Protein Production Network Sweden (PPNS):

This is a report from an inter-facility, multi-host expression screen performed in 2014, with the aim of demonstrating the advantages and possibilities with such a national, collaborative effort. We would like to offer this type of screen as a service to the Swedish research community and also investigate the possibilities for receiving support for providing such a service.

Introduction and background

The Protein Production Network Sweden (PPNS) is an informal network formed in 2013 between Swedish protein production facilities. The overall aim of the network is to make protein production methods and competence available to the Swedish academic life science community and also to share best practice and to work for best use of available resources within the network. One immediately recognized task for collaborative work was to give access to protein production in multiple expression hosts available for all researchers, something that today often is a bottleneck but a necessity for solving complex scientific questions. As different proteins are preferentially produced in different expression systems depending on their origin, cellular localization and potential modifications etc, such an expression screen could have a large impact on the success rate for each protein.

The network consists of all academic institutions in Sweden that presently have protein production facilities and includes representation from the Universities of Lund, Gothenburg and Umeå, Karolinska Institutet, Science for Life laboratory (SciLifeLab) and MAX IV Laboratory. MAX IV Laboratory participates in PPNS although it has no in-house protein production, as its protein crystallography operation is strongly dependent on reliable sources of high quality proteins. In addition, AstraZeneca participates as an industrial partner, with the aim of sharing their competence within protein production and also to provide open access to parts of its protein production facility. For full details of the centers involved in PPNS, see appendix I (p.5).

Pilot PPNS project: expression screen

As the individual facilities in the network have different expertise regarding recombinant expression systems and also have different remits (some facilities have a national remit while others are only available for researchers at their home university) a key point has been to explore how to make the combined competence of all centers available to academic researchers throughout Sweden.

As part of this effort it was agreed in February 2014 to undertake a pilot project within the network to carry out a multi-host parallel expression screen with a defined set of target proteins.

This represents a typical challenge facing researchers wanting to obtain sizeable amounts of their protein of interest for studying its structure, function or biological properties.

The objectives of the project were to:

- Test the feasibility of running this type of expression screen as a collaborative project between the facilities, specifically looking at timelines, cost, organization and documentation.
- Assess the throughput and service models of the different facilities.
- Identify the challenges and opportunities with such an approach by trying it out in practice.

The activities in the project are outlined below:

1. **Nomination of five different target proteins to be expressed.**
   Proteins were nominated by the different facilities and selected to represent several different sub-cellular localizations of human proteins, and also a bacterial protein, and are listed in Table 1. All proteins were relevant targets as there was an interest in them from different researchers in Sweden. Green fluorescent protein (GFP) was included as a control protein *(completed in May 2014)*.

Table 1. The target proteins included in the expression screen performed within PPNS in the fall of 2014.

Protein	Origin	Normal localization	Known modifications	Mw
Carbonic anhydrase IX (CAIX) ***(UniProt nr Q16790)***	Human	Secreted	Glycosylation	50 kDa
ADP-ribosyl-transferase (PARP6) ***(UniProt nr Q2NL67)***	Human	Cytosolic	Automodification, phosphorylation (?)	74 kDa
Proline dehydro-genase (PRODH) ***(UniProt nr O43272)***	Human	Mitochondrial	Not known	49 kDa (domain)
YjbH ***(UniPROT nr O31606)***	*Bacillus subtilis*	Intracellular	None	29 kDa
Zymogen granule protein 16 (ZG16) ***(UniProt nr O60844)***	Human	Secreted	None	21 kDa
Green fluorescent protein (GFP) ***(UniProt nr P42212)***	*Aquea victoria*	Intracellular	None	30 kDa

2. **Sub-cloning of targets in an agreed set of expression vectors.**
   The nominating sites were responsible for designing two variants of each protein; one full-length version and one version covering a domain within the protein. Sub-cloning and distribution of expression vectors to the respective facilities were performed by the Protein Science Facility (PSF) in Stockholm *(completed in June 2014).*

3. **Expression screen at the different facilities:**
   Each site expressed the different proteins in their expression system, according to table 2. Expressed proteins were analyzed by the individual centers by SDS-PAGE and/or Western blots *(completed in November 2014).*

Table 2. Protein production platforms and expression systems included in the expression screen.

