Expression and purification of the extracellular domains of human glycoprotein VI (GPVI) and the receptor for advanced glycation end products (RAGE) from Rattus norvegicus in Leishmania tarentolae

Thomas Langer, Carsten Corvey, Katja Kroll, Oliver Boscheinen, Thomas Wendrich, and Werner Dittrich

R&D Biologics Research, Sanofi-Aventis Deutschland GmbH, Frankfurt am Main, Germany

ABSTRACT
Glycosylation is one of the most complex post-translational modifications and may have significant influence on the proper function of the corresponding proteins. Bacteria and yeast are, because of easy handling and cost reasons, the most frequently used systems for recombinant protein expression. Bacteria generally do not glycosylate proteins and yeast might tend to hyperglycosylate. Insect cell- and mammalian cell-based expression systems are able to produce complex N-glycosylation structures but are more complex to handle and more expensive. The nonpathogenic protozoa Leishmania tarentolae is an easy-to-handle alternative expression system for production of proteins requiring the eukaryotic protein folding machinery and post-translational modifications. We used and evaluated the system for the secretory expression of extracellular domains from human glycoprotein VI and the receptor for advanced glycation end products from rat. Both proteins were well expressed and homogeneously glycosylated. Analysis of the glycosylation pattern identified the structure as the conserved core pentasaccharide Man$_2$GlcNAc$_2$.

Introduction
Recombinant protein expression is mainly done in microbial systems. The most widely used host is the prokaryotic Gram-negative bacterium Escherichia coli. Other well-established prokaryotic systems are based on different strains of the Gram-positive genus Bacillus. Despite the ease, productivity, and low cost of these systems, there are limitations concerning post-translational modifications. If not especially engineered, bacterial systems do not glycosylate proteins. To obtain glycosylated proteins, eukaryotic expression systems are required. The most commonly used systems are yeast, baculovirus-infected insect cells, and mammalian cell lines. Yeasts are eukaryotic, single-celled organisms that can easily be cultivated and manipulated. Examples for such yeast species are Saccharomyces cerevisiae, Pichia pastoris, and Kluyveromyces lactis. The main disadvantage of yeast-based systems is their bias toward hyperglycosylation. Baculovirus/insect cell-based systems can lead to high levels of protein expression, contain the full eukaryotic protein folding machinery, and are capable of post-translational modification. The main disadvantages of this system are the need for virus infection, and distinct differences between the glycosylation pattern of insect cells and human. With regard to human-like glycosylation pattern, mammalian cell lines need to be used. Probably, the most common system is based on Chinese hamster ovary (CHO) cells and human embryonic kidney (HEK) cells. CHO and HEK cells are able to produce proper folded and glycosylated proteins at high yields. Their main drawbacks are slow growth and high costs for cultivation.[1–4]

An alternative, easy-to-handle eukaryotic system using the protozoan Leishmania tarentolae as host for recombinant protein expression was developed a few years ago. L. tarentolae can be used either for intracellular or secreted protein expression. As a true eukaryotic cell, L. tarentolae is capable to glycosylate proteins in a mammalian-like fashion.[5] Since only very low levels of endogenous protein are secreted by L. tarentolae, it is attractive to use it for expression of secreted proteins[5,6] and the secreted recombinant protein accounts for the vast majority of protein in the medium. Additional advantages of this system are its ease of handling, low cost, and rapid doubling times.[6,7] The mainly used cultivation medium is brain heart infusion (BHI), but completely defined media are available as well.[8]

