Proteomics Based Process and Cell Line Development Applied to a Mammalian Therapeutic Enzyme

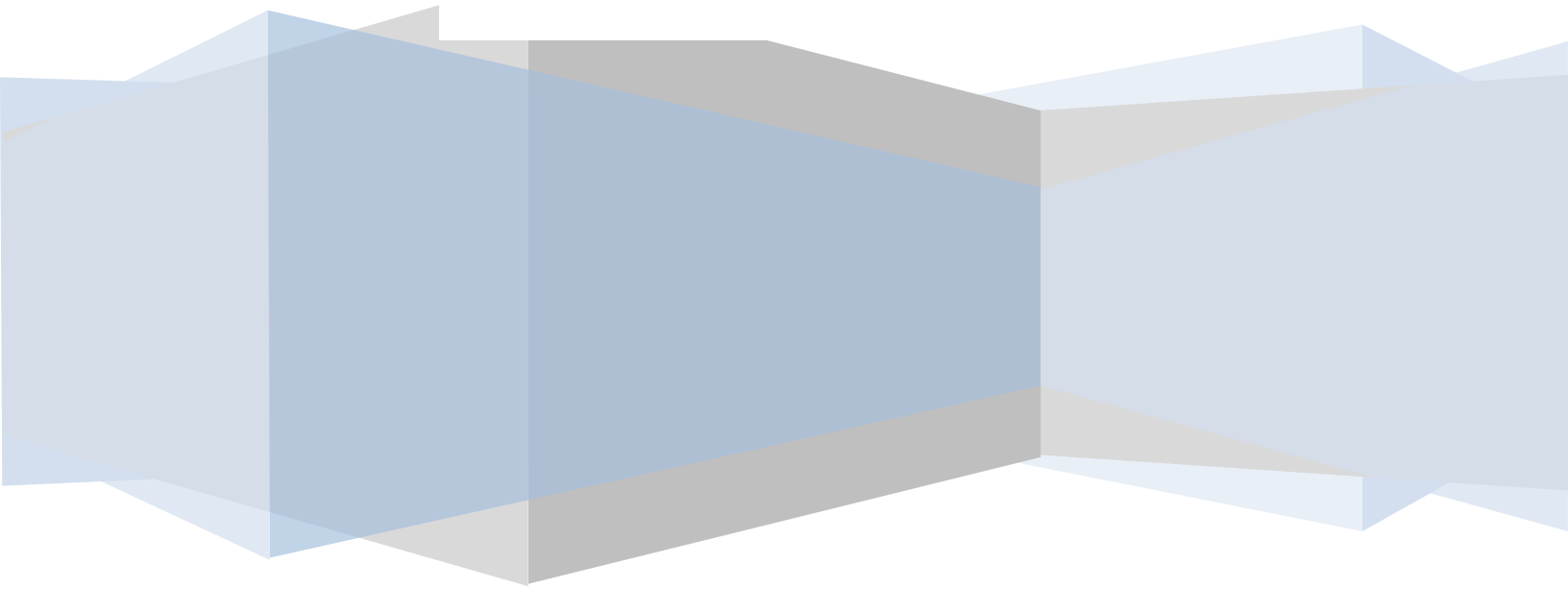
A thesis submitted to University College London for

the degree of Doctor of Biochemical Engineering

by

Damiano Migani

2017



Department of Biochemical Engineering, University College London

Bernard Katz Building, Gower Street London WC1E 6BT

Declaration

I, Damiano Migani, confirm that the work presented in this thesis is my own. Where information had been derived from other sources, I confirm that this has been indicated in the thesis.

Table of Contents

Table of Figures

Table of Tables

1 Abstract

Recombinant human Acid Alpha Glucosidase (GAA) is the therapeutic enzyme used for the treatment of Pompe disease, a rare genetic disorder characterised by GAA deficiency in the cell lysosomes. The manufacturing process for GAA can be challenging, in part due to protease degradation. The overall goal of this project was to understand the effects of GAA overexpression on cell lysosomal phenotype and host cell protein (HCP) release, and any resultant consequences for protease levels and ease of manufacture. To do this we first generated a human recombinant GAA producing stable CHO clonal cell line and then developed a two-step bioprocess based on capture chromatographic step anion exchange (IEX) and intermediate hydrophobic interaction (HIC). The purity of GAA after HIC was determined via LC/MSMS to be above 80%. We then collected images of cell lysosomes via transmission electron microscopy (TEM) and compared the resulting data with that from a Null CHO cell line. TEM imaging revealed 72% of all lysosomes in the GAA cell line were engorged indicating extensive cell stress; by comparison, only 8% of lysosomes in the Null CHO had a similar phenotype. Furthermore, comparison of the HCP profile among cell lines [GAA, mAb and Null] capture eluates, showed that while most HCPs released were common across them, some were unique to the GAA producer, implying that cell stress caused by overexpression of GAA has a molecule specific effect on HCP release. Protease analysis via zymograms showed an overall reduction in proteolytic activity after the capture step but also revealed the presence of co-eluting proteases at approximately 80 KDa, which MS analysis putatively identified as dipeptidyl peptidase 3 and prolyl endopeptidase.

2 Introduction

2.1 Impact statement

The work contained within this thesis has addressed the manufacture of therapeutic protein representative of the challenges faced in manufacturing such medicines.

These medicines are complex and labile so that bioprocess development times and costs tend to be high due to unforeseen issues that occur during scale-up of the manufacturing processes. Currently there is little scope to alter a manufacturing processes because we cannot readily predict the effects that such changes will have. This research will allow researchers in the area to gain a better prediction of the likely routes for manufacture of medicines. The tools here described, will enable the biotechnology industry to design more efficient bioprocesses that fit to available as well as emerging manufacturing platforms.

Bioprocess engineering offers a remarkably high return on investment to improve the state of manufacturing, with worldwide revenues of ~£110 billion in 2009. In comparison to other areas of manufacturing, bioprocess engineering is unusual in several respects. Typical product quantities are small (~250 kg/year), but are manufactured to extremely high purity and quality specifications (impurities < 0.001%). If the product does not meet specifications, no rework or reprocessing is allowed. Continuous improvement efforts (such as the popular 6s initiative) cannot be applied directly due to the high variability intrinsic to bioprocesses; it is estimated that the best that could be achieved is 3-4s. These sources of variability have led to highly regulated manufacturing, whose dictum is that “the process is the product”. No significant change can be made to a licensed manufacturing process without detailed and time-consuming review by the regulatory authorities. Developing and validating a bioprocess for manufacture takes ~10 years at a cost of £800M.

2.2 Literature review

2.2.1 Biopharmaceutical industry

Pharmaceutical industry has its roots in the chemists and rudimental pharmacies dating as far back as the Middle Ages, however the pharmaceutical industry as we know today generated in the middle of 19th century. Merck and GlaxoSmithKline were possibly the first trademarked names in this industry and started their business selling alkaloids (firstly isolated by Emanuel Merck in early 19th century) and other chemicals such as laxatives.

Living microorganisms had been employed for thousands of years in the food industry but it was not until the early twentieth century that one single product derived from fungi had such a big impact in global health that its founder was awarded with the Nobel Prize in 1945 for it. The product was penicillin and its discoverer, Alexander Fleming.

Eight years later in 1953, Watson and Crick arguably made the discovery of the century, the DNA structure, and the first steps in proteomics started to be made. At this point, the scientific world started realising the potential of utilising human DNA sequences or genes and host living organisms to produce recombinant therapeutic proteins. A new branch of pharmaceutical industry was born: biopharmaceuticals. However, it was not for another two decades that the first recombinant protein made it to the market. In 1976 Eli Lilly, launched recombinant human insulin using *Escherichia coli*.

In the following decades more and more products followed using genomic engineering techniques, including recombinant Human Growth Hormone, (also known as Protropin®) in 1981 by Genentech (Genentech, 1985), recombinant hepatitis B vaccine by Valenzuela in 1986 (Fisher, 1986), recombinant blood clotting factor in 1992 (RECOMBINATE® Factor VIII) by (Baxter, 2011), and more.

Beginning in 1986 a new class of therapeutic molecules started being produced: monoclonal antibodies (mAbs). These proteins have the great advantage of being highly specific to an antigen, therefore they can be used to target specific cells in the human body, such as cancer cells, and take advantage of the patient's immune system to neutralise them, minimizing side effects deriving from other therapies such as chemotherapy or radiotherapy. The first mAb to be approved by the FDA was Muronomab-CD3 in 1986 for treatment of transplant rejection (Hooks et al., 1991). In the next two decades, mAbs sales have grown exponentially, due to multiple companies entering the market. Some of the blockbusters mAbs of the last two decades include Rituximab for treatment of non-Hodgkin lymphoma in 1997 (BiogenIdec/Genentech, 2012), Natalizumab® in 2004 for treatment of Multiple Sclerosis and Crohn's disease (BiogenIdec/Elan, 2012) and more.

Biopharmaceuticals also include other proteins construct such as enzymes, nucleic acids and living microorganisms such as viruses and vaccines.

2.2.2 Biopharmaceutical global market

The pharmaceutical industry has seen an exponential growth in the last two decades; especially in the last 8-10 years, the growth rate has been in the double digits. This can be attributed mainly to two factors: research and development investment increase and market expansion.

Recent developments in recombinant protein products using mammalian cell cultures and new fermentation techniques have pushed titer levels up to 10g/l and over as described by Huang et al (Huang et al., 2010), which has had a dramatic impact on amount of product per batch produced. This figure is of particular impact when compared with early 2000s industry averages of barely 1.5 g/l for CHO cell lines.

Although the number of new molecular entities (NMEs) approved by the FDA has been steady especially in the last fifteen years, R&D investments have grown steadily (Figure 2-1) which has meant that more and more block buster drugs have reached the masses.

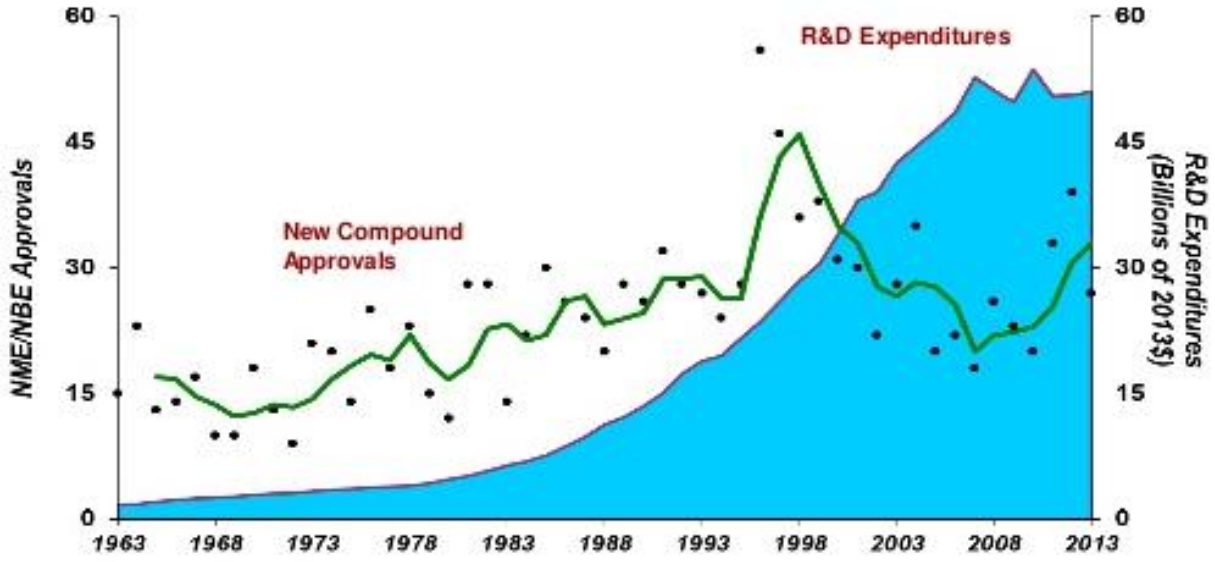


Figure 2-1 NME vs R&D expenditure. R&D expenditures are adjusted for inflation; curve is a 3-year moving average for NME/NBEs. Sources: Tuffs CSDD; PhRMA, 2014 Industry Profile, with permission.

Market penetration into booming economies has also played an important role in market growth. Reports from last year show that China alone has accounted for over $16bn in biopharmaceutical sales in 2009, number destined to increase by 4 fold by 2020 (China national Bureau of Statistics, 2012). India, Russia and Brazil also are also headed the same way.

2.2.3 Rare diseases: lysosomal glycogen storage disease type II.

2.2.3.1 Rare diseases

There is not a commonly accepted definition of rare diseases in the western world. According to the FDA, a disease is to be considered rare when “it affects less than 200 thousand people in the US” (FDA, 2013), or 1 in every 1500 people. The European Commission on Public Health

defines rare diseases as “diseases, including those of genetic origin, that are life-threatening or chronically debilitating, which are of such low prevalence that special combined efforts are needed to address them” (EuropeanCommission, 2004).

There are over 6000 rare diseases recognised by the WHO, and although each one of them might be considered rare, the effect as a whole on the world population is definitely non-negligible.

International collaboration has proven necessary in treating such diseases, as sometimes the number of people with a particular disease in a particular country is so low that it is impossible to gather enough of them to run clinical trials, “There are just three cases of progeria in France,” says Paloma Tejada, communications director at the European Organization for Rare Diseases (EURORDIS). “You cannot organize clinical trials on that basis. You have to achieve critical mass and that means international collaboration”(WHO, 2012).

2.2.3.2 Pompe disease

Pompe is a rare hereditary progressive disease caused by a mutation on the GAA gene and its symptoms include muscular weakness especially to lower limbs, heart and skeletal muscles, which get worse with time and can provoke respiratory problems. The genetic transmission is autosomal recessive, meaning that the gene is not localised on sex chromosomes, therefore it is only transmitted when both parents carry the defective gene. In this case, at each pregnancy, only 25% of the times the new-born manifests Pompe’s symptoms (affected individual); in 50% of the cases only one defective gene is transmitted (unaffected carrier), and the remaining 25% of the cases, both good copies are transmitted (unaffected non-carrier).

People that suffer from this disease lack a specific enzyme, called lysosomal acid alpha glucosidase (GAA), which function is to convert glycogen stored in the muscle cells lysosomes into glucose, used by the organism for multiple functions. When glycogen cannot be converted, it accumulates and it damages the cells, affecting muscular functionality. Since glycogen in cells

is always stored in lysosomes, this disease is also called ‘lysosomal glycogen storage disease’ (LSD).

2.2.3.3 Available treatments

Current commercial treatments are based on an enzyme replacement therapeutic approach. Recombinant human GAA enzyme produced via CHO culture is administered intravenously to the patient. At the time of writing, only two molecules are licensed by the FDA for the treatment of Pompe disease. Both drugs are manufactured by Genzyme: Lumizyme® for adults and Myozyme® for paediatric patients. According to one recent FDA news release article, Lumizyme®, reserved for late on-set of the disease, “is believed to work by replacing the deficient GAA, thereby reducing the accumulated glycogen in heart and skeletal muscle cells”. “This drug is approved with a risk evaluation and mitigation strategy (REMS). It will only be available through a restricted distribution system called the Lumizyme® ACE (Alglucosidase Alfa Control and Education) program to ensure that it is used by the correct patient group”(FDA, 2010). The safety and effectiveness of Lumizyme® are based on clinical study in 90 patients, aged 10-70 years, with late on-set Pompe disease.

The only other approved treatment for Pompe in the United States is Myozyme®. This drug has been in short supply due to limited manufacturing capacity and is currently reserved for infantile Pompe form (8 years and younger) (FDA, 2010).

Bio Marin, a global pharmaceutical company based in Novato CA, specialises in enzyme replacement therapies for a number of rare diseases (BioMarin pipeline summarised in *Table 2-1*). In 2010 they purchased Zesters Therapeutics Inc. for $22 million, and acquired rights to ZC-701, an IGF2-GAA complex at the time in early development for Pompe disease (Leuty, 2010). This molecule is formed by a complex of insulin-like growth factor 2 (IGF2) linked to

GAA enzyme. IGF2 is believed to aid in the process of attachment and assimilation of the therapeutic enzyme into muscle cells.

Compound ZC-701, renamed to BMN-701 completed a phase 1/2 open label, multicentre, multiple dose escalation clinical trial, and was expected to publish results in December 2015 (ClinicalTrials.gov, 2013b). A parallel study of long-term effects of BMN-701 was also being conducted, with results expected to be published towards the end of 2016 (ClinicalTrials.gov, 2013a).

UPDATE: In June 2016, BioMarin decided to discontinue the clinical development of the BMN 701 (reveglucosidase alfa) Pompe program (BioMarin, 2016) even though safety and effectiveness had been established. This was due to economic unprofitability of the program.

Name	Disease	Causes	Symptoms	Incidence/ status
Naglazyme®	MPS VI	lack of the lysosomal enzyme N-acetylgalactosamine 4-sulfatase	- Severe skeletal dysplasia - Short stature - Motor dysfunction	1 in 75k Approved in EU US AU
Aldurazyme®	MPS I	lack of the lysosomal enzyme alpha-L-iduronidase	- Mental retardation - Retinal degeneration - Corneal clouding - Cardiomyopathy	1 in 100k. Approved in EU US AU
Kuvan®	PKU (Cofactor to PAH enzyme)	The lack of PKU causes accumulation of Phe (Phenylketonuria)	- Mental retardation - Seizures	1 in 10k Approved in EU,US, JP
Firdapse®	LEMS (Lambert-Eaton myasthenic syndrome)	Delays the repolarisation of nerve terminals after a discharge, allowing more calcium to accumulate in the nerve terminal	- Muscular weakness in lower limbs	Approved in EU, Phase III CT in US
BMN-111	Achondroplasia	growth factor receptor, which causes a change in the way cartilage mutates into bone tissue	- Dwarfism	Early Phase II CT
poly ADP-ribose polymerase	Repair damaged DNA	detection and initiation of repair of damaged DNA in cancer patients diagnosed with mutations on gene BRCA1 and BRCA2	- Breasts and ovarian cancer	Phase I CT
BMN 701	Pompe disease	lack of **acid alpha-glucosidase (GAA)** causes accumulation of glycogen with consequent muscle cell death	- Heart - Liver - Neurons - Myocytes	1 in 140k Phase II CT

Table 2-1. BioMarin Pharmaceuticals Inc. pipeline as of 12/2013 (BioMarin_Pharmaceutical_Inc., 2013)

2.2.4 The Product

2.2.4.1 Structure

"Human acid α-glucosidase (glycan 1,4 α glucosidase) is a typical lysosomal enzyme with substrate specificity to glycogen, maltose and isomaltose" (Rosenfeld, 1975). The enzyme is "most active at low pH (4.0 to 5.0) and hydrolyses both the α-1,4 and α-1,6 glycosidic linkages" (Jeffrey et al., 1970).

As shown in *Figure 2-2*, the gene in the human genome is located on chromosome 17, q21-23 (Martiniuk et al., 1985), and it counts 3636 nucleotide pairs.

Figure 2-2 Diagram of human chromosome 17 showing GAA gene location (Martiniuk et al., 1985)

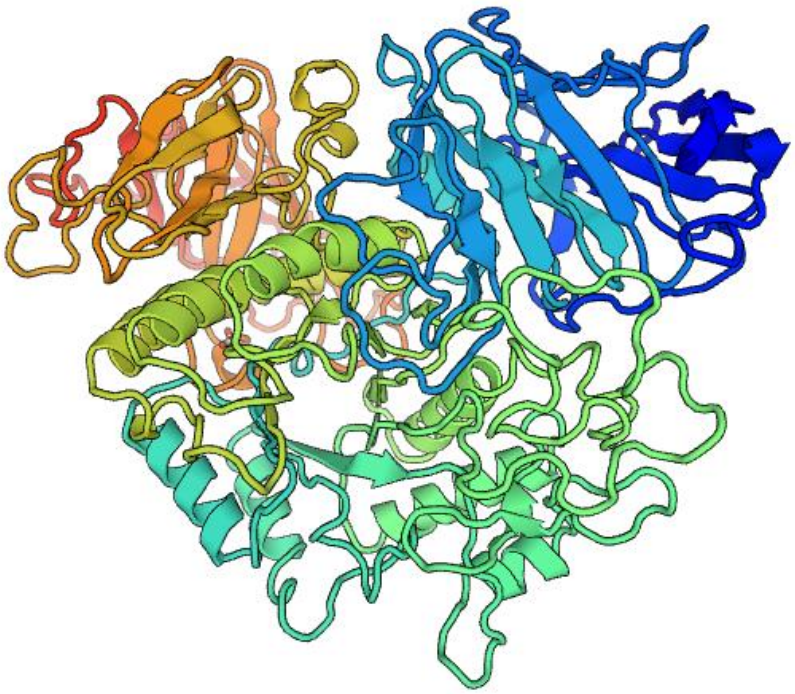
There are three common alleles of GAA in the human genome: GAA1, GAA2 and GAA4. GAA 1 is the most common, alleles GAA 2 and GAA 4 are much rarer (UniProt, 2012). The isoform of interest is the 110kDa isoform (952 aminoacids).

