cDNA cloning, bacterial expression, *in vitro* renaturation and affinity purification of the zinc endopeptidase astacin

Sabine REYDA^{*1}, Elard JACOB[†], Robert ZWILLING^{*} and Walter STÖCKER^{*2}

^{*}Zoologisches Institut, Universität Heidelberg, Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany, and [†]BASF AG, Abteilung Toxikologie, D-67056 Ludwigshafen, Germany

Astacin (EC 3.4.24.21) from the freshwater crayfish (*Astacus astacus*) is a prototype for the metzincin superfamily and for the astacin family of zinc peptidases, enzymes which are involved in hatching processes, embryonic patterning and tissue remodelling. Here we report on the cloning and overexpression in *Escherichia coli* of an astacin cDNA which was reverse-transcribed from crayfish midgut-gland mRNA. A cDNA construct based on this clone was generated which comprised the nucleotide sequence encoding mature astacin devoid of the signal and propeptide. This construct was cloned into the pET3a vector and used to transform *E. coli* BL21(DE3) cells. Recombinant astacin was purified from inclusion bodies and dissolved under reducing conditions. For folding, the protein was diluted into neutral buffer containing L-arginine, GSH and EDTA. Eventually, Zn²⁺ was added by dialysis and the fraction of active enzyme was

affinity-purified on immobilized Pro-Leu-Gly hydroxamate. As shown by superimposition of the corresponding three-dimensional structures, this inhibitor binds to a region of the active-site cleft that is conserved in most metzincins. Therefore this principle behind this affinity technique, originally introduced for fibroblast collagenase by Moore and Spilburg [Biochemistry (1986) 25, 5189–5195], is applicable throughout the metzincin superfamily of metalloproteases, despite their otherwise differing cleavage specificities. Recombinant astacin is active on gelatine zymograms and in a quenched fluorescence assay, yielding kinetic parameters comparable with those of wild-type astacin purified from crayfish stomach.

Key words: astacins, metalloproteases, metzincins, protein folding.

INTRODUCTION

Astacin (EC 3.4.24.21) from the digestive tract of the freshwater crayfish *Astacus astacus* L. is a small (23 kDa) zinc endopeptidase that is able to digest fibrillar collagen and other proteins [1,2]. The enzyme is the prototype of a family of metalloproteinases [3,4,5] (clan MA, family M12A, in the nomenclature of Rawlings and Barrett [6,7]; see also the MEROPS database: <http://www.bi.bbsrc.ac.uk/Merops/merops.htm>), whose members are involved in the cleavage of egg envelopes during hatching (for example, cf. [8]), in developmental pattern formation [9–11], in the assembly of the extracellular matrix [12–14] and in the processing of biologically active peptides and matrix proteins [5,15,16].

Generally, astacins are secreted proteinases, which require proteolytic removal of a signal peptide and a propeptide for optimal activity. The enzymic activity resides in a catalytic module of about 200 amino acid residues [17]. In addition, many astacins contain other modules attached C-terminally to the catalytic domains, which are thought to be responsible for targeting, for regulation of the catalytic activity or acting as membrane anchors [4,5].

The tertiary-structure analysis of crayfish astacin, consisting solely of the catalytic domain, revealed a typical fold, which is

valid also for the other members of the astacin family of zinc peptidases [4,18]. Furthermore, there is also a topological and sequential relationship with the corresponding catalytic modules of other protein families, like the adamalysins/reprolysins/‘ADAMs’ (A Disintegrin And Metalloproteinase-like) [19,20], the serralsins [21], the matrix metalloproteinases (MMPs) [22], the leishmanolysin-like [23] and snapalysin-like enzymes [24]. These protein families form a superfamily, the metzincins, which have in common topologically equivalent three-dimensional structures, a conserved zinc-binding motif (HEXXHXXG/NXXH/D) and an invariant methionine residue within a typical Met-turn beneath the active site [25]. There is also a more distant structural relation to the thermolysin-like enzymes (‘gluzincins’), which likewise contain a HEXXH zinc-binding motif [26,27].

Structural and functional studies have unravelled the essentials of substrate and inhibitor binding by astacin [1], which can be used to gain insight into the functions of other astacins by molecular modelling [4]. From the binding of phosphinic transition-state analogues to the crayfish enzyme a catalytic mechanism for this group of zinc proteinases has been derived that suggests the shift of a unique zinc-binding tyrosine residue for stabilization of the transition state during catalysis (‘tyrosine switch’) [28,29].

Abbreviations used: DTT, dithiothreitol; IPTG, isopropyl β -D-thiogalactoside; poly(A)⁺ RNA, polyadenylated RNA; RACE, rapid amplification of cDNA ends; LB, Luria–Bertani; ZMBH, Zentrum für Molekulare Biologie Heidelberg; His-tag, a six-histidine tail; MMPs, matrix metalloproteinases.

¹ Present address: Institut für Pharmazeutische Chemie, Universität Marburg, Marbacher Weg 6, D-35032 Marburg, Germany

² Present address and address for correspondence: Institut für Zoophysiology, Molekulare Physiologie, Westfälische Wilhelms-Universität Münster, Hindenburgplatz 55, D-48143 Münster, Germany (email wst@uni-muenster.de).

The complete cDNA sequence of astacin has been deposited in the EMBL, GenBank® and DDBJ Nucleotide Sequence Database under the accession number AJ242595.