The receptor for advanced glycation end products (RAGE) belongs to the large family of immunoglobulins. RAGE is a cell surface receptor and consists of an extracellular region, a single membrane spanning helix and a short cytoplasmic domain.[9,10] The extracellular domain is composed of one Ig-like V-type and two Ig-like C-type domains.[11,12] The cytoplasmic domain is required for intracellular signaling.[13] Initially, RAGE was identified to be the receptor for the so-called advanced glycation end products (AGEs).[14] These products are the results of nonenzymatic modifications mainly of lysyl or arginyl side chains from proteins with sugars.[15] Over the past years, other ligands than AGEs have been...
described to bind to RAGE. To these molecules belong members of the S100 protein family, amphoterin, the beta-amyloid peptide, but also nucleic acids, glycosaminoglycans, and phospholipids. Due to the broad spectrum of ligands, RAGE is considered to be a pattern recognition receptor. Upon ligand binding, RAGE is activated and intracellular signaling occurs mainly through NF-kB, members of the mitogen-activated protein kinase family (p38, c-Jun N-terminal kinase JNK) and the phosphoinositide 3-kinase (PI3K)/Akt-pathway. In contrast to other receptor types, ligand-bound RAGE is not internalized. Increased RAGE expression occurs in several chronic and acute inflammatory diseases and RAGE activation perpetuates the inflammatory response. Furthermore, RAGE signaling is also involved in aging and other chronic diseases. Targeting the RAGE signaling pathway seems to be a logic strategy for treating chronic diseases that are driven by sustained inflammation processes.

Glycoprotein VI (GPVI) was first described as a protein present in platelets running in an SDS-PAGE below glycoprotein V (hence the name) and has an apparent molecular weight of ∼60 kDa. The first clue for its function was made when a spot on a 2D gel corresponding to GPVI was absent in a patient having platelets that did not respond to collagen. The exact recognition motif for GPVI was later identified as GlyProHyp repeats, which occur within collagen. After cloning of the GP6 gene, the receptor was identified as a protein with a single membrane spanning helix and two Ig-like domains. In platelets, the receptor is strictly associated with Fcγ chains. Dimerization of GPVI is crucial for collagen binding. Collagen binding to GPVI activates platelets leading to platelet aggregation and finally to thrombus formation. Intracellular signaling upon GPVI activation occurs through the tyrosine kinases Syk and Fyn/Lyn. Targeting GPVI seems to be an attractive approach to prevent occlusive thrombus formation in diseased vessels. Since platelets can be activated in a GPVI-independent manner, blocking of GPVI signaling does not result in bleeding risks.

Blocking the signaling pathways of RAGE and GPVI by binding inhibitors to their extracellular domains seems to be an elegant pharmaceutical approach. Monoclonal antibodies (mAbs) are excellent tools for this task. The antibody–antigen interaction is a key parameter for antibody function. Hence, antigens used for generating therapeutic antibodies should be as similar as possible to the natural target, including post-translational modifications. In addition, homogeneous protein samples are highly desired for protein crystallization trials.

The aim of this work was to express extracellular domains of hsGPVI and rRAGE in the *L. tarentolae* expression system evaluating this system as an alternative to mammalian cell-based systems. The main focus of this evaluation was on ease of handling and the glycosylation pattern.

