium containing 25 μ M ¹⁴C-labeled Lamino acid (0.05 µCi). For competition studies, the uptake of 250 μ M [¹⁴C-] L-tyrosine was analyzed in the presence of a 10-fold excess of an unlabeled L-amino acid. For the determination of the $K_{\rm m}$ value, the concentration of Ltyrosine was varied between 25 μM and 2.5 mM, while for $K_{\rm I}$ determinations, the kinetics was analyzed in the presence of 500 µM of the indicated unlabeled amino acid. The uptake of amino acid was stopped by the addition of 2 ml of ice-cold 0.1 M LiCl, immediately followed by filtration through a 0.45-µm pore size nitrocellulose filter. The filters were washed with 2 ml of ice-cold 0.1 M LiCl, submerged in scintillation fluid and the retained radioactivity was measured in a scintillation counter.

2.6. Expression of arlP

For expression studies, main cultures were incubated for 48 h (early stationary phase) in a rotary shaker at 200 rpm and 25 °C, then transferred to fresh minimal medium containing either urea (0.4%, w/v), ammonium acetate (0.4%, w/v) or ammonium acetate plus 10 mM tyrosine, phenylalanine or leucine as nitrogen source. Incubation was continued for 5 h, and mycelial samples were taken at various times, washed and ground with liquid nitrogen. RNA was extracted using Trizol reagent (Gibco BRL) and specific *arlP* mRNA levels were measured by quantitative RT-PCR.

3. Results

3.1. Cloning of amino acid permeases of P. chrysogenum

To clone genes encoding the *P. chrysogenum* amino acid permeases, PCR was performed using degenerated primers based on conserved regions of fungal amino acid permeases (Fig. 1). Products that varied in length from about 350 to 500 bp were cloned, sequenced and analyzed by means of BLAST searches [30]. Of the 20 different PCR products, only one fragment of 406 bp showed significant homology (up to 37% amino acid identity) with genes encoding amino acid permeases of *S. cerevisiae* and other fungi. To clone the complete gene, a genomic library of *P. chrysogenum* in lambda ZAP II (Stratagene) was screened by plaque hybridization. One positive plaque was isolated, and by in vivo excision from the lambda ZAP II phage vector, pBluescript, containing an insert of 8 kb, was obtained. The insert was sequenced and found to contain an ORF that specifies a permease gene. The full-length sequence of *arlP* is shown in Fig. 2. The ORF of 1680 bp is interrupted by one intron of 59 bp. The encoded transport protein shows the highest homology with putative amino acid transporters of *Uromyces fabae* (41% amino acid identity), *Amanita muscaria* (39%), *Schizosaccharomyces pombe* (39%) and with characterized amino acid transporters of *S. cerevisiae* (Dip5, 37%; Put4, 36%) (BLAST search [30]).

3.2. Expression in S. cerevisiae M4054

The gene was expressed in S. cerevisiae strain M4054 $(\Delta gap 1)$ to analyze the substrate specificity of the permease. This strain lacks the general amino acid permease, Gap1. Gap1 is a nonspecific, high-capacity permease that transports most L- and D-amino acids. Due to the deletion, strain M4054 exhibits a low endogenous amino acid uptake activity [12,13]. The permease gene was amplified from a cDNA library, thereby obtaining the intron-less ORF, and inserted into the yEP352-based vector yCN, under control of the copper inducible CUP1 promoter, yielding yCA. This construct was introduced into strain M4054. For control experiments, the "empty vector" yCN was transformed into M4054. To determine if the gene was expressed upon addition of CuSO₄, mRNA levels were analyzed by RT-PCR. The presence of CuSO₄ resulted in a high transcription of the permease gene, whereas without CuSO₄, only a low level of transcript was found (see also Fig. 5).

 Fig. 1. Amino acid sequence alignment of conserved regions of fungal amino acid permeases. Sequences were searched from PubMed database at NCBI (http://www.ncbi.nlm.nih.gov/PubMed) using the program BLAST [30], except for the ArIP sequence (this work). Alignments were made with ClustalX [36]. Degenerated oligonucleotides were based on the indicated sequences (dashed lines). Put4 [37], Dip5 [15], Can1 [38], Lyp1 [39], Gap1 [13] and Tat2 [19] are S. cerevisiae permeases for proline, glutamate and aspartate, arginine, all amino acids and aromatic amino acids, respectively; Inda is a general amino acid permease from T. harzianum [40].

	1	
Put4 Dip5 Can1 Lyp1 Gap1 Tat2 Inda	GDSEPHKLKQGLQSRHVQLIALGGAIGTGILVGTSSTLHTCGPAGLFISY GKDENTRLRKDLKARHISMIAIGGSLGTGLLIGTGTALLTGGPVAMLIAY GEVQNAEVKRELKQRHIGMIALGGTIGTGLFIGLSTPLTNAGPVGALISY AHYEDKHVKRALKQRHIGMIALGGTIGTGLFVGISTPLSNAGPVGSLIAY IITAQTPLKHHLKNRHLQMIAIGGAIGTGLLVGSGTALRTGGPASLLIGW GSFDTSNLKRTLKPRHIIMIAIGGSIGTGLFVGSGKAIAEGGPLGVVIGW GMVELERPMKARHLHMIAIGGSIGTAGFFVGSGGALAKGGPGSLFVDF	$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$
ArlP	QLKEIHEFRQGLHQRHTQMIALAGIVGIGIFLSSGRATVEAGPLGAFLAY	45 - 94
	2	
Put4 Dip5 Can1 Lyp1 Gap1 Tat2 Inda	CVVVILNFSAVKVYGESEFWFASIKILCIVGLIILSFILFWGGGPNHDRL VVIVAINVVGVKFFGEFEFWLSSFKVMVMLGLILLFIIMLGGGPNHDRL VIITIMNLFPVKYYGEFEFWVASIKVIAIIGFIIYCFCMVCGAG-VTGPV VIITLMNFFPVKVYGEFEFWVASVKVIAIMGYLIYALIIVCGCS-HQGPI LAIVIINMFGVKGYGEAEFVFSFIKVITVVGFIILGIILNCGGGPT AVIVSINLFGVRGFGEAEFAFSTIKAITVCGFIILCVVLICGGGPD GVIIIINLFGALGYAEEEFWASCFKLAATVIFMIIAFVLVLGGGPKDGRY	246 - 295 219 - 268 220 - 268 242 - 290 222 - 267 213 - 258 196 - 245
ArlP	LIMISTAFVFVRVYGELEFGFSIIKIMLIIGVNIMALVITCGGAPNKSSI	187 - 236