Protein production platform	Acronym	Expression system
Protein Science Facility, Karolinska Institutet/SciLifeLab	PSF	*E.coli*
Mammalian Protein Expression Facility, University of Gothenburg	MPE	Mammalian cells (HEK293 and CHO)
Lund Protein Production Platform, Lund University	LP3	Insect cells/baculovirus (BEVS)
Swedish NMR Centre, University of Gothenburg	NMR	Cell free translation
Protein Expertise Platform, Umeå University	PEP	*E. coli*

Note: Two protein production centers within the network did not engage in the expression screen. The Protein Expression and Characterization facility within the Drug Discovery and Development Platform of SciLife Labs did not, as its remit is strictly limited to drug development activities within the projects run within the platform. Astra Zeneca did not participate due to their limited possibilities for performing services to academic researchers.

4. **Summary of data and writing of report**
In the next section, the summary and conclusions from this study are presented, followed by an appendix where the individual centers describes in detail how the work was performed (Appendix II, p. 6) *(completed in December 2014).*

Results and conclusions

In this study, the Protein Production Network Sweden (PPNS) performed a joint expression screen with the objective to demonstrate how the different protein production platforms could work together to help Swedish researchers to the best possible outcome for the expression of their protein of interest. Five proteins of varying origin were selected for the screen that included expression in *E. coli*, insect cells, mammalian cells and a cell free system. Vectors encoding either the full-length protein or one domain thereof were constructed and each gene was inserted into expression vectors suitable for the respective expression systems. The screen was then performed at the different platforms in the network.

Results were presented and summarized at a network meeting in November 2014. Details from each facility can be found in appendix II (p.6). As a conceivable deliverable a summarizing table (Table 3) was put together in which each participating site judged the possibilities of a successful outcome in a scaled-up protein production, based on the small scale screening result. Together with the detailed data and with help from PPNS, a potential user of the service can get an overview and decide how to proceed with the project. Briefly, all the selected proteins could be expressed in at least one system tested, and the screen helped to find suitable expression systems for each protein.

Table 3. Overview of results from expression screen performed by five protein production platforms within the Protein Production Network Sweden

	CA IX		PARP6		ZG16		YjbH		PRODH	
	Human secreted		Human cytosolic		Human secreted		*B. subtilis* intracellular		Human mitochondrial	
	full length	domain	full length	domain	full length	domain	full length	domain	full length	domain
Mammalian cells (MPE)	🟢	🟢	🟢	🟡	🟢	🟢	🔴	🔴	-	🟢
Insect cells (BEVS) (LP3)	🟢	-	🟢	-	🟢	-	🔴	-	-	🟢
Cell free translation (NMR)	🟢	🔴	🟡	🟡	🔴	🔴	🟢	🟡	🔴	🟡
E. coli site 1 (PSF)	🟢	🟢	🔴	🟡	🔴	🟢	🟡	🟡	🔴	🟢
E. coli site 2 (PEP)	🟢	🟢	🔴	🟡	🟢	🟢	🟢	🟢	🟡	🟢

🟢 facility happy to take on production and optimistic about outcome
🟡 facility happy to discuss continued work but less optimistic about outcome
🔴 facility pessimistic about outcome
- not performed

In this work, only a non-cleavable C-terminal 6xHis-tag was used for all proteins/domains. This is a simplification and in many cases different tags and several positions of the tags would be tested to find a construct that is well produced. This was however not tested in this screen. Further optimization of the protocols may also be needed depending on the expression system and the amounts of protein needed, before an actual production culture could be performed.

As for the evaluation of data, additional work will be required to be able to correctly compare expression levels consistently across the platforms. In addition, the verification of the functionality of the expressed protein to assure that it is fit for purpose was not included in this study. This would normally be done by the researcher/user but may need to be built in to the process.

With respect to timelines no conclusions should be drawn from the time taken for this study but it could be recognized that for an expression screen to be effective a reasonable timeline would be 6-8 weeks, including vector constructions. This would require a strict planning and prioritization of the work across all platforms.

In the summaries from the centers in appendix II, some prices are listed. There has until now not been any harmonization of prices for the different facilities as each facility has its own rules and practices depending on its mother institution and funding. The prices can therefore be hard to compare directly. Going forward, PPNS envisions price levels that are similar for all academic users, independent on whether they are “internal” or “external”. This may be accomplished by some joint external funding for the network.