In the crystal structure of the mature catalytic chain of astacin, the N-terminal alanine residue is buried and forms a salt bridge to Glu¹⁰³ close to the active site [18,26] suggesting an activation mechanism for the astacin-like proteinases [4]. Although, at least in the mouse meprin α chain, this salt bridge seems not to be indispensable for enzyme activation [30], it has been demonstrated for several members of the astacin family that proforms elongated at their N-termini are indeed inactive [31,32,33]. In the crayfish, an astacin proform has never been observed at the protein level, an observation that is explained by its mode of its biosynthesis. The enzyme is produced in the F cells of the midgut gland and stored extracellularly as active proteinase in the stomach, from where it can be isolated [34,35].

Further studies into the catalytic mechanism by site-directed mutagenesis or the investigation of the otherwise elusive proastacin require recombinant DNA technology. So far there have been no reports on the bacterial production and successful folding into the active form of members of the astacin family. Recently we published an analysis of the gene structure of astacin [36], a study which was based on a cDNA clone whose origin was not covered in this previous paper. Hence, in the present report we further describe the cloning of the astacin cDNA, we report on its overexpression in *E. coli*, and we develop a procedure for folding and activation of recombinant astacin from inclusion bodies. Furthermore, an affinity-chromatography procedure is adopted which allows for the purification of the correctly folded, proteolytically active, form of the enzyme.

EXPERIMENTAL

Chemicals, enzymes, clones

Enzymes and cloning vectors (pUC18, pUC19 pET3A) were purchased from Boehringer-Mannheim/Roche, New England Biolabs, Promega, Perkin-Elmer, Pharmacia/Amersham or ClonTech. All chemicals were of analytical grade and obtained from Merck Darmstadt, Sigma/Aldrich, Fluka, Boehringer Ingelheim/Serva, or AGS (Angewandte Gentechnologie Systeme, Heidelberg, Germany). Oligonucleotides were synthesized in the core facility of BASF AG or at the Zentrum für Molekulare Biologie Heidelberg (ZMBH).

Animals, stimulation of enzyme synthesis and isolation of polyadenylated [poly(A)⁺] RNA

Freshwater crayfish, *Astacus astacus* L. were obtained and kept as described; biosynthesis of astacin was stimulated by removal of about 200 μ l of gastric fluid from starved animals [35]. After 4 h the midgut gland was excised and shock-frozen in liquid nitrogen and ground in a mortar. Poly(A)⁺ RNA was prepared from deep-frozen hepatopancreas using a standard protocol (Stratagene).

Northern-blot analysis

Poly(A)⁺ RNA from *Astacus* hepatopancreas and marker DNA were denatured in formaldehyde and formamide according to published methods [37] and separated on a 1.2%-agarose/formaldehyde gel in Mops buffer, pH 10.5, at 75 V for 75 min. The gel was washed three times in diethyl pyrocarbonate-treated bidistilled water for 25 min and washed three times for 40 min in 0.15 M sodium citrate buffer, pH 7.0, containing 1.5 M NaCl. The nucleic acids were transferred by capillary suction to a Hybond N⁺ membrane (Amersham/Pharmacia), immobilized by UV irradiation and hybridized with a digoxigenin/dUTP-labelled astacin probe in 0.15 M sodium citrate buffer, pH 7.0, containing 0.3 M NaCl and 0.1% SDS under standard conditions. The

hybridized probe and the digoxigenin-labelled DNA marker were revealed as specified by the supplier (Boehringer-Mannheim/Roche).

cDNA synthesis

The cDNA was synthesized from poly(A)⁺ RNA by following the protocol of Capture Clone (Promega). cDNA clones encoding astacin were detected by PCR under standard conditions (35 cycles at 53 °C) using the degenerate primers primA 5'-ATGC-ARGARYTXGARGARAARAC-3' and primB 5'-TTRTTDA-TYTGRRTTXGCRTC-3', which correspond to amino-acid-residue positions 34-40/41 [MQELEEK(T)] and 186-191/192 [DANQIN(N)] respectively of the astacin protein sequence [17]. Astacin-specific cDNAs were ligated into the *Hinc*II site of pUC18 and transformed into *Escherichia coli* under standard conditions.

DNA sequencing

Sequence analysis of DNA was performed according to the Sanger dideoxy-chain-termination method [38] using standard M13 sequencing primers according to the 70770 protocol of USB.

Rapid amplification of cDNA ends (RACE)

5'-end completion of the astacin cDNA was achieved by a modification (Capture Clone, Promega, and ClonTech) of RACE/PCR [39]. First-strand synthesis was primed with a 5'-biotin-labelled oligo(dT)-*Xba*I-*Not*I primer, which allows purification of the cDNA product by streptavidin affinity binding and release of the bound product by *Not*I digestion. The biotin label further inhibits the ligation of a double-stranded PCR primer (UniAmp *Eco*RI Adaptor) to the 3'-end of the cDNA after completion of second-strand synthesis. This purified and adaptor-ligated cDNA was amplified with the ClonTech UniAmp primer primC 5'-CCTCTGAAGGTTCCAGAATCGATAG-3' and a nested set of astacin-specific internal antisense primers, primD 5'-TTGCATCAGTTTGCAGCATGTGAGC-3', corresponding to the astacin amino acid residues 180-187/188 [AHMTLQTDA(N)], and primE 5'-GTGTGCTCATGGTAG-AAGCC-3', corresponding to astacin residues 99-103/104 [GFYHE(H)] respectively.