Despite the absence of crystallographic data of human GAA, a group of researchers from the Chinese Culture University of Taiwan (Ho, 2011), was able to "construct a structural model of human maltase-glucoamylase (MGAM) through homology modelling using the structural

information as a template". Due to the fact that there is a high similarity between the two molecules (GAA and template), they were able to overlay the active site region for both GAA and MGAM and the key catalytic residues. With the aid of molecular modelling software MIFit (MIFit, 2010) and Molecular Operating Environment (MOE, 2012) it was possible to draw the 3D structure of GAA and characterise it.

Results show that GAA is composed by five domains (*Figure 2-3*):

"a trefoil type-II domain (res. 89-135), an N-terminal b-sandwich domain (res. 136-346), the catalytic (β/α) barrel domain (res. 347-723), a proximal C-terminal domain (res. 724-818) and a distal C-terminal domain (res. 819-952)" (UniProt, 2012).



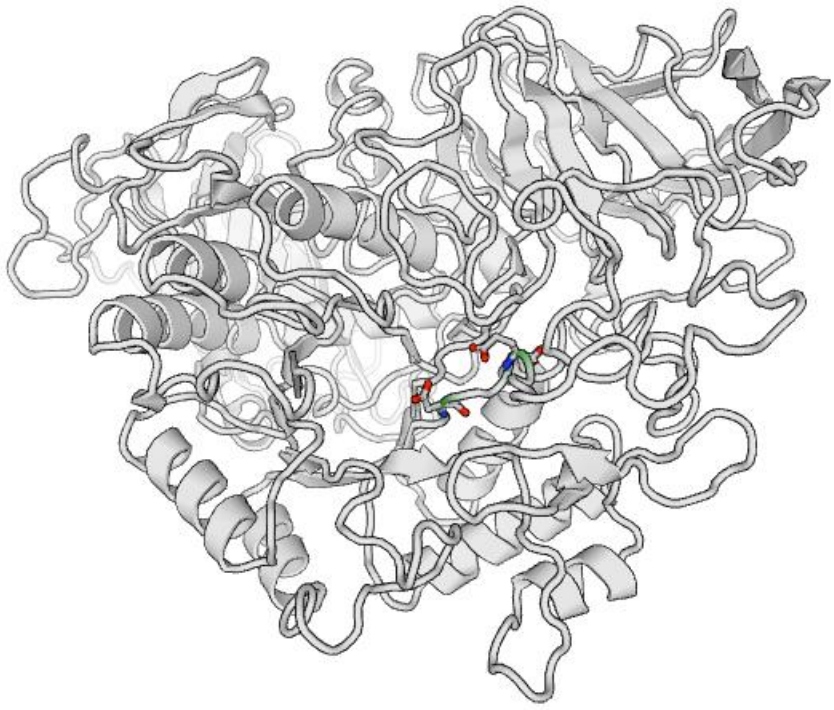


Figure 2-3. GAA 3D Structure. (Top) Different domains are identified by different colors. (Bottom) Asp-518 active site highlighted. Source (UniProt, 2012)

The GAA catalytic site was further studied in the work of (Hermans et al., 1991). The behaviour of the enzyme in presence of active site-directed inhibitor conduritol B epoxide (CBE) was observed and measured in the presence of natural (glycogen) and artificial (4-methylumbelliferyl-α-D-glucopyranoside) substrates. This is based on the fact that "the inactivation of the enzyme is time and concentration dependent and results in the covalent binding of CBE". "Since catalytic activity is required for binding to occur, CBE-labelled peptides containing the catalytic residue of the lysosomal GAA were isolated and identified by micro sequencing and amino acid analysis". It was found that "based on sequence similarity and CBE binding, Asp-518 is an essential carboxylate in the active site of lysosomal GAA", which is "situated between residue Phe-512 and Glu-521" (Hermans et al., 1991).

2.2.4.2 Synthesis

In eukaryotic cells, the genetic information of a gene present on the DNA is transcribed to a messenger RNA (mRNA), which is then transferred to the cytoplasm and translated into polypeptides via ribosomes.

RNA polymerase II pairs complementary bases to the gene of interest on the DNA and catalyses the formation of pre-mRNA, a single stranded copy of the gene that can be transferred to the cytoplasm. Before the pre-mRNA is transported to the cytoplasm, it undergoes a series of modifications that transform it to mature mRNA. Firstly, a 5` cap is added to the front (5` end) of the mRNA shortly after the start of transcription. This is critical to creating the mature version of the RNA and it ensures stability of the molecule during translation. Secondly, the mRNA molecule is spliced, which is the process by which non-coding sequences, such as introns, are removed. Lastly, mRNA undergoes polyadenylation, which is the addition at the 3` end, of a sequence of multiple adenines that prevent degradation by exonucleases. The mature 3.6kb mRNA is therefore transported outside of the nucleus through nuclear pore, where translation occurs with the help of ribosomes and tRNA.

The polypeptide precursor chain of the GAA enzyme has a size of 100kDa. This molecule undergoes a series of post translational modifications that transform it into the mature active 110kDa and other isoforms. The synthesis of GAA was thoroughly studied by (Hoefsloot et al., 1988). In his work it was describes how this precursor “enters the lumen of the RER (rough endoplasmic reticulum), where glycosylation occurs, and this results in the 110kDa molecular specie” (Vanderhorst et al., 1987). “The 110kDa specie enters the Golgi complex where high-mannose type oligosaccharide side-chains are phosphorylated” (Reuser, 1985). Additional binding of the precursor to the “mannose-6-phosphate receptor ensures further transport to the

lysosomes, where mature GAA isoforms of 70 and 76 kDa are most abundant". GAA species of "105, 100, 95 and 91kDa have also been identified as processing intermediates" (Reuser, 1985).

The trafficking of lysosomal enzymes within cellular cytoplasm was also looked into by (Kornfeld, 1987). In his work, it was described how specific receptors are responsible for the movement of the enzymes through different organelles. The overall pathway is summarised in *Figure 2-4 Table 2-2* below.

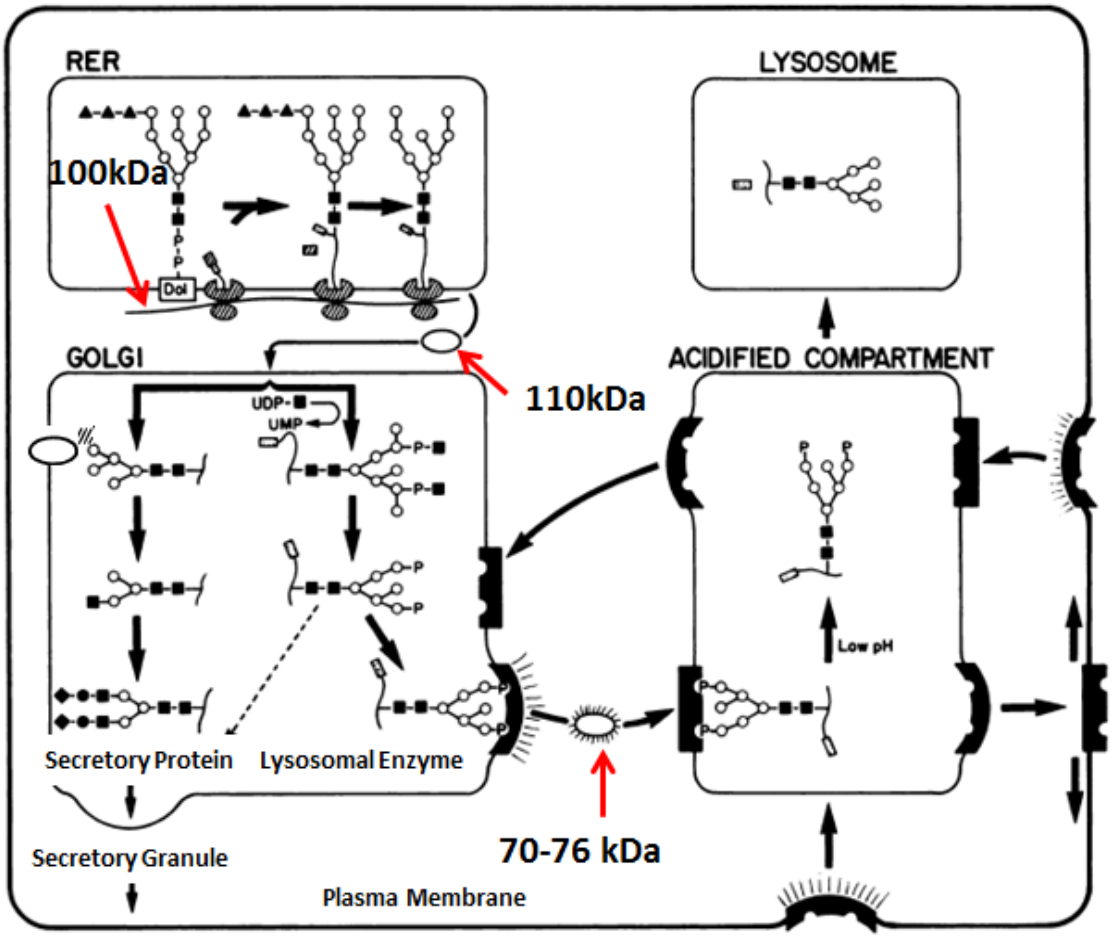


Figure 2-4 GAA Pathway with permission from The FASEB journal (Kornfeld, 1987).

Just outside the nucleus	- The hydrophobic amino-terminal region at the end of the neo-formed polypeptide interacts with a signal recognition particle, an 11S ribonucleoprotein. - Polypeptide is transported into the ER (endoplasmic reticulum).
RER	- Cotraslational glycosylation of selected asparagine residues. - A large preformed oligosaccharide [Dol*] is transferred from an intermediate to the nascent polypeptide. - The signal recognition particle is excised [ ] - Three glucoses and one mannose are excised from the oligosaccharide - The protein moves to Golgi apparatus by vesicular transport [] [Dol*]: three glucose [], nine mannose [] and 2 N-acetylglucosamine [ ] residues.
Golgi	- Two enzymes, a phosphotransferase and an acetylglucosaminidase, add phosphomannosyl [] residues to the polypeptide. - The resultant phosphomonoester serves as an essential component of a recognition marker that leads to high affinity binding to mannose-6-phosphate receptors (MPRs) in Golgi. [] - In this way, lysosomal enzymes are segregated from other proteins destined to secretion. - A small percentage of produced enzyme (10 to 20%) is secreted out of the cell through secretory granules []. - This complex exits the Golgi via coated vesicle [] and is directed to a pre-lysosomal compartment.
Prelysosomal Acidified Compartment	- The complex enzyme-MPR enters a prelysosomal acidified compartment where the low PH facilitates dissociation of the enzyme from the receptor. - The receptor returns to Golgi to pick up a new enzyme. - Some secreted enzyme renters the cell via coated pits [], and ends up in the prelysosomal compartment.
Lysosome	- The finished enzyme moves from the prelysosomal acidified compartment to the lysosome.

Table 2-2 Transport mechanism of lysosomal enzymes in human fibroblasts. Adapted from(Kornfeld, 1987) (Hoefsloot et al., 1988, Vanderhorst et al., 1987, Reuser, 1985)

2.2.4.3 Mechanism of action

The GAA molecular mechanism in human cells was studied by (Chiba, 1997). In this work the activity of α-glucosidases and glucoamylases was compared and it was observed that essentially the two are distinguished by releasing α or β glucose respectively. “These enzymes are exo-type α-glucosidic O-linkage-hydrolases releasing D-glucose from the non-reducing end of the substrate” (Chiba, 1997).

According to Chiba’s work, reaction of GAA occurs by splitting of the bond between the anomeric carbon of the glucosyl residue and glucosidic oxygen (C_1-O). The glucosyl residue is replaced by a proton from water or an acceptor, namely an exchange reaction between the glucosyl residue and the proton in both hydrolysis and transglucosylation. The reaction and mechanism of action are shown in *Figure 2-5*, *Figure 2-6* below.

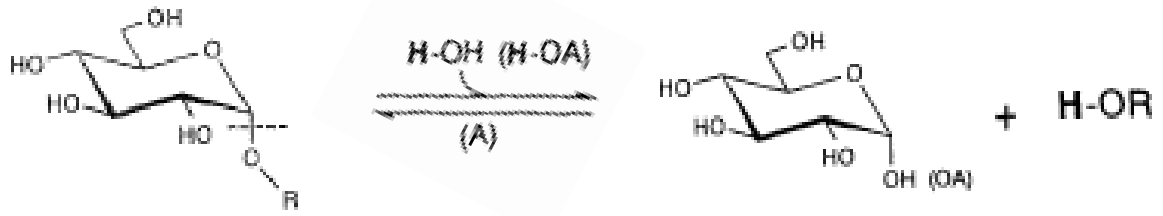


Figure 2-5 Hydrolytic and Transfer Reaction Scheme for α-glucosidase. R. reducing side sugar; H-OA. Acceptor; ----. Cleavage point. (Chiba, 1997)

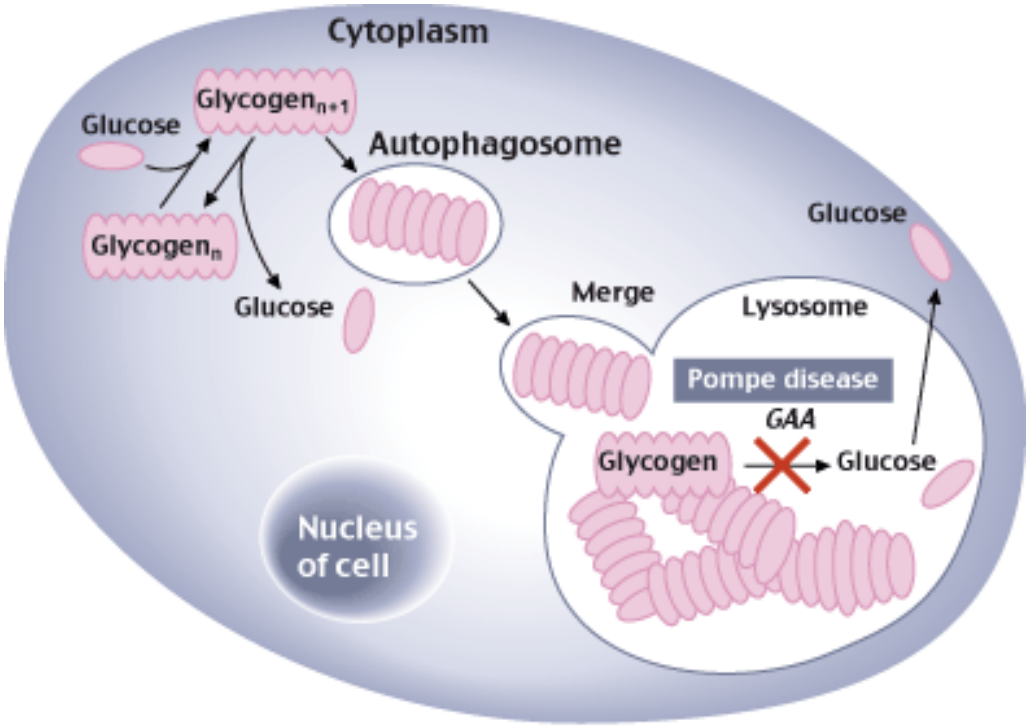


Figure 2-6. GAA mechanism of action. Source: The Swedish Information Centre for Rare Diseases

The molecule under BioMarin development (BMN-701) is a fusion protein of GAA and glycosylation-independent lysosomal targeting (GILT) tag, which contains a portion of insulin-like growth factor II, to create an active, chimeric enzyme with high affinity for the cation-independent mannose 6-phosphate receptor. This compound showed to be taken up by L6 myoblasts more efficiently that recombinant human GAA (Maga et al., 2013).

2.2.5 Production of material

2.2.5.1 Cell cloning and selection

Generation of stable recombinant protein expressing mammalian cell lines was traditionally a long and tedious task, mainly due to the bottleneck operation of clone selection, however with the advent of high throughput and cell sorting technologies this has become a much more stream

lined procedure such that stably expressing clones can typically be established in 3-4 months (Browne and Al-Rubeai, 2007). (Yoshimoto et al., 2013)

One of the first steps in recombinant GAA production is to establish a stable CHO cell line. This is a very important step as it allows the generation of product in a reproducible and consistent fashion. The choice of bioreactor size and fermentation time and conditions, and size and type of instruments required for downstream processing will largely depend on the amount of protein that the cultured cell line is able to produce. “Expression systems can be bacterial, yeast plant insect and mammalian. Mammalian cells, mainly CHO, despite lower yields is the system of choice mainly because they are more similar to human cells in regards to post translational modifications and glycosylation patterns” (Browne and Al-Rubeai, 2007)

Generally, the process consists of cloning the gene carrying the sequence to the protein of interest into a plasmid vector. The vector carries resistance to some antibiotic in order to allow for selection of clones that have not acquired the gene. Plasmid is then incorporated into the cell via the use of lipid transfection agent such as PEI (Polyethylenimine) or physical aids such as electroporation. Once the genetic material is inside the cell’s nucleus, it is replicated like the rest of genetic material. In a few occasions, the vector will actually be incorporated within the host DNA in which case a much more stable product production is obtained. All cells will have different expression levels because the way the vector is incorporated into the host DNA is random. Sometimes it will be placed into non coded regions of the host genome, some others in more active regions.

At this point cell line selection can start. In order to select the best clone among all, usually a specific antibiotic firstly is added to the mix. All cells that have incorporated the vector carrying the gene codifying for the resistance to that specific antibiotic will survive, all others will die. To now select among the surviving cells the best clone, the limiting dilution cloning (LDC)

technique is executed. This technique consists of preparing in a 96 well microtiter plate, 200µl dilutions of the cell broth in a way that in each aliquot there is chance of having one or less cells. Wells that show growth are those in which one cell carrying the resistance and the target protein was able to divide. Usually this process is done twice to ensure that the colony truly derives from one single cell. "A clone of cells derives from one single cell. This is necessary to ensure that all organisms in the culture have the exact identical genetic pool and will behave identically" (Browne and Al-Rubeai, 2007). At this point all wells are screened to find out the highest producing clone, usually using an immunoassay. The content of the well is expanded. Lastly, aliquots of this 'parent' cell line are banked for future use.

In *Table 2-3* below we summarize some of the more common cloning and selection techniques.

The process of CHO cell line development involved the use of human GAA ready transfer construct from (Origene, 2013), pcDNA/3.1 and pcDNA5/FRT vectors from Thermo Fisher *Figure 2-7* and took place at University of Kent.

There are many challenges associated with the production of a recombinant product in CHO host, one being that the transfection of a plasmid into the CHO genome is a rare random event and therefore a very low number of clones will present the desirable features of high titre and good quality product. Data shows that this procedure is very inefficient, as low as 1 in 10000 clones will have favourable features (Gorman and Bullock, 2000). The reason for this is known as the 'position effect' (Huang et al., 2007): as a rare random event, the odds that a plasmid integrates in a highly transcribed section of the host genome (away from chromatin or other less-transcribed areas) are very slim. Current practice for the selection of high producing clones involves a time-consuming and laborious screening strategy due to the fact that the random integration of the transgene of interest gives rise to very different expression levels due to

variation in the chromosomal surroundings or unpredictable cross-interactions of multiple gene copies (Wolffe, 1998)

Several strategies have been proposed to overcome the positional dependency of random integration. One of the most promising methods is the use of homologous recombination system for site-directed recombination such as the Flp-in/FRT system from *S. cerevisiae* (Li et al., 2010). Notable work in this field has been done by Laboratory of Protein Engineering of Institute of Biotechnology in Beijing, China (Zhou et al., 2010). In their paper, it was described how they used the pcDNA5/FRT vector in conjunction with pMCEscan selection system to rapidly generate recombinant proteins without cloning.

In this work we used the Invitrogen developed Flp-In FRT system (Cat# K601001, Thermo Fisher) to shorten the clonal screening process due to its 'site-directed recombination', which facilitates integration of the gene of interest into a specific highly transcribed FRT site in the host genome, and allows all clones generated to produce considerable and comparable levels of recombinant product.

