Available online at www.sciencedirect.com
SCIENCE DIRECT
Protein Expression and Purification 48 (2006) 167–172
Protein Expression & Purification
www.elsevier.com/locate/yprep

Isotopic labeling of recombinant proteins expressed in the protozoan host *Leishmania tarentolae*

Anca Niculae [a], Peter Bayer [b], Ion Cirstea [a], Tim Bergbrede [a], Regina Pietrucha [c], Mathias Gruen [c], Reinhard Breitling [c], Kirill Alexandrov [a,*]

[a] *Department of Physical Biochemistry, Max-Planck-Institute for Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany*
[b] *Structural and Medicinal Biochemistry, University Duisburg-Essen, Universitätsstraße, 45117 Essen, Germany*
[c] *JenaBioscience GmbH, Löbstedter Str. 78, 07749 Jena, Germany*

Received 24 March 2005, and in revised form 12 April 2006
Available online 25 April 2006

Abstract

Isotope labeling of recombinant proteins is a prerequisite for application of nuclear magnetic resonance spectroscopy (NMR) for the characterization of the three-dimensional structures and dynamics of proteins. Overexpression of isotopically labeled proteins in bacterial or yeast host organisms has several drawbacks. In this work, we tested whether the recently described eukaryotic protein expression system based on the protozoa *Leishmania tarentolae* could be used for production of amino acid specific ^{15}N-labeled recombinant proteins. Using synthetic growth medium we were able to express in *L. tarentolae* and purify to homogeneity (15)N-valine labeled Enchanced Green Fluorescent Protein (EGFP) with the final yield of 5.7 mg/liter of suspension culture. NMR study of isolated EGFP illustrated the success of the labeling procedure allowing identification of all 18 valine residues of the protein in the HSQC spectrum. Our results demonstrate the suitability of the *L. tarentolae* expression system for production of isotopically labeled proteins.
© 2006 Elsevier Inc. All rights reserved.

Keywords: ^{15}N-labeling; Recombinant protein; Eukaryotic expression system

Nuclear magnetic resonance spectroscopy (NMR)[1] is one of two existing methods that allow determination of protein structure at atomic resolution. A majority of NMR techniques in biology require isotopic labeling (^{2}H, ^{13}C, and/or ^{15}N) of recombinant proteins. Currently, most isotopically labeled recombinant proteins are expressed heterologously in *Escherichia coli*. Despite its obvious advantages such as rapid growth, developed methods of protein expression and cheapness of cultivation *E. coli* has a range of shortcomings that limits its utility in protein studies. The most prominent problem relates to inefficiency of *E. coli* to assist folding of eukaryotic polypeptides producing only ca. 15% of eukaryotic proteins in their active form [1]. Moreover, the prokaryotic expression systems lacks the posttranslational modifications often essential for functionality of eukaryotic proteins. To circumvent these problems several eukaryotic expression systems were tested for their ability to produce significant amounts of isotopically labeled recombinant proteins. Several authors reported successful labeling of recombinant proteins in methylotrophic yeast [2,3]. A limited number of studies report successful isotopic labeling in baculoviral expression system [4,5]. In the last 4 years much effort has been put into improving the cell-free protein expression systems to serve the needs of NMR arm of structural genomic efforts. Although this approach was quite successful when applied to prokaryotic cell-free systems no report of successful isotopic labeling of proteins system expressed in eukaryotic cell-free systems exist to date [6].

We recently described a new protein expression system based on the non-pathogenic trypanosomatid *Leishmania*

* Corresponding author. Fax: +49 0231 1331651.
E-mail address: kirill.alexandrov@mpi-dortmund.mpg.de (K. Alexandrov).

[1] *Abbreviations used:* NMR, nuclear magnetic resonance spectroscopy; EGFP, enchanced green fluorescent protein; TEA, triethanolamine base.