The 3'-end of the astacin cDNA was completed using a nested RACE/PCR reaction [36] with primA (see above) and primF 5'-ACATGCATCCGCTTTGTGCCAG-3'; residue positions 41-47/48, TCIRFVP(R)] as internal sense primers. The oligonucleotides primG 5'-CGAGGGGGATGGTTCGACGGAAG-CGACCT₁₈-3', primH 5'-CGAGGGGGATGGTTCGACGG-3' and primI 5'-GATGGTTCGACGGAAGCGACC-3' were used as external nested antisense primers.

Both the 5' and the 3' fragments were cloned into the *Hinc*II site of pUC18 and cleaved at their internal unique *Sph*I sites. Further digestion with *Eco*RI and *Sal*I resulted in fragments that could be combined by ligation into the *Eco*RI and *Sal*I sites of pUC19, yielding the full-length astacin cDNA which was named 'pSPxx1-astacin'.

To complete the 5'-part of the sequence, a statistically primed (ZAP cDNA Synthesis Kit; Stratagene) hepatopancreas cDNA library in the UniZAP vector was analysed by PCR with internal antisense and one vector sense primer. All sequences were determined on both strands.

An internal PCR product synthesized from cDNA of *Astacus* hepatopancreas using the degenerate primers primA and primK

(5'-TAYTTXCCRTARTGCATDAT-3'; antisense; residues 146–151/152 of astacin) according to the DIG DNA Labeling Kit (Boehringer-Mannheim/Roche) was utilized as hybridization probe for clone detection and for Northern-blot analysis.

Protein expression

For the expression of the recombinant protein, pSPxx1-astacin was modified by PCR using the sense primer primL 5'-GAG-AGGTACCATATGGCAGCCATCTTGGGAGAC-3' containing a 5' *KpnI* and a *NdeI* site, and the antisense primer primM5'-GAGAGTCGACGGATCCTAGTGATGGTGATG-GTGTATGCCTTAGGCTACACTC-3', which would add a tail encoding five histidine residues and a *BamHI* and a *Sall* restriction site to the 3'-end.

The PCR product was cloned in the *KpnI/Sall* site of pUC18 and sequenced. An *NdeI* and a *BamHI* fragment of the astacin cDNA was transferred into the expression vector pET3a, thereby placing the target cDNA under the control of the T7 promoter [40]. pET3a astacin was transformed into *E. coli* strain BL21(DE3) which expresses T7 polymerase upon induction with IPTG (isopropyl β -D-thiogalactoside). Appropriate transformants were identified by restriction analysis and sequencing.

BL21(DE3)/pET3a-astacin cultures were grown in 30 ml of Luria-Bertani (LB) broth at 37 °C with 100 μ g/ml carbenicillin in a shaking incubator overnight. A 90 ml portion of LB broth/carbenicillin was added and bacteria were cultivated for 1 h. A 20 ml sample of this culture was grown in 200 ml of prewarmed LB broth/carbenicillin until the A_{600} increased to about 0.7–0.9. IPTG was added to a final concentration of 0.4 mM to induce production of T7 polymerase and thus the synthesis of the metalloproteinase precursor. Cells were harvested after 4 h by centrifugation (5000 g, 10 min).

Protein concentration

Protein concentration was determined by the Bradford method employing the protein detection kit of Bio-Rad (Bio-Rad product documentation) or by using the molar absorption coefficient of astacin (ϵ 42 800 M⁻¹·cm⁻¹) [1].

SDS/PAGE

Samples for SDS/PAGE (performed according to standard procedures) were centrifuged, and cell pellets were suspended in loading buffer and heated to 100 °C for 2 min. Debris was removed by centrifugation and the supernatants were analysed by SDS/PAGE on 14% gels stained with Coomassie Brilliant Blue R250 or G250.

Immunoblotting

SDS/polyacrylamide gels were blotted according to standard protocols in a semi-dry blotting apparatus (BioTec Fischer) on to nitrocellulose (Schleicher und Schuell) in Tris-buffered saline [TBS; 50 mM Tris/HCl buffer (pH 8.0)/0.15 M NaCl] and blocked (2 × 30 min) with 3% gelatin/TBS at 37 °C. For astacin detection, a polyclonal anti-(crayfish astacin) antiserum from rabbits [35] in 0.05% Tween 80/3% gelatin/TBS was used that had been pre-absorbed with complete extract of *E. coli* BL21(DE3)/pET3a cells lacking the astacin insert. The second antibody, alkaline phosphatase-conjugated anti-(rabbit IgG) from goat, was detected as described in [41].