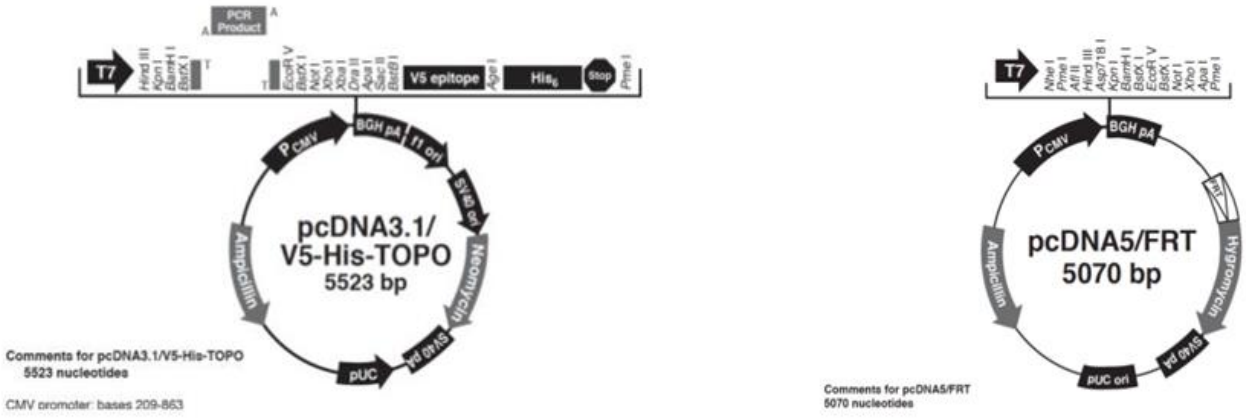


Figure 2-7. plasmid vectors pcDNA 3.1 and pcDNA5/FRT (Invitrogen, 2013a, Invitrogen, 2013b)

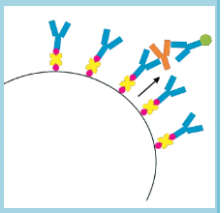
Cloning	
Limiting Dilution Cloning (LCD)	Simple, low cost. Slow (6-8 weeks to reach viable clone). Done on 96well microplate, suspension diluted so that less than 1 cell in a 200µl aliquot. Aliquots are placed in plate wells. Based on assumption that colonies in each well come from 1 single cell (not always valid). Colonies that proliferate successfully are isolated and tested for productivity using enzyme linked immunoassay (ELISA).
Surface Protein expression	Based on correlation between surface expression of recombinant proteins and levels of secreted product. Recombinant protein linked to a non-fluorescent reporter molecule (such as CD20). Proteins are lined by an internal ribosome entry site. (IRES), meaning that both are expressed by same mRNA, therefore relative expression levels but separate translation. Cells are analysed via flow cytometry for high level expression of reporter molecule with fluorescent anti reporter Ab.
Green Fluorescent protein (GFP)	Based on use of reporter molecule such as GFP. Several mammalian cell lines show correlation between GFP expression and recombinant production. Fluorescent based selection using flow cytometer.
Methotrexate	Fluorescently labelled MTX permeates cells and bind to dihydrofolate reductase (dhfr). The higher fluorescence, the higher are levels of dhfr Cells are transfected with a gene carrying dhfr, which confers resistance to MTX, and the gene for recombinant product. High producing cell lines are selected based on MTX resistance.
Gel Microdrop technology (Bohm et al, 2004)	Useful for cells that are not permeable to fluorescent molecules, or when this can affect product quality. Based on creation of viscous layer around the cell that traps expressed product. Cells are grown in biotinylated agarose medium, cell surface is covered in biotin (red dots), then with avidin (yellow crosses), and with an Ab against the product conjugated to biotin. Cells are incubated to allow them to secrete product, which is trapped on surface. Product is stained with secondary Ab conjugated with fluorescent label. 
Selection	
Laser-enabled analysis and processing (LEAP)	Cells are grown on capture matrix immobilises secreted protein, which is labelled with fluorescent Ab. Image is captured and software analyses image. Low producing colonies are eliminated using a pulsed laser beam.
Automated colony picker	ClonePix system from Genetix or CellCelector from Aviso Cells are immobilised in semi-solid medium and incubated to form colonies. Product is secreted and retained in cell vicinity. Halo of product is visualised by addition of fluorescent labelled capture Ab.
Cello system	High throughput selection of high producing clones, Automated advanced robotic system that handles medium handling, incubation and measurement of cell growth Transfected cells are fed into the machine, which seeds them automatically into plates Clonality is measured via integrated microscope Throughput 600-800 plates, thousands of cell lines can be screened.

Table 2-3 Common cloning and selection methods. Adapted from (Browne and Al-Rubeai, 2007)

2.2.5.2 Available GAA bio processes comparison

2.2.5.2.1 Generation of recombinant GAA in bacteria.

Due to the considerably faster and simpler fermentation of bacteria compared to mammalian cells, attempts have been made to produce human GAA using prokaryotic systems. A group of researchers from New York Medical Centre came to the conclusion that: The "resulting molecule had an antigenic reaction to polyclonal rabbit antibody to human GAA using ELISA and western techniques, however it was enzymatically inactive. This suggests that glycosylation may play a role in enzymatic function" (Martiniuk et al., 1992). The bioprocess is summarised in *Table 2-4* below.

Vector preparation	- **cDNA containing full human GAA gene, was subcloned at the *Hind*III site of expression fusion plasmid pMalp-RI and pMalc-RI.** - **The gene-plasmid constructs were sequenced with maltose binding protein (MBP) to generate a fusion protein.** - **TB1 and DH5 alpha bacterial strains were transformed with the constructs.**
Fermentation	- MBP-GAA fusion protein was generated in both strains that containing the vector. - Non-transfected colonies were eliminated by addition of NZCYM-ampicillin. - Induction by IPTG (0.3mM) and growth for additional 2 hr.
Primary Purification	- Harvest by centrifugation. - Localisation of MBP-GAA protein performed according to manufacturer's instructions (New England Biolabs). Protein found in insoluble fraction. - Protein was solubilised in 8M urea, followed by dialysis. - Fusion protein was digested with Xa protease to separate MBP from GAA.
Analysis	- Digested protein created a 106kDa and 40kDa bands on SDS PAGE, as expected from the 150kDa fusion protein. - GAA reacted with rabbit polyclonal antibody to human GAA by western blotting - GAA was not enzymatically active as shown by incubation with artificial substrate 4-methylumbelliferyl-α-D glucoside at pH 4.0.

Table 2-4. GAA bioprocess using prokaryotic systems. Adapted from (Martiniuk et al., 1992)

2.2.5.2.2 Generation of recombinant GAA in insect cells.

A few years after Martiniuk tried to produce human recombinant GAA in bacteria, another group of researchers tried a similar experiment using insect cells instead. Insect cells expression system was chosen because it "apparently gives post-translational modifications and folding similar to mammalian cells and relatively large quantities of enzyme can be produced in a relatively short period of time" (Wu et al., 1996). The study resulted in production of catalytically active GAA, but that was not efficiently taken up by Pompe patients' fibroblasts. This again proved that choice of host expression system has a crucial impact to correct enzyme activity. Although similar, differences exist between mammalian and insect cell post translational modifications. The bioprocess is summarised in *Table 2-5* below.

Plasmid construction	- **GAA cDNA was excised with *Eco* RI from specific vector from Reuser (Erasmus University) and ligated into pBacPAK8 plasmid, to create vecto pJW22.**
Transfection	- *Sodoptera Frugiperda* cells (IPLB-SD21) were cotransfected with pJW22 by lipofectin mediated transfection. - Cells carrying GAA gene propagate.
Purification	- Not specified.
Analysis	- Activity was assessed by measuring glucose released from maltose.
Uptake	- Fibroblasts from Pompe patients were incubated in medium with enzyme. - Poor uptake observed. This is due to differences in posttranslational glycosylation, phosphorylation and or proteolytic processing.

Table 2-5 GAA bioprocess using insect cells. Adapted from (Wu et al., 1996)

2.2.5.2.3 Generation of recombinant GAA in CHO cells

Van Hove *et al* (Van Hove, 1996) worked on the production of recombinant human lysosomal GAA in CHO cells and tested the enzyme activity in vitro on human fibroblasts and in vivo on

guinea pigs. It was demonstrated that the GAA produced using mammalian cells is active and up-taken by target tissue. This was due to the fact that post translational modifications to GAA from mammalian cells are crucial for correct enzyme activity. The bioprocess is summarised in *Table 2-5* below.

Plasmid prep	- **GAA sequence excised from cDNA of vector (pSHAF2) using Eco R1 and ligated to mammalian expression vector (pcDNA3).** - **Final plasmid contains GAA, DHFR (resistance and selection by MTX) and gene for resistance to Geneticin.**
Cell culture selection and transfection	- Cells + plasmid incubated 2 days with electroporation. - Cells cultured in Geneticin and MTX. - Cloning (limiting dilution method). - Cells lysed. Enzyme activity in supernatant tested. - Best 10 clones further grown in MTX. - Stability monitored for 3 weeks.
Primary clarification	- Harvest centrifuged - Supernatant concentrated and dialysed in Buffer A (50mM sodium phosphate, 500mM NaCl, PH6.5). - Precipitate discarded.
Chrom. 1	- Loaded on Concavalin A Sepharose 4B. - Washed with buffer A and eluted in buffer A + 1M methyl α-D glucopyranoside
Concentration and buffer exchange	- Eluate was concentrated (PM30) and dialysed with buffer B (20mMsodium acetate buffer, 25mM NaCl, 1mM EDTA, pH 5.2)
Chrom. 2 and concentration	- Clarified lysate loaded on Sephadex G-200 column. - Fractions containing different GAA isoforms collected. - Eluate concentrated using 10KDa cartridges.
Analysis	- Purity assessed by SDS-PAGE. - Activity assessed by measuring conversion rate from glycogen to glucose - Patient fibroblasts incubated with enzyme for 24 hours. - Guinea pigs were injected with GAA to observe enzyme localisation. Tissue analysed after 24 hours.

Table 2-6 GAA bioprocess using CHO. Adapted from (Van Hove, 1996). NOTE: Primary clarification, chromatography 1 and concentration/buffer exchange steps are from (Elferink et al., 1984) as described in Van Hove's article.

Enzyme activity was measured by assessing the conversion rate of glycogen and mannose to glucose. The precursor 110KDa form was found to have lower V_{max} and K_m than mixtures containing the more mature forms.

2.2.5.2.4 Isolation of GAA from human placenta.

The enzyme isolation and purification from human placenta was described in a 1985 article from (Reuser, 1985). The process is schematically summarised below in *Table 2-7* below.

Homogenisation	- **10Kg of Human placenta was thawed from -70, blended in a 1:1 (w/v) solution with 25mM NaCl, and pH adjusted to 5.2 using 2M HCl.**
Centrifugation	- Cellular debris was removed by centrifugation at 25000 g for 30min.
Filtration	- After filtering the supernatant through cotton wool, the PH was again adjusted to 6.6.
Concentration adjustment	- The solution concentration was adjusted to 10mM using 1M sodium phosphate and further clarified by centrifugation at 55000 g for 30 min.
Binding	- The supernatant was added to a 200ml Sepharose 4B in 10mM sodium phosphate and allowed to bind under continuous stirring overnight.
Wash and Elute	- The beads were collected and washed with 10mM sodium phosphate pH 6.6 and subsequently the proteins were eluted from the beads at RT with 1.5 litres of wash buffer + 1 M methylglucoside.
Concentration and buffer exchange	- The eluate was concentrated using Amicon hollow fibre and dialysed.
Chromatography	- The material was further concentrated and loaded on a Sephadex G200, in which the GAA affinity bound to the substrate.

Table 2-7 Human Placenta GAA purification

The final eluate had a 10000 fold purification factor and was active. The yield using this method however was of only 4mg of enzyme per Kg of human placenta.

2.2.5.2.5 Isolation of GAA from human urine.

The process of isolation of GAA from human urine was described in (Elferink et al., 1984) work. This is summarised in *Table 2-8* below

Collection	- **Adult male urine is collected and concentrated to less than 1% volume using hollow fibre.**
Centrifugation	- Concentrate was dialysed in Buffer A (50mM sodium phosphate, 500mM NaCl, PH6.5) and centrifuged at 15000 g for 15 minutes. - Precipitate discarded.
Chromatography 1	- Multiple batches (50) pooled and loaded on Concavalin A Sepharose 4B. - Washed with buffer A and eluted in buffer A + 1M methyl α-D glucopyranoside
Concentration and buffer exchange	- Eluate was concentrated (PM30) and dialysed with buffer B (20mM sodium acetate buffer, 25mM NaCl, 1mM EDTA, pH 5.2)
Chromatography 2	- Eluate loaded into Sephadex G-100 column. Eluted with buffer B (Q 12 ml/h).

Table 2-8 Human Urine GAA purification

Sephadex G100 allowed separation of the precursor form (110KDa) from the mature isoforms (70 and 76KDa). According to the authors, "separation is due to a combination of affinity chromatography and gel filtration", and also they affirm that "the molecular masses of the two forms differ sufficiently to allow separation by gel filtration"(Elferink et al., 1984).

The ability to separate the isoforms allowed the authors to further study and compare the properties of the two polypeptides. From this resulted that there is a "striking difference in the mannose 6-phosphate (M6P) content of the two forms. The high molecular mass form (110KDa) contains about 3.5 M6P residues/polypeptide chains. The two mature forms contain next to none M6P" (Elferink et al., 1984). This result was also confirmed by the results of (Gabel et al., 1982) work, and the more recent (Van Hove, 1996).

2.2.5.2.6 Summary table

Table 2-9 below summarises the production methods discussed in this section.

Recombinant human GAA produced in bacteria and insects cells was either inactive or not able to be taken up by human fibroblasts. This is due to substantial differences in post translational glycosylation.

Human GAA isolated from human placenta or urine was, as expected, active and functional, however this method would not be feasible for production in large scale. These two methods were reviewed as they gave important information about purification methods available.

The only really viable method involved the use of mammalian cells as hosts.

	Immunogenic	**Enzymatically active**	**Cell Up-take**	**Feasible for market production**	**Purification summary**
Bacteria cells	YES	NO	NO	YES	Not specified
Insect cells	YES	YES	NO	YES	Not specified
Human urine	YES	YES	YES	NO	Concavalin A Sepharose 4B Sephadex G100 column
Human placenta	YES	YES	YES	NO	Binding onto Sepharose 4B Eluate loaded on Sephadex G200
CHO cells	YES	YES	YES	YES	Clarified broth loaded on Concavalin A Sepharose 4B Eluate concentrated and dialysed. Eluate loaded on Sephadex G-200 column Isoforms divided in fractions and concentrated

Table 2-9 GAA production methods. Summary table.

2.2.6 Proteases

Proteases are enzymes that cleave specific peptide bonds. They are present in all living organisms and are involved in many processes, from food digestion to regulation of protein metabolism to neuropeptide processing and processing of precursors of proteins and polypeptides such as hormones and proenzymes. Proteases can be generally classified based on the specificity or catalytic mechanisms: endopeptidases hydrolyse internal peptide bonds within the protein and exopeptidases remove aminoacids sequentially from the ends of protein fragments (Agarwal, 1990).

The presence of proteases during a bioprocess for pharmaceutical production must be understood and tightly controlled, as it might affect target molecule purity and activity.

The production of complex therapeutic proteins often uses mammalian cell cultures. Cell lines such as Chinese hamster ovary (CHO), murine myeloma cell line (NS0) or human retina derived cell (PER.C6) are commonly employed for high level expression of recombinant proteins (Birch and Racher, 2006). Target protein productivity has risen vertiginously in the last decade with titer levels surpassing 10g/l, to support increasing demands, high drug dosage, and control production costs of therapeutic proteins such as antibodies, fusion proteins (Robert et al., 2009) and enzymes.

Because of the high value per batch of current large scale fermentation runs, more now than ever it is of categorical importance to be able to minimize product degradation early on in the process, by optimizing fermentation and downstream processing conditions. One important factor that affects product quality is protease levels in the broth during fermentation. As CHO cells secrete target protein together with many other proteases, it is important to have a clear understanding of

which proteases are secreted, how they interact with the product and how and if they need to be inhibited to prevent product degradation.

This area has been extensively studied in the last two decades, and overall four major families of proteases have been identified to be secreted by CHO cells: metalloproteases or gelatinases, serine proteases, cathepsins (part of cysteil or aspartyl proteases) and cysteine proteases.

2.2.6.1 Metalloproteases

In a 2003 work (Elliott et al., 2003), the authors monitored the protease activity secreted by CHO-K1 cell line grown in serum free media and found that the most common extracellular protease was a pro form of matrix metalloproteases-9 (pro-MMP-9), also known as gelatinase B. The protease was identified by western immunoblotting by running on an electrophoretic gel a solution of clarified media and MMP antiserum and a secondary antibody conjugated to HRP (Horseradish Peroxidase). After staining the gel, the bands were identified as pro and active form of MMP-9.

Metalloproteases are cation-dependent proteases that require Zn^{2+} or Ca^{2+} to maintain activity. These proteases are typically not expressed in normal, healthy, resting tissues, while in contrast they are expressed in any repair or remodeling process, in any diseased or inflamed tissue and in essentially any cell type grown in culture (Parks, 1999). This enzyme can be inhibited by specific Tissue Inhibitor MetalloProteases (TIMPs), which are substrate-based inhibitors containing a hydroxamic acid moiety that chelates the active Zn^{2+} site and renders MMPs catalytically inactive (Moore and Spilburg, 1986). As later determined, iron citrate inhibits metalloproteases by blocking the Zinc atom in the enzymatic catalytic domain (Clincke et al., 2011).

The authors concluded that MMPs are highly active proteases against their substrates. This protease cleaves collagens, gelatins, proteoglycans and elastins (Nagase and Woessner, 1999).

Heterologous protein sensitivity to MMP should therefore be assessed as a first step (Elliott et al., 2003).

2.2.6.2 Serine Proteases

Serine proteases are widely present in nature and take part in several biological activities including digestion process, immunoresponse, blood coagulation and reproduction. They are called serine proteases because they catalyse the hydrolysis of peptide bonds in polypeptides, and they all have a serine residue that plays a critical role in the catalytic process. The catalytic site is common for all serine proteases and is composed by an aspartate residue (Asp102), histidine residue (His57) and a serine residue (Ser195) (Mathews et al., 2003).

They can be divided in two major groups based on principal structural folding: trypsin and subtilisin-like as shown in *Figure 2-8*. "Trypsin-like structure (a) consists of two β barrels with the catalytic triad coming together at the interface of two domains. Members of this family in eukaryotic cells include cathepsin G, chymase, chymotrypsin, elastase, factor D, factor VIIa, thrombin and trypsin. Subtilisin-like structure (b) consists of a three layer αβα sandwich fold with catalytic serine and histidine the ends of adjacent helices" (Madala et al., 2010).

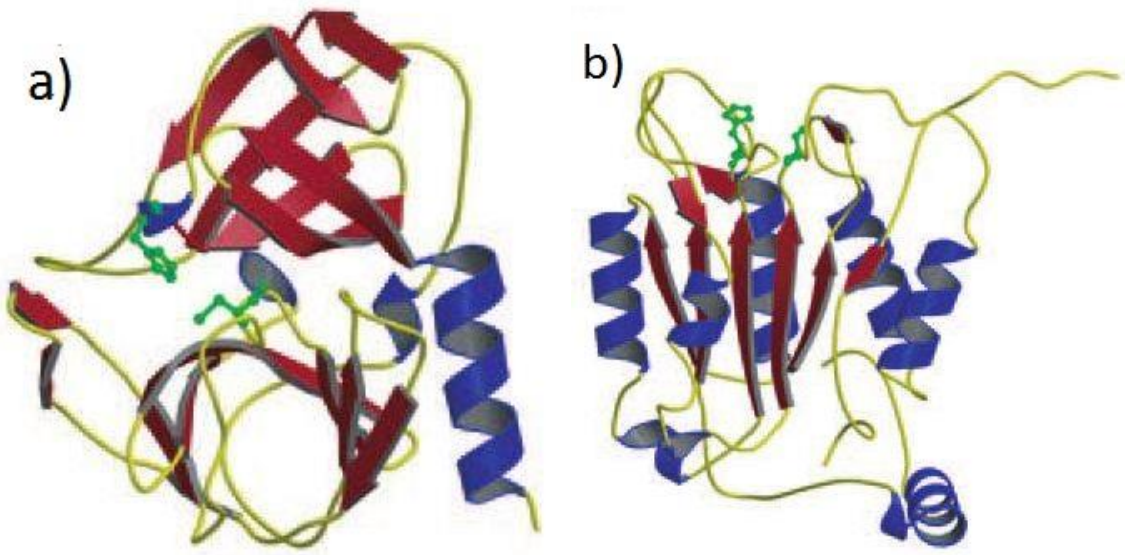


Figure 2-8 a) trypsin-line serine proteases b) subtilisin-like serine proteases. Source pubs.acs.org

This class of proteases has been identified in CHO matrix in the 2011 work of Clinke *et al.* By using specific protease inhibitors several proteases were discovered in the CHO culture clarified medium, among which serine proteases by using phenylmethylsulfonyl fluoride (PMSF) and Complete Inhibitor Cocktail (Roche). The presence of metalloproteases was also reconfirmed by using EDTA inhibitor. Similarly to Elliott's experiments, clarified medium was analyzed via western immunoblotting.

In one review paper, (Madala et al., 2010) summarised some of the most common serine proteases found, and identified inhibitors to most of them. Of these, only a small portion is secreted extracellularly. The data can be seen in *Table 2-10* below.