**Materials and methods**

**Generation of recombinant Leishmania tarentolae strains**

*Leishmania tarentolae* laboratory strain P10, expression plasmid pLEXSY-S (former designation pF4SpLmsapX1.4sat), and antibiotic nourseothricin were purchased from Jena Bioscience (Jena, Germany). All restriction enzymes were purchased from New England Biolabs (Frankfurt, Germany). DNA coding for rat RAGE amino acids 23-325 (Swiss Prot Q63495) comprising the extracellular domain was amplified with appropriate primers, which also add six consecutive histidine residues at the C-terminus, from a rat kidney DNA bank (rat kidney cDNA, Clontech, Mountain View, CA, USA) and cloned into the KasI/Xhol site of pLEXSY-S, generating pLEXSY-S-rRAGE 23-235. The N-terminal domain for human GPVI was amplified using RT-PCR from total RNA (human skeletal muscle total mRNA, Clontech) and subcloned into the Ncol/BamHI site of pET3d (Novagen, Madison, WI, USA). This vector was used as template and DNA coding for amino acids 21-204 from human GPVI (Genbank NM_016363) was cloned with seven consecutive histidine residues at the C-terminus into the KasI/NotI site of pLEXSY-S, generating pLEXSY-S-hsGPVI 21-204. Cloning into pLEXSY-S adds an N-terminal signal sequence from the gene for secreted acid phosphatase (LMSAP) of *Leishmania mexicana* to the desired expression constructs. The expression plasmids were transformed into *E. coli* DH5alpha (New England Biolabs) and sequences were checked on a 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Recombinant plasmids were purified from an overnight *E. coli* culture grown in LB medium (Invitrogen, Karlsruhe, Germany) containing 100 µg/mL ampicillin (Applichem, Darmstadt, Germany) with the aid of the Qiagen plasmid mini-kit (Qiagen, Hilden, Germany). Prior transformation of *L. tarentolae*, plasmids were digested with Swal as recommended by the manufacturer. Complete digestion was controlled on a 1.2% agarose gel (E-Gel, Invitrogen). *L. tarentolae* was cultured in BHI medium (35 g/L, Fluka, Germany) supplemented with 0.25% hemin (Fluka). A total of 400 µL of a culture with drop-like shaped *Leishmania* cells (checked with the aid of a microscope) were put in a 2-mm electroporation cuvette (Bio-Rad, München, Germany) and 10 µg of Swal-digested DNA was added. One culture was also subjected to the electroporation procedure but without addition of DNA. This culture served as a control during the selection process. Cells were chilled on ice for 10 min and electroporation was accomplished with a 450 µF/450 V pulse (GenPulser II equipped with the Capacitance Extender Plus device, Bio-Rad). Cells were immediately placed on ice for another 10 min and subsequently transferred to 4 mL of prewarmed (26°C) BHI medium in a 24-deep-well plate (Greiner Bio-One, Frickenhausen, Germany). Cultures were maintained at 26°C and 200 rpm in an incubation shaker and after 24 hr 100 µg nourseothricin (Jena BioScience) per milliliter medium was added. Every 2 days, cultures were diluted with fresh BHI medium supplemented with 100 µg/mL nourseothricin 1:10. After several passages, the cells in the control culture died, whereas the cells in the transformed cultures grew normally in the presence of nourseothricin.

**Expression and purification of human GPVI and rat RAGE**

Recombinant protein expression was checked by analyzing the culture supernatant. From a small-scale shaker flask, samples...
were withdrawn at different time points. Cells were removed by centrifugation (4,000× g, 15 min, 4°C), the supernatant was concentrated (Vivaspin 6, 5 kDa cutoff, Sartorius-Stedim, Göttingen, Germany) and an aliquot of the concentrated supernatant was mixed with reducing SDS-PAGE sample buffer and loaded on a 4–12% Bis-Tris gel (NuPAGE gel with Mops-running, Invitrogen) and run for 45 min at constant 200 V. Gels were stained with Coomassie blue (Instant blue, Expeideon, Harston, Cambridgeshire, UK).

For preparative protein expression, L. tarentolae strains were grown in 3-L Fernbach flasks (Greiner Bio-One) containing 1 L of BHI medium supplemented with hemin. The main culture was inoculated with 50 mL of an overnight culture (OD_{600} ∼ 1.1). The initial OD_{600} of the main culture was 0.1. After 48 hr of incubation at 26°C and 140 rpm (ISF 1 W incubator, Kühner, Basel, Switzerland), the OD_{600} of the cultures reached ∼2.0 and cells were removed by centrifugation (15 min, 5,000 rpm, 4°C). The supernatant was filtered through a 0.2-µm membrane (Durapore membrane filter, Millipore, Schwalbach, Germany) and pH was adjusted to 8.0 with 1 M HCl, pH 9.0. The supernatants were loaded on a 5-ML HisTrap Crude FF column (GE Healthcare, München, Germany) equilibrated in 50 mM Tris, pH 8.0, 500 mM NaCl. Proteins were eluted from the HiTrap Crude FF column with a linear gradient over 10 column volumes to 50 mM Tris, pH 8.0, 500 mM NaCl, 500 mM imidazole on an ÄKTA-explorer system (GE Healthcare). Fractions were analyzed by SDS-PAGE and fractions containing the protein at the expected molecular size were pooled. rRAGE and hsGPVI were further purified by gel filtration using a Superdex 75 pg column (GE Healthcare), equilibrated in 50 mM HEPES, pH 7.5, 300 mM NaCl (rRAGE) or 25 mM Tris, pH 7.5, 150 mM NaCl (hsGPVI). Fractions containing the corresponding proteins were pooled and concentrated (Vivaspin 20, 5 kDa cutoff, Sartorius-Stedim) to 2.9 mg/mL (rRAGE) and 6.5 mg/mL (hsGPVI). Total yield from 1 L of culture was 1.5 mg of purified protein for rRAGE and 1.0 mg for hsGPVI.