Protein purification

Harvested cell paste from 1 litre of bacterial cell culture was suspended in 24 ml of 50 mM Tris/HCl buffer, pH 8.0, containing 1.0 M NaCl, frozen to –20 °C overnight, and centrifuged after thawing. The pellet was suspended and incubated (1 h, 37 °C) in 15 ml of 50 mM Tris/HCl buffer, pH 8.0, containing 1.7 mg/ml lysozyme and 1.7 mM EDTA. Subsequently cells and DNA were disrupted by sonication using a Branson sonifier. The mixture was extracted three times with 5% Triton X-100 in aliquots of 25 ml of 50 mM Tris/HCl buffer (pH 8.0)/1 mM EDTA, including sonication, incubation (30 min, 37 °C) and centrifugation (5000 g, 10 min) respectively. The pellet was suspended by sonication in 24 ml of 50 mM Tris/HCl buffer (pH 8.0)/1 M urea/1 mM EDTA, incubated (37 °C, 30 min) and centrifuged (5000 g, 10 min). The resulting pellet consisted predominantly of inclusion bodies, which were solubilized in 24 ml of 50 mM Tris/HCl buffer, pH 8.0, containing 8 M urea, 0.1 M β -mercaptoethanol and 1 mM EDTA for 1 h at 37 °C and cleared by centrifugation (15800 g, 30 min).

Folding

Inclusion bodies were dissolved in 8 M urea/0.1 M β -mercaptoethanol/1 mM EDTA/50 mM Tris/HCl, pH 8, at a protein concentration of 10 mg/ml. This solution was diluted into 50 mM Tris/HCl buffer, pH 8, containing 1 mM GSH, 0.1 mM GSSG and 0.8 M L-arginine, to a final concentration of 10 μ g/ml, incubated for 24 h at 4 °C and dialysed against 50 mM Tris/HCl buffer, pH 8, containing 0.1 μ M ZnSO₄, and eventually against zinc-free 50 mM Tris/HCl buffer, pH 8.

Affinity chromatography.

The peptide hydroxamate inhibitor Pro-Leu-Gly hydroxamate [28,42] was coupled to CH-Sepharose 4B (Pharmacia/Amersham) according to the manufacturer's recommendations. Affinity chromatography was performed as described in [42]. The filtered folding mixture was loaded on to the affinity column and equilibrated with 50 mM Tris/HCl buffer, pH 8. Elution was achieved by applying 0.1 M Tris base without adjustment of pH, after thoroughly washing with 0.5 M NaCl/Tris/HCl pH 8.

Structure comparison

X-ray crystal structures were superimposed and analysed on a Indigo II Workstation (Silicon Graphics) using the INSIGHT II software (Version 97.0; MSI, San Diego, CA, U.S.A.). Three-dimensional co-ordinates of leishmanolysin, snapalysin and serralysin were retrieved from the Brookhaven PDB database.

Protein sequencing

N-terminal Edman degradation was performed in the core facility of the ZMBH.

Enzyme-activity assays

Gelatine zymography

Gelatine zymography [43] was performed on SDS/PAGE gels containing 0.014% gelatin. Samples were prepared omitting reducing agents and incubated for 30 min at 20 °C. Electrophoresis was carried out at 25 mA and 4 °C and then the gels rinsed in 2.5% Triton X-100/50 mM Tris/HCl, pH 8, for 1 h, and washed with 50 mM Tris/HCl (pH 8)/0.1 μ M ZnSO₄ for 1 min. After 15 h incubation in 50 mM Tris/HCl, pH 8 at 25 °C, the gels were stained with Coomassie Brilliant Blue. Proteo-

lytically active bands are recognized as unstained spots on the blue gel.

Fluorimetric assay

Fluorimetric analysis of astacin activity was performed as described in [1] using the quenched fluorescent substrate dansyl-Pro-Lys-Arg-Ala-Pro-Trp-Val [44] on a Perkin-Elmer LS50 fluorimeter.

RESULTS

cDNA cloning

Poly(A)⁺ RNA was purified from the midgut glands of crayfish that had been stimulated 4 h before by removal of about 200 µl of gastric fluid [35]. Northern-blot analysis showed a major band corresponding to a size of about 1.0 kb, which was in the expected range of a complete mRNA encoding astacin, including 5' end elongations coding for a signal and a propeptide respectively (Figure 1).

The population of cDNAs synthesized from the poly(A)⁺ RNA was screened for astacin specific cDNAs by PCR, using the degenerate oligonucleotides primA and primB whose sequences correspond to peptide sequence positions 34–40 and 186–192 respectively, which flank the core of the astacin protein sequence (Figure 2). PCR produced a fragment of approx. 500 bp, which was cloned into pUC18 and sequenced.

The 5'-end of this cDNA was completed using a nested set of antisense primers (primE and primD) complementary to protein sequence positions 180–187/8 and 99–103/4 respectively, and an adaptor sense primer construct based on primC (see the Experimental section). The resulting PCR product of approx. 460 bp was cloned into the *Hinc*II site of pUC18. To complete the 5'-part of the sequence, a statistically primed hepatopancreas cDNA library in UniZAP was analysed for astacin sequences by PCR with internal antisense primers and external sense primers, followed by sequencing on both strands.

The complete 3'-end was obtained using primA and primF as internal sense primers and primG, primH, primI as external nested antisense primers (Figure 2). The resulting construct was likewise cloned into the *Hinc*II site of pUC18.

Figure 1 Northern-blot analysis of poly(A)⁺ RNA from *Astacus* hepatopancreas

Detection on a 1.2%-agarose/formaldehyde gel was achieved by hybridization with an internal astacin-specific digoxigenin-labelled probe subsequent to transfer on HyBond membrane. Lane M, size markers are indicated. Two different concentrations [3 (lane 3) and 1 (lane 1) µg of total poly(A)⁺ RNA] were loaded. In both lanes there is only a single mRNA species detectable. Size in kb ('kB') is shown on the right.