Serine Protease	EC code	Inhibitor PDB Code
brachyurin C	3.4.21.32	1azz
cathepsin G	3.4.21.20	1kyn, 1au8, 1cgh
chymase	3.4.21.39	1klt, 1pjp
α-chymotrypsin	3.4.21.1	1acb, 1cgi, 1cgj, 1cho, 1gl0, 1gl1, 1hja, 1mtn, 2cha, 6cha
γ-chymotrypsin	3.4.21.1	1ab9, 1afq, 1ca0, 1cbw, 1gcd, 1gct, 1gg6, 1ggd, 1gha, 1ghb, 1gmc,1gmd, 1gmh, 1k2i, 1n8o, 1vgc, 2gct, 2gmt, 2vgc, 3gch, 3gct,3vgc, 4gch, 4vgc, 6cha, 6gch, 7gch
δ-chymotrypsin		1dlk
complement factor D	3.4.21.46	1bio, 1dfp, 1dic
cytomegalovirus assemblin	3.4.21.97	1jq7, 2wpo, 1njt, 1nju, 1nkk, 1nkm
human neutrophil elastase	3.4.21.37	1b0f, 1h1b, 1hne, 1ppf, 1ppg
pancreatic elastase I	3.4.21.36	1b0e, 1bma, 1btu, 1e34, 1e35, 1e36, 1e37, 1e38, 1eai, 1eas, 1eat,1eau, 1ela, 1elb, 1elc, 1eld, 1ele, 1elf
pancreatic elastase II	3.4.21.71	1bru
enteropeptidase	3.4.21.9	1ekb
coagulation factor VIIa	3.4.21.21	1cvw, 1dan, 1fak, 1qfk
coagulation factor IXa	3.4.21.22	1pfx, 1rfn
coagulation factor Xa	3.4.21.6	1ezq, 1f0r, 1f0s, 1fax, 1fjs, 1fxp, 1g2l, 1g2m, 1ioe, 1iqe, 1iqf, 1iqg,1iqh, 1iqi, 1iqj, 1iqk, 1iql, 1iqm, 1iqn, 1j17, 1kig, 1ksn, 1kye,1lqd, 1lpg, 1lpk, 1lpz, 1mq5, 1mq6, 1nfu, 1nfw, 1nfx, 1nfy, 1p0s,1xka, 1xkb
granzyme B (natural killer cell protease)	3.4.21.79	1fi8, 1iau
hepatitis C NS3 4a protease(hepacivirin)	3.4.21.98	1dxw, 1dy8, 1dy9
kallikrein	3.4.21.35	1hia, 1l2e, 1lo6, 2kai, 2pka
kexin	3.4.21.61	1ot5, 1r64
kumamolisin	3.4.21.-	1gtj, 1gtl
lysyl endopeptidase	3.4.21.50	1arc
R-lytic protease	3.4.21.12	1gbb, 1gbc, 1gbd, 1gbf, 1gbh, 1gbi, 1gbk, 1gbl, 1gbm, 1p01, 1p02,1p03, 1p04, 1p05, 1p06, 1p10, 1p11.
matripase	3.4.21.-	1eaw
Mesentericopeptidase	3.4.21.62	1mee

plasmin	3.4.21.7	1bui
t-plasminogen activator	3.4.21.68	1bda, 1rtf, 1a5i, 1a5h
prolyl oligopeptide	3.4.21.26	1e5t, 1e8m, 1e8n, 1h2y, 1h2z, 1o6f, 1o6g, 1qfm, 1qfs
proteinase K	3.4.21.64	1bjr, 1pek, 1pfg, 1pj8, 1oyo, 3prk
protein C (activated)	3.4.21.69	1aut
sedolisin (pseudomona-pepsin)	3.4.21.100	1ga1, 1ga4, 1ga6, 1kdv, 1kdy, 1kdz, 1ke1, 1ke2
signal peptidase I	3.4.21.89	1b12
streptogrisin A (proteinase A)	3.4.21.80	1sgc, 3sga, 4sga, 5sga
streptogrisin B (proteinase B)	3.4.21.81	1cso, 1ct0, 1ct2, 1ct4, 1ds2, 1sgd, 1sge, 1sgn, 1sgp, 1sgq, 1sgr,1sgy, 2sgd, 2sge, 2sgf, 2sgp, 2sgq, 3sgb, 3sgq, 4sgb
streptogrisin E (glutamyl endopeptidase II)	3.4.21.82	1hpg
subtilisin BPN	3.4.21.62	1a2q, 1gns, 1gnv, 1lw6, 1sbn, 1sib, 1sua, 1sue, 1sup, 1ubn, 2sic,2sni, 3sic, 5sic
subtilisin Carlsberg	3.4.21.62	1a10, 1av7, 1avt, 1be6, 1be8, 1cse, 1oyv, 1r0r, 1scn, 1vsb, 3vsb
subtilisin DY	3.4.21.62	1bh6
thermitase	3.4.21.14	1tec, 2tec, 3tec
thrombin	3.4.21.5	1a46, 1abj, 1ay6, 1bth, 1de7, 1fph, 1k22, 1nrn, 1nro, 1nrp, 1nrq,1nrs, 1tbq
tricorn	3.4.21.-	1n6d, 1n6e, 1n6f
trypsin	3.4.21.4	1aut, 1ezx, 1sbw, 1tyn, 3btg
urokinase plasminogen	3.4.21.73	1c5w, 1c5x, 1c5y, 1ejn, 1f5k, 1f5l, 1f92, 1fv9, 1gi7, 1gi8, 1gi9,1gj7, 1gj8, 1gj9, 1gja, 1gjb, 1gjc, 1gjd
venom plasminogen	3.4.21.-	1bqy

Table 2-10 Crystal structure Listing for 220 Inhibitors Complexed with 44 Serine Endoproteases. EC code: Enzyme Commission number; PDI code: Protein Data bank Code. Adapted from (Madala et al., 2010)

2.2.6.3 Cysteine Endopeptidase

Among proteases found to be secreted by cultured CHO cells, cysteine proteases are also found. From one review paper written by Satoh *et al* in 1990, the activity of proteases present in serum free conditioned medium of DHFR-deficient CHO cells was analysed using Methyl Cumaryl Amide (MCA) substrates at pH 7.0. They found that some substrates were degraded by the serum consistently with increment of non-viable cells, indicating the presence of intracellular proteases released by the dying cells. There was also observed to be peptidase activity towards peptidyl substrate, Boc-Leu-Arg-Arg and Z-Phe-Arg in particular. This activity was found to be inversely proportional to the number of viable cells, indicating the secretion by intact CHO (Satoh et al., 1990). In the paper, the effects of various inhibitors were also assessed. It was found that known "cysteine protease inhibitors had measurable effects on the activity of the medium, indicating the presence of cysteine proteases". In particular the inhibitors used were leupeptin, E-64 and chymostatin. The pH optima for the endopeptidases was found to be between 6.0 and 9.5.

Overall this study showed how exopeptidases were mainly secreted intracellularly, and released outside during cell lysis, whereas endopeptidases were actively secreted by viable CHO cells.

2.2.6.4 Cathepsins

Most cathepsins are cysteine proteolytic enzymes mainly localised in the lysosome. With a few exceptions, the majority of cathepsins belong to the endopeptidases family (Agarwal, 1990). They catalyse the hydrolysis of a variety of protein substrates with different specificities and are involved in a wide range of processes. In particular they have been identified responsible of intracellular protein turnover and post-translational processing of some precursors such as

endorphin and insulin and are involved in several pathological processes such as tumour metastasis, muscular dystrophy and arthritis (Agarwal, 1990).

In one review paper, Agarwal *et al* describes how just like GAA, cathepsins are synthetised on membrane bound ribosomes, then transferred through the RER and finally into the Golgi apparatus. After modifications to the tertiary structure, they are transported into the lysosomes via MPRs, and also that just like GAA, they are active at acidic pH (4 to 5).

It has been pointed out that although rarely secreted, in some cases, the precursors of these lysosomal enzymes escape the normal pathway and continue along the secretory route, entering the storage granules and finally getting released into the extracellular space. This occurs in activated macrophages, osteoclasts and interestingly in fibroblasts from patients with I-cell disease (lysosomal storage disorder), thereby extending the pattern of physiological and pathological conditions in which these enzymes may be involved (Santamaria et al., 1998, Kornfeld and Mellman, 1989).

There are many different cathepsins and the relevant features are summarised in *Table 2-11* below.

Cathepsin type	PH optimum	Involved in	Inhibitors
A	5.2-5.7	Muscular dystrophy, Skeletal muscle. Cleaves sequentially C-terminal hydrophobic residues.	- di Isopropyl fluorophosphates (DIPF) - phenylmethane-sulphonylfluoride (PMSF)
B	3.5-6.0	Degradation of tissue proteins in the lysosomal system. Possesses dipeptidyl carboxypeptidase activity.	- 4-Chloromercuribenzoic acid (PCMB) - iodoacetamide (IAM) - N-ethylmaleimide (NEM) - leupeptin
C	7.7	Muscular dystrophies and polymyositis. Splits dipeptide naphthylamides and removes dipeptides from amino terminus	- iodoacetic acid (IAA)
D	2.8-5	degradation of intracellular and endocytosed proteins (myelin)	- pepstatin
E	2.5	found in bone marrow, polymorphonuclear leukocytesand macrophages	- pepstatin
F	4.5	degrades cartilage proteoglycan at pH 4.5	- α_2-macroglobulin - whole egg white
G	7.5	digests a number of proteins such as prothrombin and elastin, and cleaves some synthetic chymotrypsin substrates	- soybean trypsin inhibitor (STI) - α_1 proteinase inhibitor (α_1PI) - α_1 antychimotrypsin, chymostatin - diisopropylphosphofluoridate (DIPF)
H	6.8	hydrolyses blocked synthetic substrates such as benzoylarginine naphthylamide (BANA) and substrates with free oL-amino groups such as arginine naphthylamide	- 4-Chloromercuribenzoic acid (PCMB) - iodoacetamide (IAM) - N-ethylmaleimide (NEM)
I	6.5-7	BANA hydrolase	- leucine chloromethane
J	6.8	hydrolyzes only a synthetic substrate, benzyloxycarbonyl-Phe-Arg-aminomethylcoumarin (Z-Phe-Arg-AMC)	- 4-(hydroxymercuri)BenzoicAcid (PHMB) - α_1 proteinase inhibitor (α_1PI) - ATP - Ca^{++}
K	6.2	Bone remodelling. Hydrolyses various synthetic	- leupeptin

		substrates such as Z-Arg-Phe-AMC, Z-Arg-Arg-AMC and Bz-Val-Lys-Lys-Arg-AMC.	- potassium ions
L		splits cytosolic proteins, collagen, myosin, actin, tubulin, calmodulin, insulin, glucagon and many enzymes	- leupeptin - benzyloxycarbonyl-Phe-Phe-diazomethyl ketone
M	7	associated with lysosomal membranes	- leupeptin
N	3.5	Cleaves N-terminal peptides of native collagen	- N/A
P	5-6	processing of propolypeptides	- 4-Chloromercuribenzoic acid (PCMB) - leupeptin
R	7-8	hydrolyses casein and most of the ribosomal proteins	- di Isopropyl fluorophosphates (DIPF) - soybean trypsin inhibitor (STI) - α_1 proteinase inhibitor (α_1PI)
S	3.5	antigen presentation: digests haemoglobin and other proteins.	- z-Phe-Phe-diazomethane
T	7	catalyses the conversion of multiple forms of tyrosine aminotransferase	- thiol blocking reagents
W		immune response: cytotoxic effector cell compartment and natural killer cells	- peptide based inhibitor
Z (or X)		Intracellular protein turnover	- Gly-Phe-Gly-semicarbazone - Proteinase Inhibitor E 64

Table 2-11 Cathepsin A to T data adapted from (Agarwal, 1990). Cathepsin W data adapted from (Ondr and Pham, 2004). Cathepsin Z adapte from (Santamaria et al., 1998)

Cathepsin Z

The behaviour and monitoring of this specific protease is of particular interest in this project as it is believed that it might be responsible for target protein (GAA) degradation during the bioprocess.

Cathepsin Z (or X) is a cysteine endopeptidase part of the papain family. It is well established that the different cysteine proteases in this family are synthetized as preproezymes, which are processed to the corresponding proenzymes and targeted to the lysosomes by the M6P signal attached to them. The enzymes are further processed to mature forms consisting of either a single polypeptide chain or a two chain form composed of heavy and light chains linked by a disulphide bond (Berti and Storer, 1995). However in some cases, the precursors of these lysosomal enzymes escape from this processing pathway and continue along the secretory route, entering storage granules and being finally released into the extracellular space (Kornfeld and Mellman, 1989).

It is believed that Cathepsin Z expression is linked to overexpression of lysosomal GAA. This could be due to the transport mechanism, mannose–6–phosphate receptor being overloaded with GAA, which induces CatZ and other lysosomal proteins being secreted as well. The commercial form of GAA, having the GILT tag is likely more sensitive to CatZ and other proteases due to potential clipping. Therefore overexpression of GAA/GILT results in CatZ expression which impacts stability of the target molecule.

2.2.7 Detection of product and impurities

Downstream purification is by definition the process of removal of impurities from the process stream in order to purify the target molecule to levels established by appropriate regulatory agencies. In the biopharmaceutical industry, this level is set extremely high, and failure to meet regulatory requirements on protein purity and stability can have serious consequences on patient

safety and drug effectiveness and overall business reputation, which will have a tremendous economic impact on the manufacturer.

Host cell protein (HCP) release during production of therapeutic proteins is a critical quality attribute that must be monitored and reduced to acceptable levels during the production of a biotherapeutic protein (Bracewell et al., 2015). Being able to control HCP release during harvest and their removal via downstream processing is thus an important component of the manufacturing of such molecules and assurance of the ability of processes to achieve safe levels of HCP must be reported to regulatory authorities. Typical purity targets for a mAb type include <100 ppm HCP, <10 ng/dose DNA and <5% product aggregates (Chon and Zarbis-Papastoitsis, 2011). Some HCPs are known to be co-purified with the recombinant protein target, whilst others, specifically proteases, are reported to lead to product degradation (Gao et al., 2011).

The physical properties of particulates in the bioprocess stream are exploited during the downstream process to achieve separation and thus purification. During the first stages of primary clarification of CHO culture, usually the cells are separated from the media via centrifugation and filtration. These techniques are not very selective and separate particles only based on mass. In this instance for example, cells are easily removed via continuous centrifugation based on the great difference in mass between them and the proteins suspended in the media.

As purification moves forward and the process stream is constantly more and more clarified there might not be as many radical differences between proteins and nucleic acids in solution and finer separation techniques are required. One of the most common and powerful separation techniques used in biopharmaceutical industry is chromatography.

Chromatography is a powerful tool for the purification of biomolecules. “It aims at exploiting every little difference in the chemical and physical properties of molecules to effect the

separation of a mixture into individual compounds. These include differences in chemical reactivity, polarity and molecular size" (Ravindranath, 1989). There are several techniques used to separate molecules based on properties such as molecular charge, hydrophobicity, size and affinity. By using a combination of two or more of these techniques it is possible to achieve very high levels of protein purity. Before moving forward however, it is important to briefly define such properties:

Isoelectric point. It is the pH at which the molecule has a neutral charge. This quality allows separation through ion exchange chromatography (IEC). The pH of the mobile phase plays a very important role for sample retention as a change in pH over the isoelectric point (pI) would shift the charge of the protein. As explained later in greater detail, this technique separates large proteins, small nucleotides and aminoacids on the basis of their charge.

Hydrophobicity. "it's the property of a molecule to repel from a mass of water" (Ben-Naim, 1980). Separation based on this property takes advantage of hydrophobic interactions between molecules. "Hydrophobic regions (which are non-covalent and non-electrostatic) of a molecule may be attracted to non-polar regions of another molecule" (Ravindranath, 1989). Chromatography technique based on this property is hydrophobic interaction chromatography (HIC). When a sample containing a mixture of proteins is applied to a HIC column in high salt, hydrophobic and hydrophilic regions interact with the resin and bind to it. The salt interacts with water molecules to reduce solvation of the protein molecules in solution (Porath, 1986). As salt is gradually decreased, hydrophobic regions become exposed and are absorbed by the mobile media and therefore start eluting. The more hydrophobic the molecule, the less salt is needed to promote binding.

Size. Protein size is determined by the number of amino acids forming the polypeptide and by the way the protein is folded. This quality allows separation in size exclusion chromatography

(SEC), in which molecules larger than the beads average pore size travel around them come out of the column first. Smaller molecules diffuse into the gel structure and are retained by the stationary phase and elute in order of decreasing molecular size.

Affinity. Affinity it's a method of separating biochemical mixtures based on a highly specific interaction between antigen and antibody, enzyme and substrate, or receptor and ligand. It's the most selective type of separation technique as it takes advantage of 'highly specific affinities that certain molecules naturally have for certain substrates, mainly enzyme-substrate, antibody-antigen and receptor-hormone' (Ravindranath, 1989). The process is reversible and it is possible to break the disulphate covalent bonds to release the antibody during elution.

2.2.7.1 The issue

It is hypothesized that the issue in this particular process of purification of recombinant human acid alpha glucosidase (rhGAA) is that some proteases naturally produced by CHO cells in culture might have similar physical properties in common with GAA such as size and isoelectric point and therefore could be harder to separate than most other impurities. The presence of such proteases in one or more stages of purification and possibly even in the drug product could have a detrimental effect on the target protein quality.

During GAA production using a CHO culture, eventual proteases present in the process stream will be screened using an array of analytical techniques described below. Known proteases will be recognised using immunoassays, unknown ones will have to be identified using mass spectrometry.

2.2.7.2 Analytical assay development

Detection, quantification and identification of the components of the matrix at each different stage of the purification process will be achieved with ad-hoc analytical assays described below.

Immunoassays

In order to ensure that the developed cell line is producing the target molecule, one of the first assays that will have to be performed is for GAA detection. As this is a well-known protein, GAA antibody is commercially available and running an immunoassay such as ELISA or AlphaLISA would be quite straightforward. This assay takes advantage of the immunogenic reaction that naturally occurs between antigen and antibody. Generally, microtiter plate wells are coated with immobilised antibody specific to GAA. Sample is added and eventually present antigen binds to the plate. A detection antibody specific to the antigen is added to the plate, a secondary antibody specific to the primary antibody is then added. The secondary antibody carries an enzyme bound to it. Enzyme substrate is finally added to the solution, which makes the solution change colour. The colour wavelength can be measured by a spectrophotometer and concentration of original antigen is therefore determined. This technique reveals presence and amount of a particular protein in a solution, and it's useful when the protein of interest is known.

In order to determine if other unknown proteases are present however, it is necessary to screen the solution using other techniques such as 2D gels or western blot analysis.

Electrophoresis

2D gel electrophoresis is a useful technique used to separate a diverse mix of proteins. An SDS-PAGE gel is loaded with a mix of proteins, and are separated as they move through it. The advantage of this technique over regular SDS-PAGE is that it allows separation based on two different molecular properties such as mass and charge, which is handy when diverse molecules are hard to separate because for example have all the same mass. With this technique, a mix of peptides could be separated by charge first and by size on the second axes. Generally, gels are then stained to make proteins visible using Silver or Comassie Blue staining. Proteins are usually treated with a reducing agent prior to electrophoresis in order to break disulphide bonds and

favour the formation of linear structures. The other advantage of this technique over immunoassays is that it allows for recovery of material for further analysis. The protein of interest can in fact be 'cut-out' of the gel and identified using other techniques such as mass spectrometry (MS).

Western blot

Western blot is usually used in conjunction with a developed electrophoretic gel. The gel containing the mix of proteins separated by a specific property is subject to a perpendicular electric field, which moves the proteins outside of the gel. A membrane usually made of nitrocellulose or polyvinylidene difluoride (PVDF) is placed just next to it, and proteins bind non-specifically to it in a process called 'blocking'. This membrane then is subject to a process similar to ELISA. The membrane is incubated in a solution with a primary antibody specific to a particular protein, washed and finally incubated with a secondary antibody specific to the primary antibody, linked to an enzyme specific to a substrate. Once the substrate is added, a change in colour reveals where the specific protein of interest is located. This process can be repeated several times with different antibodies to identify all proteins present. The material that does not react with any known antibody can then be collected and analysed via MS. Information source (Abcam, 2012)

Enzymatic activity

Enzyme activity is another metric that needs to be assessed as this measurement gives important information on enzyme kinetics with a particular substrate and is related to enzyme purity, as the value becomes larger as the enzyme is purified. Reaction rate however stays the same or might at times increase due to removal of eventual inhibitors from the process stream. The activity is generally measured in $\mu mol\ mg^{-1}min^{-1}$

In an article by (Reuser et al., 1978) GAA activity towards substrates such as glycogen and maltose was measured. 10µl of cell homogenate was mixed with maltose (50µl of 10g/l, Baker, Phillipsburg, N.J.) or glycogen (50µl of 50g/l Gibco, Grand Island, N.Y.) as substrate. Both were dissolved in 0.2M potassium phosphate-citrate buffer, pH4.4. After incubation and heating for 2 min at 100^0C the amount of released glucose was determined by adding 200µl glucose reagent according to (Koster et al., 1976). Results showed that enzyme activity for infantile Pompe disease was 1%, and late onset Pompe disease was between 7 and 22% compared to healthy patients GAA activity.

This assay will be used to determine the presence and activity of multiple proteases in different stages of purification by measuring conversion rates of specific substrates.

GAA diagnostic assay

Most enzymatic activity assays work by measuring colorimetric change due to the enzymatic degradation of a substrate. This approach is not feasible for GAA quantification as the presence of unrelated α-glucosidases (predominantly maltase-glucoamylase) can interfere with absolute GAA quantification. Acarbose, an amylase inhibitor, was used in a study by Okumiya *et al* to quench activity of other amylases and was proven capable of measuring accurately GAA activity. This method is now diagnostically used in the clinics to test suspected Pompe patients' blood. Due to its reproducibility and rapidity, it was adapted to measure GAA amount in in-process samples.

Zymography

Zymographic techniques allow detection of proteases and some other hydrolytic enzymes following electrophoresis in various types of gel matrices" (Lantz and Ciborowski, 1994). Zymography is a versatile two stage technique involving protein separation by electrophoresis

followed by detection of proteolytic activity. It is useful as it allows the determination of enzyme activity using a SDS PAGE electrophoretic gel. In this process, polypeptides are not treated with a reducing agent neither are boiled in order to retain original active structure. After electrophoresis SDS is removed from the gel with the use of a detergent such as Triton X-100 and the gel is stained using Comassie Blue. Enzyme activity indicated by "zones of lysis, where proteolytically active bands have degraded the substrate in the indicator gel, are visualized as clear zones against an opaque background on dark-field illumination or as clear zones against a dark blue background after staining of the gels with Comassie Brilliant Blue R-250" (Lantz and Ciborowski, 1994).