1046-5928/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2006.04.006

tarentolae [7]. Although parasitic in vivo, this organism can be indefinitely grown in vitro with doubling times of 5 h and is amenable to large scale fermentation reaching densities of 10^9 cells/ml [8] (Breitling unpublished). The molecular biology of *Leishmania* is well studied due to its public health relevance, which led to development of a number of potentially useful expression architectures based on circular episomes and chromosomal integrations [7,9–12]. Due to the unique organization of the transcriptional and RNA processing machineries *Leishmania* sp. can efficiently process and translate mRNA generated by RNA polymerase I (PolI) or foreign polymerases. We used this features to develop two expression systems based either on constitutive transcription of heterologous genes by endogenous PolI or on regulated transcription mediated by T7 RNA polymerase [7,13]. We demonstrated that these systems can produce heterologous proteins at the levels of ca. 4% of total cellular protein and can be used for co-expression of multiple genes [7]. Since, *Leishmania* are naturally rich in glycoproteins we used this host for overexpression of human glycoproteins. We demonstrated that recombinant human erythropoietin isolated from the culture supernatants was biologically active, natively processed at the N-terminus, and N-glycosylated. The N-glycosylation was exceptionally homogeneous, with a mammalian-type biantennary oligosaccharide and the Man(3)GlcNAc(2) core structure accounting for >90% of the glycans present [7]. Since, these features auger well for broader use of the system for expression of eukaryotic proteins we decided to investigate whether it could be used for production of isotopically labeled proteins for structural studies. Here we describe NMR analysis of Enhanced Green Fluorescent Protein (EGFP) as a model protein overexpressed and purified from recombinant *L. tarentolae* strain cultivated in a synthetic medium for selective ^{15}N-labeling.

Materials and methods

Expression strain construction and protein expression

Constitutive expression

The construction of the *L. tarentolae* strain with three *egfp* gene expression cassettes integrated into the host chromosomal rDNA locus transcribed by RNA polymerase I was described earlier [7]. Initial cultivation of this expression strain was performed in Complex LEXSY broth BHI (Jena Bioscience) to the $OD_{600} = 2$ (10^8 cells/ml) as static suspension culture at 26 °C. For isotope incorporation the culture was diluted to OD = 0.04 into Synthetic LEXSY Broth (Jena Bioscience) lacking valine and cultivated in agitated 2 L Fernbach Flasks for one passage to $OD_{600} = 0.8$–1. This valine starved culture was then diluted to OD = 0.04 into valine-free Synthetic LEXSY Broth supplemented with 25 mg/l ^{15}N-valine (Spectra stable isotopes, Columbia, USA), cultivated at 26 °C with agitation to the density of 10^7 cells/ml and harvested by centrifugation.

Inducible expression

The construction of *L. tarentolae* host strain constructively expressing NLS-tagged T7 polymerase and TET repressor (HI_NLS_T7_TET) was described earlier [13]. To achieve inducible expression of *egfp* gene this strain was transformed with a modified version of the pTUB_APegfp1.4neo#802 vector [13]. In this vector, the 3′ and 5′ fragments of tubulin gene were deleted to prevent vector's integration into the genome and the resulting vector was termed pUC LAC2-EGFP#01226 [19]. The vector was transformed into the HI_NLS_T7_TET strain of *L. tarentolae* and the recombinant clones were selected as described earlier [13]. For protein production the recombinant strain was grown as described above in Complex LEXSY broth BHI (Jena Bioscience) to $OD_{600} = 2$ (10^8 cells/ml). At this stage the cells were pelleted by centrifugation for 10 min at 2500*g* and washed in Synthetic LEXSY Broth (Jena Bioscience) lacking valine. Finally cells were resuspended to an $OD_{600} = 1$ in valine-free Synthetic LEXSY Broth supplemented with 25 mg/l ^{15}N-valine. The protein production was induced by addition of 5 mg/l tetracycline 1 h following the transfer and the cultures were incubated at 26 °C with agitation for additional 24 h until the OD reached ca. 1.8.

Protein purification

The cells expressing ^{15}N-valine labeled EGFP were pelleted by centrifugation 10 min at 2500*g* and resuspended in 20 mM Tris, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1 mM PMSF, and disintegrated by passing the cell suspension through fluidizer (Microfluidics). The homogenate was cleared by ultracentrifugation for 1 h at 30,000 rpm and the ^{15}N-Val labeled EGFP was purified from the supernatant by organic extraction as described [14]. In this procedure triethanolamine base (TEA) and dry ammonium sulfate were added to final concentrations of 100 mM and 1.6 M (40% saturation), respectively. After incubation on ice for 1 h, the precipitated proteins were removed by centrifugation for 20 min at 6000 rpm at 4 °C. Dry ammonium sulfate was added to the green supernatant to a final concentration of 2.8 M (70% saturation) and this suspension was twice extracted with a one-fourth and a 1/16th volume of ethanol, respectively, by vigorous shaking for 1 min. The aqueous and ethanol phases were separated by centrifugation for 5 min at 4000 rpm at room temperature and EGFP extraction into the organic phase was apparent. A one-fourth volume of *n*-butanol was added to the combined ethanol extract. After vigorous shaking for 30 s the formed phases were separated by centrifugation as indicated above. At this step EGFP shifted almost completely into the lower aqueous phase. The upper organic phase was discarded and an equal volume of chloroform was added to the aqueous phase. After extraction for 30 s, the phases were separated as previously performed. The upper aqueous phase containing EGFP was collected and the interphase was extracted with an equal volume of 30% ammonium sulfate solution in water. Both aqueous