-240TTTACGCGA AAGCCGCGCT GTAAAGGTGG CCCAGTTAT CAGCCACAA TA-180ATTGCAAATT GGCCTTTCC AAGTAGCTG GATTGTCAA CTATGCCTGG GG-120TGCGCAATAG GGGGGAGAAT TAGTACGTG AAAAAAAAG TACGCGCGAA AA-60TATTCCTTAG CTTAAGTCTT TTCCTCCCA AAGATTTGT GTCACGCCAC GG+ 1ATGCCACTT TATAACGAC CCATGTTGG AAGACCATG GCAGACAATCA-61TCCAACCCAA ATGTCCCAAT CTACGGAGAC GGGAAATGA CGGCACAATGA CGAGTCAA-61TCCAACCCAA ATGTCCCAAT CTACGGAGAC GGGAAATGA CGACAATGA CGACAATGA GG CGACCCTGCA-71N S P F I T T H V W K D D G S T T-61TCCAACCCAA ATGTCCCAAT CTACGGAGAC GGGAAATGA CGACAATGA CGACAATGA GG CGCACTGCA TATGCCCACA TGGCGCGC TCTTCGCAA CGACACATG GG L H121AATGCGGCCG GTCAGCTCA GGAGATCAC GAGTTCAGAC AGGCCCTCCA TN A A G Q L K E I H E F R Q G L H181ATTCAGATGA TTGCTCTCGC CGGTACGGTG GGGACGGGTA TTTCCTCAG TT1 Q M I A L A G T V G T G I F L S241GCCATTGTGG AAGCTGGACC ATTGGGCGC TTCTGGCAT ATACAATCAT TG301GCTGCCAGTG CGTCTTACGG AGTTGTGAG ATGGGAGCCC TGGTACCCT CA- Y A S V Y Y G V G E M G A L V P L361GTTATTCGCT ACGCAGAGAT CTTCTGTGGC TCGCGCGCAG TG- Y I R Y A E I F C D P A L A F A N421CAGATATACT CCTACTGTG TCCGATTCCG TCGGAATAC TCGGCTGCG TT- Q I Y S Y C V S I P S E I V A A A481GAATTCTGGA TCACGGCAC ACACCGAATA TGGACACCG TGGCGCGC TT- S T A F V F V R V Y G E L E F G F601AAAATTATGC TCATTATCGG CTGGCGAAACCAG AGGCCCC CT- S T A F V F V R V Y G E L E F G F601AAAATTATGC TCATTATCGG ATTGCTTTG TTGGAAGCCC CTACGGCCC CT- CCCACACCC TTGTATGG ATTGCGAA CAACGCAAC CTACG	
 -180 ATTECAANTT GGCCTTTCC ANGTAGCTGT GATTGTTCAA CTATGCCTGG GG -120 TGCGCAATAG GGGGGAGAAT TAGTTACGTG AAAAAAAAG TACGAGCAAT AA -60 TATTCCTTAG CTTAAGTCTT TTCCTTCCCA AAGATTTTG GTCACGCCAC GG + 1 ATGCCACCTT TTATAACGAC CCATGTTTGG AAGACGATG GCGACAATAG CG S N P N V P I Y G D G E M S D N S 121 AATGCGGCCG GTCAGCTCAA GGAGATCAC GAGTTCAGCAC AGGCCTCCA TC N A A G Q L K E I H E F R Q G L H 181 ATTCAGÀTGA TIGCTCTCGC CGGTACGGTG GGGACGGGTA TTTTCCTCAG TT I Q M I A L A G T V G T G I F L S 241 GCCATTGTGG AAGCTGGAC ATTGGCGCG TTCTGGCAT ATACAATCAT TG A I V E A G P L G A F L A Y T I I 301 GTTATCGCT ACGCAGAGT GTTCTGGCAG CGGGACCC TGGTACCCTC TA Y A S V V Y G V G E M G A L V P L 361 GTATTCGCT ACGCAGAGT CTCTGGGAG ATGGAGCCC TGCTCGCGCC TG V I R Y A E I F C D P A L A F A N 421 CAGATATACT CCTACTGTG GTCGATTCGC TCGGAGATG TCGCGCGCA TG V I R Y A E I F C D P A L A F A N 421 CAGATATACT CCTACTGTG GTCGATTCCG TCGGAGATG TCGCGCGCG TG 421 CAGATATACT CCTACTGTG TCGCGATTCG TCGGAGATG TCGCTGGGCC TG 421 CAGATATACT CCTACTGTG TCGGCATTCG TGGAGACTG TGCCTGGGCC TG 421 CAGATATACT CCTACTGTG TCGCGATTCG TGGAGACTG TGCTGGGCG TG 421 CAGATATACT CCTACTGTG TCGGCATAT TGGACTCACT GTCTGGGCT TC E F W I T V N N A I W I T V L G L 541 TCCACCGCCT TTGTCTTTGT TCGCGCTATA TGGAGAGCTC ACTCGGCCT TT S T A F V F V R V Y G E L E F G F 661 AAAATTAGC TCATTATCGG CGTAAACTCT TTTTGGCTC TCTGGAAAACCC TT K I M L I I G V N L M A L 661 GGAACATGTC CTTCTTGGGT TATGCTTGG TTTTGGCCT CCGGAAAACC CT Y L G V G G P L G R F L G F W K T 840 GCCCTCTCG CTTACTCTG TATCGGC TATGCCGAA CATCTCG CCGCATGA CATCTCG CCGCATGA CATCTCG CCGCATGA CATCTCG CCGCATGA CATCTTGG CTGACACCA TCTTACTGG CGCCAAAAAC CT Y L G V G G P L G R F A Y W K A P Y G P 780 TACTTGGGGG TGGCCGGCC CTTGGCCGAA CTTACTCTG CCGCAAAAAC CT Y L G V G G C P L G R F L G F W K T 84	CAATGTGA
-120 TGCGCANTAG GGGGGAGANT TAGTTACGTG AAAAAAAAG TACGAGCAAT AA - 60 TATTCCTTAG CTTAATAGTCT TTCCTTCCCA AAGATTTGT GTCACGCCAC GG + 1 ATGTCACCTT TTATAACGAC CCATGTTGG AAGGACGATG GCAGTACTAC CC M S P F I T T H V W K D D G S T T 61 TCCAACCCAA ATGTCCCAAT CTACGGAGAC GGCGAAATGA GCGACAATAG CG S N P N V P I Y G D G E M S D N S 121 AATGCGGCCG GTCAGCTCAA GGAGATTAC GAGTTCAGAC AGGCCTCGCA TC N A A G Q L K E I H E F R Q G L H 181 ATTCAGATGA TTGCTCTCGC CGGTACGGTG GGGACGGGTA TTTCCTCAG TT I Q M I A L A G T V G T G I F L S 241 GCCATTGTGG AAGCTGGACC ATTGGGCGCG TTCTTGGCAT ATACAATCAT TG A I V E A G P L G A F L A Y T I I 301 GTTGCGAGTG TCGTCTCAG AGTTGGTGAG CCCGCGCTAG ATACAATCAT TG V A S V V Y G V G V G E M G A L V P L 361 GTTATTCGCT ACGCAGAGAT CTTCTGTGAC CCCGCGCAG CGTTCGCCAA TG V I R Y A E I F C D P A L A F A N 421 CAGATATACT CCTACTGTGT GTCGATTCCG TCGGAGATTG TCGCTGCGCG TG Q I Y S Y C V S I P S E I V A A A 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTGTGGTCT TC E F W I T V N N A I W I T V L G L 541 TCCACCGCCCT TTGTCTTTGT TCGCGTGTAT GGAGAGCTG AGTTGGGCT TT S T A F V F V R V Y G E L E F G F 601 AAAATTATC CTACTGTGG GCTAAACCTG ATGGACCCC GTTGGGCTT TT S T A F V F V R V Y G E L E F G F 601 AAAATTATGC TCATTATCG CGTAAACCTG ATGGACCCC AGTTCGCCCAT TT S T A F V F V R V Y G E L E F G F 601 AGAAATTATGC CTACTATTGG ATTGCTTAC TGGAGACTCG TCTGGGGCTT TT C T C CCGCAATAGT CGTCCATTGG ATTGCTTAC TGGAAAGCAC CTTCGGCCC TT S T A F V F V R V Y G E L E F G F 601 AAAATTATGC CTACTATCGG CGTAAACCTG ATGGACGCCC AGTTCGCCC TT K I M L I I G V N L M A L 661 GGAACAATGT GTCCATTGG ATTGCTTAC TGGAAAGCAC CTTCGGAGCCC CT P N K S S I G F A Y W K A P Y G P 780 TACTTGGGGG TGGCCGGGCC CTTGGCGGGC TTTTGGGCT TCTGGAAAACC CT Y L G V G G P L G R F L G F W K T 840 GCGCTCTCC CTTACTCCGG TATGCTCGG CTGACATCTC TCCGCAAAAAC CT Y L G V G G I P L G R F L G F W K T 840 GCGCTCTCG CTTACTCCG TACCCCATGG CTTGTCTCG CCGCATAAAA F 90 CCCCGACATC TTACCCCAT GGCCGCTGAA CGTTACTCCG CCGCATAAAA C V I T I F M Y G C L U V S S A D K R 1020 TCGTCGGGGA AACCCCCCATGA CGTCCCCTTC CTTACTCCC	CCACGATC
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	TGACTTGT