Analysis of protein glycosylation using mass spectrometry

Protein deglycosylation was performed according to manufacturer’s recommendation by PNGaseF treatment (New England Biolabs). Subsequently, the samples were subjected to reducing SDS-PAGE on a 4–12% Bis-Tris gel (NuPAGE gel with Mops-running, Invitrogen). The lanes of interest were punched out from the SDS-PAGE and washed thrice with 50% acetonitrile/25 mM NH4HCO3, pH 8.0. Gel pieces were soaked in 100% acetonitrile for 5 min followed by reduction and alkylation using a solution of 10 mM 1,4-dithiothreitol (DTT), 25 mM NH4HCO3, pH 8.0, and further incubated at 56°C for 45 min. After drying with 100% acetonitrile, the supernatant was removed and a 55 mM iodoacetamide solution (25 mM NH4HCO3, pH 8.0) was added. The samples were further incubated in the dark for 45 min and washed thrice with 50% acetonitrile/25 mM NH4HCO3, pH 8.0. After drying with 100% acetonitrile, the in-gel digest was performed using a diluted trypsin solution (5 ng/µL) in 25 mM NH4HCO3/10% acetonitrile, pH 8.0. Samples were incubated at 37°C overnight in a convection oven. Peptide extraction was performed by adding 50% acetonitrile/5% trifluoroacetic acid (TFA) to the gel pieces. Finally, samples were dried in a SpeedVac for 30 min. For consecutive analysis, 10 µL 2% acetonitrile/0.1% TFA was added per 50 µL sample. For LC–MS/MS analysis, 4 µL was injected per run. For matrix-assisted laser desorption ionization (MALDI), preparation, 0.5 µL 2,5-dihydroxybenzoic acid (DHB) matrix was placed on the target, 0.5 µL of the sample and (when dried) 0.5 µL matrix were added. The obtained LC–MS/MS and MALDI spectra were compared with an in-house user-defined database SwissProt, to which the given constructs were added for identification.

Edman sequencing

For identification of the first amino acids, the proteins were N-terminal sequenced. Protein samples were transferred onto a polyvinylidenedifluoride (PVDF) membrane (Immobilon 0.45 µm, Invitrogen) and analyzed using Edman degradation on an Applied Biosystem 494 Procise cLC device (Applied Biosystems, Foster City, CA, USA). All chemicals and the HPLC column were provided by the manufacturer.

Analytical gel filtration and static light scattering

Analytical gel filtration runs were done using a Superdex 75 10/30 HR column (GE Healthcare) with 50 mM HEPES, pH 7.5, 300 mM NaCl (rRAGE) or 25 mM Tris, pH 7.5, 150 mM NaCl (hsGPVI) as running buffer on an ÄKTA purifier system (GE Healthcare). To calculate the molecular mass of the proteins, the system was equipped with a static light scattering detector (miniDawn, Wyatt, Dernbach, Germany) and an RI detector (Shodex RI-101, Showa-Denko, München, Germany). Analysis of the light scattering data and calculating the molecular mass were performed using the ASTRA software (Wyatt).