phases were combined and dialyzed over night against 50 mM K_2HPO_4, pH 7.0. Alternatively, dialysis was substituted by gelfiltration on Superdex 200 (16/60) (GE Healthcare) equilibrated with 15 mM Hepes, pH 6.7, 10 mM LiCl, and 5 mM DTT). Finally the samples were concentrated using Viva Spin filtration unit to a final concentration of 18 mg/ml. The resulting protein was snap frozen in liquid nitrogen and stored at −80 °C.

NMR spectra acquisition

The ^{15}N-HSQC experiment was performed on a Bruker DRX-500 spectrometer, equipped with shielded Z gradients. The temperature for the experiment was set to 300 K. Uniformly ^{15}N-Val labeled protein with a concentration of 0.6–0.7 mM dissolved in phosphate buffer solution (50 mM potassium phosphate, pH 7) was used as sample.

The spectrum was collected with 40 transients and 2048 × 256 data points were used for acquisition resulting in a measurement time of 5 h. The water resonance was suppressed by applying a WATERGATE sequence. Prior to Fourier transformation, the spectrum was multiplied by a cosine bell square window function. For ^{1}H-calibration the water resonance was set to 4.75 ppm. For ^{15}N calibration we followed the IUPAC procedure. The HSQC spectrum was processed using the standard Bruker software XWINNMR. Analysis and visual representation of the two-dimensional spectrum was performed using the NDEE program package (SpinUp Inc., Dortmund, Germany) on an Octane workstation (Silicon Graphics Inc.).

Results and discussion

To assess the suitability of *L. tarentolae* for production of isotopically labeled proteins we chose an enhanced variant of Enhanced Green Fluorescent Protein (EGFP) as a model polypeptide. The choice was due to availability of high yield EGFP expression strains of *L. tarentolae* as well as the simplicity of purification protocols [14,7]. Due to the natural auxotrophy of *L. tarentolae* for valine (in addition to R, H, M, W, F, S, Y, T, L, and K [15]), and the presence of 18 valine residues in EGFP we choose ^{15}N-valine for labeling.

We initially used the established strain of *L. tarentolae* constitutively expressing EGFP. The cells were starved for one passage in the synthetic valine-free medium and were then diluted into valine-free medium supplemented with 25 mg/l ^{15}N-valine. At a density of 10^7 cells/ml the cells were harvested by centrifugation, resuspended in lysis buffer and lysed. Although this procedure yielded 5 mg of ^{15}N-Val labeled EGFP-N1 which was sufficient for NMR experiments the overall yield was low (Fig. 1) (Table 1).

The main limitation of the constitutive expression approach was the need for continues cultivation of the EGFP producing strain in the defined medium which resulted in the reduced growth rates and yields of recombinant protein. To circumvent these problems we chose to use an inducible variant of *L. tarentolae* expression system. In principle, this should allow expansion of the expression clone in the rich complex medium followed by transfer into the defined medium shortly before the induction. To this end we employed a modified version of *L. tarentolae* inducible expression system described earlier [13]. In contrast to the previously described system where all heterologous elements were chromosomally integrated, we placed the EGFP expression cassette onto the episomaly maintained circular plasmid [19]. This modification was introduced to enhance the expression yields by increasing the amount of the template available for T7 transcription as well as to simplify the expression strain construction. For production of the ^{15}N-valine labeled EGFP the recombinant clone was grown in complex LEXSY Broth BHI in the absence of tetracycline and than transferred into the defined medium supplemented with ^{15}N-valine (see Materials and methods for details). One hour following the transfer the protein production