In conclusion, we have demonstrated the concept of a multi-host expression screen across the protein production platforms in Sweden. This leads to an opportunity to, with a relatively small effort, give scientists easier access to complementary techniques in general and more specifically to additional expression hosts. We believe that this type of service would increase the competitiveness of Swedish research groups. All sites participating in this pilot project are positive to continuing this type of joint activity.

Should the PPNS concept be pursued, future work should focus on addressing the comparability of data, demonstrate relevant timelines and to adjust pricing, in order to make the PPNS a viable resource to Swedish academic researchers by providing access to a protein supply network at a national level.

Appendices:

Appendix I. List of current protein production platforms within PPNS

Platform[a]	Acro-nym	Host institution	Present funding	Expression system	Open for users (presently)
Protein Science Facility	PSF	Karolinska Institutet	Karolinska Institutet/SciLifeLab/ SWEDSTRUCT	*E. coli*	National
Mammalian Protein Expression Facility, University of Gothenburg	MPE	University of Gothenburg	University of Gothenburg	Mammalian (CHO and HEK293)	National
Lund Protein Production Platform	LP3	Lund University	University of Lund	*E. coli* and BEVS	National
Swedish NMR Centre	NMR	University of Gothenburg	University of Gothenburg/KAW/ SWEDSTRUCT	In vitro	National
Protein Expertise Platform	PEP	Umeå University	Umeå University	*E. coli*	Mainly local
The Protein Expression and Characterization facility within the Drug Discovery and Development Platform of SciLifeLab	DDD	Karolinska Institutet	SciLifeLab	Several	National but only for drug discovery projects within the platform
AstraZeneca		AstraZeneca			No external services

[a]**Web addresses of the PPNS Participating platforms:**
PSF: psf.ki.se
MPE: cf.gu.se/mammalian_protein_expression
LP3: www.lu.se/lp3
NMR: nmr.gu.se
PEP: www.pep-umu.se
DDD: www.scilifelab.se/platforms/drug-discovery-and-development/

Appendix II. Detailed information and results from the PPNS expression screen, from the different centers

Sub-cloning of genes into expression vectors suitable for the different expression systems included in the screen

Sub-cloning of the genes used in the expression screen into appropriate vectors was performed in parallel at PSF, using ligation independent (LIC) cloning. Starting from user-provided DNA templates, complementary ssDNA overhangs were generated on PCR products and cloning vectors by T4 DNA polymerase. After annealing and transformation into *E. coli* the presence of a correctly sized insert was verified by colony PCR. The method is robust and enables parallel cloning of many constructs in a 96 well format.

Four different expression vectors were used; pNIC-CH2 for work in *E. coli* and cell free, two plasmids for insect cells, pVL1392-CH2 (compatible with OET's flashBAC system) and pFB-CT10HF-LIC (compatible with Invitrogen's Bac-to-Bac system), and finally pcDNA3-CT10HF-LIC for mammalian hosts. The cloning vectors were all adapted to LIC cloning and add a C-terminal 6xHis-tag and in some cases also a FLAG-tag for affinity purification or for detection in Western blots.

40 out of 44 constructs were successfully cloned and confirmed by sequencing at Eurofins MWG using appropriate primers before distribution to the respective facilities. Three of the failed constructs corresponded to full-length variants of target PRODH and one to a GFP control vector for the insect cell system.