Identification Assay

One of the workhorses of analytical chemistry is Mass Spectrometry (MS). This technique has multiple applications in drug development ranging from protein characterisation to glycan analysis. There is a wide array of MS variations and it is not the intention of this work to go through all of them.

MS is based on the principle that charged particles that travel through a magnetic field, are subject to a deviation of their path that is dependent on their mass. The mass spectrum produced of a polypeptide acts as a sort of fingerprint as it is unique to each one of them.

A mass spectrometer is generally composed by four modules: an ion source, which excites the atoms in the molecule, an extraction system, a mass analyser and a detector.

There are two approaches in protein characterisation using MS. The most common one is to digest the protein in fragments using specific protease and analyse them in a process called 'bottom up'. The information gathered by single peptides can be used to gain knowledge on the

full peptide. The other one is to run through the MS intact proteins and this process is nicknamed ‘top down’.

Peptide Mapping is a good example of the bottom up technique and although time consuming it is frequently utilised to analyse the primary structure of complex therapeutic proteins such as monoclonal antibodies. “Biomolecules of interest are subject to a protease, usually trypsin or Lys-C, that enzymatically cleaves the protein in specific peptides, which are then separated by LC and analysed by MS. During this process proteins are denatured (unfolded), reduced (to break disulfide bonds) and carboxymethylated (to prevent reformation of bonds). Mixture is then washed and digested and injected into the MS” (Dick et al., 2009).

TEM imaging

Transmission Electron Microscopy technique has been used for the past 80 years to acquire high resolution images of biological samples. It works similarly to a light microscope, although it uses electrons as "light source" and their much lower wavelength makes it possible to get a resolution a thousand times better than with a light microscope. In this work this technique will be used to acquire images of GAA producing CHO cells and compare them with a Null strain. Due to its high resolution it will be possible to visualise cells lysosomes to assess whether GAA over-expression has an impact on lysosomes phenotype.

2.3 Objectives

2.3.1 Production of stable CHO cell line for material generation.

The production of a stable CHO cell line expressing GAA gene was the first goal of this project, and constituted the bulk of the first year MRes research project. This cell line was useful during the entire course of the EngD project as it was used to produce material to develop and optimize the purification process. This process took place in conjunction with University of Kent School of Biosciences labs.

2.3.2 Development of assays for detection of product in the bioprocess stream

Analytical assays had to be developed aimed at the detection and quantification of the product (GAA) levels at the different process stream steps. These assays were immuno, gel-electrophoresis, western blots, zymography and enzymatic activity based. Tandem Mass Spec identity assays were also used to confirm identity of the product.

2.3.3 Bioprocess mimic

The development of the fermentation and downstream purification process had to closely mimic the process already in place at BioMarin, in order to ensure the maintenance of same conditions that cause CHO cells to produce specific proteases during the bioprocess. The downstream bioprocess was constituted of a series of two chromatography steps (IEX and HIC) after clarification of harvest cell culture fluid.

2.3.4 Characterization of cell stress and its impact on protease production

It is believed that the overexpression of GAA in recombinant CHO cells causes cell stress and over production of proteases such as cathepsin B, Z and others. A series of assays were used to compare recombinant vs. its original Null (untransfected) counterpart to identify differences in lysosomal phenotype, amount and type of proteases and their localisation in the cell structure. Images of lysosomes were be taken with Transmission Electron Microscope (TEM) and different

tandem MSMS techniques and zymography were used to identify and quantify HCP presence throughout different bioprocess steps.

2.4 Project estimated timeline / Gantt chart

An estimated project timeline is included for reference in *Table 2-12* below:

Objective	Workplan Task	Year 1				Year 2				Year 3				Year 4			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Training Courses	Pilot Plant unit operations training, MBI courses																
1. Production of stable CHO cell line for material generation	Literature Review																
	Cloning human GAA gene into vectors																
	Express product protein in E.Coli, detect it with WB																
	Transfect CHO cells, select high producing clones.																
	MRes Thesis write up																
2. Assays for detection of proteases in bioprocess	Development of immunoassay based technique to detect presence of product and proteases																
	Quantification of HCP																
	Mass Spectrometry																
3. Bioprocess mimic	Anion Exchange chromatography using AKTA																
	Hydrophobic interaction chromatograohy using AKTA																
	In house mimic of existent Biomarin bioprocess.																
4. Characterisation of cell stress and its impact on protease production	Zymograms																
	LC-MS																
	TEM imaging with immunogold staining																
Write Up / contingency	Writing of EngD thesis, contingency for further work which might be required																

Table 2-12. Gantt chart of estimated project timeline.

3 Materials and methods

3.1 Cell line generation

3.1.1 Cloning

This section describes the process involved in cloning the human GAA gene into pcDNA 3.1 and pcDNA5/FRT vectors. This constitutes the first step of generation of stable CHO cell cultures. A list of all materials and reagents used in plasmid cloning is included in Table 3-1 below.

Materials and Reagents	Product code	Manufacturer	Description
Abe 1 RE	n/a		Restriction enzyme
Agar	A5054-250G	Sigma	
Agarose	A9539500G	Sigma	
Ampicillin	A9518-25G	Sigma	Ampicillin sodium salt
Ase 1 RE	ER0911	Thermo Scientific	Restriction enzyme
DH5 competent cells	n/a	n/a	
DNA stain	S33102	Invitrogen	For electrophoresis
FreeStyle MAX CHO	K900020	Thermo Fisher	Expression system
Gel Extraction Kit	28704	Qiagen	Qiaquick Gel Extraction Kit
Human GAA gene	SC125512	Origene	Homo Sapiens glucosidase, alpha, acid (GAA) transcript variant 1 as transfection ready DNA in vector pcmv6-XL6
Hyperladder	1KB plus	Bioline	For electrophoresis
Ligase enzyme	M180A	Promega	T4 ligase
Loading buffer	bio-37045	Bioline	For electrophoresis
MiniPrep kit	27104	Qiagen	Qiaprep Spin Miniprep Kit
MaxiPrep kit	A2393	Promega	Pureyield™ Plasmid Maxiprep System
NaCl	S3105-65	Fisher Scientific	
Not 1 RE	10889026	Roche	Restriction enzyme
pcDNA 3.1 vector	V790-20	Invitrogen	
pcDNA 5/FRT vector	V6010-20	Invitrogen	
Sal1 RE	11745-635	Roche	Restriction enzyme
TAE	A1416	AppliChem	Tris base, acetic acid and EDTA
Top10 Competent cells	C404010	Invitrogen	Chemically competent
Triptone	211705	BD	
TSAP enzyme	M991A	Promega	Thermosensitive alkaline phosphatase

Yeast Extract	09182-1KG	Fluka	

Table 3-1. Materials and reagents list.

3.1.1.1 Media preparation

The following media were scaled up or down as needed for different experiments.

200ml LB Agar. Mix:

2.0 g Triptone, 3.0 g NaCl, 3.0 Agar, 1.0 g Yeast Extract, 200ml milliQ water.

Autoclave, let cool to about 50^0C, add ampicillin, pour aseptically into new petri dishes

100ml LB media. Mix:

0.5 g Yeast Extract, 1.0 g NaCl, 1.0g Triptone

Autoclave, let cool to about 50^0C, add ampicillin.

3.1.1.2 Digestion

The purpose of digestion is to allow the vector to open at a determined restriction site point, where the GAA gene is to be inserted.

Both vectors are digested with Not1 RE, however only the GAA-XL6 construct is digested with both Not1 and Ase1 REs. This allows for easier purification of the gene later on from the electrophoretic gel, as it cuts the XL6 vector (vector used by Origene for shipment of the GAA gene) in segments of less than 1000 bp, which do not locate closely to the GAA DNA band.

Reaction mix *Table 3-2* below is placed in a 0.5ml tube, vortexed, incubated 2hr in a 37^0C water bath followed by 20 minutes at 65^0C to inactivate the enzymes. The amount of DNA was

calculated by measuring absorbance at 260nm using a Nanodrop-N1000.

(µl)	pcDNA 5/FRT	pcDNA3.1	GAA-XL6	pOG44
DNA	9.52	9.34	20	11.11
Buffer O	2	2	2	2
RE	1.5	1.5	1+1*	1.5
H_2O	7.2	7	0	5.4

Table 3-2. Digestion reaction mix volumes in (µl), Buffer O 50 mM Tris-HCl (pH 7.5 at 37°C), 10 mM MgCl, Thermo Fisher.

3.1.1.3 DNA Purification

This step allows the separation of DNA segments, which are then purified and ligated. A 1% agarose gel is made (1 g agarose in 100 ml water), microwaved until boiling, allowed to cool down to about 50^0C and mixed with 5µl of DNA stain. 10µl of sample is mixed with 10µl of loading buffer and pipetted into the gel alongside the 1kb hyper-ladder. The gel is run for 75 minutes at 80V.

Gene and vector bands are purified out of the gel using the Qiagen Gel Extraction kit (product code 28740).

The protocol developed by the manufacturer was used, a copy of which is attached into the bibliography (Qiagen, 2013). DNA was eluted into 20µl H_2O, concentration checked by spectrophotometry (A260 - NanoDrop N1000) and frozen to -20^0C.

3.1.1.4 Dephosphorylation

Dephosphorylation is conducted in presence of thermo sensitive alkaline phosphatase (TSAP) enzyme, which "catalyses the removal of 5′ phosphate groups from DNA, thus preventing the re-circularization and re-ligation of linearized cloning vector DNA during ligation" (Promega, 2013b). This is only carried out with vectors (pcDNA3.1 and pcDNA5/FRT).

In 0.5µl tube 1µg of vector DNA, 5µl of 10X buffer, 1µl TSAP were mixed, vortexed and incubated 15 minutes at 37^0C, followed by 15minutes at 74^0C to inactivate the enzyme. The samples were ready for ligation.

3.1.1.5 Ligation

In this step each vector is ligated with the GAA gene. Two dilutions per ligation were used, a 1:1 and 1:3, and also a negative control 1:0 (vector with no insert) to ensure ligation happened correctly.

The reactions take place in 0.5ml tubes. The volumes are reported in Table 3-3 below.

(µl)	**1:1**	**1:3**	**1:0**
	pcDNA3.1 + GAA		
Vector	5	5	5
Insert	5	15	0
Buffer	2	3	2
Enzyme	1	1	1
Water	7	6	12
	pcDNA5/FRT + GAA		
Vector	5	5	5
Insert	5	15	0
Buffer	2	3	2
Enzyme	1	1	1
Water	7	6	12

Table 3-3. Ligation reaction volumes in µl.

Tubes were vortexed, spun and incubated overnight at 4^0C.

3.1.2 Genetic material characterization

3.1.2.1 Via restriction enzyme (RE) digestion

Both constructs (pcDNA3.1+GAA and pcDNA5/FRT + GAA) were digested with RE Sal1 to check for the direction of insertion. 1 µg of DNA was mixed with 0.5 µl of Sal1 enzyme, 2.5 µl SuRE/Cut Buffer H (cat# 11417991001 Sigma Aldrich USA) and water to reach 15 µl total

reaction volume. Reaction mixtures were incubated at 37^0C 1h and 20min 65^0C. Digested reaction mixtures were run on 1% agarose gel at 100 mV for 1 h.

3.1.2.2 Via sequencing

The DNA samples to be sequenced (two pcDNA 3.1 and two pcDNA 5/FRT) were diluted in water to100 ng/ml. 10 µl of sample and the purchased primers were submitted to the Wolfson Institute for Biomedical Research at UCL for sequencing.

3.1.3 Transformation

The selected plasmids were amplified using One Shot® TOP10 Chemically Competent *E. coli* (cat#C404010 Thermo Fisher). This procedure allowed using *E. coli* cells as highly an efficient cloning and plasmid propagation system. The amplified genetic material was then extracted and purified using PureYield™ Plasmid Maxiprep System (Promega).

The manufacturer's protocol was followed (cat#C404010 Thermo Fisher): LB agar Petri dishes containing 100 µg/ml ampicillin were plated aseptically. Prior to transformation Top 10 cells were thawed on ice and aliquoted. Ligation reaction mix was added at different ratio (1:0, 1:1 and 1:3), aliquots were then incubated on ice for 30 minutes followed by thermal shock (30 sec at 42^0C in water bath) to allow inclusion of plasmid. Finally 200µl was plated and incubated overnight at 37^0C.

The next day, single colonies were picked aseptically and put in 5ml LB media at 37^0C agitated at 250rpm for 8 hours and eventually transferred to 400ml LB ampicillin+ media in the same condition overnight. The next day growth was observed, cells were spun down and DNA purified via Maxiprep. The extracted DNA was diluted in 1 ml RNAse free water and stored at -20^0C. This material was used for transfection into mammalian cells.

3.1.4 Transient Transfection

A FreeStyle MAX CHO expression system kit was acquired from Thermo Scientific. The manufacturer protocol was followed: a vial of FreeStyle™ CHO-S® (cat# R80007 ThermoFisher) cells was thawed out in a water bath and transferred aseptically into a 125ml flask containing 30ml of FreeStyle™ CHO Expression Medium with 8mM L-glutamine. The flask was incubated at 37^0C 8%CO_2 overnight. The next day cell viability was assessed using a Beckman Coulter ViCell.

CHO-S cells were sub cultured for five passages (approximately 2 weeks) before proceeding with transfection to allow complete recover from thawing. As they reached 1.5 10^6 cells/ml they were passaged to 0.3 10^6 cells/ml. Viability always stayed above 95%. Approximately 24h before transfection cells were passaged to 10^6 cell/ml.

On transfection day, purified DNA material (from transformation 3.1.3) was filter sterilised, concentration was determined via NanoDrop and 37.5 µg were mixed with OptiPRO SFM media according to manufacturer.

The plasmid preparation in media was gently mixed and added to the transfection reagent, incubated 15 minutes and slowly added to the cells. The flask was again incubated at 37^0C 8%CO_2 for one week, 1ml culture samples were taken every day and frozen for later analysis. On day of western blot analysis, frozen samples were thawed, spun at 13000rpm 10 min, supernatant the decanted and cell pellet sonicated to release the cell content. Target protein production was assessed via western blot (materials and methods 3.2.3)

3.1.5 Stable transfection

The pcDNA5/FRT-GAA construct was used to transfect the CHO Flp-In (cat# R75807, Thermo Fisher Scientific USA) commercially available cell line that had been previously adapted to grow

in chemical defined CD-CHO media (ThermoFisher) + 8mM glutamine, using PEI (Polyethylenimine linear, cat#9002-98-6 Sigma Aldrich USA) as a transfection agent and the pOG44 Flp-Recombinase Expression Vector (cat# V600520 Thermo Fisher Scientific USA) in a (1:9):3 ratio (3µg of plasmid DNA added to 27µg of pOG44, incubated 5 minutes at RT, followed by 90µg of linear PEI).

The cells were incubated in CD-CHO media (Thermo-Fisher Scientific, USA) and passaged to 0.3 10^6 cell/ml when they reached 3 million cells per ml, this was done three times over the course of 18 days to ensure the cells were fully recovered from thawing. Viability was always maintained above 95%. 24 hours before transfection the cells were seeded at 0.5 10^6 cell/ml so that they would be at the desired concentration of 10^6 cell/ml at time of transfection. On the day of transfection the concentration of plasmid pcDNA/FRT A and B (from 3.1.3 Transformation) and pOG44 was assessed via NanoDrop at 260nm. As per protocol, the ratio of plasmid DNA to pOG44 to PEI was (1:9):3. In this case 3µg of DNA were added to 27µg of pOG44, incubated for 5 minutes at RT, followed by 90µg of linear PEI added, and vial vortexed and incubated for 5 min at RT. This solution was then added drop by drop to culture in flasks. Freshly transfected cultures were incubated at 37^0C 8% CO_2 125rpm in a spin tray incubator. The following day cells were spun at 300g 5 min and media changed with fresh media containing selection agent (500 µg/ml HygromycinB). Cell viability and count was assessed daily for three weeks, media was replaced by fresh HygB+ media every four days.

3.1.6 Description of the Flp-In™ System

The Flp-In™ System streamlines the generation of stable mammalian expression cell lines by taking advantage of a Saccharomyces cerevisiae-derived DNA recombination system. This DNA recombination system uses a recombinase (Flp) and site-specific recombination (Craig, 1988,

Sauer, 1994) to facilitate integration of the gene(s) of interest into a specific site in the genome of mammalian cells.

In the Flp-In™ System, three different vectors are used to generate isogenic stable mammalian cells lines expressing your gene(s) of interest. The first major component of the Flp-In™ System is the pFRT/lacZeo target site vector that is used to generate a Flp-In™ host cell line. The vector contains a lacZ-Zeocin™ fusion gene whose expression is controlled by the SV40 early promoter. A FRT site has been inserted just downstream of the ATG initiation codon of the lacZ-Zeocin™ fusion gene. The FRT site serves as the binding and cleavage site for the Flp recombinase. The pFRT/lacZeo plasmid is transfected into the mammalian cell line of interest and cells are selected for Zeocin™ resistance. Zeocin™-resistant clones are screened to identify those containing a single integrated FRT site. The resulting Flp-In™ host cell line contains an integrated FRT site and expresses the lacZ-Zeocin™ fusion gene (see the *Figure 3-1* below). Note: Integration of the pFRT/lacZeo plasmid into the genome is random.

The second major component of the Flp-In™ System is the pcDNA5/FRT expression vector into which the gene of interest will be cloned. Expression of the gene of interest is controlled by the human CMV promoter. The vector also contains the hygromycin resistance gene with a FRT site embedded in the 5' coding region. The hygromycin resistance gene lacks a promoter and the ATG initiation codon.

The third major component of the Flp-In™ System is the pOG44 plasmid which constitutively expresses the Flp recombinase (Broach et al., 1982, Broach and Hicks, 1980, Buchholz et al., 1996) under the control of the human CMV promoter.

The pOG44 plasmid and the pcDNA5/FRT vector containing your gene of interest are cotransfected into the Flp-In™ host cell line. Upon cotransfection, the Flp recombinase

expressed from pOG44 mediates a homologous recombination event between the FRT sites (integrated into the genome and on pcDNA5/FRT) such that the pcDNA5/FRT construct is inserted into the genome at the integrated FRT site (see the figure below). Insertion of pcDNA5/FRT into the genome at the FRT site brings the SV40 promoter and the ATG initiation codon (from pFRT/lacZeo) into proximity and frame with the hygromycin resistance gene, and inactivates the lacZ-Zeocin™ fusion gene. Thus, stable Flp-In™ expression cell lines can be selected for hygromycin resistance, Zeocin™ sensitivity, lack of ß-galactosidase activity, and expression of the recombinant protein of interest.

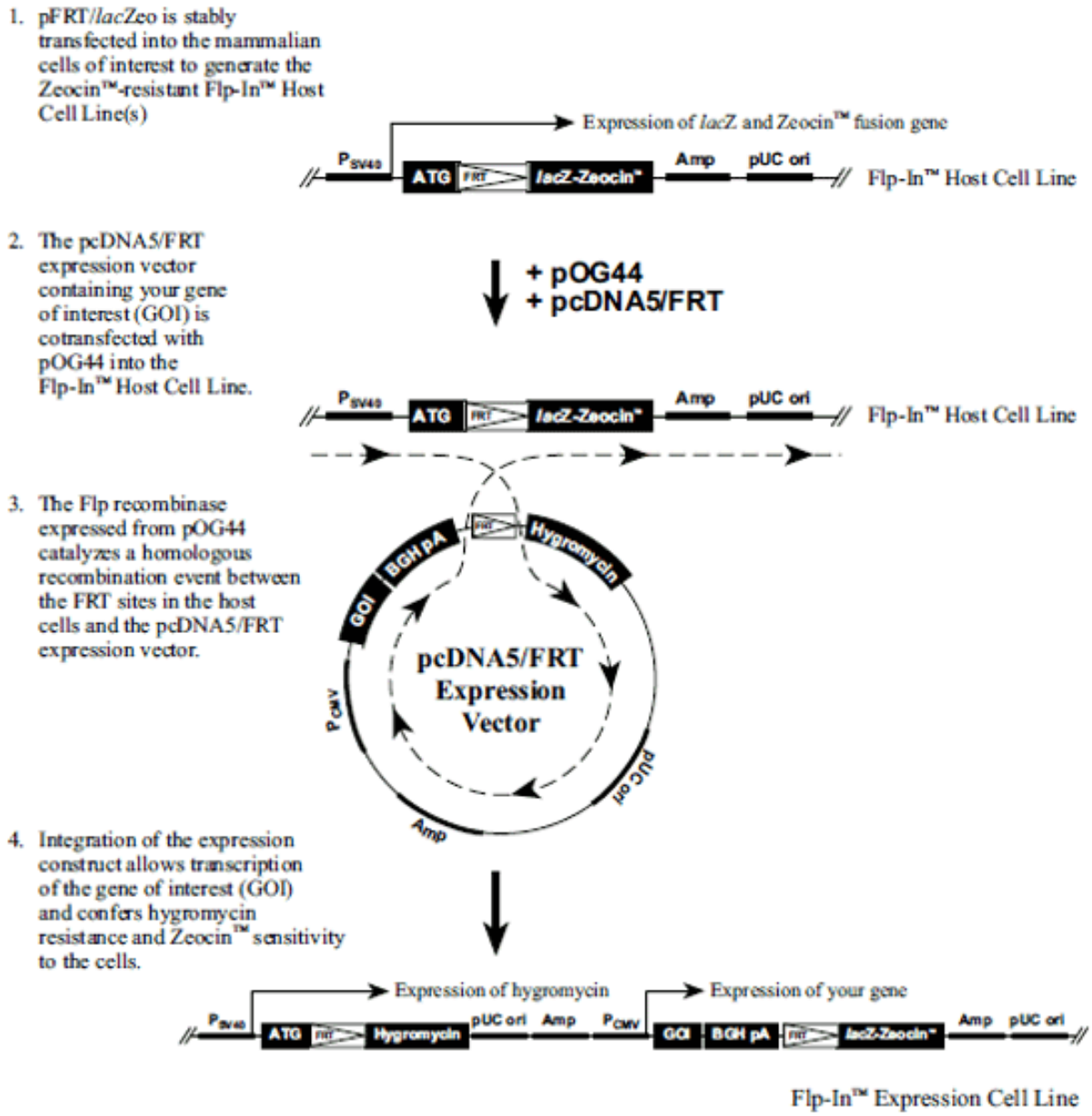


Figure 3-1 Diagram of the Flp-In™ System. The figure illustrates the major features of the Flp-In™ System as described. For a brief description about FRT sites and the mechanism of Flp-mediated recombination, see published reviews (Craig, 1988, Sauer, 1994). With permission.