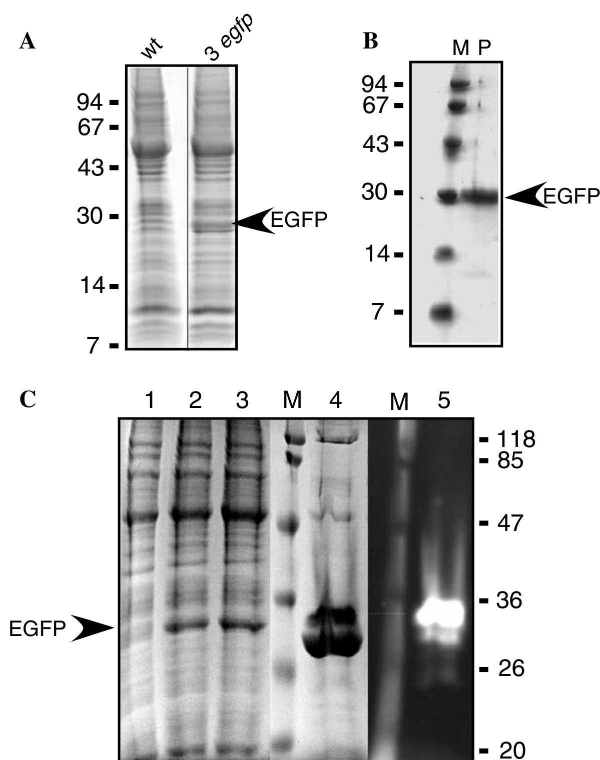

Fig. 1. Expression and purification of ^{15}N-Val EGFP. (A) SDS–PAGE analysis of total cellular lysate of wild type *L. tarentolae* strain and the transgenic strain transformed with 3 *egfp* expression cassettes. The position of EGFP is indicated by arrows. The prominent band at 47 kDa represents *L. tarentoale* tubulin. (B) ^{15}N-Val EGFP protein isolated from the constitutive expression strain resolved on SDS–PAGE and stained with Coomassie. (C) Expression of ^{15}N-Val EGFP in the inducible expression strain. Lanes 1–3, total cell extracts at the time of induction, 9 and 24 h, respectively. Lane 4, purified ^{15}N-Val EGFP-N1 after aqueous $(NH_4)_2SO_4$ extraction. The double band represents denatured and non-denatured EGFP protein. The lane 5 is the same as lane 4 but photographed in the UV light to identify folded and unfolded EGFP protein.

Table 1
Purification of ^{15}N-Val EGFP from the constitutive *L. tarentolae* expression system

Purification step	Total protein (mg)	EGFP-^{15}N-Val (mg)	Yield (%)
Supernatant after ultracentrifugation	920	21	100
40% $(NH_4)SO_4$	61	17	81
Ethanol extraction	17.5	14	66
Aqueous $(NH_4)_2SO_4$ extraction	8	7	33
After dialysis	6.5	6.5	31
After concentration	5	5	24

Ten liters of culture were used for purification.

was induced by addition of tetracycline and the cells were cultivated for additional 24 h. In this case, the expression yield reached ca. 14 mg of protein/liter of culture (Table 1).

For isolation of the ^{15}N-Val labeled EGFP cells obtained form both preparations were lysed and EGFP was purified from the supernatant by organic extraction as described [14]. The main steps of this procedure include extraction of EGFP-N1 with ethanol and subsequent protein re-extraction into the aqueous phase (see Materials and methods for details). The results of expression and purification of ^{15}N-Val labeled EGFP-N1 using both expression systems are summarized in Tables 1 and 2 and in Fig. 1. We conclude that the inducible version of the *L. tarentolae* expression system is more suitable for production of the isotopically labeled proteins.

Protein analysis and NMR spectroscopy

Since, N1 variant of EGFP used in this study contains 18 valine residues, we expected to find a total of 18 NH

Table 2
Purification of ^{15}N-Val EGFP from the inducible *L. tarentolae* expression system

Purification step	Total protein (mg)	EGFP-^{15}N-Val (mg)	Yield (%)
Supernatant after ultracentrifugation	99	9.5	100
40% $(NH_4)SO_4$	n.d.	n.d.	
Ethanol extraction	n.d.	n.d.	
Aqueous $(NH_4)_2SO_4$ extraction	4	4	42
After gel filtration	1.2	1.2	13
After concentration	1.2	1.2	13

n.d., not determined.