Figure 2 Nucleotide sequence of the cDNA and deduced amino acid sequence of crayfish astacin

In the nucleotide sequence, start and stop codons, and in the protein sequence, Ala1, the zinc-binding region and the Met-turn, are shown in **bold** respectively. The prepro part is *italics*. The predicted signal peptide is underlined and the pro part is underlined twice. The arrow (†) indicates the predicted cleavage spot of the signal peptidase. The mature protein sequence is boxed. Primers are indicated as horizontal arrows (→); the sequence of the antisense primer primM, which contains the His-tag, is indicated in parentheses.

The cDNA inserts of both clones were sequenced, and each was found to contain an internal unique *Sph*I site. Furthermore, the 3'-part has a unique *Sal*I site which resides in the 3' external antisense primer and the 5'-part has a unique *Eco*RI site which resides in the 5' external sense primer (UniAmp primer). Hence, both clones were digested with *Eco*RI, *Sph*I and *Sal*I and the derived 5' *Eco*RI/*Sph*I and 3'-*Sph*I/*Sal*I fragments were purified and ligated into *Eco*RI/*Sal*I cleaved pUC19. From the resulting clones the one that contained the complete astacin cDNA was selected and dubbed 'pSPxx1-astacin'.

The cDNA differs in three nucleotides from the genomic sequence [36]: two of these changes are silent and one leads to an exchange from Ala to Val in position -41, which corresponds to the signal peptide.

Expression of recombinant astacin

Among different strategies to obtain a strong and reproducible expression of the recombinant protein, the pET3a T7-based expression system was chosen, which allows protein expression upon induction of the T7-polymerase in *E. coli* BL21(DE3) cells by IPTG [40].

Figure 3 Purification of recombinant astacin

(a) Affinity chromatography. Inclusion bodies from pET3a-astacin-transformed *E. coli* BL21(DE3) cells were first dissolved in 8 M urea under reducing conditions, subsequently submitted to renaturation, dialysed against urea-free 50 mM Tris/HCl buffer, pH 8.0, and loaded on to a column with immobilized Pro-Leu-Gly hydroxamate ($K_i = 16 \mu\text{M}$). The recombinant enzyme binds at pH 8.0 (0.05 M Tris) and is desorbed at pH 9.8 (0.1 M Tris base). (b) Coomassie Blue-stained SDS/14%-PAGE gels (anode at the bottom) of purified recombinant astacin. 'crayfish 5μg', 5μg of wild-type astacin from crayfish stomach; 'E.coli 1.4μg', 1.4 μg of recombinant astacin after folding and affinity chromatography on Pro-Leu-Gly hydroxamate-Sepharose (i.e. material from fraction 22 in a); the recombinant enzyme (23.8 kDa) migrates slightly behind the wild-type enzyme (22.6 kDa) owing to the existence of a tail of seven additional residues, i.e. the five-residue His-tag and two additional residues (Arg and His; see Figure 2) encoded by the cDNA but absent in the wild-type enzyme; 'IB', 100 μg of inclusion-body protein dissolved in 8 M urea in the presence of β-mercaptoethanol. 'M', marker proteins (Dalton Mark VII-L; Sigma/Aldrich) from the top, 67, 45, 36, 29, 24, 20 and 13 kDa. (c) Immunodetection of recombinant astacin compared with the wild-type enzyme from crayfish stomach (loaded on to a SDS/10%-PAGE gel in concentrations of 4–30 μg/ml as indicated). The recombinant enzyme migrates as a single band slightly above the wild-type enzyme. The quantity of the recombinant enzyme in this particular diluted sample was estimated to be about 6–8 μg/ml; (d) Gelatin zymography of recombinant (fraction 22 in a) and wild-type astacin on a SDS/10%-PAGE gel. Proteolytic activity due to digestion of immobilized gelatin (1%) is shown by Coomassie Blue staining. Wild-type astacin from crayfish stomach migrates slightly further than recombinant His-tagged astacin harvested from *E. coli* inclusion bodies. M, proteolytically inactive marker proteins.

By PCR with sense primer primL and antisense primer primM, the astacin cDNA was amplified from the vector pSPxx1. primL introduces *KpnI* and *SalI* restriction sites and anneals to the start of the catalytic domain and adds an ATG start codon 5' to the GCA encoding Ala1 (Figure 2). primM adds five codons for histidine, a stop codon and *BamHI* and *SalI* sites immediately to the 3' end of the cDNA. Together with the C-terminal histidine residue this resulted in a six-histidine tail (His-tag) at the C-terminus that could be utilized for nickel chelate affinity chromatography (Qiagen product documentation). The resulting fragment, which codes for the catalytic domain of astacin, was then cloned into pUC18 by using its *KpnI* and *SalI* sites and designated 'pUC18-astacin'. Clones with the correct insert were identified by sequencing.

In a second step utilizing the *NdeI* and *BamHI* sites the cDNA of mature His-tag astacin was excised from pUC18-astacin and cloned in-frame into the *NdeI/BamHI*-treated IPTG-inducible T7 expression vector pET3a [36]. The resulting plasmid, pET3a-

astacin, was transformed into the *E. coli* strain BL21(DE3).