Results and discussion

The pLEXSY-S plasmids were used to generate L. tarentolae strains recombinantly expressing the soluble domains of either rRAGE (Aa 21-204) or hsGPVI (Aa 23-325) as described in Figure 1. The extracellular domains of both hsGPVI and rRAGE consist of immunoglobulin (Ig)-like domains containing each one disulfide bond. In addition, both hsGPVI and rRAGE are N-glycosylated. The N-glycosylation sites were conserved in the corresponding expression constructs (Figure 1). Upon transformation, the expression cassettes from the pLEXSY-S plasmids are integrated in the rrnB1 locus of the genome. Transcription is driven by RNA polymerase I in a constitutive manner. Secretion of the proteins into the medium is ensured by the signal sequence from the gene for secreted acid phosphatase (LMSAP) of L. mexicana. Correct integration of the expression cassette was investigated indirectly by detection of the secreted recombinant protein. Cell growth and expression were followed and analyzed by measuring the OD_{600} and SDS-PAGE during expression (Figure 2a and 2b). Then, 48 hr after inoculation of the main culture, the stationary phase was reached and no additional protein was expressed as judged by SDS-PAGE (Figure 2a).
For hsGPVI, additional bands in the samples are clearly visible, whereas this was not so pronounced for rRAGE. The secreted proteins were clearly visible in the SDS-PAGE (Figure 2a). The main cultures were harvested 48 hr after inoculation. The doubling time during the log phase was estimated to be ∼17 hr (Figure 2b).

The recombinant proteins were purified directly from the cleared culture supernatant through a C-terminal six consecutive histidine residues tag on an IMAC column. A subsequent gel filtration step led to >90% pure protein preparations (Figure 3a and 3b). Without any additional optimization step, 1.5 mg of purified rRAGE and 1.0 mg hsGPVI were obtained per liter culture with this two-step procedure.

To analyze whether the isolated soluble domains underwent N-glycosylation by L. tarentolae, the proteins were digested with PNGaseF and consecutively analyzed using SDS-PAGE. After deglycosylation, the proteins migrated with a slightly reduced apparent molecular weight in the SDS-PAGE (Figure 3a and 3b; lane 3). To further analyze the glycosylation sites and structure of the glycan pattern, the recombinantly expressed proteins were in-gel digested and a peptide mass fingerprint (PMF) analysis was performed. The total sequence coverage obtained for hsGPVI using trypsin was 64%. Comparing the results obtained before and after deglycosylation indicated that the peptide is glycosylated at the consensus site...
at residue Asn92 (… CSYQNGSLWSL …). After deglycosylation, a mass shift of 891.27 Da was observed (Figure 4a). Intact mass analysis of hsGPVI has shown that only one N-glycosylation site per protein was observed. For rRAGE, the total sequence coverage using trypsin was 56%. rRAGE contains two glycosylation sites at Asn25 and Asn80 (UniProt entry Q63495). Only one peptide containing the glycosylation site at Asn80 (… ILPNGSSLL …) could be identified. Here, a mass shift of approximately 891.31 Da was observed after deglycosylation with PNGaseF (Figure 4b). These data suggest that the glycan structure on both proteins comprises only the core glycan structure.[Figure 5 core glycan structure (892 Da)].

N-terminal sequencing was done to further characterize the proteins. In the case of rRAGE, two sequences with similar intensities were obtained (Figure 6). The second sequence lacks the first two amino acids (note that the first two amino acids GlyAla are added due to the cloning procedure). The amino acid corresponding to Asn25 in the sequence could not be identified. Presumably, this amino acid corresponds to the glycosylated Asn25. For hsGPVI, however, Edman sequencing was not possible since the N-terminus was blocked. To analyze whether the proteins are monomeric or form higher oligomeric structures, analytical gel filtration runs in combination with static light scattering were done. Light scattering in addition with the measurement of the protein concentration in solution allows to calculate the molecular mass of the protein independently of calibration curves.[33] For hsGPVI and rRAGE, molecular weights of 19.7 and 36.0 kDa, respectively, were obtained (Figure 7). The calculated molecular weights of the proteins (not considering any post-translational modification) are 21.129 Da (hsGPVI) and 33.411 Da (rRAGE). These results suggest that the proteins are monomeric in solution.

Targeting the extracellular domains of hsGPVI and rRAGE is of pharmacological interest, but they do not have a catalytic activity per se. The proteins contain Ig-like domain with internal disulfide bridges and are N-glycosylated. The presence of N-glycans seems to be essential for proper function of the proteins.[34,35] The material used in these experiments are from mouse neuroblastoma cell lines and bovine lung[34] or an engineered human hematopoietic HEL cell line[35] suggesting that they contain the proper glycosylation pattern.