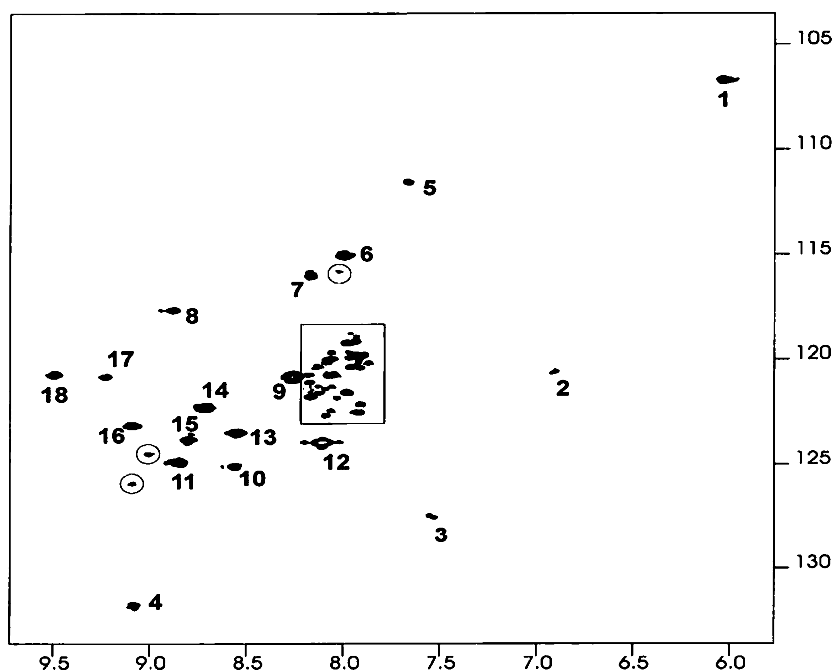

Fig. 2. ^{15}N-HSQC NMR spectrum of ^{15}N-Val EGFP-N1. The 18 major resonances are labeled with numbers. Minor resonances are surrounded by an open circle. Resonances marked by a frame exhibit chemical shift close to the random coil shift of valine amide groups observed in small unstructured peptides and are likely to arise from protein degradation or an unfolded protein species.

signals in the corresponding ^{15}N-HSQC spectrum. As can be seen in Fig. 2 18 major resonances (labeled 1–18) were observed in the spectrum in addition to some minor peaks (circles) and a group of resonances (see rectangle) around 8 ppm (^{1}H-axis) and 120 ppm (^{15}N-axis). Both preparations of EGFP produced similar spectra (data not shown). The dispersion of resonances along both frequency axes indicates a folded state of the protein. The chemical shifts of resonances found in the spectrum are characteristic for a protein having fewer valine moieties in helical elements, with more in β-strands and loop or turn regions, which is in accordance with the structure of a crystallized $GFP_{Phe64\text{-}Leu,\ Ser65\text{-}Thr,\ Val163\text{-}Ala}$ mutant (GFPuv; (PDB entry1B9C)). The latter protein, which contains 14 valine residues and a N-terminal eight amino acid deletion when compared to the original published sequence of GFP, is known to comprise 11 antiparallel β-strands and only one α-helix [16].

Due to the sequence identity between GFPuv and EGFP-N1 of 96%, we were able to assign 12 of the major resonances of EGFP-N1 based on the previously published assignments for the GFPuv variant [17]. With respect to the sequence notation of GFPuv (addition of 1 yields number in EGFP-N1 sequence) resonances belonging to residues Val11 (9), Val12 (4), Val22 (13), Val29 (16), Val61 (2), Val93 (18), Val112 (10), Val120 (11), Val16 (14), and Val219 (3) could be assigned unambiguously in the HSQC spectrum of EGFP-N1. Resonance 6 in Fig. 2 represents the amide group of Val176 and resonance 7 that of Val193 (or vice versa). The four out of six unassigned major resonances 1, 5, 8, 12, 15, and 17 in the HSQC spectrum are thus belonging to the amide groups of residues Val150, Val55, Val68, and Val224 while the origin of two is still unclear. In addition to the 18 major resonances three minor peaks (circles) were observed, which could not be assigned on the basis of the published data. They were either belong to non-valine residues labeled with ^{15}N by metabolic transfer of amide groups or they indicate valine residues present in an alternative conformation in the protein. A second group of small resonances in the frequency region of 7.8–8.2 ppm (^{1}H) and 118–124 ppm (^{15}N) (see box in the HSQC spectrum) are likely to arise from degradation of the protein during measurement, as the chemical shifts of these resonances are close to the *random coil* shift found for valine residues in short peptides [18]. Alternatively, those represent a minor fraction of co-purified unfolded EGFP-N1.