+ 1 ATGTCACCTT TTATAACGAC CCATGTTTGG AAGGACGATG GCAGTACTAC CC M S P F I T T H V W K D D G S T T 61 TCCAACCCAA ATGTCCCAAT CTACGGAGAC GGCGAAATGA GCGACAATAG CG S N P N V P I Y G D G E M S D N S 121 AATGCGCCCG GCCAGCTCAA GGAGATTCAC GAGTTCAGAC AGGGCCTGCA TC N A A G Q L K E I H E F R Q G L H 181 ATTCAGATGA TGGCTCTCGC CGGTACGGTG GGGACGGGTA TTTTCCTCAG TT I Q M I A L A G T V G T G I F L S 241 GCCATTGTGG AAGCTGGACC ATTGGGCGCG TTCTTGGCAT ATACAATCAT TG A I V E A G P L G A F L A Y T I I 301 GTTGCGAGTG TCGTCTACGG AGTTGGTGAG ATGGGAGCCC TGGTACCTCT CA V A S V Y G V G V G C E M G A L V P L 361 GTTATTCGCT ACGCAGAGAT CTTCTGTGAC CCGCGCTAG CGTTCGCCAA TG V I R Y A E I F C D P A L A F A N 421 CAGATATACT CCTACTGTG GTCGATTCCG TCGGAAGTG TCGCTGCGCC TG Q I Y S Y C V S I P S E I V A A A 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTGTGGTCT TC <u>E F W I T V N N A I W I T V L G L</u> 541 TCCACCGCCT TGTCTTTGT TCGCGTGTAT GGAGAGCTCG AGTTCGGCTT TT <u>S T A F V F V R V Y G E L E F G F</u> 601 AAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTCG TGATTCACC CT <u>K I M L I G V N L M A L</u> 661 <u>GGAACATGTC TACTTTTGCT TGCGTGTCT TTTGGGCT TCTGGGAAAGCCC CTCGGGGCC CTTGGCCCATGG CTTTTGCCT TCTGGGG TGGCGGGCC CTTGGCCCT TTTGCGCT TCTGGAAAGCAC CTTACGG GGAAAGCAC CTCAGGGCCC CT <u>F N K S S I G F A Y W K A P Y G P</u> 780 TACTGGGAG TGGCGGGCC CCTTGGCGAG CTTTGCGCT TCCGGAAAACC CT <u>Y L G V G G P L G R F L G F W K T</u> 840 GCGCCTCTG CTTACTCTG TACGGGCAGAC TTTCCGCCT TCTGGAAAAC CT <u>Y L G V G G P L G R F L G F W K T</u> 840 GCGCCTCTG CTTACTCTG TACGAGAAC TTTACTCGC CTCCGGCGCCA GA <u>A L F A Y S G I E N F T L A A A A E</u> 900 CCTCGGAACAC TCTTTTTTTTTTTTTTGGCT TCTCGGCATTCG CCGCCTGG CTGCCCATGG AGCGCGAGAC CTTACTCGC CGCCTATC GCGCCTGAA CGTTCTGC CCGCCTGG CGGCCGAAC CTTACCCCG CTGCCCGG TGGCCGGGCG CCTTGCCCGG TGTCCGCCTG CTCCGCATGG CGGCCCATGC CTTCCGCCTG CTCCGCATGG CGGCCCATGC TGCCCATCC CT P N K S S I G F A Y W K A P Y G P 780 TACTTGGGAG TGGCCCCTTGC CTTGCCGGA CGTACTCTG CCGCATAAAG AC <u>V I T I F M I G L I V S S A D K R</u> 1020</u>	GGTTCTCA
61 TCCAACCCAA ATGTCCCAAT CTACGGAGAC GGCGAAATGA GCGACAATAG CG S N P N V P I Y G D G E M S D N S 121 AATGCGGCCG GTCAGCTCAA GGAGATCAC GAGTTCAGAC AGGGCCTGCA TC N A A G Q L K E I H E F R Q G L H 181 ATTCAGATGA TTGCTCTCGC CGGTACGGTG GGGACGGGTA TTTTCCTCAG TT I Q M I A L A G T V G T G I F L S 241 GCCATTGTGG AAGCTGGACC ATTGGGCGC TTCTTGGCAT ATACAATCAT TG A I V E A G P L G A F L A Y T I I 301 GTTCCGAGTG TCGTCTACG AGTTGGTCAG ATGGGAGCCC TGGTACCTCT CA V A S V V Y G V G E M G A L V P L 361 GTTATTCGCT ACGCAGAGAT CTTCTGTGAC CCGCGCCTAG CGTTCGCCA TG V I R Y A E I F C D P A L A F A N 421 CAGATATACT CCTACTGTG GTCGATTCG TCGGAGATTG TCGCTGGGCG TG Q I Y S Y C V S I P S E I V A A A 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTGGGCT TC <u>E F W</u> I T V N N A I W I T V L G L 541 TCCACCGCCT TTGTCTTTGT TCGCGTGTAT GGAGAGCTCG AGTTCGGCTT TT <u>S T A F V F V R V Y G E L E F G F</u> 601 AAAATTATGC TCATTATCGG CGTAAACCTT TTGGAAGGCCC CTACGGCCT TT <u>C CCGAATAAGT CGCTCATTGG ATTTGCTTAC TGGAAAGCCC CTACGGCCT TT S T A F V F V R V Y G E L E F G F 601 AAAATTATGC TCATTATCGG CGTAAACCTT TTTGATACCT CT CCGAATAAGT CTACTTGGT GTCGCGCGAGAC CCTACGGCCT TT S T A F V F V R V Y G E L E F G F 601 AAAATTATGC TCATTATCGG CGTAAACCTT TTTGAAAGC CTT K I M L I I G V N L M A L 661 GGAACATGTC TACTTGGG TATTGCTTAC TGGAAAGCC CCTACGGCCC TT P N K S S I G F A Y W K A P Y G P 780 TACTTGGGGG TGGGCGGGCC CCTTGGCCGG TTTTGGGGCT TCTGGAAAAC CT Y L G V G G P L G R F L G F W K T 840 GCGCTCTTCG CTTACTCGG TATCGAGAC TTTTGCTTCG CTGCGGCTAG GA A L F A Y S G I E N F T L A A A P 900 CCTCGAACAT CTTTATGAT CGGCCTAGA CGTTCTTCG CCGCAATACT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACA TCTTATGAT CGCCCTAGT GTTCGTCTG CCGAATACT GC P R H S I F M A A R R <u>I F V R I L</u> 960 GTGATCACA TCTTATGAT CGCCCCATGC TGTTTCGTCTG CCGAATAAAA K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCTTC GTTTGCGCAG CGGCCAAGC TG</u>	CCATCGTG PIV
121 AATGCGGCCG GTCAGCTCAA GGAGATTCAC GAGTTCAGAC AGGCCTGCA TC N A A A G Q L K E I H E F R Q G L H 181 ATTCAGATGA TTGCTCTCGC CGGTACGGTG GGGACGGGTA TTTTCCTCAG TT I Q M I A L A G T V G T G I F L S 241 GCCATTGTGG AAGCTGGACC ATTGGCGCG TTCTTGGCAT ATACAATCAT TG A I V E A G P L G A F L A Y T I I 301 GTTGCGAGTG TCGTCTACG AGTTGGTGAG ATGGGAGCCC TGGTACCTCT CA V A S V V Y G V G E M G A L V P L 361 GTTATTCGCT ACGCAGAGAT CTTCTGTGAC CCCGCGCTAG CGTTCGCCAA TG V I R Y A E I F C D P A L A F A N 421 CAGATATACT CCTACTGTG GTCGATTCCG TCGGAGATTG TCGCTGCGGC TG Q I Y S Y C V S I P S E I V A A A 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTGTGGTCT TC E F W I T V N N A I W I T V L G L 541 TCCACCGCCT TTGTCTTTGT TCGCGTGTAT GGAGAGCTCG AGTTCGGCTT TT S T A F V F V R V Y G E L E F G F 601 AAAATTATGC TCATTATCGG CGTAACCCT ATGGCATCGG TGATTCACCC CT K I M L I I G V N L M A L 661 GGAACATGTC TACTTTGCATTGC ATTGCTTAC TGGAAAGCAC CTTGCGCCC TT 720 CCGAATAAGT CGTCCATTGG ATTGCTTAC TGGAAAGCAC CTTGCGCCC TT N K S S I G F A Y W K A P Y G P 780 TACTGGGGG TGGCCGGCC CCTTGGCCGG TTTTGGCGT TCTGGAAAAC CT Y L G V G G P L G R F L G F W K T 840 GCCCTCTTG CTTATCCG TATCGAGAAC TTTACTCTC CGCGGCTGA A A L F A Y S G I E N F T L A A A E 900 CCTCGACACT CTTATCCGA TGGCCTGAT GTTTCGTCTG CCGATAAA A 1220 TCGTCGGGGA CAGCCTCGA GTCCCTTTC GTTGTCTTCG CCGATAAAAA A 1220 TCGTCGGGGA CAGCCTCGCA GTCCCTTC GTTTGTCTTCG CCGATAAAAA A 1220 TCGTCGGGGA CAGCCTCGCA GTCCCTTC GTTTGTCTCG CCGATAAAAA A 1220 TCGTCGGGGA CAGCCTCGCA GTCCCTTC GTTTGTCTCG CCGATAAAAA A 1220 TCGTCGGGGA CAGCCTCGCA GTCCCTTC GTTTGTCTCG CCGATAAAAA A 1220 TCGTCGGGA CAGCCTCGCA GTCCCTTC GTTTGTCGCAG CCGATAAAAA A 1220 TCGTCGGGGA CAGCCTCGCA GTCCCCTTC CTTTTGCAACG CCGCCATGC TG	AAATAGAG E I E
