YPREP 1630
DISK / 29/5/02
ARTICLE IN PRESS
No. of pages: 10
DTD 4.3.1/SPS

ACADEMIC PRESS

Protein Expression and Purification xxx (2002) xxx–xxx

Protein Expression & Purification

www.academicpress.com

Non-pathogenic trypanosomatid protozoa as a platform for protein research and production

Reinhard Breitling,[a] Susanne Klingner,[a] Nico Callewaert,[b] Regina Pietrucha,[a] Anett Geyer,[a] Gunter Ehrlich,[a] Regina Hartung,[a] Angelika Müller,[c] Roland Contreras,[b] Steven M. Beverley,[d] and Kirill Alexandrov[e,*]

[a] *JenaBioscience GmbH, Löbstedter Str. 78, 07749 Jena, Germany*

[b] *Department of Molecular Biomedical Research, Ghent University and Flanders Interuniversity Institute for Biotechnology, K.L.-Ledeganckstraat 35, B-9000 Ghent, Belgium*

[c] *Klinikum der FSU, Kinderklinik, Kochstr. 2, 07745 Jena, Germany*

[d] *Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, USA*

[e] *Department of Physical Biochemistry, Max Planck Institute for Molecular Physiology, Otto Hahn Strasse 11, 44227 Dortmund, Germany*

Received 25 October 2001, and in revised form 21 January 2002

Abstract

All currently existing eukaryotic protein expression systems are based on autonomous life forms. To exploit the potential practical benefits associated with parasitic organisms we have developed a new protein expression system based on *Leishmania tarentolae* (Trypanosomatidae), a protozoan parasite of lizards. To achieve strong transcription, the genes of interest were integrated into the small subunit ribosomal RNA gene. Expression levels obtained were up to 30 mg of recombinant protein per liter of suspension culture and increased linearly with the number of integrated gene copies. To assess the system's potential for production of post-translationally modified proteins, we have expressed human erythropoietin in *L. tarentolae*. The recombinant protein isolated from the culture supernatants was biologically active, natively processed at the N-terminus, and N-glycosylated. The N-glycosylation was exceptionally homogenous, with a mammalian-type biantennary oligosaccharide and the $Man_3GlcNAc_2$ core structure accounting for >90% of the glycans present. *L. tarentolae* is thus the first described biotechnologically useful unicellular eukaryotic organism producing biantennary fully galactosylated, core-α-1,6-fucosylated *N*-glycans.

Heterologous expression of proteins is an important alternative to protein isolation from the native sources. Protein expression systems include bacteria, yeast, fungi, insect cells, mammalian cells, transgenic animals, and more recently, transgenic plants. Despite the variety of systems to choose from, none is universally applicable. Notably, all expression systems developed so far are based on autonomous organisms. The biology of obligatory parasites has many salient features that so far have not been exploited for biotechnological purposes. Some of the best-characterized parasitic organisms are the pathogenic protozoans from the family Trypanosomatidae. Representatives of this group colonize a broad range of hosts ranging from plants to higher mammals [1]. Due to their public health significance, most of the research has concentrated on the human parasites *Trypanosoma* and *Leishmania* sp. To the unique features of these species belong RNA editing, arrangement of genes in tandem arrays, polycistronic transcription followed by trans-splicing, and regulation of gene expression almost exclusively at the post-transcriptional level [2,3]. Trypanosomatidae are rich in glycoproteins which can account for more than 10% of total protein [4]. Probably due to their parasitic lifestyle, the oligosaccharide structures of their glycoproteins are often similar to those of mammalians and in some cases can include complex-type oligosaccharides with α-linked galactose, fucose, and sialic acid residues [5].

* Corresponding author. Fax: +49-231-1331651.
E-mail address: kirill.alexandrov@mpi-dortmund.mpg.de (K. Alexandrov).

104-6592/02/$ - see front matter
PII S104-6592(82)00001-3

Methods of genetic manipulation of Trypanosomatidae are well established through the efforts of molecular parasitologists and several attempts were made to express heterologous genes in Trypanosomatidae. Although in all cases reported the heterologously expressed proteins were biologically active, the yields for recombinant mammalian proteins were measured in micrograms per liter of culture [6–8]. Moreover, these experiments were carried out in species that are pathogenic to humans. Apart from safety considerations, these species typically display low growth rates and require serum for their cultivation, thus limiting their utility as expression systems.

The availability of genetic engineering methods in combination with advanced post-translational modifications makes members of the Trypanosomatidae family potentially interesting for biotechnological applications [3,9]. Several Trypanosomatidae members display rapid growth rates and simple nutrient requirements. For our work, we selected *Leishmania tarentolae*, a parasite of the gecko *Tarentolae anularis* [10]. We have developed an expression vector, which can mediate high-level production of heterologous proteins both intracellularly and extracellularly. Furthermore, we demonstrated that secreted proteins were modified by N-linked oligosaccharides and that these modifications are more similar to the ones incorporated by mammals than has been found in any other lower eukaryotic host.

Experimental procedures

Plasmid construction and molecular biology

To construct pJBS expression plasmids a *Swa*I restriction site was introduced into pBS II KS(-) vector by PCR mutagenesis, replacing nucleotides 655–759. Subsequently nucleotides 143–2139 of the *L. tarentolae* small subunit rRNA gene (*ssu*) were PCR amplified from genomic DNA with *Swa*I containing primers and inserted into this *Swa*I site of modified pBS II KS(-). Upon deletion of an internal *Nco*I site, the central *ssu* part between nt 682 and 1068 was replaced by a PCR-generated fragment of the pIRsat vector containing the ORF for nourseothricin resistance genes fused to *Leishmania major* DNA of dihydrofolate reductase region providing the *sat*3′ processing site (Hubel and Beverley et al, manuscript in preparation). Next, we transferred into this plasmid a PCR generated 0.4 kb *Sal*I × *Bgl*II fragment containing the splice acceptor site of *L. tarentolae* calmodulin A gene (*camA*) and a 1.8 kb *Not*I × *Bam*HI trimmed PCR fragment spanning the entire *camBA* intergenic region. Finally, the multiple cloning site (MCS) *Bgl*II, *Nco*I, *Xho*I, *Pac*I, and *Not*I was inserted as *Bgl*II × *Not*I cassette generated from annealed oligonucleotides, yielding expression plasmid pJBS1.8sat. Plasmids with alternative selection markers were constructed by exchanging the *Bam*HI × *Spe*I *sat* cassette by a PCR derived *Bam*HI × *Spe*I cassette encompassing the ORF of the alternative marker gene such as phleomycin–bleomycin–zeocin resistance gene or hygromycin resistance gene [11].