Cultures of pET3a-astacin-transformed BL21(DE3) cells were grown to a D_{660} of 0.7–0.9, induced with 0.4 mM IPTG, and harvested after 4 h. As a control, cultures were bred without induction. Under these conditions, protein expression of astacin as inclusion-body deposits was only observed in the IPTG-treated cultures. The inclusion-body fraction was purified by a series of extractions and centrifugations to remove membrane lipids, soluble proteins, nucleic acids and cell debris. This procedure yielded about 20 mg of inclusion-body protein per litre of culture broth, the proportion of astacin being about 10% (2 mg/l). Using 0.8 M IPTG the yield could be increased to about 50% (10 mg/l).

Folding

For folding, the inclusion bodies were dissolved in urea solution under reducing conditions at a protein concentration of

10 mg/ml. This solution was diluted into Tris buffer, pH 8.0, at 4 °C. Optimal folding was achieved in the presence of 0.8 M L-arginine, 1 mM GSH, 0.1 mM GSSG, 1.7 mM Pro-Leu-Gly hydroxamate and by adjusting the final protein concentration to 10 µg/ml.

Affinity chromatography

To separate folded astacin from unfolded forms and other proteins, an affinity column with Pro-Leu-Gly hydroxamate was employed. This peptide hydroxamate, originally designed [42] as an inhibitor of fibroblast collagenase, binds as a substrate analogue inhibitor of astacin ($K_i = 16 \mu\text{M}$) [28]. It only binds to correctly folded astacin and therefore is a highly efficient tool to distinguish and purify active from inactive enzyme (Figure 3a).

The yield of affinity-purified astacin from 20 mg of inclusion-body protein with a content of about 2 mg of astacin was about 0.4 mg in the preparation shown in Figure 3(a) (fractions 20–26) as judged from SDS/PAGE (Figure 3b), immunodetection and quantification (Figure 3c). This means that the recovery from folding and affinity chromatography is about 20% of the initial amount of astacin in the inclusion bodies.

The purified recombinant protein was submitted to Edman degradation resulting in the sequence Ala-Ala-Ile-Leu-Asp, which is identical with the N-terminal sequence of wild-type astacin from crayfish. Hence, as expected, the initial methionine has been cleaved off by the intrinsic methionine aminopeptidase of *E. coli* [45]. Finally, this was an additional proof that the recombinant astacin was not contaminated with an *E. coli* protein.

Enzymic activity of recombinant astacin

Astacin purified from *E. coli* inclusion bodies proved to be active on gelatin zymograms (Figure 3d) and comparable with the wild-type enzyme. The recombinant enzyme is a bit larger than the crayfish enzyme, which is expected, owing to the existence of the His-tag (Figure 3b and Figure 3d). However, the histidine residues and the C-terminal arginine residue (see Figure 2) can be removed by the use of carboxypeptidase B, yielding a recombinant enzyme of a size identical with that of the wild-type form (results not shown).

To compare kinetic data of recombinant astacin with that of crayfish astacin, the cleavage of dansyl-PKRAPWV (one-letter code) [44] was examined. Recombinant astacin cleaves this substrate with $k_{\text{cat}}/K_m = 0.85 \times 10^6 \text{ M}^{-1} \cdot \text{s}^{-1}$, which is close to the value of $1.0 \times 10^6 \text{ M}^{-1} \cdot \text{s}^{-1}$ published previously for the wild-type enzyme from crayfish stomach [44].

DISCUSSION

The cDNA the cloning and expression of which have been described here has previously been used for the detection and analysis of the astacin gene, a topic which has been the subject of a previous publication in which details of the cDNA structure have also been anticipated and will be not further discussed here [36]. However, the sequence of the complete astacin cDNA provided first evidence that the enzyme is indeed translated as a preproprotein, that requires N-terminal trimming to acquire its native mature structure. It had been stipulated that the first methionine residue in the cloned cDNA was the start codon, though no formal proof for this assumption was available at that time. Subsequent analysis of the genomic astacin DNA showed that this assumption was indeed correct [36].

From denaturation and renaturation experiments with crayfish astacin *in vitro*, it turned out that the enzyme could be refolded

after disruption of its tertiary structure in 8 M urea in the presence of dithiothreitol. This also demonstrated that the propeptide of astacin is not required for proper folding into the native enzymically active conformation. This issue seemed important for the strategy of expression for astacin, since there are cases, as in the subtilisin family of serine proteinases, where the propeptide is required as an initial nucleus for correct folding [46].

The yield of overexpressed protein in pET3a-transformed *E. coli* BL21(DE3) cells varied due to the amount of IPTG inducer between 2 mg and 10 mg of astacin protein/litre of culture. The purified inclusion bodies were dissolved in 8 M urea, reduced with β -mercaptoethanol and submitted to the folding procedure. The addition of EDTA in the folding buffer should help to prevent the interaction of the free thiol groups of the unfolded protein with heavy metals and also to protect the unfolded protein from the attack of metalloproteinases, including astacin itself. On the basis of its cleavage specificity [44] there are at least six potential cleavage sites in the polypeptide chain of astacin that are exposed in the unfolded state, but inaccessible in the native conformation. The dilution of the protein to about 10 µg/ml turned out to optimize the yield of correctly folded protein and minimize the formation of aggregates. The chosen ratio for GSH/GSSG of 1 mM:0.1 mM and the inclusion of 0.8 M L-arginine in the folding buffer considerably improved the yield of folded astacin [47]. For the formation of active recombinant astacin, arginine, EDTA and glutathione were removed by dialysis under concomitant addition of zinc ions that are essential for the function of astacin [1].