---

**Figure 4.** MALDI-ToF analysis of the trypsin-digested peptide from (a) hsGPVI and (b) rRAGE before and after deglycosylation with PNGaseF. Shown are the spectra of the glycosylated peptides with their corresponding amino acid sequence. The consensus glycosylation site is highlighted. *Note: GPVI, glycoprotein VI; RAGE, receptor for advanced glycation end products.*
Figure 5. Schematic structure of N-glycosylation pattern. The specified masses are average masses corresponding to the respective residues. As an example for mammalian N-glycosylation, a complex, diantennary type is shown. The glycosylation pattern described by Breitling et al.⁵¹ is similar to the mammalian type. The only difference is the absence of terminal sialic acids. The core structure consists of two acetylglucosamine and three mannose residues. The masses identified for hsGPVI and rRAGE expressed in L. tarentolae fit to the Man3GlcAc2 core structure. Note: GPVI, glycoprotein VI; RAGE, receptor for advanced glycation end products.

Figure 6. Results from N-terminal sequencing of rRAGE. The upper sequence corresponds to the sequence from Swiss Prot entry Q63495 starting at amino acid 23. The additional two N-terminal amino acids (shown in grey) are due to cloning. Two sequences were obtained with similar signal intensities (specified as pmol). Note: RAGE, receptor for advanced glycation end products.

Figure 7. Size exclusion chromatography (Superdex 75 10/30 HR column, GE Healthcare) of purified hsGPVI (a) and rRAGE (b). The UV 280 nm absorbance curves are shown in blue. The calculated molar masses of the peak area (shown in red) are plotted as a function of elapsed time. Note: GPVI, glycoprotein VI; RAGE, receptor for advanced glycation end products.
We used the *L. tarentolae* expression system as a possible, easy-to-handle, and cheap alternative for expression of proteins with complex mammalian N-glycosylation. For both proteins, the extracellular domains of hsGPV1 and rRAGE, generating the recombinant *L. tarentolae* expression strains was straightforward. We observed reasonable secretion of the proteins in the culture supernatants. The doubling time for the *L. tarentolae* expression culture was determined to be ∼17 hr. This is much slower than the 6.7 hr of doubling time reported.[36] However, the reason for this might be that our culture were grown in shaker flasks and not in a fermentor and limited aeration is probably the limiting factor. Both proteins were glycosylated as shown by PNGaseF digestion and analysis by SDS-PAGE and PMF. The N-glycosylation is remarkably homogeneous but consists only of the Man3-GlcNAc2 core structure. Other N-glycans could not be identified. These findings are in accordance with the results from Klatt et al.[37] who have produced a soluble amyloid precursor protein in *L. tarentolae*. Successful production of glycoproteins in *L. tarentolae* has been reported elsewhere, but a detailed analysis of the glycosylation pattern was not done.[38–40] Analysis of the N-terminus of rRAGE revealed two sequences. Since it has been reported that *L. tarentolae* secretes only very limited amounts of endogenous proteins,[5] the truncated species might be a result from imprecise cleavage of the signal peptide by the signal peptidase. We used the signal sequence from secreted acid phosphatase (LMSAP) of *L. mexicana*, which was included in the commercial available expression vector but other signal sequences might be more appropriate.

**Conclusion**

*Leishmania tarentolae* is an easy-to-handle eukaryotic expression system that is able to produce homogeneously glycosylated proteins. The glycosylation pattern was identified as the Man3,GlcNAc2 core structure. Complex human-like glycosylation pattern was not detected. Further studies using *L. tarentolae* as expression host are needed to elucidate whether the glycosylation pattern may vary depending on the type of expressed protein or on the cultivation conditions.

**Conflict of interest**

The authors are all employees of Sanofi-Aventis Deutschland GmbH; however, they declare no conflict of interest.

**Acknowledgments**

The authors gratefully acknowledge the technical assistance of Katja Böhler and Karin Jouvenal.

**ORCID**

Carsten Corvey http://orcid.org/0000-0003-2614-0960

**References**


Immune Complexes is Mediated by HMGB1 an RAGE. *Nat. Immunol.* 2007, 8, 487–496.