The EGFP expression plasmids pJBSegfp1.8sat, pJBSegfp1.8ble, and pJBSegfp1.8hyg were generated by insertion of the *Nco*I × *Not*I *egfp* cassette from pEGFP-N1 (Clontech Labs) into MCS of the respective pJBS1.8 plasmids.

The EPO expression plasmids were constructed by inserting the cDNA of the human erythropoietin precursor gene derived from pcDNA3.1/GSepo (Invitrogen) upon correction of the K143E mutation present in the original construct with or without the C-terminal V5 hexa His tag. For optimization of secretion the native signal sequence of human EPO was replaced with the signal sequence derived from acid phosphatase of *Leishmania mexicana* [12].

Transfection and cultivation of L. tarentolae

Leishmania tarentolae parrot was cultivated as described [10]. Transfections were performed by electroporation of in vitro cultivated promastigotes as described [9] with selection of single colonies on solidified M199 media containing 50 μg/ml nourseothricin (WERNER BioAgents), 25 μg/ml zeocin (Invitrogen), or 25 μg/ml hygromycin B (Calbiochem) or by the limited dilution method in liquid cultures with twice the indicated drug concentrations. Approximately 5 μg of the expression plasmids was digested with *Swa*I for each electroporation. Integration of the expression constructs into the *ssu* locus was confirmed by genomic PCR.

For expression assays and protein purification, *L. tarentolae* strains were cultivated at 26 °C in suspension cultures in BHI medium (DIFCO) supplemented with 5 μg/ml hemin (Sigma).

EGFP expression analysis

Fluorometric assays of EGFP expression were performed with 0.2 ml samples of normalized cell lysates (10 mM Tris, pH 8.0, l% Triton) in 96-well format on FLUOstar Galaxy (BMG Labtechnologies) at E485 nanometer excitation and E520 nanometer emission. In Western blots of cell extracts EGFP was detected with rabbit anti-GFP antibodies (Clontech).

Purification and analysis of rhEPO

Concentrated culture supernatants were loaded onto a ConA column (Pharmacia) using the protocol supplied by the manufacturer. Epo containing fractions were pooled, applied onto a phenyl Sepharose column

(Pharmacia) equilibrated with 20 mM Hepes, 1 M $(NH_4)_2SO_4$, pH 7.4, and eluted by a gradient of 1–0 M $(NH_4)_2SO_4$ followed by a gradient of 0–1% CHAPS. Epo containing fractions were pooled and concentrated by ultrafiltration (Millipore). Fractions were analysed by SDS–PAGE/silver staining (BioRad) and Western blotting with rabbit anti-human Epo antibodies (GENZYME). Quantification of EPO was performed with an ELISA kit (MEDAC). N-terminal sequencing of purified rhEPO was performed by automated Edman degradation with a Model Procise 494 apparatus (Applied Biosystems) using the standard blot procedure.

The biological activity of Epo preparations was determined by the colony forming unit assay CFU-GEMM (granulocytes–erythrocytes–monocytes–macrophages, and megakaryocytes).

Five hundred CD34-positive selected peripheral blood stem cells from healthy donors were seeded into 1 ml culture medium composed of IMDM (Bio-Whittaker), 15% fetal calf serum, 1% bovine serum albumin (Terry Fox Labs), 0.2 mg iron-saturated transferrin (Sigma), 6 µl 200 mM *N*-acetyl-L-alanyl-L-glutamine (Biochrom), 10 µM β-mercaptoethanol, 100 U IL-3, 100 U IL-6 (Sigma), 10 U SCF (Amgen), 28 ng GM-CSF (TEBU), 0.3% agar (Bacto agar, DIFCO). The positive control was supplemented with 5 U erythropoietin (ERYPO, Fresenius AG). The 12-well culture plate was incubated at 37 °C for 14 days in an incubator at 5% CO_2. The colonies were identified by inspection under an inverted light microscope.

N-glycosylation analysis

N-Glycan analysis was performed essentially as described [13]. In brief, 1 ml of the rhEPO preparation and of the control sample was concentrated by vacuum evaporation to 100 µl and 250 µl of a solution containing 8 M urea, 360 mM Tris, pH 8.6, and 3.2 mM EDTA was added. Subsequently, the solutions were incubated for 1 h at 50 °C to denature the protein. Then the proteins were bound to the PVDF membrane at the bottom of a 96-well plate. The rest of the procedure for deglycosylation, labeling, and cleanup of the glycans was performed as described [13].

After sample cleanup the APTS-derivatized[1] glycans were dissolved in 5 µl water and 1 µl batches were subjected to the exoglycosidase mixtures described in Fig. 4. All exoglycosidases were obtained from Glyko (Novato, CA, USA) and were used at the concentrations suggested by the manufacturer. Enzyme digests were performed in a total volume of 5 µl in a 20 mM sodium acetate buffer, pH 5.0, by incubating samples for 16 h at 37 °C with subsequent evaporation. The A2F reference glycan (sialylated biantennary core-fucosylated) was obtained from Glyko.

Prior to loading on a 36 cm 10% polyacrylamide sequencing gel, the samples were reconstituted in 1 µl H_2O. To this, 0.5 µl deionized formamide and 0.5 µl of the Genescan 500 ROX-labelled oligonucleotide standard were added (Perkin–Elmer, Foster City, CA). On every gel, one lane was reserved for a malto-oligosaccharide size reference ladder [14]. Data were analysed using the Genescan 3.0 (Perkin–Elmer) package and electropherogram aligning and editing were done in CorelPhotopaint 9.0.