The inhibitor affinity chromatography with immobilized Pro-Leu-Gly hydroxamate [42], proved to be an efficient tool for the separation of correctly from incorrectly folded enzyme and of other inclusion-body proteins. The final yield after folding and affinity chromatography was about 20% of the total astacin from inclusion bodies. Thus this affinity technique turned out to be advantageous for other enzymes too, in addition to the MMPs. It may be of general use for the large variety of proteases of the metzincin superfamily. In the metzincins, substrates align antiparallel with the typical edge strand of the β -sheet in the N-subdomain under formation of a short ladder of hydrogen bonds between backbone amide and carbonyl groups [25]. As shown by X-ray crystallography for Pro-Leu-Gly hydroxamate complexes of both astacin [28] and neutrophil collagenase [48], the C-terminal hydroxamic acid moiety chelates the active-site Zn^{2+} , while the proline ring slots in a very similar manner into a cavity lined by the two aromatic side chains in the edge-strand sequences -Trp-Ser-Tyr- (astacin) and -His-Ala-Phe- (collagenase) (to view this the reader is referred to the multimedia adjunct located at <http://www.BiochemJ.org/bj/344/bj3440851add.htm>). Other metzincin prototypes likewise carry -Tyr-Ala-Phe- (serralyisin) [49], -Leu-Ala-Tyr- (tumour necrosis factor- α convertase, 'TACE') [50], -Trp-Ala-Thr- (leishmanolysin) [23] or -Tyr-Ala-Gln- (snalysin) [24] sequences in their edge strands. An overlay of the corresponding three-dimensional structures reveals that these edge-strand triplets are perfectly superimposable. The short residue in the middle (Ala or Ser) is always directed toward the protein body, whereas the flanking (most frequently aromatic) side chains protrude from the upper rim of the active-site cleft in the same orientation as in astacin and collagenase. Hence, in all subfamilies of the metzincin superfamily there seems to be a conserved pattern determining similarities in the specificity on the non-primed side of the substrate-binding cleft which may be utilized for purification and identification of these enzymes [according to the nomenclature of Schechter and Berger [52], the 'non-primed side' and the 'primed side' of the substrate-binding

cleft face the substrate side chains N-terminally (P_1 , P_2 , P_3 , etc.) and C-terminally (P_1' , P_2' , P_3' , etc.) of the cleavage point respectively}. On the other hand, the enzymes also offer sites which allow to discriminate between them, since they differ basically in their substrate-recognition structures on the C-terminal side of the point of cleavage (the 'primed subsites') [25].

Because of the good yield of astacin in the inclusion bodies, it was not necessary to make use of the attached His-tag for nickel chelate column purification of the recombinant enzyme. However, this option has to be exploited for enzymically inactive proforms or mutant forms of astacin which fail to bind Pro-Leu-Gly hydroxamate. It turned out that the His-tag can easily be removed by incubation with carboxypeptidase B if necessary. This presumably simulates the *in vivo* situation, where the two C-terminal residues of arginine and histidine are most likely removed by the intrinsic carboxypeptidase B activity of the crayfish stomach [51], because these residues are encoded on the mRNA (Figure 2), but absent from the protein purified from the crayfish [17]. This is also consistent with the three-dimensional structure of astacin, in which the C-terminal end following the disulphide bond between Cys⁴² and Cys¹⁹⁸ is freely accessible to the bulk solvent [18,26].

This work was supported by grants from the Deutsche Forschungsgemeinschaft to W.S. and R.Z. (Sto185/3-1) and to W.S. (Sto 185/3-2), respectively. We thank Monika Dörhöfer (Heidelberg), Susanne Petersen-Werner (Ludwigshafen) and Eva Grosse Berkhoff (Münster) for excellent technical assistance.