181 ATTCAGATGA TGCTCTCGC CGGTACGGTG GGGACGGGTA TTTTCCTCAG TT 1 Q M I A L A G T V G T G I F L S 241 GCCATTGTGG AAGCTGGACC ATTGGGCGCG TTTTGGCAATCAT TG I F L S 301 GTGCGAGTG TCGTCTCACGG AGTGGGGAGA ATTCTGGGAGAGAT CTGGGAGGAGCC TGGTACCTCT CA 301 GTTATTCGCT ACGCAGAGAAT CTTCTGTGAC CCCCGCGCCAA G V V V P L A A L V P L 361 GTTATTCGCT ACGCAGAGAAT CTCGTGGAG ATGCGAGAGATCG CCCGGGGGCC CTGGTCGGGGCC TG Q V V A <td< td=""><td>AGCGACAC Q R H</td></td<>	AGCGACAC Q R H
I = Q = M = I = A = C = C = C = C = C = C = C = C = C	CCCCCCC
241 GCCATTGGG AAGCTGGACC ATTGGGCGCG TTCTTGGCAT ATACAATCAT TG <u>A I V E</u> A G P <u>L G A F L A Y T I I</u> 301 GTTGCGAGTG TCGTCTACGG AGTTGGTGAG ATGGGAGCCC TGGTACCTCT CA <u>V A S V V Y</u> G V G E M G A L V P L 361 GTTATTCGCT ACGCAGAGAT CTTCTGTGAC CCCGCGCTAG CGTTCGCCAA TG V I R Y A E I F C D P A L <u>A F A N</u> 421 CAGATATACT CCTACTGTG GTCGATTCCG TCGGAGATG TCGCTGGCGC TG <u>Q I Y S Y C V S I P S E I V A A A</u> 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTTGGTCT TC <u>E F W I T V N N A I W I T V L G L</u> 541 TCCACCGCCT TTGTCTTGT TCGCGGTGTAT GGAGAGCTCG AGTTCGGCTT TT <u>S T A F V F V R V Y G E L E F G F</u> 601 AAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTGG TGTCGCCC CT <u>K I M L I I G V N L M A L</u> 661 GGAACATGTC TACTTTGCGT GCTAATCCTT TTTGCAVACG TAATTACCTG CG 720 CCGGAATAAGT CGTCCATTGG ATTTGCTTAC TGGAAAGCAC CCTACGGCCC CT <u>P N K S S I G F A Y W K A P Y G P</u> 780 TACTTGGGGG TGGCCGGCC CCTTGCCCGG TTTTTGGCCT TCGGAAAAC CT <u>Y L G V G G P L G R F L G F W K T</u> 840 GCGCTCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTG CGCGGCTGA GA <u>A L F A Y S G I E N F T L A A A</u> 900 CCTCGACACT CTATCCCCAT GGCCGCAGAC CTTACTCTG CCGCATTCT GC <u>P R H S I P M A A R R I F V R I L</u> 960 GTGATCACCA TCTTATAGAT CGGCCTGATT GTTCGTCTG CCGCATAGA AC <u>V I T I F M I G L I V S S A D K R</u> 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTC CTTATCGCAG CGCCCATGC TG	S G R
301 GTTECGAGTG TCGTCTACGG AGTTGGTGAG ATGGGAGCCC TGGTACCTCT CA V A S V V Y G V G E M G A L V P L 361 GTTATTCGCT ACGCAGAGAT CTTCTGTGAC CCCGCGCTAG CGTTCGCCAA TG V I R Y A E I F C D P A L <u>A F A N</u> 421 CAGATATACT CCTACTGTGT GTCGATTCCG TCGGAGATTG TCGCTGCGGC TG Q I Y S Y C V S I P S E I V A A A 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTGGTCT TC E F W I T V N N A I W I T V L G L 541 TCCACCGCCT TTGTCTTGT TCGCGTGTAT GGAGAGCTCG AGTTCGGCTT TT S T A F V F V R V Y G E L E F G F 601 AAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTGG TGATTCCCC CT K I M L I I G V N L M A L 661 GGAACATGTC TACTTTGGCT GCTAATCCTT TTTGATACG TAATTACCTG CG 720 CCGAATAAGT CGTCCATTGG ATTTGCTTAC TGGAAAGCAC CCTACGGCCC CT P N K S S I G F A Y W K A P Y G P 780 TACTTGGGGG TGGCGGGCC CCTTGGCCGG TTTTTGGGCT TCTGGAAAAC CT Y L G V G G P L G R <u>F L G F W K T</u> 840 GCGCTCTTCG CTTATCTGG TATCGAGAAAC TTTACTCTCG CTGCGGCTGA CA A L F A Y S G I E N F T L A A A E 900 CCTCGACACT CTATTCTCCAT GGCCGCTAGA CGTATTCTCG CCGCATTCT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACCA TCTTTATGAT CGGCCTGAT GTTCGTCTG CCGATAAAAG AC V I T I F M I G L I V S S A D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCTTTC GTTATCGCAG CGCCCATGC TG	GTGCTACT G A T
361 GTTATTCGCT ACGCAGAGAT CTTCTGTGAC CCCCCGCTAG CGTTCGCCCAA TG V I R Y A E I F C D P A L <u>A F A N</u> 421 CAGATATACT CCTACTGTGT GTCGATTCCG TCGGAGATTG TCGCTGCGCG TG <u>Q I Y S Y C V S I P S E I V A A A</u> 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTTGGTCT TC <u>E F W I T V N N A I W I T V L G L</u> 541 TCCACCGCCT TTGTCTTGT TCGCGTGTAT GGAGAGCTCC AGTTCGGCTT TT <u>S T A F V F V R V Y G E L E F G F</u> 601 AAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTG <mark>G TGATTTCACC CT <u>K I M L I I G V N L M A L</u> 661 <u>GGAACATGTC TACTTTGCCT GCTAATCCTT TTTGATACG</u> TAATTACCTG CG 720 CCGAATAAGT CGTCCATTGG ATTTGCTTAC TGGAAAGCAC CCTACGGCCC CT P N K S S I G F A Y W K A P Y G P 780 TACTTGGGGG TGGGCGGGCC CCTTGGCCGG TTTTTGGGCT TCTGGAAAAC CT Y L G V G G P L G R <u>F L G F W K T</u> 840 <u>GCGCCTTCG CTTACTCTGG TATCGAGAAC</u> TTTACTCTCG CGCGCGGCG AG A L F A Y S G I E N F T L A A A E 900 CCTCGACACT CTATCCCAT GGCCGCTAGA CGTATCTTCG CCGCATTCT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACA TCTTTAGAT CGGCCTGATT GTTTCGTCTG CCGATAAAAG AC <u>V I T I F M I G L I V S S A D K R</u> 1020 TCGTCGGGGGA CAGCCTCGCA GTCCCTTTC GTTATCGCAG CGCCCATCC TG</mark>	ATGGTGGT N G G
421 CAGATATACT CCTACTGTGT GTCGATTCG TCGGAGATTG TCGCTGCGGC TG Q I Y S Y C V S I P S E I V A A A 481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTTGGTCT TC E F W I T V N N A I W I T V L G L 541 TCCACCGCCT TTGTCTTTGT TCGCGTGTAT GGAGAGCTCG AGTTCGGCTT TT S T A F V F V R V Y G E L E F G F 601 AAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTGG TGATTTCACC CT K I M L I I G V N L M A L 661 GGAACATGTC TACTTTGCGT GCTAATCCTT TTTGATAGG TAATTACCTG CG V I T C 720 CCGAATAAGT CGTCCATTGG ATTGCTTAC TGGAAAGCAC CCTACGGCCC CT P N K S S I G F A Y W K A P Y G P 780 TACTTGGGGG TGGGCGGGCC CCTTGGCCGG TTTTTGGGCT TCTGGAAAAC CT Y L G V G G P L G R F L G F W K T 840 GCGCTCTTCG CTTACTCTG TATCGAGAAC TTTACTCTCG CGCGCGCA GA A L F A Y S G I E N F T L A A A 840 CCCCGAACAT CTTATCCCCAT GGCCGGCTAGA CGTACTCTCG TCCGCGCTGA GA A L F A Y S G I E N F T L A A A 900 CCTCGACACT CTATCCCCAT GGCCGGCTAGA CGTACTCTCG CCGCATTCT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACA TCTTATGAT CGGCCTGATT GTTCGTCTG CCGATAAAGA C V I T I F M I G L I V S S A D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCTTCC GTTACCCAG CGCCCATGC TG	GGTGGAAC G W N