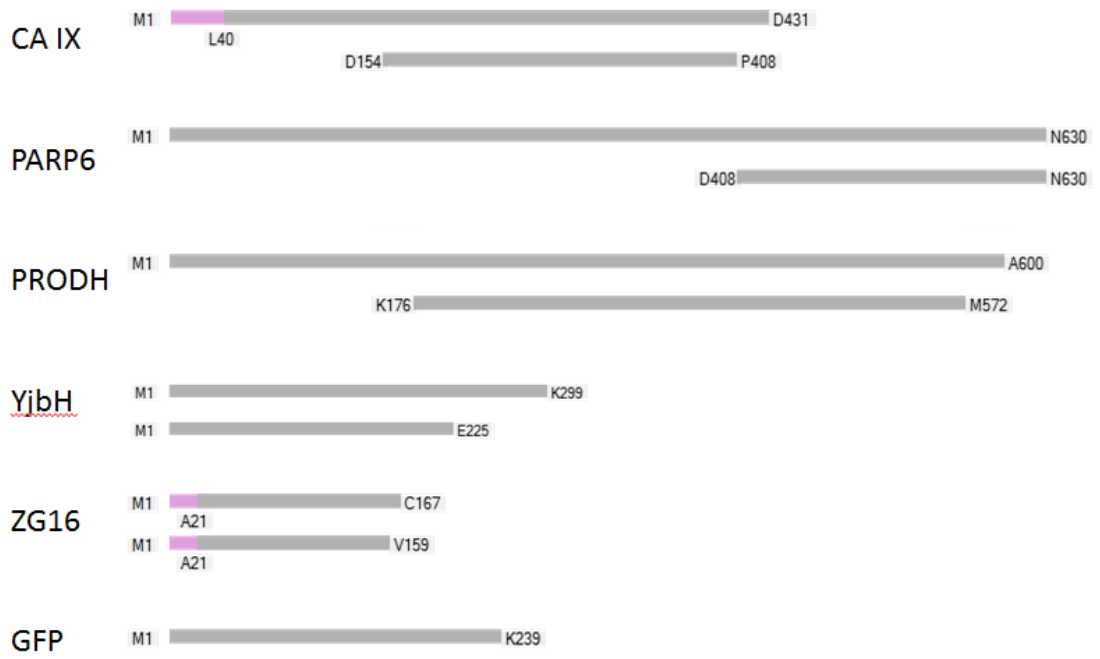

Figure 1. Graphical representation of the recombinant proteins included in the expression screen. Signal peptides are shown in purple.

Expression screen in *E. coli,* performed at PSF, Karolinska Insitutet/SciLife Lab

At the Protein Science Facility the constructs in plasmid pNIC-CH2 were transformed into competent E. coli BL21(DE3) T1R pRARE2 cells. Expression clones were cultured at 37°C in 1 ml Terrific Broth, down tempered to 18 °C and induced with 0.5 mM IPTG. Expression continued over-night. Harvest of the cells was followed by lysis and IMAC affinity purification utilizing the 6xHis-tag. Finally the levels of both totally expressed and purified protein were analyzed on Coomassie stained SDS-PAGE. The bands on the SDS-PAGE were judged according to a well-established scoring system in order to predict the success in a production scale experiment and also compared to the internal PPNS standard (a His-tagged TEV protease).

The short versions, the domains of PRODH, ZG16, and CAIX all qualify for a standard protein production run. PARP6 and YJHB can also be tested with methods taking the low abundance into account and with lower expectations on the result. In addition a promising result was found for the full-length version of CAIX when re-running the screen in 1 ml auto-induction media. Surprisingly only one more full length construct was detected, FL-YJHB was found in the total fraction but was not seen in the soluble fraction. For the un-successful variants different optimization methods as well as other expression systems should be considered, all depending on the downstream application and the user's needs.

The time line for expression and solubility screens at PSF is one week if suitable expression plasmids are available. In November 2014 the cost for testing one construct as part of a standard monthly screen is SEK 600.

PSF also offers the sub-cloning into a range of expression plasmids as a service. Cloning sessions are usually initiated monthly and take approximately three weeks. The cost for a single construct performed in triplicate as part of a standard cloning session is SEK 2500 and for a package of 10 variants (e.g different truncations) of one gene SEK 7500 (November 2014).

Expression screen in *E. coli,* performed at PEP, Umeå University

The constructs received from PSF were transformed into Rosetta (DE3) and Arctic Express (DE3)_RIL (contains copies of the argU, ileY, and leuW tRNA genes). All Rosetta (DE3) constructs were cultivated and induced in auto-induction medium with the appropriate antibiotics, the cultures were left at 37˚C for 3h and thereafter transferred to 20˚C over-night. The Arctic Express (DE3) _RIL were grown in LB at 30˚C for 4h, then the temperature was lowered to 12˚C and the cells induced with 1mM IPTG. The bacteria were collected and lysed by sonication. IMAC was performed on the lysate utilizing the 6xHis-tag. Expression levels and quantities of soluble protein were evaluated with SDS-PAGE.

Soluble proteins could be detected at large quantities for full length YJBH, ZG16 and CAIX when expressed in Rosetta (DE3). Lesser amount of the full length PRODH could also be expressed and partially purified. All the short domains were expressed in Rosetta (DE3), and with the exception of the PARP6, all could be purified as soluble protein. The expression and solubility pattern in Arctic Express (DE3)_RIL cells was similar to what was observed in Rosetta (DE3). However, in Arctic cells it was possible to express and purify the short version of PARP6 as a soluble protein.