Results

Construction of pJBS1.8 expression vectors

We sought to construct a vector for high level protein expression in *L. tarentolae*. One of the advantages of Trypanosomatidae species is that the protein-coding genes can be expressed when transcribed by RNA polymerase I or a foreign polymerase since RNA processing is uncoupled from DNA transcription [15]. To generate large amounts of specific transcripts, we chose to integrate our expression cassette into the small ribosomal subunit RNA gene, which is strongly transcribed by RNA polymerase I [16]. Since gene regulation in Trypanosomatidae occurs predominantly post-transcriptionally through intergenic untranslated regions (UTRs), the choice of such UTRs is crucial for construction of an efficient vector [15]. To obtain UTRs that would mediate high level expression, we cloned a part of the *L. tarentolae* calmodulin cluster containing three tandemly arranged calmodulin genes (*camCBA*) separated by intergenic regions and mapped the splice acceptor sites (Klingner et al., unpublished). For construction of pJBS1.8 (Fig. 1) the 1.8 kb *camBA* intergenic region was used. The 5′ UTR of the multicloning site was a 0.4 kb fragment with the *camA* splice acceptor whereas its 3′ UTR consisted of the entire 1.8 kb *camBA* intergenic region fused to an exchangeable antibiotic resistance gene such as streptothricin acetyltransferase (*sat*), phleomycin–bleomycin binding protein (*ble*), or hygromycin phosphotransferase (*hyg*). The antibiotic resistance gene was fused to *L. major* DNA of the dihydrofolate reductase–thymidylate synthase locus (1.7k-IR) providing the 3′ processing site. The expression cassette was flanked by two fragments of the small subunit rRNA locus for double homologous recombination and was supplied with the ampicillin resistance gene linked to a bacterial origin of replication.

[1] *Abbreviations used:* APTS, 8-amino-l,3,6-pyrenetrisulfonic acid; BHI, brain heart infusion broth; cam A, calmodulin A gene; Con A, concanavalin A; DSA-FACE, DNA-sequencer-aided fluorophore-assisted carbohydrate electrophoresis; EGFP, enhanced green fluorescent protein; EPO, erythropoietin; sat, streptothricin acetyltransferase; ssu, small subunit RNA gene; UTR, untranslated region.

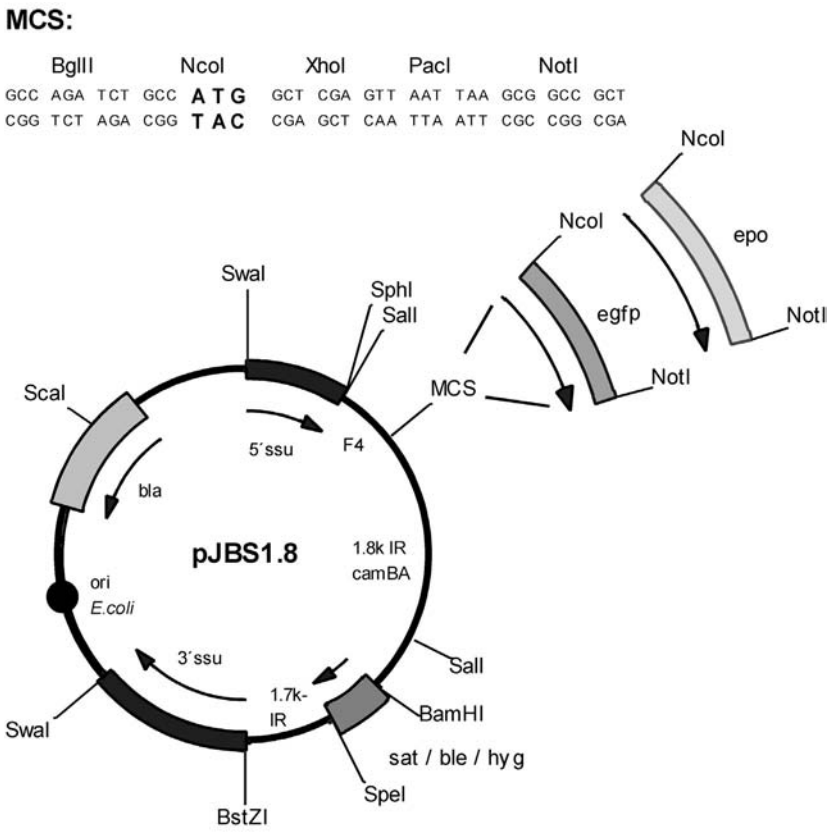

Fig. 1. Derivatives of pJBS*1.8* vector used in the present study. *Abbreviations: bla*, β-lactomase, 5′ *ssu*, 5′ portion of the small subunit *L. tarentolae* rRNA gene, *MCS*-multiple cloning site, *1.8k IR CamBA*, a portion of the intragenic region between calmodulin A and B genes of *L. tarentolae*. *sat/ble/hyg*, nourseothricine, bleomycin, or hygromycin resistance genes. *1.7k-IR*, a portion of 3′ untranslated region of the *L. major* dihydrofolate reductase gene, 3′ *ssu*, 3′ portion of the small subunit rRNA gene.

Expression of EGFP in L. tarentolae

For the initial test, we chose the EGFP protein, whose expression can be easily monitored by its fluorescence in living cells or in lysates. The pJBSegfp1.8sat construct was linearized and electroporated into *L. tarentolae* cells as described under experimental procedures. The transformants were selected by plating on nourseothricin-containing solid medium. The expression of EGFP was readily evident from the green fluorescence of the colonies (Fig. 2A). Nearly all of the drug resistant clones were fluorescent and PCR and Southern blotting analysis confirmed that a single copy of the *egfp* gene was integrated into the 18S RNA locus (not shown). To evaluate the amounts of expressed protein, we disrupted the cells by sonication and analysed the supernatants by SDS–PAGE and Western blotting. Fig. 2B shows a dominant band of 32 kDa that was immunoreactive with anti-GFP antibodies and absent in extracts of untransfected cells. We estimated that EGFP was expressed at a level of ca. 10 mg/L or 1% of total cellular protein.