481 GAATTCTGGA TCACGGTCAA CAACGCAATA TGGATCACTG TGCTTGGTCT TC <u>E F W</u> I T V N N <u>A I W I T V L G L</u> 541 TCCACCGCCT TTGTCTTTGT TCGCGTGTAT GGAGAGCTCG AGTTCGGCTT TT <u>S T A F V F V R V Y G E L E F G F</u> 601 AAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTG <mark>C TGATTTCACC CT <u>K I M L I I G V N L M A L</u> 661 <u>CCAACATGTC TACTTTGCGT GCTAATCCTT TTTCGATAG</u> 720 CCGAATAAGT CGTCCATTGG ATTTGCTTAC TGGAAAGCAC CCTACGGCCC CT <u>P N K S S I G F A Y W K A P Y G P</u> 780 TACTTGGGGG TGGGCGGGCC CCTTGGCCGG TTTTTGGGCT TCTGGAAAAC CT <u>Y L G V G G P L G R F L G F W K T</u> 840 <u>GCGCTCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTCG CTGCGGCGCG GAAAAC CT Y L G V G G P L G R F L G F W K T 840 <u>CCGCCCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTCG CTGCGGCGCG GAAAAC CT</u> <u>Y L G V G G P L G R F L G F W K T</u> 840 <u>CCGCCCTTCG CTTACTCCTG TACCCCATGG TATCGAGAAC CT</u> <u>Y L G V G G P L G R F L G F W K T</u> 840 <u>CCGCTCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTCG CTGCGGCTGA GA</u> <u>A L F A Y S G I E N F T L A A A E</u> 900 CCTCGACACT CTATCCCCAT GGCCGCTAGA CGTATCTTCG TCCGCATTCT GC <u>P R H S I P M A A R R I F V R I L</u> 960 <u>GTGATCACCA TCTTTAGAT CGGCCTGAT GTTTCGTCTG CCGATAAAAG AC</u> <u>V I T I F M I G L I V S S A D K R</u> 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTC GTTATCGCAG CGCCATGC TG</u></mark>	TTATCATT V I I
541TCCACCGCCTTTGTCTTTGTTCGCGTGTATGGAGAGCTCGAGTTCGGCTTTTSTAFVFVRVYGELEFGF601AAAAATTATGCTCATTATCGGCGTAAACCTGATGGCTCTGGTGATTTCACCCTKIMLIIGVNLMAL661GCAACATGTCTACTTTGCGTGCTAATCCTTTTTCATGGGTAATTACCTGCG720CCGAATAAGTCGTCCATTGGATTTGCTTACTGGAAAAGCACCCTACGGCCCCTPNKSSIGFAYWKAPYGP780TACTTGGGGGTGGGCGGGGCCCCTTGGCCGGTTTTTGGGCTTCTGGAAAACCTTYLGVGPLGRFLGFWKT840GCGCTCTTCGCTTACTCTGGTATCCCCATGGCCGCTAGACTTACTCTCGCTGCGGCGCGAAAAACCTTAAAE900CCTCGACACTCTATCCCCATGGCCGCTAGACGTATCTTCGTCCGCATTCTGCCGATAAAAGACVILLLPNKR960GTGATCACCATCTTTATGATCGGCCTGATGTTTCGTCTGCCGATAAAAGACVITLPNKRIL960GTGATCACCATCTTTATGATCGCCCCTTTC<	TTATGCTT L M L
$ \begin{array}{c ccccc} S & T & A & F & V & F & V & R & V & Y & G & E & L & E & F & G & F \\ \hline & & AAAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTGG TGATTTCACC CT \\ \hline & & I & M & L & I & I & G & V & N & L & M & A & L \\ \hline \hline & & & & & & & & & & & & & & & & &$	CTATTCTC
601AAAATTATGC TCATTATCGG CGTAAACCTG ATGGCTCTGG TGATTTCACC CT K I M L I I G V N L M A L661GGAACATGTC TACTTTGCGT GCTAATCCTT TTTGATAGG GCAATAGT CGTCCATTGG ATTTGCTTAC TGGAAAGCAC CCTACGGCCC CT P N K S S I G F A Y W K A P Y G P720CCGAATAAGT CGTCCATTGG ATTTGCTTAC TGGAAAGCAC CCTACGGCCC CT P N K S S I G F A Y W K A P Y G P780TACTTGGGGG TGGGCGGGGCC CCTTGGCCGG TTTTTGGGCT TCTGGAAAAC CT Y L G V G G P L G R F L G F W K T840GCGCTCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTCG CTGCGGCTGA GA A L F A Y S G I E N F T L A A A E900CCTCGACACT CTATCCCCAT GGCCGCTAGA CGTATCTTCG TCCGCATTCT GC P R H S I P M A A R R I F V R I L960GTGATCACCA TCTTTATGAT CGGCCTGATT GTTTCGTCTG CCGATAAAAG AC V I T I F M I G L I V S S A D K R1020TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCCATGC TG	SIL
661 $\begin{array}{c c c c c c c c c c c c c c c c c c c $	CTAGGTCG
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	GTGGAGCC
720 CCGAATAAGT CGTCCATTGG ATTTGCTTAC TGGAAAGCAC CCTACGGCCC CT P N K S S I G F A Y W K A P Y G P 780 TACTTGGGGG TGGGCGGGCC CCTTGGCCGG TTTTTGGGCT TCTGGAAAAC CT Y L G V G G P L G R <u>F L G F W K T</u> 840 GCGCTCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTCG CTGCGGCTGA GA <u>A L F A Y S G I E N F</u> T L A A A E 900 CCTCGACACT CTATCCCCAT GGCCGCTAGA CGTATCTTCG TCCGCATTCT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACCA TCTTTATGAT CGGCCTGAT GTTTCGTCTG CCGATAAAAG AC <u>V I T I F M I G L I V S S A</u> D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCCCATGC TGC	G G A
780 TACTTGGGGG TGGGCGGGGC CCTTGGCCGG TTTTTGGGCT TCTGGAAAAC CT Y L G V G G P L G R F L G F W K T 840 GCGCTCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTCG CTGCGGCTGA GA <u>A L F A Y S G I E N F</u> T L A A A E 900 CCTCGACACT CTATCCCCAT GGCCGCTAGA CGTATCTTCG TCCGCATTCT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACCA TCTTTATGAT CGGCCTGATT GTTTCGTCTG CCGATAAAAG AC <u>V I T I F M I G L I V S S A</u> D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTC GTTATCGCAG CGCCCATGC TG	TCGTCCAG F V Q
840 GCGCTCTTCG CTTACTCTGG TATCGAGAAC TTTACTCTCG CTGCGGCTGA GA <u>A L F A Y S G I E N F</u> T L A A A E 900 CCTCGACACT CTATCCCCAT GGCCGCTAGA CGTATCTTCG TCCGCATTCT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACCA TCTTTATGAT CGGCCTGATT GTTTCGTCTG CCGATAAAAG AC <u>V I T I F M I G L I V S S A</u> D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCGCCATGC TG	TCGACAAC F D N
900 CCTCGACACT CTATCCCCAT GGCCGCTAGA CGTATCTTCG TCCGCATTCT GC P R H S I P M A A R R <u>I F V R I L</u> 960 GTGATCACCA TCTTTATGAT CGGCCTGATT GTTTCGTCTG CCGATAAAAG AC <u>V I T I F M I G L I V S S A</u> D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCGCCATGC TG	CTCGAAAC T R N
900 CCTCGACACT CTATCCCCAT GGCCGCTAGA CGTATCTTCG TCCGCATTCT GC P R H S I P M A R R I F V R I L 960 GTGATCACCA TCTTTATGAT CGGCCTGATT GTTTCGTCTG CCGATAAAAG AC V I T I F M I G L I V S A D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCGCCATGC TG TGTATCGCAG CGCGCCATGC TG TG	
960 GTGATCACCA TCTTTATGAT CGGCCTGATT GTTTCGTCTG CCGATAAAAG AC V I T I F M I G L I V S S A D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCGCCATGC TG	TCTTTTAC
V I T I F M I G L I V S S A D K R 1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCGCCATGC TG	TCCTGGGC
1020 TCGTCGGGGA CAGCCTCGCA GTCCCCTTTC GTTATCGCAG CGCGCCATGC TG	LLG
SSG TAS QSPF VIA ARHA	GCATCAAG GIK