Based on the presented data we conclude that it is possible to efficiently label a recombinant 28 kDa protein expressed in *L. tarentolae* with ^{15}N-Val. Amount of labeled protein sufficient for NMR analysis was isolated from the *L. tarentolae* culture and all 18 valine residues of the target protein were observed in the HSQC spectrum. We expect that the presented system can be also used for labeling of recombinant proteins with 10 other essential ^{15}N-amino acids. Although the several isotopically labeled amino acids are quite expensive this still can provide a viable approach for selected target proteins. The uniform labeling of proteins could be potentially achieved using medium with isotopically labeled carbon source. Therefore it is expected that the *Leishmania* expression system can be used for a complete assignment strategy of a variety of proteins that are not accessible to NMR structure determination when expressed in other expression systems. The fact that the described system expresses isotopically labeled proteins inducible and with a high yield also suggests that the system might be suitable for in vivo NMR studies where no eukaryotic expression system was successfully used so far.

Acknowledgments

We thank Roger S. Goody for advice and critical reading of the manuscript. We are grateful to Lukas Trantirek for fruitful discussions on in vivo NMR. This work was supported by the Grant 2001 FE243 of the Thüringer Aufbaubank to Jena Bioscience and by a Grant of the Deutsche Forschungs-Gemeinschaft AL 484/7-1/2 to K.A.

References

[1] A.M. Edwards, C.H. Arrowsmith, D. Christendat, A. Dharamsi, J.D. Friesen, J.F. Greenblatt, M. Vedadi, Protein production: feeding the crystallographers and NMR spectroscopists, Nat. Struct. Biol. 7 (Suppl.) (2000) 970–972.

[2] E. Rodriguez, N.R. Krishna, An economical method for (15)N/(13)C isotopic labeling of proteins expressed in *Pichia pastoris*, J. Biochem. (Tokyo) 130 (2001) 19–22.

[3] A.R. Pickford, J.M. O'Leary, Isotopic labeling of recombinant proteins from the methylotrophic yeast *Pichia pastoris*, Methods Mol. Biol. 278 (2004) 17–34.

[4] A. Strauss, F. Bitsch, B. Cutting, G. Fendrich, P. Graff, J. Liebetanz, M. Zurini, W. Jahnke, Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies, J. Biomol. NMR 26 (4) (2003) 367–372.

[5] M. Bruggert, T. Rehm, S. Shanker, J. Georgescu, T.A. Holak, A novel medium for expression of proteins selectively labeled with 15N-amino acids in *Spodoptera frugiperda* (Sf9) insect cells, J. Biomol. NMR 25 (4) (2003) 335–348.

[6] S. Yokoyama, Protein expression systems for structural genomics and proteomics, Curr. Opin. Chem. Biol. 7 (2003) 39–43.

[7] R. Breitling, S. Klingner, N. Callewaert, R. Pietrucha, A. Geyer, G. Ehrlich, R. Hartung, A. Muller, R. Contreras, S.M. Beverley, K. Alexandrov, Non-pathogenic trypanosomatid protozoa as a platform for protein research and production, Protein Expr. Purif. 25 (2002) 209–218.

[8] J.D. Alfonzo, O.H. Thiemann, L. Simpson, Purification and characterization of MAR1. A mitochondrial associated ribonuclease from *Leishmania tarentolae*, J. Biol. Chem. 273 (1998) 30003–30011.

[9] 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.

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

[11] 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.

[12] J.H. LeBowitz, C.M. Coburn, D. McMahon-Pratt, S.M. Beverley, Development of a stable *Leishmania* expression vector and application to the study of parasite surface antigen genes, Proc. Natl. Acad. Sci. USA 87 (1990) 9736–9740.
[13] S. Kushnir, K. Gase, R. Breitling, K. Alexandrov, Development of an inducible protein expression system based on the protozoan host *Leishmania tarentolae*, Protein Expr. Purif. 42 (2005) 37–46.
[14] A.V. Yakhnin, L.M. Vinokurov, A.K. Surin, Y.B. Alakhov, Green fluorescent protein purification by organic extraction, Protein Expr. Purif. 14 (1998) 382–386.
[15] S.M. Krassner, B. Flory, Essential amino acids in the culture of *Leishmania tarentolae*, J. Parasitol. 57 (1971) 917–920.
[16] M. Ormo, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, S.J. Remington, Crystal structure of the *Aequorea victoria* green fluorescent protein, Science 273 (1996) 1392–1395.
[17] F. Khan, K. Stott, S. Jackson, 1H, 15N and 13C backbone assignment of the green fluorescent protein (GFP), J. Biomol. NMR 26 (3) (2003) 281–282.
[18] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, J. Biomol. NMR 5 (1995) 67–81.
[19] I. Cirstea, K. Alexandorv, Construction of transgenic strains of *Leishmania tarentolae* inducibly expressing heterologouse proteins, unpublished work.