Both variants of ZG16, PRODH, CAIX and YJBH, all qualified for a standard protein purification from Rosetta (DE3) although PRODH need to be purified at a larger scale than the others. The short domain of PARP6 could be purified from Arctic Express (DE3) _RIL. An

approach to improve expression and solubility of the full length PARP6 would be to add a cleavable fusion partner such as MBP or GST. This strategy might also improve the expression of the more poorly expressed constructs.

The time line for an expression and solubility screen is one week if the customer provide us with a ready to go vector. Additional optimizations using fusion partners and different *E. coli* expression strains are available to further enhance expression/solubility. Scaled up expression and purification is also available.

A small scale expression screen costs 100 SEK/setup for an internal user, with a higher price for external users.

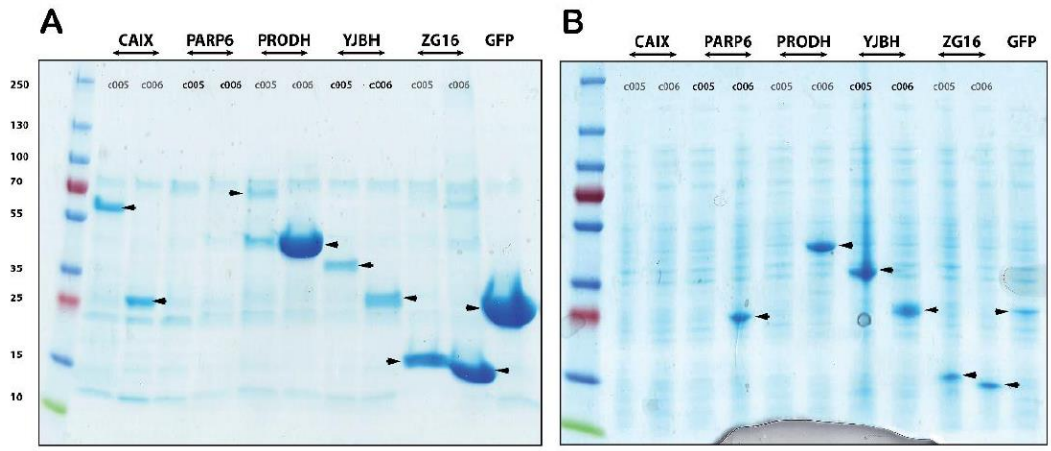

Figure 2. Example of a result from the expression screen, performed in *E. coli* and analyzed by SDS-PAGE with Coomassie stain. A. Purified protein (IMAC enriched) and B. Total expression. Expressed proteins at the expected target size are indicated with arrows.

Expression screen in a cell-free system, performed at the Swedish NMR centre, Göteborg

Constructs in pNIC-CH2 was received from PSF, consequently transformed into *E. coli* OmniMax 2 T1^{R} cells and then isolated from 50 ml LB ON culture with a QIAGEN PlasmidPlus kit. Plasmid was thoroughly washed with PE buffer while on-column before eluting with DEPC-treated mQ water. Cell-free reactions of the PSF constructs, a negative control (–DNA) and our usual positive control were set up in 200 µl scale. Reactions were run at 30°C, 800 rpm for 2h on an Eppendorf thermomixer. Total and soluble samples for Western blot were taken for each reaction. Western blot band intensities were estimated with ImageJ using the built-in gel analysis procedure, integrating the relevant band peak for the samples with visible intensity in the soluble fraction.

Expression was detected for all but the GFP construct, but the full-length PRODH, truncated CAIX and both ZG16 constructs barely produced detectable proteins. The reason for low expression of ZG16 was likely the presence of an N-terminal signal peptide in the constructs – translated protein is potentially stuck on ribosomes in the absence of a functional membrane-bound signal peptidase activity in the cell-free system. The full-length constructs of YjbH and CAIX should be suitable for scale-up expression. For projects chosen to move forward, a temperature and additive screen is usually performed before scaling up.

To implement cell-free protein synthesis more readily into a national protein production setting, it would be desirable to add our current first-choice expression vector pCPR0012 to the repertoire of PSF. pCPR0012 is a high copy number vector designed for ligation

independent cloning (with the same LIC overhangs as *e.g.* pNIC-CH2), encoding a 6xHistag followed by a TEV protease site N-terminal to the gene of interest.