3.1.7 Cell line cloning and selection

360 clones were set up in 6 x 96 well plates (60 usable wells per plate). Transfected stable culture was subjected to serial dilution in order to reach a cell concentration of approximately

360 cells in 36ml total volume (1 cell per 100 µl aliquot per well) of CD-CHO media containing 500 µg/ml HygromycinB. 100µl aliquots were dispensed aseptically into the plate wells and a sterile cover placed over the wells. Plates were incubated at 37^0C 5% CO_2 and monitored daily by optical microscopy for five weeks. Wells where doubling was observed, were transferred into increasingly larger volumes to 24 well plate, then to 6 well plate and finally to 125 ml flasks. Emerging clones were compared based on target product production and growth performance. Titer levels were assessed via GAA diagnostic assay (3.2.5.2) and western-blot analysis (3.2.3) Viability and growth performance were assessed via Vi-cell™ XR Cell Viability Analyser (Beckman Coulter, High Wycombe, UK).

3.1.8 Lab scale fermentation

The selected clone was thawed using standard protocol and resuspended in 30ml prewarmed CD-CHO under moderate agitation at 37^0C 5%CO_2 using single use sterile bottom baffled cell 250ml cell flasks After overnight incubation 250µg/ml of Hygromycin B was mixed into the broth and cell viability and count assessed via ViCell. Cells were left in incubation conditions until they reached a cell number > 12 10^6 cell/ml (8 days circa). Harvest would then follow.

3.1.9 Cell banking

Cell banking was performed according to standard procedure as per (Stacey and Masters, 2008) under sterile conditions.

An actively growing cell line in CD-CHO Hyg+ media was checked via ViCell for viability and cell count. After ensuring viability was > 95% and cell count was above 10^7 cells/ml, the following protocol was followed.

- Determine volume of freezing medium based on number of aliquots to be prepared with each aliquot containing 1ml of cells in freezing medium.

- Freezing medium composition: 10% cryoprotectant dimethyl sulfoxide (DMSO), 90% CD-CHO.
- Determine volume of cell broth needed to achieve required cell number.
- Vol needed (ml) = desired concentration (cell/ml) x total volume (ml) / starting concentration (cell/ml)
- Transfer measured volume needed into centrifuge tube and spin 10min 2000g, discard supernatant.
- Resuspend pellet gently into freezing medium, aliquot into prelabelled cryo tubes.

To achieve the highest levels of survival, vials were placed into Mr. Frosty (Thermo Scientific) containing isopropyl alcohol for controlled cooling overnight into -70^{o}C freezer, and transferred into liquid nitrogen the next day for indefinite storage.

3.2 Product characterization

3.2.1 SDS-PAGE

NuPAGE® Novex® Bis-Tris Gels (ThermoFisher Scientific USA), 4-12%, 1mm 12well precast gels were used for protein separation.

Materials/equipment needed:

- 1X MOPS running buffer (prepare from 20X MOPS stock)
- Appropriate ladder to cover range of expected protein size
- Pre-set 70^{0}C water bath
- Vortex
- XCell SureLock™ Mini-Cell.
- 0.5ml capped vials
- Deionized water
- NuPAGE LDS sample buffer

For non-reduced sample preparation mix 0.5µg of protein was added to 2.5µl 4X LDS Sample buffer in plastic vials. Up to 7.5µl of deionized water to reach 10µl reaction volume. Briefly and vortex for few seconds. Samples were placed in a water bath for 10min. Material was then transferred to the gel well and the protein ladder loaded.

Assemble XCell SureLock™ Mini-Cell following manufacturer specifications, add 600ml 1X MOPS to lower buffer chamber and 200ml to space in between gels (use dummy gel if running 1 single gel). Place lid over gel chamber, attach electrodes and set gel to 200V constant, 50min run time.

At the end of the run, the gel was removed by opening casing gently, then placed in weight boat and stain by pouring 50ml readymade Comassie blue solution over the gel and agitate gently for 4h up to overnight. The gel was destained by multiple washes in de-stain solution (40% Methanol, 10% Acetic Acid). Destained gel was then scanned for image analysis or stored for further use. Alternatively SYPRO® Ruby Protein Gel Stain (Thermo Fisher Scientific, USA) could be used. SYPRO is a highly sensitive, ready-to-use fluorescent stain for the detection of proteins separated by polyacrylamide gel electrophoresis (PAGE), no destaining is required.

3.2.2 Protein absorbance – NanoDrop

Protein concentration reading was needed at multiple stages throughout the project. Two instruments were used for this purpose. Thermo Fisher Scientific Nanodrop ND1000 and Tecan multimode microplate reader.

NanoDrop absorbance range goes from 220nm to 750nm. Its main benefit is sample volume requirement as it needs only 1µl of material to produce reproducible and fast result.

3.2.3 Detection of product – Western blot

GAA presence in samples was assessed via western blot analysis

Invitrogen XCell II blot module was used for western blot analysis. This kit includes the cassette for gel electrophoresis and membrane protein transfer which connects to power block via dedicated electrodes. Before the transfer of proteins from SDS-PAGE to membrane can happen, the SDS-PAGE gel must be run under standard conditions to separate protein mixtures based on charge and size. For the materials and protocol concerning gel running, refer to 3.2.1 SDS-PAGE. This section assumes the gel is already ready for transfer (gel staining-destaining not required for protein transfer to membrane).

Day 1:

SDS_PAGE

- Thaw proteins and ladder on ice.
- Prepare MOPS Running Buffer (1x) (dilute from stock), store at room temperature (RT).
- Prepare master mix of SDS sample buffer (3X) and dilute 1:10 DTT (1DTT: 10SDS) in it (DTT from 30X to 3X) Do not need to add B-MEtOH, is already in DTT (prepare in excess).
- Pipette X µL protein (50µg) and X µL of SDS/DTT3X mix (mix: total→ 1:3) into Eppendorf tube.
- Vortex samples, quickly spin and heat samples at 70^0C for 10 min.
- Vortex and spin one more time to collect volume at the bottom of the vial
- Pipette samples in NuPAGE® Novex® Bis-Tris Gels (Thermofisher scientific USA), 4-12%, 1mm 12well precast gels, add appropriate prestained ladder
- Fill cassette with 0.6L MOPS running buffer (1x).
- Run gel at 200 V for 50min

Transfer

- Keep gel in cassette soaked in running buffer until ready for transfer
- Prepare 1X Transfer buffer from 20X Nupage (for 1L: 50ml 20X Nupage, 1ml antioxidant, 100ml MeOH, 849ml RO water)
- Soak blotting pads in transfer buffer and push out air bubbles with a spatula
- Prepare PVDF membrane: cut membrane to match gel size. Never touch membrane directly: use wrapping paper and gloves. Pre-soak membrane 30sec in 100% MeOH and place in transfer buffer together with blotting pads.

- Follow manufacturer protocol instructions from page 10 to page 16 of XCell II™ blot Module, Manual part no. IM9051. Rev. Date: 14 December 2009. Transfer at 30V 1h. (https://tools.thermofisher.com/content/sfs/manuals/blotmod_pro.pdf)

Blocking

- Prepare 100 mL (per gel) blocking solution (non-fat milk 5% in PBS-T) and store at 4^0 C.
- Discard gel
- With protein side facing up, cut upper right corner of the membrane to mark membrane orientation, cut membrane to size.
- Place each blot in 50 mL blocking solution in a weight boat, agitate for 1h at RT.

Primary Antibody incubation

- Primary anti-GAA rabbit monoclonal antibody was purchased from Abcam (cat.# ab137068 (Abcam, 2013), working dilution (1:1000)
- Prepare primary antibody solution in milk 5% in PBS-T(10ml;10µl)
- Place blot into solution with protein side facing up, shake at 4^0C overnight.

Day 2:

- Wash blots in PBS-Tween 20 (0.2%) for 3x 10 min changing buffer between intervals

Secondary Antibody incubation

- Secondary anti rabbit polyclonal HRP-linked antibody was purchased from New England Biolabs (NewEnglandBiolabs, 2013)
- Prepare secondary antibody solution in milk 5% in PBS-T(10ml;10µl)
- Incubate membrane 1h RT with secondary antibody mix under moderate agitation
- Wash blots in PBS-Tween 20 (0.2%) 3x 10 min changing buffer between intervals.

Detection

- Use SigmaFast Diaminobenzidine (Sigma-Aldrich, 2014) detection reagent (DAB).
- Mix reagent pills (gold and white) in 5ml RO water.
- Drop reaction mix onto membrane

3.2.4 Identification of product – MALDI TOF MSMS

Mass Spectrometry was used to identify the product in the cell culture supernatant. Digested samples were analysed using a Bruker UltrafleXtreme MALDI-TOF-TOF mass spectrometer (Bruker Coventry UK)

Samples were first run on a 12%Bis-Tris SDS gel in order to separate proteins based on molecular weight. Bands were then excised one by one using a scalpel and subjected to in gel tryptic digestion overnight at $4^{O}C$ in digestion buffer (25µl sequencing grade modified trypsin (cat# V511 Promega USA), 470µl H_2O, 5 µl N_4HCO_3 1M, 50µl acetonitrile).

Sample excision and digestion protocol:

Spot Excision

- Wash the gel with water (2 x 10 mins).
- Excise spot/band of interest with clean scalpel cutting as close to the edge of the spot as possible. (It is important to reduce the amount of background gel).
- Cut the excised spot into small squares (~ 1 x 1 mm) and transfer into a 0.5 m1 microfuge tube.

Reduction and Alkylation

Always use freshly prepared NH4HCO3 (prepare a 100 mM stock and dilute as necessary).

- Wash the gel particles with 100 ml of 50 mM NH4HCO3:acetonitrile (1:1) for 15 min. Spin down and remove the liquid.
- Add 100 ml acetonitrile and leave for 15 min. until the gel pieces have shrunk (they become white and stick together).
- Spin and remove all liquid.
- Swell the gel pieces in 10 mM DTT in 50 mM NH4HCO3, adding enough liquid to cover the gel (ca 50 ml). Incubate for 30 min. at 56 °C.
- Spin down the gel pieces and remove excess liquid. Shrink the gel pieces briefly in acetonitrile.
- Remove acetonitrile and add 55 mM iodoacetamide in 50 mM NH4HCO3, adding enough liquid to cover the gel (ca 50 ml). Incubate for 20 min. at room temp. in the dark.
- Spin down the gel pieces and remove iodoacetamide solution.
- Wash gel pieces with 100 ml of 50 mM NH4HCO3 for 15 min. Spin gel pieces and remove all liquid.
- Repeat step 8.
- Shrink gel pieces with acetonitrile. Spin and remove all liquid.
- Dry gel pieces in vacuum centrifuge.

In-gel Digestion

- Rehydrate the gel pieces in 20 ml digestion buffer (10 mM NH4HCO3, 10% acetonitrile) containing 10 ng/ml of trypsin at 4 °C (on ice) for 30 min. (After 15 min. check samples and add more trypsin solution if all the liquid has been absorbed by the gel pieces).
- Remove the remaining supernatant. Add 10 ml of the digestion buffer without trypsin to cover the gel pieces and keep wet during enzyme cleavage.
- Leave samples at room temperature overnight.

Extraction of Peptides

- Add 5 ml acetonitrile and sonicate for 15 mins.
- Spin down gel pieces and collect the supernatant in a 0.5 ml microfuge tube.
- Add 10 µl 50% acetonitrile with 5% formic acid and sonicate for 15 mins.
- Spin down gel pieces and collect the supernatant.
- Pool the supernatants.
- Retain the gel pieces until the supernatant has been analysed by mass spectrometry.

0.5µl of sample and standard was placed on the MALDI plate and allowed to dry, followed by a droplet of 1ul of Super-DHB matrix. As the matrix dries, it forms a clean crystalized layer over the sample which immobilizes the peptides. MALDI plate was inserted in the instrument and appropriate method selected. MALDI-TOF-TOF was conducted using an UltrafleXtreme MALDI-TOF instrument (Bruker, Coventry, UK) in positive ion reflector mode and 50% laser power and MS-MS was conducted on the ten most intense peaks for each target spot. Generated peptide masses with an ion score exceeding the threshold set for $p < 0.05$ were interrogated using the Mascot algorithm (matrix-science.com) to search all taxonomies in the SwissProt database. Instrument run setup options: fixed modifications, carbidomethyl (C); variable modifications, oxidation (M); mass values: monoisotopic; protein mass: unrestricted; peptide mass tolerance: +/− 100 ppm; fragment mass tolerance +/− 0.5 Da. Instrument MALDI-TOF-TOF

3.2.5 Quantification of product

3.2.5.1 Via AlphaLISA

This assay allows identification of a specific antigen in a relatively quick and reproducible way. When the antigen of interest is recognised by its specific antibody, it forms a complex in which an acceptor and a donor beads are present and adjacent. Only when this complex is formed, excitation of donor bead by fluorescent light at 680nm, causes the attached acceptor bead to emit light at 615nm which is detected by the AlphaScreen analyser. Reaction mixture composed of samples containing antigen, biotinylated antibody and acceptor beads is added to wells of a 96 well OptiPro plate, and incubated 1h at RT and then donor beads are added. After 30min incubation, the plate is read.

Since the manufacturer does not provide already made detection kit for GAA, the assay had to be developed and optimised. This involves choosing the right antibody pair, biotinylating one of the two antibodies, and calibrating analyte sensitivity and response. The AlphaLISA development guide was followed (PerkinElmer, 2012).

Materials:

Item	Manufacturer	Catalogue
AlphaLISA Acceptor beads	Perkin Elmer, Inc.	(1mg) 6772001
AlphaScreen Streptavidin-coated Donor beads	Perkin Elmer, Inc.	(1mg) 6760002S
Antibody 1: anti-GAA rabbit monoclonal	Abcam	ab137068
Antibody 2: anti-GAA rabbit polyclonal	Sigma-Aldrich Co.	SAB2100872
Microplates: ½ AreaPlate-96	Perkin Elmer, Inc.	6005560
TopSeal-A Adhesive Sealing Film	Perkin Elmer, Inc.	6005185
ChromaLink™ Biotin Antibody Labelling Kit	SoluLink Inc.	B-9007-105K
NHS activated biotinylation reagent	SoluLink Inc.	B1001-105
Carboxymethylamine hemihydrochloride (CMO)	Sigma-Aldrich Co.	C13408

Sodium cyanoborohydride	Sigma-Aldrich Co.	152159
Zeba desalting columns	Pierce (ThermoFisher Scientific Inc.)	89889 (2 mL)
Proclin-300	Sigma-Aldrich Co.	48912-U
Tween-20	Pierce (ThermoFisher Scientific Inc.)	28320
Casein 5% Alkali-soluble solution	Novagen (EMD Chemicals Inc.)	70955
Dextran 500	Sigma-Aldrich Co.	D1037
Triton-X100	Pierce (ThermoFisher Scientific Inc.)	28314
Streptavidin-Sepharose beads	GE Healthcare, Inc.	17-5113-01

Table 3-4 PerkinElmer AlphaLISA reagents

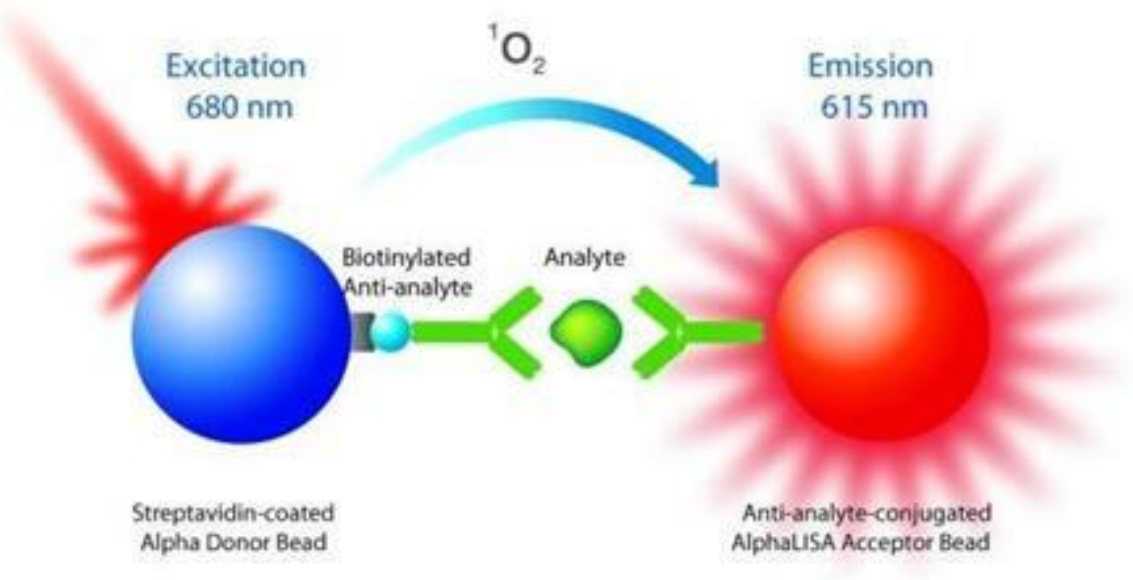


Figure 3-2 AlphaLISA principle. In presence of analyte, the complex donor bead/acceptor bead forms and excitation of donor bead causes release of O_2 which transfers to acceptor bead and emit light at 615nm. Adapted from PerkinElmer (PerkinElmer, 2014)

3.2.5.2 GAA diagnostic assay

This method was used in this work to detect and measure amount of GAA in different clonal lines.

The GAA diagnostic assay developed by the University of Kumamoto Japan (Okumiya et al., 2006) is used in the medical field to detect presence of GAA in leukocytes of people suspected of being affected by Pompe disease. This method employs glycogen and 4-methylumbelliferyl-α-D-glucopyranoside as substrates for measuring GAA activity and incorporates acarbose to eliminate the interference of unrelated α-glucosidases (predominantly maltase-glucoamylase). After allowing GAA to degrade the substrates present in solution, an alkaline buffer is added to the reaction mix to stop GAA activity. This causes 4-methylumbelliferone to fluoresce at different wavelengths from unhydrolysed substrate, thereby permitting its measurement in presence of vast excess of unhydrolysed substrate.

Reagents.

- Buffer: 0.15M citrate/phosphate buffer pH 4.0
- Substrate: 4-methylumbelliferyl-α-D-glucopyranoside 250mg (MW 338.31, Sigma Aldrich). Dissolve 250mg substrate in 503ml of buffer by warming to 80^{o}C. Store stock substrate at -20^{o}C.
- Inhibitor: Acarbose 1g (MW 645.6, Sigma Aldrich). Dissolve 1g acarbose in 442.6 ml buffer to give 3.5mM solution.
- Standard: Stock standard solution. Dissolve 176mg 4-methylumbelliferone (MW 176 Sigma Aldrich) in 1L H_2O. Dissolve 100ul stock standard solution in 19.9ml distilled water to give working standard (1nmol per 200µl).
- Stopping reagent: 1M glycine buffer pH10.4. Dissolve 75g glycine and 58g sodium chloride in 1l H_2O. Add 55.7ml of this solution to 44.3ml 1M sodium hydroxide.
- Substrate preparation: Dilute 250µl 3.5mM acarbose in 249.75ml stock substrate solution to give substrate working solution containing 3.5µM acarbose.

Method.

- In 2ml Eppendorf vials prepare standard, blank and samples for analysis in triplicate.
- Blank: 5µl H_2O + 100µl substrate working solution
- Sample/standard: 5µl sample/standard + 100µl substrate working solution
- Cover plate and incubate at 37ºC for 15-120min (time must be optimized)
- Add 1ml stopping reagent, mix.
- Transfer 500ml from each tube in 96well flat bottom polypropylene plate and read fluorescence at 360-480nm.