Next, we wanted to learn whether this expression level could be further increased. We generated two additional pJBSegfp1.8 vectors where the *sat* resistance marker was exchanged for the *ble* or *hyg* gene. Using these vectors one or two additional copies of the *egfp* gene were integrated into the 18S rRNA locus. Upon analysis of the resulting clones, a direct correlation between the number of integrated copies and the EGFP fluorescence was observed (Fig. 2B). The increase appeared to be linear although the variability between the individual clones increased with the number of copies integrated. The clone with the highest expression produced recombinant EGFP at ca. 30 mg/L of suspension culture. We successfully expressed several other proteins in *L. tarentolae* using the same approach. These included both cytosolic proteins (T7 RNA polymerase, Cu/Zn superoxide dismutase, proto-oncogene Myc, etc.) and membrane-associated proteins such as the small GTPase Rab7. The obtained expression levels were of the same order of magnitude as EGFP. All proteins analysed were biologically active and in the case of Rab7 modified with isoprenoid moieties (not shown).

Expression of human erythropoietin in L. tarentolae

Post-translationally modified proteins remain to be the most problematic for recombinant production.

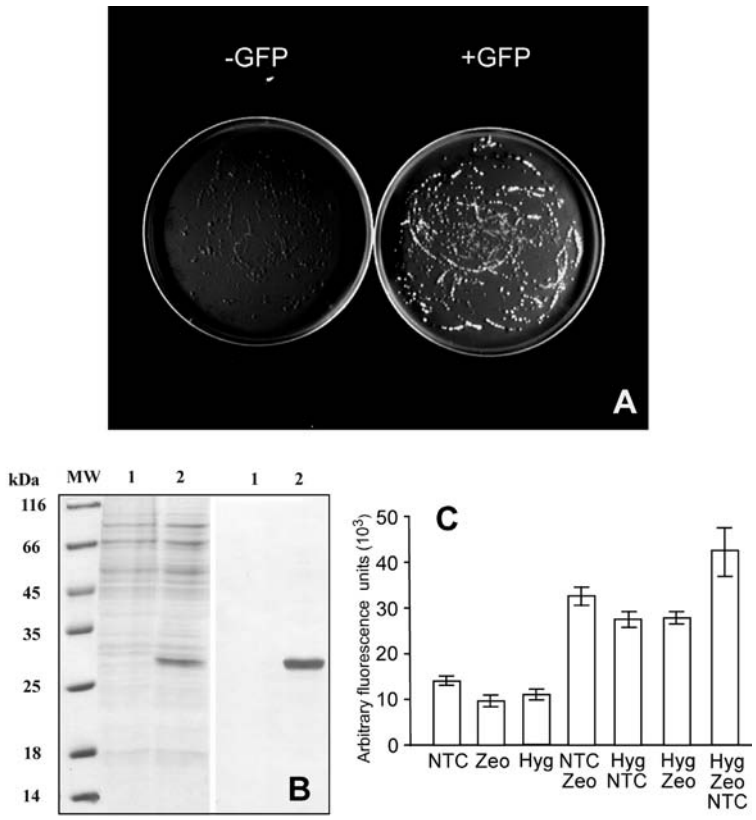

Fig. 2. (A) Colonies of *L. tarentolae* bearing a single copy of *egfp* gene integrated into the small subunit rRNA gene. Cells were plated on the nourseothricin containing solid medium after transformation and emerged colonies were photographed under UV light. (B) SDS–PAGE (left panel) and Western blot (right panel) analysis of 20 μg of total cytosolic fraction of wild-type *L. tarentolae* cells (lane 1) EGFP expressing *L. tarentolae* cells (lane 2). (C) Fluorescence of *L. tarentolae* lysates expressing 1 (bars 1–3), 2 (bars 4–6), or 3 (bar 7) copies of the *egfp* gene. The respective antibiotics used for construct integration are abbreviated as follows NTC, nourseothricin; Hyg, hygromycin B; Zeo, zeocin. Correlation between the number of integrated gene copies and the level of protein expression was confirmed by quantitative Western blotting (data not shown).

At the same time the activity of many biotechnologically important proteins keenly depends on glycosylation. To assess the ability of *L. tarentolae* to process proteins of mammalian origin, we attempted expression of human erythropoietin (hEPO) in this system. hEPO is a hemopoietic hormone specific to cells of erythroid lineage and is one of the most widely used protein therapeutics [17]. In the initial experiments the coding sequence of the human EPO, including its native signal peptide, was inserted into the pJBSsat vector together with C-terminal hexa-histidine and V5 tags. Transformation and clone selection were performed as described under experimental procedures. The presence of recombinant hEPO (rhEPO) in the culture supernatants and in the cells was determined by Western blotting. Culture supernatants of transformed cells contained a single protein of ca. 35 kDa that reacted with the anti-V5 tag antibody (Fig. 3A). A band of lower molecular weight could also be detected in the cells, suggesting that the protein underwent some modification during secretion (data not shown). In mammalian cells, EPO is processed by cleavage of the signal peptide and addition of N- and O-linked oligosaccharides, with the former modification being crucial for its biological activity [17]. To determine the presence and the type of the glycosylation, we treated the culture supernatant with *N*-glycosidase, *O*-glycosidase, and neuraminidase. As can be seen in Fig. 3A, only treatment with *N*-glycosidase resulted in an increase of the electrophoretic mobility indicating that the protein was N-glycosylated. To perform further analysis, hEPO was expressed without tags and purified from the culture supernatants by a combination of hydrophobic and affinity chromatography. The resulting protein preparation appeared as a single band in SDS–PAGE (Fig. 3B). This is in contrast to recombinant hEPO from CHO cells, which appears as a stack of bands due to the heterogeneity of the *N*-glycan chains attached. To analyse the structure of the *N*-glycans, we performed a series of analytical digestions with specific glycosidases followed by DSA-FACE analysis [13]. In accordance with the homogenous behaviour of the *Leishmania* produced rhEPO on SDS–PAGE, we found that the *N*-glycan structures present on this purified protein were exceptionally homogenous. Fig. 4A, panel 2, shows the electropherogram of a control sample, containing all of the components of the EPO-containing