The cost for screening expression entails only the consumables. The cost of a small-scale reaction screen as described above with up to 8 constructs/condition, if suitable plasmids are supplied, is 750 SEK. The price includes solubility assessmentbased on Western blot intensities as above. The timeline for delivery of screening data is three weeks.

Expression screen in BEVS & insect cells performed at LP3, Lund University

All plasmids were received from PSF in suitable amounts to be used directly.

The gene of interest had been either subcloned into pFB-CT10HF-LIC (plasmid received from Addgene) and compatible with the Bac-to-Bac System (Invitrogen) or pVL1392-CH2 (pVL1392 from BD and adapted for LIC-cloning by PSF. Not all constructs were available in both sets of donor vectors.

As planned, first the pFB-CT10HF-LIC based donor vectors were used to make bacmids. However no virus was generated after transfection of these bacmids into Sf9 cells. As a fast alternative, a smaller number of pVL1392-CH2 based donor vectors (1 vector per construct and full-length constructs if available) were then used to create virus using flashbac viral DNA (OET). That means the expression of the domains of CA IX, PARP6, ZG16 and YjbH has not been tested.

All full-length proteins and the domain of PRODH (no full-length construct available for this protein) could be expressed, with the exception of full-length YjbH. All expressed proteins were found soluble and in the predicted localization. For the successfully expressed proteins we would recommend further work in the BEVS.

Overall conclusions from the screening:

- The pFB-CT10HF-LIC did not work in our hands.
- The screening protocol was designed to test if a particular protein could be made or not. Additional experiments are needed to be able to estimate expression levels.
- The costs for a parallel screen of 10 constructs would be 14 kSEK for Lund University researchers and 47 kSEK for scientists outside Lund University.
- The timelines would be 5 weeks if using the Bac-to-Bac system and 3 weeks using the co-transfection with flashbac viral DNA.
- LP3 could test up to 30 constructs in parallel.
- LP3 will work further on the development of more suitable donor vectors and an improved screening process in the BEVS with the aim to shorten times, costs and allowing a prediction about expression levels.

Expression screen in mammalian cells, performed at MPE core facility, University of Gothenburg

The expression of the ten different genes in the vector pcDNA3-CT10HF-LIC (that were obtained from PSF) was tested in suspension-adapted cells of both Chinese Hamster Ovary (CHO) cells and Human Embryonic Kidney (HEK) 293 cell lines. For this, more DNA was prepared by Macherey-Nagel Miniprep kit. The expression test was performed in small scale (5 ml; 1x10E6 cells/ml) in 50 ml bioreactor tubes. The cultures and transfections were performed in a shaker CO_2-incubator at 37°C. The transfection reagents used were Nova CHOice (Merck) for CHO cells and PEI for 293 cells. Supernatants from and lysates of

transfected cells were prepared after 48h and were analyzed for the presence of the different protein constructs by Western blots using anti-His.

As expected, only the proteins containing a signal peptide for secretion (CAIX full-length and the two ZG16 constructs) were found in the supernatants. These proteins were also found in the cell lysates, and so were also the constructs for PARP6 full-length, the domain of CAIX and the domain of PRODH. YjbH could not be detected at all or at very low levels and would probably not be suitable for scaling up. For the proteins expressed within the cells (not secreted), 293 cells gave better expression levels than CHO and would most likely be the choice for scaling up a production.

The time needed for performing this test was one week and a scale-up for production could be performed within a few weeks, depending on the scale needed. Transfections ≤300 mL can be performed in shake cultures in flasks, whereas larger scale productions would preferentially be performed in bioreactors with more controlled culture conditions. In all cases, additional time will be needed for subsequent purification of the proteins, which was however not tested in this pilot screen.

A small scale test transfection of one vector in two cell lines presently costs 3000 SEK including analysis, for an internal user (Nov 2014), with a higher price for external users from other universities. This is higher than for prokaryotic expression mainly due to higher reagent costs. Mammalian cell expression is mainly used for proteins that are known to be modified with for example a large number of glycans or human proteins rich in disulfide bonds, or other proteins that may be difficult to make, and the higher cost is then justified. Proteins with these modifications were however not included in this test screen.