3.3 Fermentation and downstream processing

3.3.1 Harvest

Cells were grown under batch culture conditions in 500ml bottom-baffled shake flasks containing 125ml of CD-CHO media (with 250µg/ml hygromycin B). Flasks were incubated under moderate agitation at 37^0C 5% CO_2, until they reached at least 10 million cells/ml (8 days). Cell number and viability was checked daily using a ViCell (Beckman Coulter USA). On harvest day, cell broth from each flask was transferred to multiple 50ml conical tubes and spun at 2000g for 10 minutes to pellet the cells. The supernatant was pooled and it was filtered through a 0.22µm Stericup® (Merck Millipore USA). Harvest Cell Culture Fluid (HCCF) was then concentrated 4 fold using multiple 10000MWCO (up to approximately 30ml)

3.3.2 Capture step – Anion Exchange chromatography

A GE AKTA Avant 25 (GE Healthcare Life Sciences USA) system was used to set up the anion exchange chromatography capture step. The column used was a GE CaptoQ HiTrap 1ml, equilibration buffer 50mM Sodium Acetate pH6.0 (Buffer A), elution buffer 1M Sodium Chloride in 50mM Sodium Acetate pH6.0 (Buffer B) at a flowrate of 1 ml/min. Concentrated HCCF was diluted 1:4 into equilibration buffer to adjust to optimal conductivity and favour binding of target material to resin. 30 ml loading volume at a time was loaded onto column using

AKTA sampling port, followed by 14CV wash (100% buffer A), 40CV linear gradient elution (0-35% buffer B), 5 CV strip (100% Buffer B) and regeneration (10 CV 0.5 M NaOH, 5 CV buffer A, 10 CV 20% EtOH). 2ml fractions of column wash and eluate were collected and absorbance at 280nm monitored throughout the whole run.

3.3.3 Intermediate step – Hydrophobic interaction chromatography

3.3.3.1 Binding condition screening

This step was performed at BioMarin Pharmaceutical Inc. laboratories with Dr. John Henstrand's group assistance. GAA reference material was used (BMN 103, 2.09 mg/ml).

Resins

B) Butyl Sepharose 4FF (GE Healthcare, USA cat# 17098001)

P) Phenyl Sepharose 6 Fast Flow (High Sub) (GE Healthcare, USA cat# 17-0973-03)

Buffer A

a) 50 mM sodium acetate 1.0 M Sodium Chloride pH 5.0

b) 50 mM sodium acetate 1.5 M Sodium Chloride pH 5.0

c) 50 mM sodium acetate 2.0 M Sodium Chloride pH 5.0

d) 50 mM sodium acetate 1.0 M Ammonium Sulphate pH 5.0

e) 50 mM sodium acetate 1.5 M Ammonium Sulphate pH 5.0

f) 50 mM sodium acetate 2.0 M Ammonium Sulphate pH 5.0

Buffer B

50 mM sodium acetate pH5 .0

Procedure

2mg reference protein was bound to two resins in six equilibration buffers (Buffer A). Binding and recovery was assessed by spinning reaction mixture in micro centrifuge 5min 12000g and determining protein concentration of supernatant via NanoDrop N1000. Reactions were performed in 1.5ml Eppendorf tubes.

- Binding

T = 0h

Mix 200µl of resin (80% slurry) in 1ml Buffer A, add 200µl 10mg/ml protein (2mg) in 1.5 ml Eppendorf tube.

Incubate 4^0C under moderate agitation overnight.

T = 12h

Spin and determine supernatant concentration

- Wash

Wash 3 times with 1 ml Buffer A, vortex, spin and collect supernatant. Determine concentration after each wash.

- Elution

Resuspend beads in 1 ml of Elution buffer 1 (75% A + 25% B), vortex, spin, determine concentration of supernatant. Repeat with Elution buffer 2: (50% A + 50% B) and Elution buffer 3: (25% A + 75% B).

- Strip

Resuspend beads in 1 ml Buffer B, vortex, spin and determine protein concentration of supernatant. Repeat the strip step three times total.

3.3.3.2 Separation method details

This step was run on AKTA Avant 25 (General Electrics) using an in-house packed (At BioMarin laboratories) column containing Butyl Sepharose 4FF (Fast Flow) resin. The packed column measured 8cm in bed height, and 0.34cm radius giving a total bed volume (V) = 8cm (h) x 0.34^2cm (r) x π = 2.94ml. Equilibration buffer 1M Ammonium Sulphate pH 5.0 (Buffer A), Elution buffer 50mM Sodium Acetate pH 5.0, flowrate 1ml/min.

The anion exchange eluate fraction was buffer exchanged in Buffer A using Amicon 10000MWCO spin filters, then 4CV loaded on the column using AKTA sampling port followed by 14 CV wash (100% buffer A), 40CV linear gradient elution (0-35 % buffer B), 5CV strip (100 % Buffer B) and regeneration (10 CV 0.5 M NaOH, 5CV buffer A, 10 CV 20 % EtOH). 2 ml fractions of column wash and all elution peaks were collected and absorbance at 280 nm monitored throughout the whole run.

3.4 Impurities characterization and cell stress

3.4.1 Zymography

Novex™ 12 % Zymogram Casein Protein Gels (cat# EC6405BOX Thermo Fisher Scientific USA) were used to visualise protease presence pre- and post- capture step in Null (CHO-S cat#R80007 Thermo Fisher) and GAA CHO cells. Samples were first denatured in SDS buffer and separated based on size similar to a standard SDS-PAGE gel, then they were renatured by incubating the gel in non-ionic detergents (renaturing buffer, Invitrogen) to re-establish enzymatic activity. The gel was then incubated overnight at RT in developing buffer (Invitrogen USA) which adds divalent metal cations required for enzymatic activity. The following day the gel was stained with Comassie blue and de-stained in 40 % MeOH, 10 % acetone. Regions of

protease activity appear as clear bands against a dark blue background where the protease has digested the substrate.

Sample and system preparation:

After determining samples' protein concentration by absorbance at 280 nm, 10 µg are mixed with 5µl of provided Tris-Glycine SDS Sample Buffer 2X (LC2676), and the sample mix is brought to 10 µl volume using HPLC grade water. Heat is not required, as it would permanently denature proteases in samples.

1L of 1X Tris-Glycine SDS running buffer is prepared by diluting 100ml of 10X buffer (LC2675) into 900ml of di-ionized water.

Place empty gel in XCell SureLock® Mini-Cell (Thermo Fisher Scientific, USA), lock in place and add about 700ml of 1X Tris-Glycine running buffer according to manufacturer. Add 10ul of samples to gel wells and appropriate molecular weight standard.

Run gel at 125V constant for 90 minutes. Expected current should be 30-40 mA/gel.

As gel is running prepare 1X renaturing and developing buffers from stock 10X solution.

Remove gel from casing and place it in 1X renaturing buffer for 30min at room temperature with gentle agitation, then decant buffer and replace with 1X developing buffer. Incubate at 37 ^{O}C overnight with gentle agitation. Next day proceed to Comassie blue staining and de-staining (40 % MeOH, 10 % Acetic Acid). Bands containing proteases are displayed in white over a blue background.

3.4.2 CHO HCP - AlphaLISA

Commercially available CHO HCP quantitation kit (cat# AL301C Perkin Elmer, Waltham MA USA) was used to quantify total HCP. Kit components

Materials:

Item	Specifications
AlphaLISA Anti-CHO HCP Acceptor beads stored in PBS, 0.05% Proclin-300, pH 7.2	50 µL @ 5 mg/mL (1 brown tube, white cap)
Streptavidin (SA)-coated Donor beads stored in 25 mM HEPES, 100 mM NaCl, 0.05% Proclin-300, pH 7.4	200 µL @ 5 mg/mL (1 brown tube, black cap)
Biotinylated Antibody Anti-CHO HCP stored in PBS, 0.1% Tween-20, 0.05% NaN3, pH 7.4	50 µL @ 500 nM (1 tube, black cap)
AlphaLISA CHO HCP (1 µg), lyophilized analyte *	1 tube, clear cap
AlphaLISA HiBlock Buffer (10X)	10 mL, 1 small bottle
Microplates: ½ AreaPlate-96	Cat# 6005560
TopSeal-A Adhesive Sealing Film	Cat# 6005185

Table 3-5. Perkin Elmer CHO HCP AlphaLISA reagent list

Protocol:

Prepare 1X AlphaLISA HiBlock Buffer: add 2.5 mL of 10X AlphaLISA HiBlock Buffer to 22.5 mL H2O.

Prepare CHO HCP analyte standard dilutions: reconstitute lyophilized CHO HCP (1 µg) in 100 µL H2O. Prepare standard dilutions as follows (change tip between each standard dilution):

Tube	Vol. of CHO HCP (µL)	Vol. of diluent (µL) *	[CHO HCP] in standard curve	
			(g/mL in 5 µL)	(pg/mL in 5 µL)
A	10 µL of reconstituted CHO HCP	90	1E-06	1 000 000
B	60 µL of tube A	140	3E-07	300 000
C	60 µL of tube B	120	1E-07	100 000
D	60 µL of tube C	140	3E-08	30 000
E	60 µL of tube D	120	1E-08	10 000
F	60 µL of tube E	140	3E-09	3 000
G	60 µL of tube F	120	1E-09	1 000
H	60 µL of tube G	140	3E-10	300
I	60 µL of tube H	120	1E-10	100
J	60 µL of tube I	140	3E-11	30
K	60 µL of tube J	120	1E-11	10
L	60 µL of tube K	140	3E-12	3

*Table 3-6 CHO HCP standard curve preparation. *Standards diluted in cell culture medium.*

Prepare 2.5X AlphaLISA Anti-CHO HCP acceptor beads + biotinylated Antibody Anti-CHO HCP mix (25µg/ml). Add 50 µl of 5mg/ml AlphaLISA Anti-CHO HCP acceptor beads and 50 µl of 500mM Biotinylated Antibody Anti-CHO HCP to 9900 µl 1X AlphaLISA HiBlock buffer

Prepare 2X Streptavidin coated Donor beads. Under subdued light add 200 µl of 5 mg/ml Donor beads to 12300 µl of 1X AlphaLISA HiBlock Buffer

In 96well ½ area microplate add:

- 5 µl of each analyte standard dilution or sample in triplicate. Use 5 µl of sample diluent as blank.
- Add 20 µl of the 2.5X mix (freshly prepared)
- Incubate 60 min at RT
- Add 25 µl of a 2X streptavidin coated Donor bead (subdued lighting)
- Cover with black TopSeal-A Adhesive Sealing Film
- Incubate 30min at RT
- Read in EnVision AlphaLISA reader

Data analysis:

Calculate the average count value for the blank wells. Generate a standard curve by plotting the AlphaLISA counts versus the concentration of analyte. A log scale can be used for either or both axes. No additional data transformation is required. Analyse data according to a nonlinear regression using the 4-parameter logistic equation (sigmoidal dose-response curve with variable slope) and a $1/Y^2$ data weighting (the values at maximal concentrations of analyte after the hook point should be removed for correct analysis). Read from the standard curve the concentration of analyte contained in the samples. If samples have been diluted, the concentration read from the standard curve must be multiplied by the dilution factor.

3.4.3 HCP identification and quantification via tandem MSMS

3.4.3.1 MALDI TOF LC-MS/MS

For digestion, GAA harvest HCCF and IEX fractions were suspended in 50 µl 8 M urea, 100 mM ammonium carbonate. Next, 2 µl 450 mM dithiothreitol (DTT) was added and the mixture incubated at room temperature for one hour. This was followed by incubation with 20 µl of 100 mM iodoacetamide (IAM) for 15 min at room temperature to achieve carbidomethylation of the sample. Urea was diluted by the addition of 128 µl MilliQ water prior to incubation at 37◦C for 1 h with 5 µg modified sequence grade trypsin (Promega, Southampton, UK).

For separation via liquid chromatography (LC), samples were applied to a C18 Acclaim PepMap100 (Thermo, UK) 75 µm internal diameter x 15 cm (C18, 3 µm, 100A) for on-line reverse phase HPLC (NanoLC Ultimate3000, Thermo UK). Elution was performed with a linear acetonitrile gradient (solvent A = 0.05% TFA, solvent B = 0.05% TFA, 90% acetonitrile) in a 40 min cycle at a flow rate of 300 nl/min. LC was coupled to a fraction collector (Proteineer fcII, Bruker, Coventry, UK), dividing eluates into 120 fractions and mixing with matrix solution (α-cyano-4-hydroxycinnamic acid (CHCA), Sigma UK, reconstituted according to manufacturer's

guidelines in 89% acetonitrile, 0.1% TFA) prior to being spotted onto a MALDI-TOF target plate (MTP AnchorChip600384 T F, Bruker, Coventry, UK).

MALDI-TOF-TOF was conducted using an UltrafleXtreme MALDI-TOF instrument (Bruker, Coventry, UK) in positive ion reflector mode and 50% laser power and MS-MS was conducted on the ten most intense peaks for each target spot. Generated peptide masses with an ion score exceeding the threshold set for $p < 0.05$ were interrogated using the Mascot algorithm (matrix-science.com) to search all taxonomies in the SwissProt database. Acquisition software was set up as follow: fixed modifications, carbidomethyl (C) and variable modifications, oxidation (M); mass values: monoisotopic; protein mass: unrestricted; peptide mass tolerance: +/− 100 ppm; fragment mass tolerance +/− 0.5 Da. Instrument MALDI-TOF-TOF.

3.4.3.2 High resolution nano LC-MS/MS

Eluate fractions post anion exchange from three cell lines (GAA Cell line1, CHO-S Null and mAb producing) were screened via nanoLC tandem MS to determine HCP profile and presence of product. LC-MS/MS analysis was undertaken by first digesting samples with trypsin as described below, separating digested peptides on UPLC and lastly injecting them into the mass spectrometer.

For digestion, post-IEX column eluates fractions were concentrated via 10000 MWCO concentrators (Millipore USA) by a factor of four. 50 µl of sample were mixed with 8 M urea in 100 mM NH_4HCO_3 + 1 µl 450 mM DTT. This was allowed to incubate at RT for 1 h and followed by 15 min incubation at RT with 10 µl 100 mM iodoacetamide. Final protein concentration was determined with NanoDrop at 280 nm. The urea concentration was then diluted with 64 µl of HPLC grade water. Digestion with trypsin was performed for 1 h at $37^{O}C$ by adding to each sample 4 µg of sequencing grade Trypsin (Promega V5111 USA). A 2 mg/ml BSA (Thermo Fisher, USA) stock solution was used as a quantitation standard. 50 fmol of BSA

was prepared by serial dilution and spiked into each sample prior to separation to allow for label free quantitation.

For peptide separation, 10ul of sample was injected in triplicate onto a HSS T3 Acquity column (Waters, USA) 75 μm internal diameter x 15 cm (1.8 μm, 100 A) for on-line reverse phase UPLC (Acquity M-Class, Waters, USA). Elution was performed with a linear gradient from 3 to 40% B over 40 mins; (buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile) at a flow rate of 300 nl/min. The M-class mass spectrometer was coupled via a nanospray source to a Synapt G2-Si (Waters, USA) and data was collected in HDMSe mode. Data was analysed using Progenesis QIP and searched against SwissProt using MSe Search and a false detection rate of 4%. Progenesis QIP software provided resulting data in the form of normalised average abundance (molar equivalent) of each specie matched in SwissProt database in triplicate readings. Average of each triplicate was calculated eliminating eventual outliers. Average abundances were converted to mass by multiplying them by the respective molecular weights. Score and probability values were used to avoid identification of false positives. In particular Anova (p) probability factor threshold was set at 0.1max, which means that at p of 0.1, there is a 10% chance of false positive being identified.

Species were checked individually against UniProt database using accession IDs to determine biological functions (proteolytic activity). Isoelectric point was calculated using ExPASy pI calculator online tool (ExPASy. Compute pI/Mw tool). Hydropathicity index (GRAVY) was also calculated based on amino acidic sequence. Relative percentage of each protein out of the total were also calculated and used for resulting bubble graph diagram.

3.4.4 Lysosomal protein production pathway: TEM imaging

We know that only a small portion of produced GAA is secreted, while the vast majority is directed into the cell preacidified compartment and eventually ends up in lysosomes where it

fulfills its role of converting glycogen into glucose. In order to assess the morphological differences of lysosomal structure and content in GAA overexpressing cells, highly detailed images of its lysosomes were taken with transmission electron microscope and compared to lysosomes from the originating Null cell line.

Cell preparation. One vial of GAA Clone 1 and Null CHO were thawed into 30ml of prewarmed CD-CHO under sterile conditions following standard thawing protocol. After 24 incubation at 37^0C 5% CO_2 mild agitation, 250 mg/ml hygromycin B was added to the GAA Clone 1 flask but not to the Null. Flaks were placed back in incubator for 3 days and growth monitored daily. Cells were eventually passaged back to 0.3 10^6 cell/ml in fresh media and re incubated. Cell viability and number was monitored for the following 7 days, after which a small aliquot (150ul) was taken for TEM sample prep

Sample preparation. Cell culture aliquot was spun at 15k g for 15 minutes, supernatant decanted and pellet resuspended in primary fixing solution.

Primary fixation

- Resuspend pellet in 2% paraformaldehyde, 1.5% glutaraldheyde in 0.1M cacodylate buffer pH7.3 for 1h
- Wash in 0.1M cacodylate buffer
- Resuspend in 1% Osmium tetroxide in 0.1M cacodylate buffer 7.3pH, incubate 1h at 4^OC
- Wash in 0.1M cacodylate buffer 5 min
- Incubate in 0.5% uranyl acetate for 20min
- Wash in H_2O for 5 min
- Wash in 0.1M cacodylate buffer for 5 min

- Wash in H_2O for 5 min

Dehydration

- Wash in 25% EtOH 5 min
- Wash in 50% EtOH 5 min
- Wash in 70% EtOH 5 min
- Wash in 90% EtOH 5 min
- Wash 4 times in 100% EtOH 5 min in Analar and dry

Epoxy Resin

- Prepare resin (12 g agar, 8 g dodecenylsuccinic anhydride, 5 g methyl nadic anhydride, 0.65 ml N-benzyldimethylamine, Sigma Aldrich USA)
- Mix 2 part polypropylene oxide and 1 part resin for 1h
- Apply to sample for a minimum of 4h. Overnight incubation preferred
- Let harden for 24 h at 60^0C

Ultra-thin sections are cut at 70-80 nm using a diamond knife on a Reichetr ultracut E microtome.

Sections are collected on a 200 mesh copper grids.

Sections are stained with uranyl acetate and lead citrate.

Viewing and photography done on a Joel 1010 transmission electron microscope using Orius Gatan camera system.

RESULTS

4 Chapter one

Production of a stable CHO cell line for target product generation.

4.1 Introduction

This chapter focuses on the generation of a stable Chinese Hamster Ovary (CHO) cell line expressing the target protein GAA. This initial step was necessary in order to establish a method of target protein production that would supply material for investigation for the whole length of the doctorate project and beyond. The generation of an independent cell line producing stably GAA was necessary to provide material for further investigation and optimization of the specific downstream process that the manufacturer BioMarin runs in their facility with product BMN 701 (GAA).

The process of generation of a stable CHO involved multiple steps. Initially the human recombinant GAA gene was purchased and cloned into a vector plasmid using standard molecular biology techniques. This involved cutting the plasmid with appropriate restriction enzymes, cloning the gene into the vector and transforming competent *E. coli* cells in order to generate multiple copies of the genetic material. Antibiotic selection was then used to ensure that only the cells that inherited the gene were brought forward. Nucleic acid from competent cells was then purified using a Miniprep kit. The material extracted was tested via agarose gels to ensure gene insertion had happened correctly. Base sequencing was also performed to establish that no errors were present on the newly formed plasmid.

The purified genetic material was then used to transfect CHO cells. Before attempting stable cell line transfection and selection, we decided to transiently transfect CHO cells to test whether they could produce functional GAA. As explained in detail later on, stable clone generation is a far lengthier process compared to transient transfection as it involves integration of recombinant

material into host genome, however it is also a necessary step for long term studies as a stable clone produces target product indefinitely. After confirming via western blot that transiently transfected CHO cells were able to secrete GAA, we proceeded with stable transfection. The process took three attempts, each one taking between 6 to 8 weeks.

The last step in this process was to select the high producer cell line to bring forward. This was done via one round of limiting dilution cloning (LDC) and the generated lines production and growth performance was compared.

4.2 Plasmid generation

The generation of a CHO cell line producing the therapeutic enzyme GAA involved multiple steps ranging from plasmid cloning to bacterial transformation to genetic material amplification, multiple rounds of transient and stable transfection and finally cell clone selection. Early in the project we decided to use site-specific gene integration expression system Flp-In/FRT to overcome location effect of random recombinant DNA integration into host genome, considerably shortening the clone selection process. pcDNA/3.1 plasmid was also used as a back-up transfection strategy as it allows for efficient forward-orientation multiple cloning site via selectable marker (Geneticin®). Human GAA gene was purchased from Origene and cloned into the plasmids. The newly generated genetic material was used to transform TOP10 chemically competent cells to amplify it. Mini prep allowed for genetic material purification and transfection into CHO host could begin. The data presented here below shows a series of agarose gels and gene sequencing to characterize the genetic material. It ultimately confirms that only pcDNA/FRT plasmid successfully received the gene without errors. This plasmid was then used to transfect mammalian cells.