REFERENCES

- Stöcker, W. and Zwilling, R. (1995) *Methods Enzymol.* **248**, 305–325
- Stöcker, W. (1998) in *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D. and Woessner, Jr., F., eds.), pp. 1219–1222, Academic Press, New York
- Dumermuth, E., Sterchi, E. E., Jiang, W., Wolz, R. L., Bond, J. S., Flannery, A. V. and Beynon, R. J. (1991) *J. Biol. Chem.* **266**, 21381–21385
- Stöcker, W., Gomis-Rüth, F.-X., Bode, W. and Zwilling, R. (1993) *Eur. J. Biochem.* **214**, 215–231
- Bond, J. S. and Beynon, R. J. (1995) *Protein Sci.* **4**, 1247–1261
- Rawlings, N. D. and Barrett, A. J. (1993) *Biochem. J.* **290**, 205–218
- Barrett, A. J., Rawlings, N. D. and Woessner, J. F. (eds.) (1998) *Handbook of Proteolytic Enzymes*. Academic Press, New York
- Geier, G. and Zwilling, R. (1998) *Eur. J. Biochem.* **253**, 796–803
- Piccolo, S., Agius, E., Lu, B., Goodman, S., Dale, L. and DeRobertis, E. M. (1997) *Cell* **91**, 407–416
- Marques, G., Musacchio, M., Shimell, M. J., Wunnenberg-Stapleton, K., Cho, K. W. Y. and O'Connor, M. B. (1997) *Cell* **91**, 417–426
- Blader, P., Rastegar, S., Fischer, N. and Strähle, U. (1997) *Science* **278**, 1937–1940
- Kessler, E., Takahara, K., Biniaminov, L., Brusel, M. and Greenspan, D. S. (1996) *Science* **271**, 360–362
- Suzuki, N., Labosky, P. A., Furuta, Y., Hargett, L., Dunn, R., Fogo, A. B., Takahara, K., Peters, D. M. P., Greenspan, D. S. and Hogan, B. L. M. (1996) *Development* **122**, 3587–3595
- Li, S.-W., Sieron, A. L., Fertala, A., Hojima, Y., Arnold, W. V. and Prockop, D. J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5127–5130
- Walker, P. D., Kaushal, G. P. and Shah, S. V. (1998) *Kidney Int.* **53**, 1673–1680
- Lottaz, D., Maurer, C. A., Hahn, D., Büchler, M. W. and Sterchi, E. E. (1999) *Cancer Res.* **59**, 1127–1133
- Titani, K., Torff, H.-J., Hormel, S., Kumar, S., Walsh, K. A., Rödl, J., Neurath, H. and Zwilling, R. (1987) *Biochemistry* **26**, 222–226
- Häse, C. C. and Finkelstein, R. A. (1994) *Microbiol. Rev.* **57**, 823–837
- Birkedal-Hansen, H. (1995) *Curr. Opin. Cell Biol.* **7**, 728–735
- Schlagenhauf, E., Elges, R. and Metcalf, P. (1998) *Structure* **6**, 1035–1046
- Kurusu, G., Kinoshita, T., Sugimoto, A., Nagara, A., Kai, Y., Kasai, N. and Harada, S. (1997) *J. Biochem. (Tokyo)* **121**, 304–308
- Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Rüth, F.-X., McKay, D. B. and Bode, W. (1995) *Protein Sci.* **4**, 823–840
- Gomis-Rüth, F.-X., Stöcker, W., Huber, R., Zwilling, R. and Bode, W. (1993) *J. Mol. Biol.* **229**, 945–968
- Hooper, N. M. (1994) *FEBS Lett.* **354**, 1–6
- Grams, F., Dive, V., Yiotakis, A., Yiallouris, I., Vassiliou, S., Zwilling, R., Bode, W. and Stöcker, W. (1996) *Nat. Struct. Biol.* **3**, 671–675
- Yiallouris, I., Yiotakis, A., Vassiliou, S., Zwilling, R., Stöcker, W. and Dive, V. (1998) *Biochem. J.* **331**, 375–379
- Johnson, G. D. and Bond, J. S. (1997) *J. Biol. Chem.* **272**, 28126–28132
- Dumermuth, E., Eldering, J. A., Grünberg, J., Jiang, W. and Sterchi, E. E. (1993) *FEBS Lett.* **335**, 367–375
- Corbeil, D., Milhiet, P.-M., Simon, V., Ingram, J., Kenny, A. J., Boileau, G. and Crine, P. (1993) *FEBS Lett.* **335**, 361–366
- Marchand, P., Tang, J. and Bond, J. (1994) *J. Biol. Chem.* **269**, 15388–15393
- Zwilling, R. and Neurath, H. (1981) *Methods Enzymol.* **80**, 633–664
- Vogt, G., Stöcker, W., Storch, V. and Zwilling, R. (1989) *Histochemistry* **91**, 373–381
- Geier, G., Jacob, E., Stöcker, W. and Zwilling, R. (1997) *Arch. Biochem. Biophys.* **337**, 300–307
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998–9002
- Studier, W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89
- Blake, M. S. (1984) *Anal. Biochem.* **136**, 175–179
- Moore, W. M. and Spilburg, C. A. (1986) *Biochemistry* **25**, 5189–5195
- Heussen, C. and Dowdle, E. B. (1980) *Anal. Biochem.* **102**, 196–202
- Stöcker, W., Ng, M. and Auld, D. S. (1990) *Biochemistry* **29**, 10418–10425
- Ben-Bassat, A., Bauer, K., Chang, S. Y., Myambo, K., Boosman, A. and Chang, S. (1987) *J. Bacteriol.* **169**, 751–757
- Zhu, X., Ohta, Y., Jordan, F. and Inouye, M. (1989) *Nature (London)* **339**, 483–484
- Creighton, T. E. (ed.) (1984) *Protein Function, A Practical Approach*, IRL Press, Oxford
- Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S. and Tschesche, H. (1994) *EMBO J.* **13**, 1263–1269
- Baumann, U. (1994) *J. Mol. Biol.* **242**, 244–251
- Maskos, K., Fernandez-Catalan, C., Huber, R., Bourenkov, G. P., Bartunik, H., Ellestad, G. A., Reddy, P., Wolfson, M. F., Rauch, C. T., Castner, B. J. et al. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3408–3412
- Titani, K., Ericsson, L. H., Kumar, S., Jakob, F., Neurath, H. and Zwilling, R. (1984) *Biochemistry* **23**, 1245–1250
- Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162

Received 21 June 1999/8 September 1999; accepted 14 October 1999