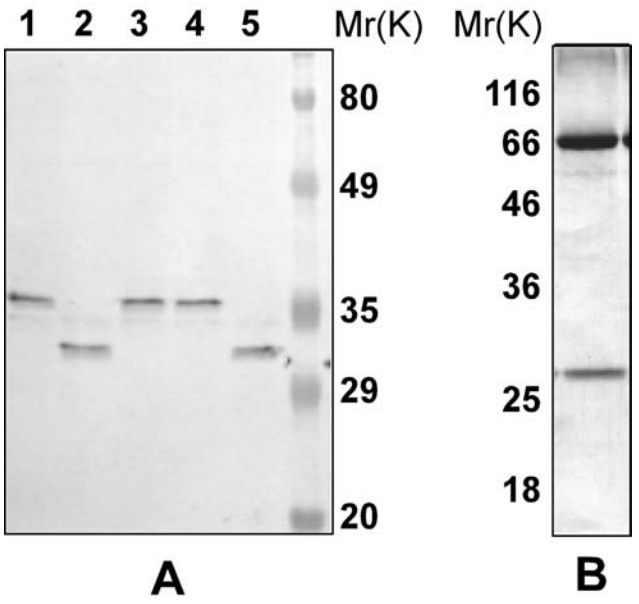

Fig. 3. (A) Western blot analysis of culture supernatants of *L. tarentolae* strain expressing human EPO bearing C-terminal V5 and 6×His tags. Lanes: 1, untreated; 2, treated with *N*-glycosidase; 3, treated with *O*-glycosidase; 4, treated with neuraminidase; 5, treated with mixture of all three glucosidases. (B) Silver stained SDS–PAGE with purified untagged rhEPO. Purification procedure is described in the text; ca. 1 μg of total protein was loaded. Band at 66 kDa corresponds to human serum albumin added in the last purification step as a carrier.

sample, excluding EPO itself. As can be seen in Fig. 4A, panel 3, only two glycan peaks were detected in a DSA-FACE analysis upon release of the glycans from the rhEPO with PNGase F. The peak with the higher electrophoretic mobility has an identical behaviour as the $Man_3GlcNAc_2$ core structure of *N*-glycans as determined by digestion with jack bean α-mannosidase (the glycan loses 1.9 glucose units in mannose residues; compare Fig. 4, panels 3 and 4). This glycan also co-migrates with the $Man_3GlcNAc_2$ product of combined digestion with sialidase, fucosidase, galactosidase, and β-*N*-acetylhexosaminidase of the sialylated form of the reference glycan with the structure shown in panel 4B (upper structure; compare Fig. 4, panels 3 and 12). The glycan with the lower electrophoretic mobility in panel 3 was sequenced with a panel of exoglycosidase mixtures. Fucosidase digestion induces a shift of 1.1 glucose units (compare panels 4 and 5), consistent with loss of one core-α-1,6-linked fucose residue (fucose residues in other linkages, p.e. to the outer branches of the glycan, result in smaller shifts in electrophoretic mobility, as determined in our laboratory). β-1,4-Galactosidase digestion results in a shift of 1.7 glucose units (compare panels 3 and 6), consistent with the loss of two galactose residues. Triple digestion with both enzymes and the jack bean β-*N*-acetylhexosaminidase converts the structure to $Man_3GlcNac_2$ (compare panels 3 and 7). To confirm the deduced non-sialylated, biantennary, bi-β-1,4-galactosylated, core-α-1,6-fucosylated structure, we performed the same digestion series on the desialylated reference glycan A2F, which has this structure. The reference glycan has exactly the same behaviour throughout exoglycosidase sequencing (Fig. 4A, compare panels 3–7 for the EPO preparation with panels 8–11 for the reference glycan). This is very strong evidence for the identity of the EPO glycan with this reference glycan, since the analytical technology used resolves most isomeric glycan structures. The probability that two isomers exactly co-migrate throughout five digestions is extremely small. To obtain exhaustive evidence for the linkage of the fucose residue, we finally performed a digestion with almond fucosidase, which only removes α-1,3/4-linked fucoses. Under conditions, which result in total removal of the α-1,3/4-linked side-branch fucose on human serum glycoprotein triantennary glycans (Callewaert et al., unpublished), the EPO-glycan-linked fucose residue was not removed (compare panels 3 and 8). Thus, we conclude that the *N*-glycans making up over 90% of the total *N*-glycan pool of rhEPO have near-native structures that have no known immunogenic potential in humans. To our knowledge, this is also the first report demonstrating the presence of the biantennary, fully galactosylated, core α-1,6-fucosylated *N*-glycan structure in a lower eukaryote.

To address the influence of the signal sequence on the levels of rhEPO production and integrity of glycosylation, the native signal sequence of human EPO was replaced with the signal sequence derived from secreted acid phosphatase of *L. mexicana*. The clones of *L. tarentolae* transfected with this construct secreted significantly larger amounts of rhEPO than the previously described clones. We purified the rhEPO from cultural supernatants according to the procedure described above and subjected it to *N*-glycan sequencing. The results of the *N*-glycan analysis of rhEPO with *Leishmania*-optimized signal sequence are shown in the bottom 3 panels of Fig. 4A. The same two glycans as described above are present, but here, the complex-type glycan has the largest abundance, accounting for 52% of the total *N*-glycan pool. The $Man_3GlcNac_2$ glycan represents 31%, and a number of smaller peaks (see panel 13) account for the remaining 17%. The peaks running before $Man_3GlcNac_2$ are present in the blanc sample (panel 2), and should not be taken into account for this reason. The structures of the smaller peaks will be the subject of further study, but triple digestion with fucosidase, galactosidase, and β-*N*-acetylhexosaminidase converts virtually all of them to $Man_3GlcNac_2$ (see panel 15), indicating that the large majority of them do not contain modifications, which are not found in mammalian-type *N*-glycans.

The biological activity of the EPO preparation was measured in a cell proliferation assay using human CD34-positive selected peripheral blood stem cells [18].