Fig. 2. Nucleotide sequence and deduced amino acid sequence of the genomic clone of *P. chrysogenum* ArlP. Sequences with gray background indicate the position of the conserved regions of AAP family permeases, on which degenerated primers were based. The 59 bp intron is boxed in black. The 12 predicted transmembrane segments are underlined.

1080	GTGG	TTC	CTT	CGA	TCA	TCAA	CG	CAG	TTG	TG	CTCA	CAI	CAG	CAT	GGT	CCTC	CG	GGA	ACA	GC
	v	v	Ρ	S	I	I	N	Α	v	v	\mathbf{L}	т	S	Α	W	S	S	G	Ν	S
1140	AATA	TCT	TGG	GTG	GCT	CCAG	AA	TCC	TAT	AC	GGTA	TGG	CAA	CCC	AGG	GTCA	тG	CCC	CCG	CC
	N	I	L	G	G	S	R	I	L	Y	G	М	А	т	Q	G	Η	А	Р	А
1200	GTCT	TCA	CAC	GTA	TCA	ACCG	СТ	TCG	GTA	TC	CCTI	GGG	STTG	CTG	TCG	CGCT	СТ	ATG	GCG	тС
	v	F	т	R	Ι	Ν	R	F	G	Ι	Р	W	V	Α	v	A	L	Y	G	V
1260	TTCA	TGT	CAC	TTG	GAT	ACAT	GA	GTC	TTT	CC	AGCI	CGG	CCA	GCA	CGG	TATT	CA	CAT	GGC	TA
	F	М	S	L	G	Y	М	S	L	S	S	S	Α	S	т	<u>v</u>	F		W	L
1220	a a a	100		mam		mama	~			-			поп		пот			— —		
1320	CAGA	ACC	TTG	TCT	CTA	TCTC	GA	CTC	TAG	TG	AACC	TGA	TGT	GCA	TCT	GTAT	TG	TCT	ACC	TT
	<u>Q</u>	N	Ц	V	5	T	5	т	Ц	V	N	Ц	м	C	Т	C	T	V	Y	Ц
1200	NCN T	mom	N.C.T.	NCC	<u>с</u> лт	CCNN	C N	200	ACC	C N	አመመር	אשר	COM	T C 7	ACC	NCCT	cc	CCT	ccc	~ 7
1300	AGAI	TCT	v	v	GAI	C	GA V	NGC.	AGG	C	T		D. D.	TCF	w W	F	T	D	w W	
	к	г	T	1	G	C	К	к	Q	G	1	D	к	Ľ	к	Б	ш	r	vv	А
1440	GCAC	Стт	mee	ACC	СЪТ	ልሮልሞ	22	ሮሞሞ	CCA	ሞሮ	тсст	יידיכיז	יידירים	ጥጥባ	סידעי	TTCT	CC	ጥጥጥ	ጥርጥ	ሞሮ
1110	A	P	F	0	P	v	Т	T T	W	т	S	т.	F	T	v	V	v	т.	F	F
		-	-	×	-	-	-	-		-			-	-	-	•	•		-	-
1500	ACCG	GAG	GGT	тта	CGA	CATT	ΤА	TGC	GCG	GТ	САТТ	GGA	GCA	СТС	CAA	CCTT	CG	TTT	CGA	ст
	т	G	G	F	т	Т	F	M	R	G	Н	W	S	т	A	т	F	v	S	т
1560	TATT	TCA	ACC	TGC	САТ	TCAT	CG	тса	TTG	TG	TATT	TTG	CCT	ACA	AGT	TCTG	GG	CAA	ААА	CG
	Y	F	N	\mathbf{L}	Р	F	I	v	I	v	Y	F	А	Y	K	F	W	А	к	т
1620	AAGA	TCA	TCC	CGC	TGG	CAGA	AA	TTC	CCA	TΤ	CGGC	CTI	TCA	TCO	SAAA	GCTG	GC	ACA	AAA	AT
	K	I	I	Ρ	L	А	Е	I	Ρ	I	R	Ρ	F	I	Е	s	W	н	ĸ	Ν
1680	CCCG	AGC	CAG	AGC	CCA	AGCC	GA	AGC	GAG	GC	CTCA	GCA	AAC	TTA	ATA	TTTT	GΤ	GGT	CAT	AA
	Р	Е	Р	Е	Р	К	Ρ	к	R	G	\mathbf{L}	s	к	L	N	I	\mathbf{L}	W	S	-
1740	TTTG	TAC	CGT	GTA	CAG	TTTC	AC	CAT	GAA	CC.	TGAC	GCI	GGA	CCA	CCC	GACA	CC	GGC	CTG	TC
T800	CGGT	TTT	AGG	GCT	'I'AG	TGGC	CG	GGT	GAT	GC	CGGG	CCC	GCT	TTC	ATG	CATC	GA	AAA	CCG	ΤG
1000				.	-	a. a=								.			-			
T800	ACAT	ΤGΤ	GCT	TGC	'I'GA	CACT	AA	ATA	AAA	AG	TTTC	TTT.	GCT	TCA	A.L.I.Y	TTATT	тG	AGA	A'I'A	CT

Fig. 2 (continued).



Fig. 3. Time-dependent uptake of L-tyrosine (A), L-phenylalanine (B), L-tryptophan (C), L-leucine (D), L-isoleucine (E) and L-alanine (F) in *S. cerevisiae* M4054 expressing ArlP (solid symbols) or with empty vector (open symbols). Uptake experiments were performed with cells grown to exponential phase in minimal proline medium. Expression of ArlP was induced with 2 mM $CuSO_4$ 5 h before harvesting.

3.3. Transport characteristics

To determine the substrate specificity of the permease gene expressed in S. cerevisiae M4054, uptake was analyzed of 15 different [¹⁴C]-radiolabeled L-amino acids that represent all groups of amino acids. Cells were grown to mid-exponential phase in minimal medium with proline as sole nitrogen source to prevent nitrogen catabolite inhibition known to elicit the degradation of some amino-acid permeases in *S. cerevisiae* [31]. The uptake of L-[¹⁴C]tyrosine in M4054 expressing the permease was at least 10 times higher than the background uptake level (Fig. 3). A significant increase in the uptake rate was also observed for L-[¹⁴C]phenylalanine, L-[¹⁴C]leucine and L-[¹⁴C]tryptophan. For the other tested amino acids (L-[¹⁴C]alanine, L-[¹⁴C]isoleucine, L-[¹⁴C]lysine, L-[¹⁴C]glutamine, L-[¹⁴C]asparagine, L-[¹⁴C]cysteine, L-[¹⁴C]glutamate, L-[¹⁴C]histidine, L-[¹⁴C]threonine, L-[¹⁴C]serine and L-[¹⁴C]proline), the uptake rates were indistinguishable for the control and permeasecontaining cells.

The apparent $K_{\rm m}$ of tyrosine uptake was about 360 μ M (Table 1). The substrate specificity of the permease was further tested by determining the inhibition of uptake of L-¹⁴C]tyrosine by a 10-fold excess of unlabeled amino acids (Fig. 4). Interestingly, some of the amino acids that inhibited tyrosine uptake by more than 60%, such as alanine, arginine and methionine, appeared not to be transported by the permease in direct transport assays. Leucine, tryptophan and phenylalanine were strong competitive inhibitors of L- $[^{14}C]$ tyrosine with K_i values ranging from 150 to 750 μ M (Table 1). None of the amino acids that contained a branched β -carbon (i.e. isoleucine, value and threonine) inhibited tyrosine uptake. The acidic amino acids exerted low inhibition, which might be related to their negative charge. Based on these data, the transporter that is now termed ArlP appears to be involved in the uptake of aromatic amino acids and leucine.

The activity of the permease upon Cu²⁺-induced overexpression of *arlP* in *S. cerevisiae* was tested not only after growing cells in minimal medium with proline as sole nitrogen source but also with NH_4^+ . Uptake of tyrosine in cells grown in the presence of $(NH_4)_2SO_4$ was much lower than in cells grown with proline as sole N-source. Since the transcript levels of the *arlP* gene were not

Table 1

Kinetic parameters for tyrosine uptake and competitive inhibition of tyrosine uptake by phenylalanine, tryptophan and leucine

Amino acid	<i>K</i> ₁ (μM			
Tyrosine	361	,		
Phenylalanine		355		
Tryptophan	743			
Leucine		169		

The values were obtained from linear regression analysis of data in Lineweaver-Burk plots for competitive inhibition.



Fig. 4. Competition of unlabeled amino acids for uptake of $[1^{4}C]$ tyrosine by *S. cerevisiae* strain M4054 expressing ArlP. $[1^{4}C]$ tyrosine was present at a final concentration of 25 μ M, and the indicated unlabeled amino acids were added in a 10-fold excess. The rate of uptake was determined after 3 min and results shown are the means of three independent experiments.

drastically different with both nitrogen sources, this suggests a negative regulation of the amino acid permease at a posttranscriptional level in response to ammonia (Fig. 5). Posttranscriptional regulation by rapid ammonia inactivation has been shown for the proline permease PrnB in *A. nidulans* [22], and for Gap1 in *S. cerevisiae* [32].



Fig. 5. Expression (upper panel) and tyrosine uptake activity (lower panel) of ArlP in *S. cerevisiae* strain M4054 transformed with yCA (containing *arlP*) or yCN (empty vector) grown in the presence of Cu^{2+} or without Cu^{2+} , in minimal proline medium (MP, left) or minimal ammonia medium (MA, right). *ArlP* mRNA levels were measured by a quantitative RT-PCR. Tyrosine uptake was measured at an extracellular concentration of 25 μ M in minimal medium without nitrogen source (MM). ArlP mediated uptake of tyrosine was measured in cells grown in MA (with Cu^{2+} ; closed circles; without Cu^{2+} ; open circles) and MP (Cu^{2+} ; closed circles).

3.4. Expression of arlP in P. chrysogenum

The two most common mechanisms involved in regulation at the transcriptional level of AAP family transporters are substrate induction [33] and nitrogen catabolite repression [3]. To test if the expression of *arlP* is regulated by either one of these mechanisms, *P. chrysogenum* was grown in minimal media with proline (poor nitrogen source) or ammoniaoxaloacetate (rich nitrogen source) as the sole nitrogen source, as well as in presence of millimolar concentrations of tyrosine, phenylalanine or leucine. Messenger RNA levels of the permease gene were measured by RT-PCR. At none of the tested conditions, a significant level of expression could be detected during logarithmic growth, while low expression levels were found at late stationary phase, irrespective of carbon or nitrogen source.