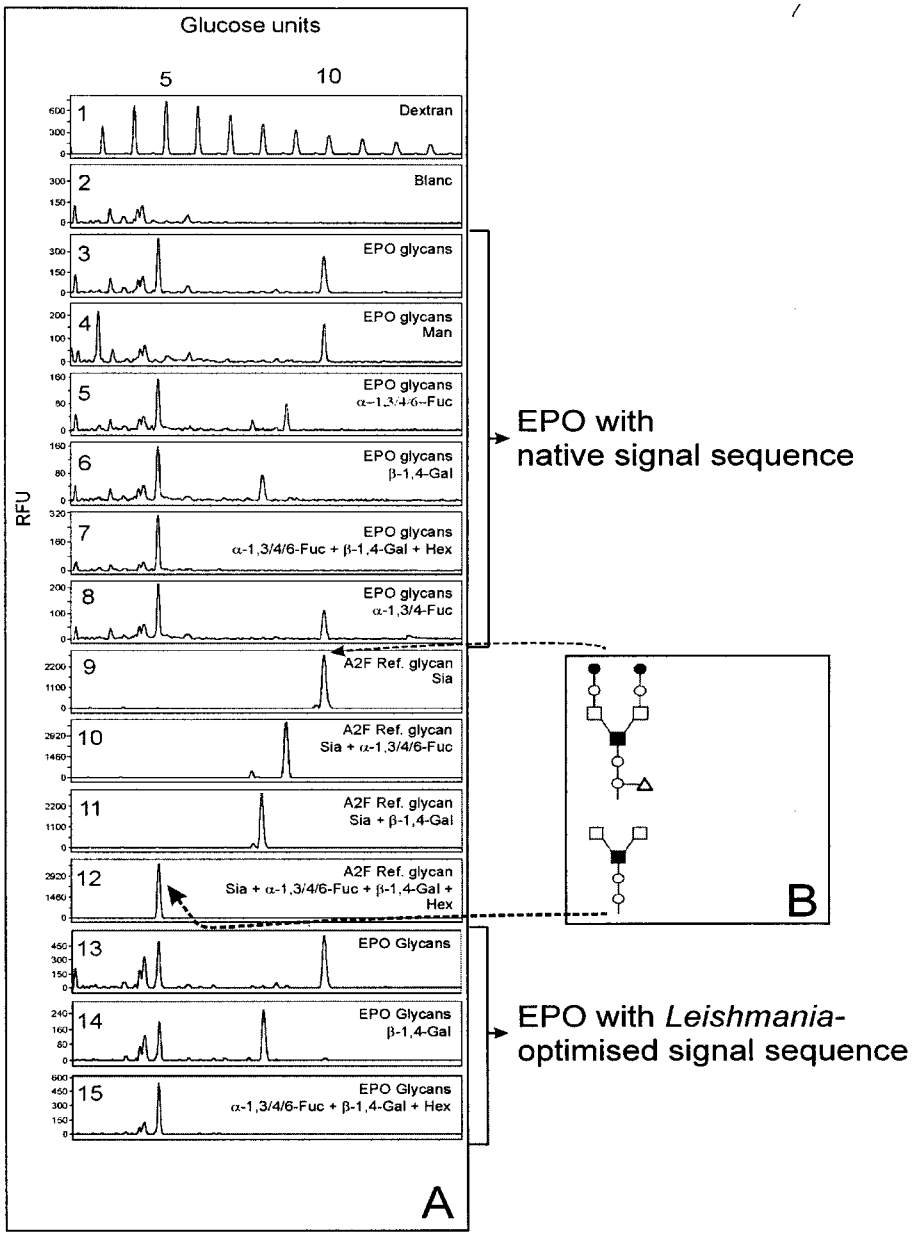

Fig. 4. (A) Analysis of PNGaseF-released glycans present on *L. tarentolae*-produced EPO. Panel 1 shows the electropherogram of an APTS-labelled malto-oligosaccharide mixture. The size of the glycans in this reference standard is indicated in glucose units. Panel 2: electropherogram of the buffer in which the final purified preparation of EPO was obtained. Panel 3: analysis of the EPO preparation. Enzyme name abbreviations: Man, jack bean α-mannosidase; α-1,3/4/6-Fuc, fucosidase from bovine kidney; β-1,4-Gal, galactosidase from *Diplococcus pneumoniae*; Hex, β-*N*-acetylhexosaminidase from jack bean; α-1,3/4-Fuc, fucosidase from almond tree. Each subsequent panel (4–7) in this figure shows the electropherogram obtained after digestion of the rhuEPO *N*-glycans shown in panel 3 with the exoglycosidase mixtures mentioned in the upper right corner of the panel. The results are detailed in the text. Panels 8–11 represent the results of the same exoglycosidase sequencing digestions of a reference glycan of the structure shown in (B), lower structure. The taller EPO glycan and the reference glycan behave identically throughout the digestions. Finally, panel 12 shows the analysis of a digest of the labelled EPO glycans with almond α-1,3/4-fucosidase. In (B), (•) denotes β-1,4-linked galactose, (○) β-linked GlcNAc, (□) α-linked mannose, (■) β-linked mannose, and (Δ) α-1,6-linked fucose.

Application of the *L. tarentolae* derived rhEPO resulted in a dose-dependent cell proliferation and differentiation of red cell precursors into mature end cells producing haemoglobin which allowed us to calculate the specific activity as 1.19×10^5 U/mg protein. This is very close to the activity of the commercial preparations of rhEPO [19]. Remarkably, the rhEPO expressed with the signal sequence derived from *L. mexicana* had a specific activity ca. 4×10^5 U/mg protein, indicating the direct connection between homogeneity of the glycan structures and biological activity of EPO. Since the secretion of rhEPO from *L. tarentolae* cells was mediated by the native

human signal sequence, we wanted to investigate how that sequence was processed. N-terminal sequencing of the rhEPO preparation from *L. tarentolae* by Edman degradation yielded the sequence APPRLI, thus indicating that the first 27 amino acids of the EPO precursor were removed. This shows that the mammalian signal peptide not only could mediate the ER-translocation and secretion in *Leishmania*, but was also natively cleaved off during this process. The N-terminus of rhEPO expressed with the signal peptide of *L. mexicana* was blocked and could not be sequenced by Edman degradation.

Discussion

We have used an in vitro cultured parasitic Trypanosomatidae species, *L. tarentolae*, as a host for protein expression. We were able to obtain a high level protein expression by integration of the expression cassettes into the small ribosomal subunit rRNA gene. The obtained expression levels were reproducibly high and could be further increased by integration of additional gene copies. Alternatively, higher levels of protein expression in Trypanosomatidae can be achieved by placing the gene of interest under the control of a foreign promoter in a host stably expressing a foreign RNA polymerase such as T7. This polymerase was shown to be more than 10 times more powerful than Pol I in Trypanosomatidae [11,20]. Moreover, both the rRNA and the T7 promoter can be controlled by a strong repressor such as TET [11,21]. Such a system developed for *L. tarentolae* would be invaluable for a multitude of applications including the expression of toxic or unstable proteins. Our initial experiments indicate that protein production can be driven by T7 RNA polymerase in *L. tarentolae* (Ehrlich et al., unpublished).

Molecular mimicry is an accoutrement of parasitic organisms that often manifests itself in tailoring of the surface proteins to resemble those of the host. Since Trypanosomatidae species naturally produce large amounts of glycoproteins we speculated that this feature could by advantageous for the production of heterologous glycosylated proteins. Indeed, the recombinant human EPO protein produced in *L. tarentolae* cells was efficiently secreted and the signal peptide was homogeneously cleaved off at the native position. The recombinant protein was biologically active, indicating proper folding. This was also in accordance with earlier observations that *Trypanosoma* and *Leishmania* species are able to produce active mammalian cytokines such as IL-2 and γ-IFN [6,8].

From the *N*-glycan structural analysis, we concluded that the N-glycosylation pathway of *L. tarentolae* was able to produce higher-eukaryote-like biantennary *N*-glycans. Only the sialylation of the glycans is missing. However, efficient in vitro sialylation procedures have been developed and are available. Moreover, it has been demonstrated that trans-sialidase of *Trypanosoma cruzi* can be expressed in active form in *Leishmania* cells [22]. This provides an alternative method for production of sialylated heterologous glycoproteins.

Another potentially beneficial feature of the *N*-glycan repertoire produced by *L. tarentolae* is its homogeneity. This is particularly important in cases where the recombinant protein needs to be crystallized for structural studies and for pharmaceutical protein production, where consistency of glycosylation is a subject of strict regulation. Moreover, it is striking that hEPO purified from the blood or purified from CHO cells has an extremely complex N-glycosylation profile, consisting of high-mannose, bi-, tri-, and tetraantennary structures, with or without one or two polylactosamine repeats [23,24]. Especially the sialylation of the *N*-glycans is important for the in vivo bio-activity [25]. In contrast to this, just two structures account for >90% of the *N*-glycans on the *L. tarentolae* produced protein. To some extent, this might be influenced by the use of a lectin in the purification procedure. However, Western blotting of the culture supernatants with specific antibodies indicated that the starting material is also exceptionally homogeneous. More importantly, the use of the ConA lectin during purification may have led to over-estimation of the abundance of the $Man_3GlcNAc_2$ glycan, relative to the complex-type structure, as ConA would preferentially bind the glycoforms substituted with mannose-terminated *N*-glycans. So, the share of the complex-type glycan observed in Fig. 4A, panel 3, is probably a conservative estimate and might have been higher in the total EPO pool.

The origin of the $Man_3GlcNAc_2$ core *N*-glycan remains so far unclear. It can be indicative of the presence of a strong β-*N*-acetylhexosaminidase activity in the secretory pathway, as has been observed with insect cells [26], or a weak *N*-acetylglucosaminyltransferase I activity or a combination of both. Alternatively, it could also be a result of the exoglycosidase activity in the culture supernatant. These issues are currently under investigation, but it is remarkable that the EPO produced with use of the *Leishmania*-optimized signal sequence favoured glycosylation of the complex type. Since this signal peptide also improved production levels, this observation would favour the hypothesis of a highly active β-*N*-acetylhexosaminidase in the secretory pathway. The higher substrate availability linked to increased glycoprotein production might outpace a limited β-*N*-acetylhexosaminidase activity in the secretory system, thus preserving a larger percentage of the complex-type glycans. Interestingly, the resulting EPO appears to have specific activity several folds higher than the commercial preparations.

Another striking feature is the absence of higher-branched *N*-glycans on the rhuEPO. The most

straightforward explanation for the lack of tri- and tetraantennary glycans would be a lack of *N*-acetylglucosaminyltransferase IV-activity in *Leishmania*. In the light of these observations, it is of interest that methods for gene disruption and integration are available for Trypanosomatidae, enabling the engineering of the glycosylation or other metabolic pathways.

Expression in *L. tarentolae* borrows from many well-developed and tested methods of molecular parasitology. It has been demonstrated that *L. tarentolae* can be cultivated on a cheap medium in a volume of 100 L reaching densities of 4×10^8 cells/ml with 6–8 h doubling time [10]. The cells can be disrupted by mild detergents or sonication, which offers a significant advantage over yeast systems. Cells can also be grown on fully defined media so that recombinant biopharmaceutical proteins are less likely to be contaminated with prions or pathogenic viruses which may accompany mammalian cell cultures dependent on components of animal origin. Furthermore, given the natural auxotrophy of *L. tarentolae* for methionine, such a system could also prove invaluable for the production of seleno-methionine-labelled proteins. This is particularly important considering the currently ongoing efforts for large-scale expression and crystallization of proteins of the human genome.