4. Discussion

In order to amplify gene fragments of amino acid permeases of *P. chrysogenum*, PCR was performed using a set of degenerated primers that were based on conserved regions of fungal amino acid permeases, forming the AAP family [2,3]. The PCR fragment was used to clone the entire gene from a genomic library. The cloned permease gene showed 30-39% amino acid sequence identity with members of this family. When the cDNA clone was expressed in *S. cerevisiae* strain M4054 (*Agap1*), which has a decreased amino acid uptake [18], a significantly increased rate of tyrosine, phenylalanine, leucine and tryptophan uptake was observed. For other amino acids, the increase in uptake was only marginal. Based on this observation, the permease was identified as an aromatic amino acid and leucine transport system, termed ArIP.

An interesting feature of ArlP is its selectivity towards the configuration of the β -carbon of amino acids. When the side chain is branched at this position, the amino acid does not compete with tyrosine for uptake. This suggests that the active site has an unusual specificity that has not been observed before for aromatic amino acid permeases. However, many of the competing amino acids are not transported by ArlP as the uptake of these amino acids is not elevated upon expression of *arlP* in *S. cerevisiae* M4054.

The physiological function of ArIP is not understood. The expression of the *arIP* gene was tested under conditions of nitrogen limitation/starvation and in the presence of millimolar concentrations of its substrates. No significant level of the *arIP* mRNA could be detected under these conditions. Some expression was observed at late stationary phase, irrespective of carbon or nitrogen source, but this was not further analyzed. Some fungal amino acid permeases have been shown to be expressed in differentiating cells. For instance, the expression of an amino acid permease of the rust fungus *U. fabae*, which is homologous to the amino acid permeases of the AAP family, is induced only in haustoria, i.e. specialized hyphae that are involved in plant infection [34]. The expression of *arlP* in *P. chrysogenum* may also be restricted to a morphologically or developmentally differentiated state.

We also noted that ArlP activity after overexpression was strongly decreased in *S. cerevisiae* when grown with ammonia as nitrogen source. Posttranscriptional regulation by rapid ammonia inactivation has been shown for the proline permease PrnB in *A. nidulans* [22], and in more detail, for Gap1 in *S. cerevisiae*, where NH₄⁺ induces ubiquitination, endocytosis and vacuolar degradation [31,35].

In summary, we have described the cloning and characterization of a permease of *P. chrysogenum*, specific for aromatic amino acids and leucine. *S. cerevisiae* proved to be a suitable host for overexpression and characterization.

Acknowledgements

S. cerevisiae strain M4054 was kindly provided by M.C. Kielland-Brandt (Department of Physiology, Carlsberg Laboratory, Copenhagen, Denmark). We thank Dr. M. van der Kamp and Dr. J. Kiel for valuable discussions. This work was support by the European Community (grant no QLK3-CT-1999-00729).

References

- [1] J. Horak, Biochim. Biophys. Acta 864 (1986) 223-256.
- [2] V. Sophianopoulou, G. Diallinas, FEMS Microbiol. Rev. 16 (1995) 53-75.
- [3] B. Andre, Yeast 11 (1995) 1575-1611.
- [4] A. Seaston, C. Inkson, A.A. Eddy, Biochem. J. 134 (1973) 1031-1043.
- [5] D.J. Hillenga, H.J. Versantvoort, A.J.M. Driessen, W.N. Konings, J. Bacteriol. 178 (1996) 3991–3995.
- [6] W.D. Stuart, K. Koo, S.J. Vollmer, Genome 30 (1988) 198-203.
- [7] K. Koo, W.D. Stuart, Genome 34 (1991) 644-651.
- [8] D. Dillon, D. Stadler, Genetics 138 (1994) 61-74.
- [9] G.B. Young, D.L. Jack, D.W. Smith, M.H. Saier Jr., Biochim. Biophys. Acta 1415 (1999) 306–322.
- [10] I.T. Paulsen, M.K. Sliwinski, B. Nelissen, A. Goffeau, M.H. Saier Jr., FEBS Lett. 430 (1998) 116–125.
- B. Nelissen, R. De Wachter, A. Goffeau, FEMS Microbiol. Rev. 21 (1997) 113-134.
- [12] B. Regenberg, L. During-Olsen, M.C. Kielland-Brandt, S. Holmberg, Curr. Genet. 36 (1999) 317–328.
- [13] J.C. Jauniaux, M. Grenson, Eur. J. Biochem. 190 (1990) 39-44.
- [14] J.C. Jauniaux, M. Vandenbol, S. Vissers, K. Broman, M. Grenson, Eur. J. Biochem. 164 (1987) 601–606.
- [15] B. Regenberg, S. Holmberg, L.D. Olsen, M.C. Kielland-Brandt, Curr. Genet. 33 (1998) 171–177.
- [16] M. Vandenbol, J.C. Jauniaux, M. Grenson, Mol. Gen. Genet. 222 (1990) 393–399.
- [17] M. De Boer, J.P. Bebelman, P.M. Goncalves, J. Maat, H. Van Heerikhuizen, R.J. Planta, Mol. Microbiol. 30 (1998) 603–613.
- [18] M. Grauslund, T. Didion, M.C. Kielland-Brandt, H.A. Andersen, Biochim. Biophys. Acta 1269 (1995) 275–280.
- [19] A. Schmidt, M.N. Hall, A. Koller, Mol. Cell Biol. 14 (1994) 6597–6606.

- [20] J. Tanaka, G.R. Fink, Gene 38 (1985) 205-214.
- [21] V. Sophianopoulou, C. Scazzocchio, Mol. Microbiol. 3 (1989) 705-714.
- [22] U.H. Tazebay, V. Sophianopoulou, B. Cubero, C. Scazzocchio, G. Diallinas, FEMS Microbiol. Lett. 132 (1995) 27–37.
- [23] P.V. Benko, T.C. Wood, I.H. Segel, Arch. Biochem. Biophys. 122 (1967) 783-804.
- [24] G.E. Skye, I.H. Segel, Arch. Biochem. Biophys. 138 (1970) 306-318.
- [25] P.V. Benko, T.C. Wood, I.H. Segel, Arch. Biochem. Biophys. 129 (1969) 498-508.
- [26] D.R. Hunter, I.H. Segel, Arch. Biochem. Biophys. 144 (1971) 168-183.
- [27] F. Lara, M.R. del Carmen, G. Vazquez, S. Sanchez, Biochem. Biophys. Res. Commun. 105 (1982) 172–178.
- [28] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989).
- [29] J.E. Hill, A.M. Myers, T.J. Koerner, A. Tzagoloff, Yeast 2 (1986) 163-167.

- [30] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. 215 (1990) 403–410.
- [31] C. Hein, J.Y. Springael, C. Volland, R. Haguenauer-Tsapis, B. Andre, Mol. Microbiol. 18 (1995) 77–87.
- [32] C. Hein, B. Andre, Mol. Microbiol. 24 (1997) 607-616.
- [33] T. Didion, B. Regenberg, M.U. Jorgensen, M.C. Kielland-Brandt, H.A. Andersen, Mol. Microbiol. 27 (1998) 643-650.
- [34] M. Hahn, U. Neef, C. Struck, M. Gottfert, K. Mendgen, Mol. Plant Microb. Interact. 10 (1997) 438–445.
- [35] J.Y. Springael, B. Andre, Mol. Biol. Cell 9 (1998) 1253-1263.
- [36] F. Jeanmougin, J.D. Thompson, M. Gouy, D.G. Higgins, T. Gibson, Trends Biochem. Sci. 23 (1998) 403–405.
- [37] M. Vandenbol, J.C. Jauniaux, M. Grenson, Gene 83 (1989) 153-159.
- [38] W. Hoffmann, J. Biol. Chem. 260 (1985) 11831-11837.
- [39] H. Sychrova, M.R. Chevallier, Yeast 9 (1993) 771-782.
- [40] V. Vasseur, M. Van Montagu, G.H. Goldman, Microbiology 141 (1995) 767–774.