Acknowledgments

We thank Christina Clayton, Ullrich Göringer, Thomas Ilg, Roger S. Goody, and David Campbell for advice and stimulating discussions. Horst Pfützner is acknowledged for excellent technical assistance. Karl-Heinz Gührs is acknowledged for help with mass spectrometry. Vyacheslav Yurchenko and Andrei Iakovenko are gratefully acknowledged for intellectual contribution at the initial stages of the project. Martin Wiese and George Cross are acknowledged for sharing reagents. Cathrine Katzka, Vadim Sidorovitch, and Joachim Reinstein are acknowledged for comments on the manuscript. Authors are very thankful to Mathias and Manfred Gruen for various organizational assistance. This work was in part supported by a grant from the Thuringian Minister of Sciences, Research, and Culture to Jena Bioscience GmbH. Nico Callewaert is a research assistant of the Fund for Scientific Research Flanders (FWO). N.C. and R.C. are supported by a GOA project of Ghent University (Project No. 12052299) and by the FWO (Project No. G.0050.97).

References

[1] J.R. Stevens, W. Gibson, The molecular evolution of trypanosomes, Parasitol. Today 15 (1999) 432–437.

[2] S.M. Teixeira, Control of gene expression in Trypanosomatidae, Braz. J. Med. Biol. Res. 31 (1998) 1503–1516.

[3] C.E. Clayton, Genetic manipulation of kinetoplastida, Parasitol. Today 15 (1999) 372–378.

[4] M.A. Ferguson, The surface glycoconjugates of trypanosomatid parasites, Philos. Trans. Roy. Soc. London B 352 (1997) 1295–1302.

[5] A.J. Parodi, N-glycosylation in trypanosomatid protozoa, Glycobiology 3 (1993) 193–199.

[6] A.C. La Flamme, F.S. Buckner, J. Swindle, J. Ajioka, W.C. Van Voorhis, Expression of mammalian cytokines by *Trypanosoma cruzi* indicates unique signal sequence requirements and processing, Mol. Biochem. Parasitol. 75 (1995) 25–31.

[7] W.W. Zhang, H. Charest, G. Matlashewski, The expression of biologically active human p53 in *Leishmania* cells: a novel eukaryotic system to produce recombinant proteins, Nucleic Acids Res. 23 (1995) 4073–4080.

[8] J.F. Tobin, S.L Reiner, F. Hatam, S. Zheng, C.L Leptak, D.F. Wirth, R.M. Locksley, Transfected *Leishmania* expressing biologically active IFN-gamma, J. Immunol. 150 (1993) 5059–5069.

[9] S.M. Beverley, C.E. Clayton, Transfection of *Leishmania* and *Trypanosoma brucei* by electroporation, Methods Mol. Biol. 21 (1993) 333–348.

[10] L. Simpson, G.C. Frech, D.A. Maslov, RNA editing in trypanosomatid mitochondria, Methods Enzymol. 264 (1996) 99–121.

[11] E. Wirtz, S. Leal, C. Ochatt, G.A. Cross, A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*, Mol. Biochem. Parasitol. 99 (1999) 89–101.

[12] M. Wiese, T. Ilg, F. Lottspeich, P. Overath, Ser/Thr-rich repetitive motifs as targets for phosphoglycan modifications in *Leishmania mexicana* secreted acid phosphatase, EMBO J. 14 (1995) 1067–1074.

[13] N. Callewaert, S. Geysens, F. Molemans, R. Contreras, Ultrasensitive profiling and sequencing of N-linked oligosaccharides using standard DNA-sequencing equipment, Glycobiology 11 (2001) 275–281.

[14] A. Kobata, Size fractionation of oligosaccharides, Methods Enzymol. 230 (1994) 200–208.

[15] C.E. Clayton, Genetic manipulation of kinetoplastida, Parasitol. Today 15 (1999) 372–378.

[16] A. Misslitz, J.C. Mottram, P. Overath, T. Aebischer, Targeted integration into a rRNA locus results in uniform and high level expression of transgenes in *Leishmania amastigotes*, Mol. Biochem. Parasitol. 107 (2000) 251–261.

[17] C. Lacombe, P. Mayeux, Biology of erythropoietin, Haematologica 83 (1998) 724–732.

[18] S. Grafte, C. Leveque, J. Paysant, M. Vasse, C. Soria, J.P. Vannier, IL-12 indirectly enhances proliferation of 5-FU-treated human hematopoietic peripheral blood $CD34^+$ cells, Am. J. Hematol. 58 (1998) 183–188.

[19] L. Annable, P.M. Cotes, M.V. Mussett, The second international reference preparation of erythropoietin, human, urinary, for bioassay, Bull. World Health Organisation 47 (1972) 99–112.

[20] J.H. LeBowitz, H.Q. Smith, L. Rusche, S.M. Beverley, Coupling of poly(A) site selection and trans-splicing in *Leishmania*, Genes Dev. 7 (1993) 996–1007.

[21] S. Yan, P.J. Myler, K. Stuart, Tetracycline regulated gene expression in *Leishmania donovani*, Mol. Biochem. Parasitol. 112 (2001) 61–69.

[22] C.M Belen, W. Gao, M. Herrera, J. Alroy, J.B. Moore, S.M. Beverley, M.A. Pereira, Heterologous expression of *Trypanosoma cruzi trans*-sialidase in *Leishmania major* enhances virulence, Infect. Immun. 68 (2000) 2728–2734.

[23] R.S. Rush, P.L. Derby, D.M. Smith, C. Merry, G. Rogers, M.F. Rohde, V. Katta, Microheterogeneity of erythropoietin carbohydrate structure, Anal. Chem. 67 (1995) 1442–1452.

[24] C.H. Hokke, A.A. Bergwerff, G.W. Van Dedem, J.P. Kamerling, J.F. Vliegenthart, Structural analysis of the sialylated N- and

O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric *N*-acetyllactosamine units, Eur. J. Biochem. 228 (1995) 981–1008.

[25] L.C. Wasley, G. Timony, P. Murtha, J. Stoudemire, A.J. Dorner, J. Caro, M. Krieger, R.J. Kaufman, The importance of N- and O-linked oligosaccharides for the biosynthesis and in vitro and in vivo biologic activities of erythropoietin, Blood 77 (1991) 2624–2632.

[26] S. Watanabe, T. Kokuho, H. Takahashi, M. Takahashi, T. Kubota, S. Inumaru, Sialylation of *N*-glycans on the recombinant proteins expressed by a bculovirus–insect cell system under beta-*N*-acetylglucosaminidase inhibition, J. Biol. Chem. (in press).