4.2.1 Cloning GAA into plasmid vectors

This first experiment was designed to open the two vectors (pcDNA3.1 and pcDNA5/FRT) using specific REs to allow cloning of GAA into both pcDNA/FRT and pcDNA/3.1. After digesting both vectors with Not1 and the GAA carrying vector (pCMV6-XL6) with both Not1 and Ase1 to linearize them, samples were loaded onto a 1% agarose gel and run for 75minutes at 80V. Results are shown in Figure 4-1 below.

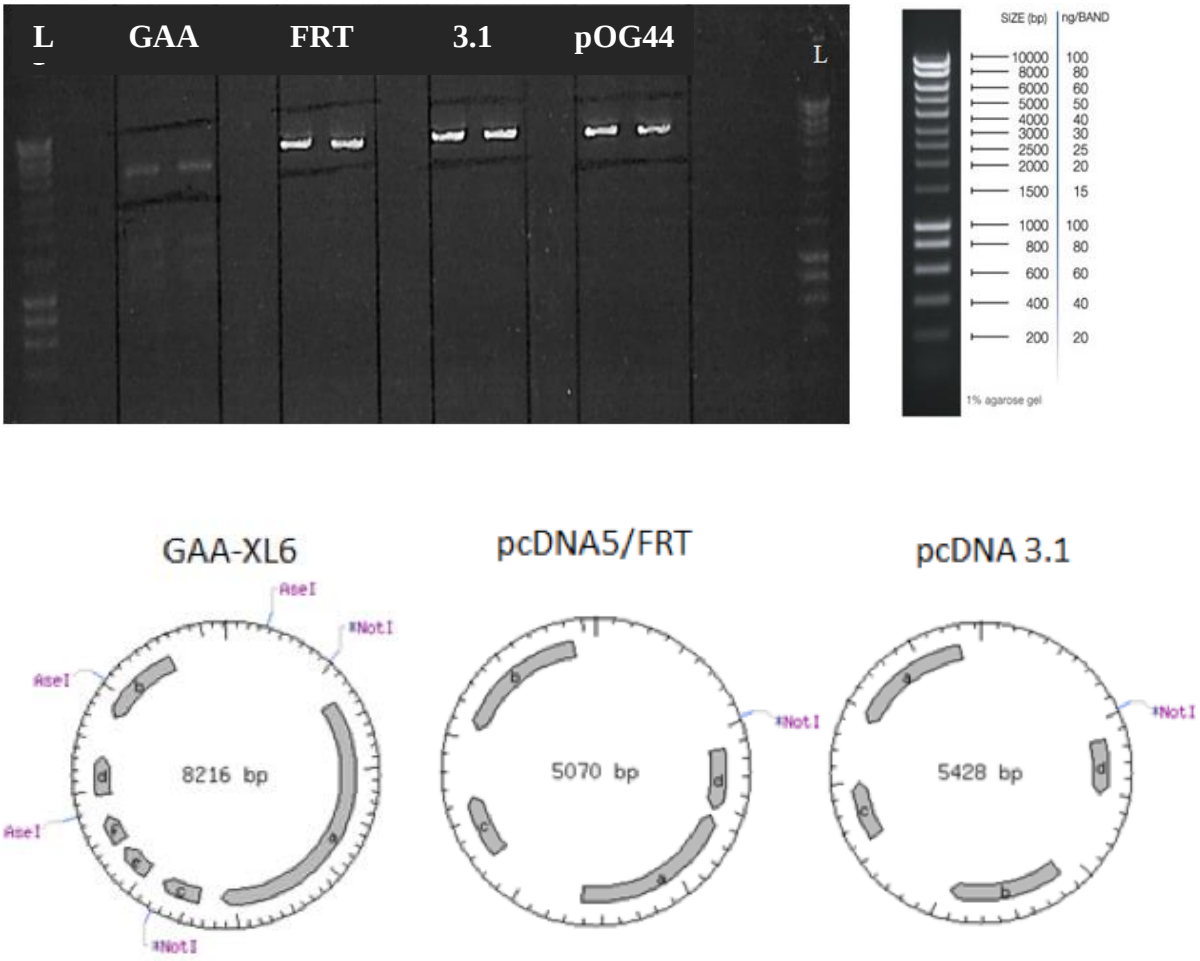


Figure 4-1: Linearized plasmids on agarose gel, from left: 1kd hyperladder, GAA, pcDNA 5/FRT, pcDNA 3.1, pOG44, 1kd hyperladder. (notebook 106: 22/04/13). GAA construct was digested with Not1 and Ase1 REs. Not1 excised the gene out (top band) and Not1/Ase1 combination cut the rest of the vector in smaller bands (four lower bands). Gel shows one clear band at around 4k bp and four smaller bands between 0.8k and 1.5k bp. This is consistent with what was

expected out of the digestion of the GAA-XL6 construct from Origene. Both vectors and pOG44 were cut once by Not1 RE to achieve linearization. The gel shows bands at around 5k, also consistent with expected results (pcDNA3.1: 5428bp, pcDNA5/FRT: 5070bp).

The DNA was purified with gel extraction kit, dephosphorylated using TSAP and ligated (pcDNA3.1 + GAA and pcDNA5/FRT + GAA) using T4 ligase enzyme as explained in section 3.1.1.3 DNA Purification, 3.1.1.4 Dephosphorylation and 3.1.1.5 Ligation in materials and methods above.

4.2.2 Transformation

Transformation is the process by which genetic information, in this case the construct (vector + GAA gene), is introduced into competent cells for replication. Upon successful transformation, competent cells replicate the nucleic acid of interest. The cells were grown on selective medium (ampicillin), because vector carried resistance gene this antibiotic, therefore cells that were able to grow had inherited a copy of the construct.

Different competent cells were used for this experiment, however only Top10 from ThermoFisher worked. The membrane of competent cells is made chemically leaky, so that under specific conditions (heat shock of 30 seconds at 42^0C) they allow genetic material to cross it.

OneShot Top 10 chemically competent cells were transformed using genetic material produced with ligation step. Transformed cells were plated on LB agar + ampicillin plate and incubated overnight at 37^0C. Next day the plates were taken out of the incubator and cell colonies were clearly visible for both constructs. Alongside the two samples of interest, positive and negative controls were also plated.

Positive control was provided by the manufacturer in the form of a plasmid (pUC19 Control DNA), which is used to verify the efficiency of the competent cells. Plates showed growth.

As a negative control the competent cells were plated without genetic material. No cells grew on selective media proving the fact that cells need to inherit vector carrying antibiotic resistance in order for them to grow.

Two single colonies per plate were aseptically picked and incubated into 5ml of LB-Amp media, at 250rpm 37^0C overnight. Next morning tubes were harvested and DNA purified via Miniprep for subsequent testing (sequencing and size confirmation).

4.2.3 Direction of insertion

During ligation it is not possible to control the direction of insertion of the genetic material into the vector. As a result, during transformation, some cells will acquire the plasmid with the gene inserted the correct way, and some will receive the plasmid with the gene inserted the other (wrong) way around. As only the gene inserted the right way will produce a functional protein, direction of insertion must be assessed. This was done by selecting a RE capable of cutting the genetic material within the gene (preferably not in the middle) and elsewhere in the vector (outside the gene), to avoid creating bands of similar sizes which are hard to distinguish on an electrophoretic gel. By checking band size after digestion, it was possible to determine if insertion happened correctly.

According to NEBcutter online digestion tool (Vincze et al., 2003) few REs cut both constructs twice in convenient locations:

pcDNA3.1 + GAA	**Band sizes (bp)**	**pcDNA5/FRT + GAA**	**Band sizes (bp)**
Pfl M1	7.3/1.8	Pfl M1	7.3/1.8
Sal I	5.7/3.5	Sal I	5.3/3.5
Ahd I	7/2.1	Ahd I	6.7/2.1
Stu I	n/a	Psh AI	n/a

Sac II	n/a	n/a	n/a

Table 4-1 Restriction enzymes compatible with both constructs. NOTE: Pfl MI cuts twice in the gene, therefore not usable.

Both Sal I and Ahd I cut the plasmid in convenient locations and create bands of different enough molecular sizes. Sal I was selected as the more economical between the two.

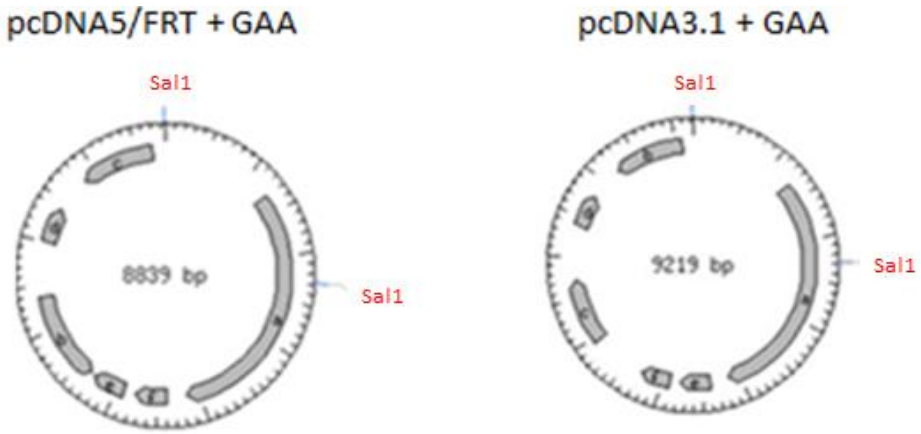


Figure 4-2. Sal 1 digestion of constructs to check for direction of insertion.

As seen in *Figure 4-2*, in pcDNA 3.1 + GAA construct, Sal I produces two bands of 3.5 and 5.6 bp if direction of insertion is right, and two bands at 4.6 and 4.5 bp if direction is wrong. In pcDNA 5/FRT + GAA construct, Sal I produces two bands of 3.5 and 5.3 bp if direction of insertion is right, and two bands at 4.6 and 4.2 bp if direction is wrong. So essentially we expect to see two separate bands if right or two overlapping bands if wrong.

Construct	**3.1 + GAA**		**FRT + GAA**	
direction	right	wrong	right	wrong
Sal I	3.5	4.6	3.5	4.6
	5.6	4.5	5.3	4.2

Table 4-2 Estimated construct sizes (in kb) after digestion with Sal I

Before checking for direction of insertion however, it was important to check that the re-linearized plasmid molecular weight matches what expected. To do this, a simple 2 point digestion with Not1 RE was performed to open the plasmid (Figure 4-3 below), and digestion product was run on a 1% agarose gel.

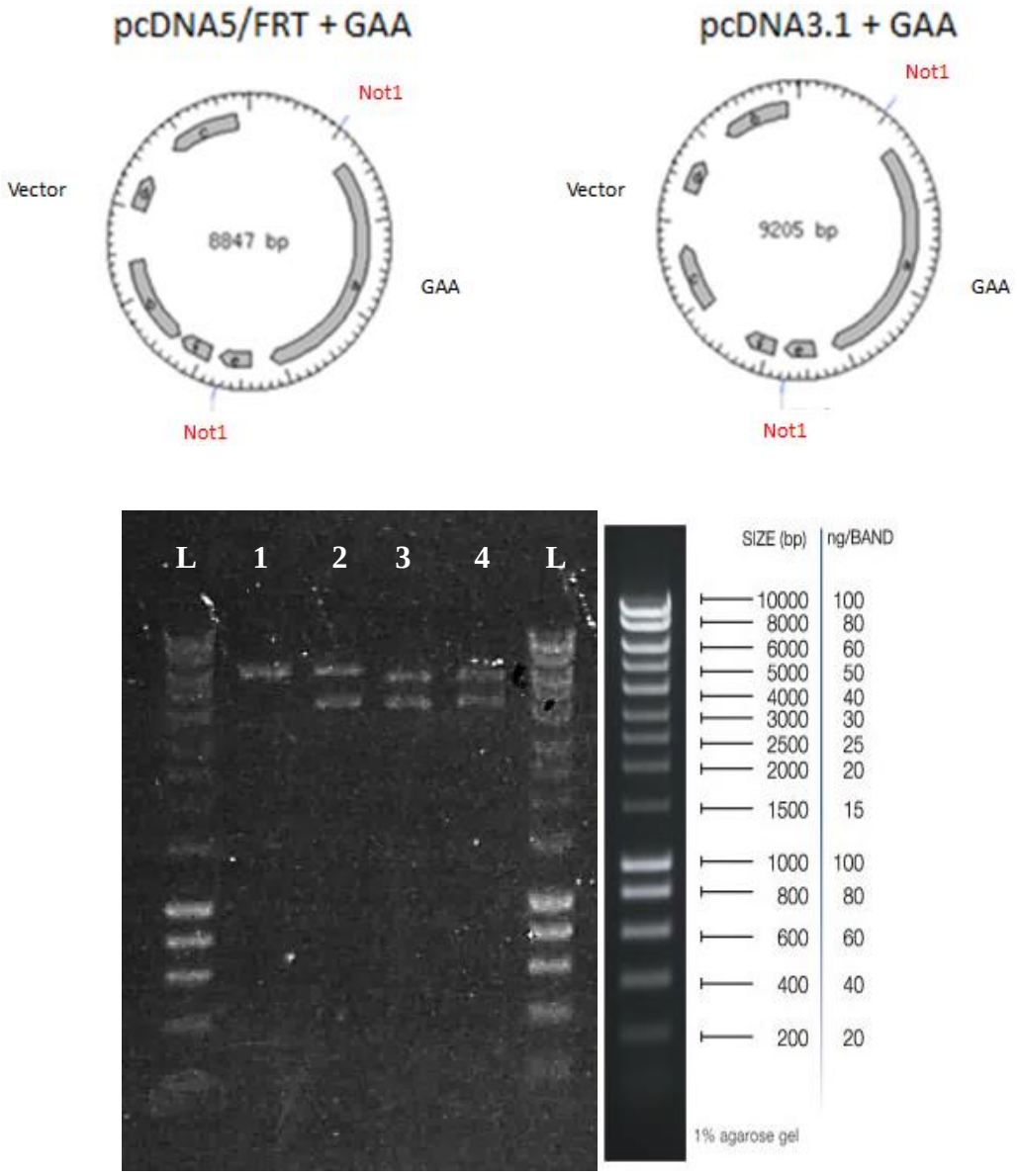


Figure 4-3. Size confirmation. L 1 kb ladder; 1 pcDNA3.1 A, 2 pcDNA3.1 B, 3 pcDNA 5FRT A; 4 pcDNA 5FRT B. (notebook 106: 07/05/13). Data shows: pcDNA3.1 A: one band at around 6k bp. pcDNA3.1 B: two bands at around 6k bp and 4k bp. pcDNA5/FRT A: two bands at around 6k bp and 4k bp. pcDNA5/FRT B: two bands at around 6k bp and 4k bp.

The result in Figure 4-3 above was consistent with what expected and proved that the construct contains both vector and gene. Digestion of pcDNA3.1 A did not work as only one band was observed. The reason for this became clear with sequencing later on.

At this point, we digested the two constructs with Sal I restriction enzyme to check for direction of insertion. Results are in the Figure 4-4 below.

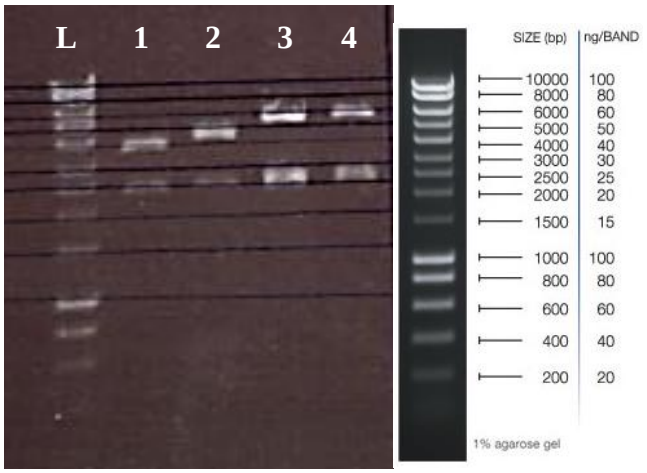


Figure 4-4 Gene Insertion direction. left to right: (L) hyperladder; (1) pcDNA/3.1A, (2) pcDNA/3.1B, (3) pcDNA/FRT A, (4) pcDNA/FRT B. (notebook 106: 12/05/13). Data shows pcDNA3.1 A: one band at around 2.5k bp and one band at around 4k bp; pcDNA3.1 B: one band at around 2.5k bp and one band at around 4.5k bp; pcDNA5/FRT A: one band at around 2.5k bp and one band at around 5.5k bp; pcDNA5/FRT B: one band at around 2.5k bp and one band at around 5.5k bp

Although the numbers are slightly lower than what expected (Table 4-2), the fact that all constructs show two clearly separated bands is a clear indication that all four constructs received the gene in the correct direction. This was however confirmed with sequencing below.

4.2.4 Gene sequencing

The last step of ensuring that the gene was in fact inserted the correct way is via sequencing. The Wolfson Institute for Biomedical Research at UCL provided this service to students and

members of staff on campus. The resulting data was provided on .ab1 files and analysed BioEdit Sequence Alignment Editor (Hall, 2013)

In order to run a sequencing reaction a primer or a set of primers had to be created. Primers are short DNA sequences (15-30 nucleotides) that are used to align the DNA polymerase onto the section of the DNA of interest. In each reaction it is possible to sequence around 1000 bases, and generally only the section between base 50 and base 800 are accurately sequenced. For a GAA gene of around 3500 bases, 5 primers were therefore necessary. Primers were designed to align to the sequence on the vector just before the beginning of the gene and ending close enough to the end of the gene, so that part of the adjacent vector sequence is included. This was done to ensure the gene was inserted the right way and presented no errors.

When multiple primers are to be used in one sequencing reaction they need to have similar melting temperature (T_m), which is defined as "the temperature at which one half of the DNA duplex will dissociate to become single stranded and indicates the duplex stability". (PrimerBioSoft, 2013). In other words T_m represents the temperature at which the double DNA filament opens up to allow DNA polymerase to start replication. T_m can be calculated using one of the many online tools (Promega, 2013a) or by the following formula:

T_m (^{0}K) = [ΔH / ΔS + R ln(C)], Or Tm (^{0}C) = [ΔH/ ΔS + R ln(C)] - 273.15, where:

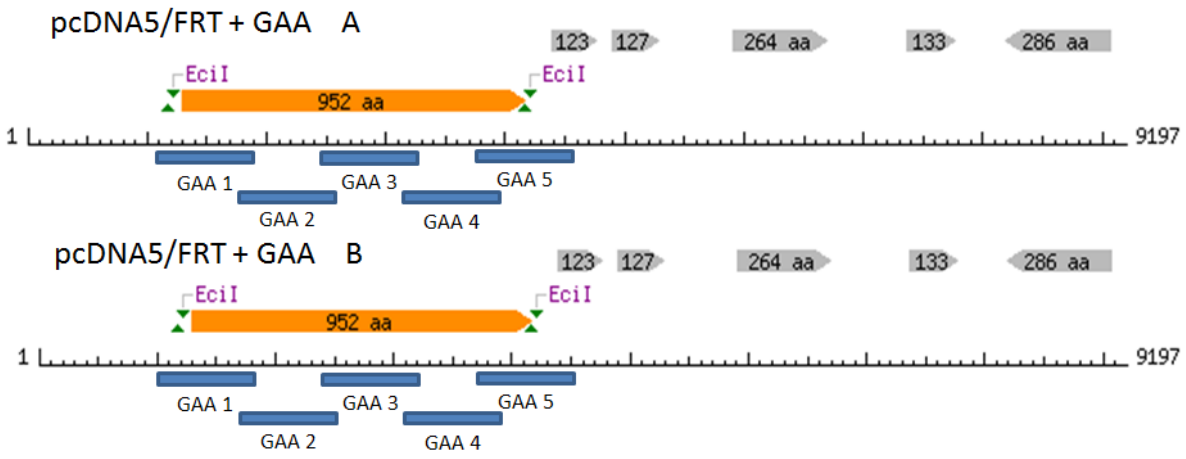
ΔH (kcal/mole) change in Enthalpy. In the above formula the ΔH is obtained by adding up all the di-nucleotide pair enthalpy values of each nearest neighbouring base pair; ΔS (kcal/mole) change in Entropy obtained by adding up all the di-nucleotide pair entropy values of each nearest neighbouring base pair (PREMIER Biosoft, USA)

The following custom primers were designed and purchased from Invitrogen. Details are listed in Table 4-3 below.

Name	GAA1	GAA2	GAA3	GAA4	GAA5
Sequence	TAATACG ACTCACT ATAGGG	TACAAGC TGGAGAA CCTGAG	AGTGGA ACGACCR GGACTAC	ACAACAG CCTGCTC AGTCTG	TGTCTCC AACTTCA CCTACA
Yield (OD)	2	2	2	2	2
Mol. Weight	6126	6176	6152	6063	5973
Tm C (at 5pmoles/µl)	48	52	54	54	50
Extinction coefficient	230.4	235.4	227	212.7	203.3
Percent CG	0.4	0.5	0.6	0.6	0.4
µg per OD	26.6	26.2	27	28.5	29.4
nMols per OD	4.3	4.2	4.4	4.7	4.9

Table 4-3 Custom primers specifications.

In *Figure 4-5* below there we present a graphical representation of the resulting analysis. Blue bars represent regions coded by respective primers; orange sections represent GAA gene.



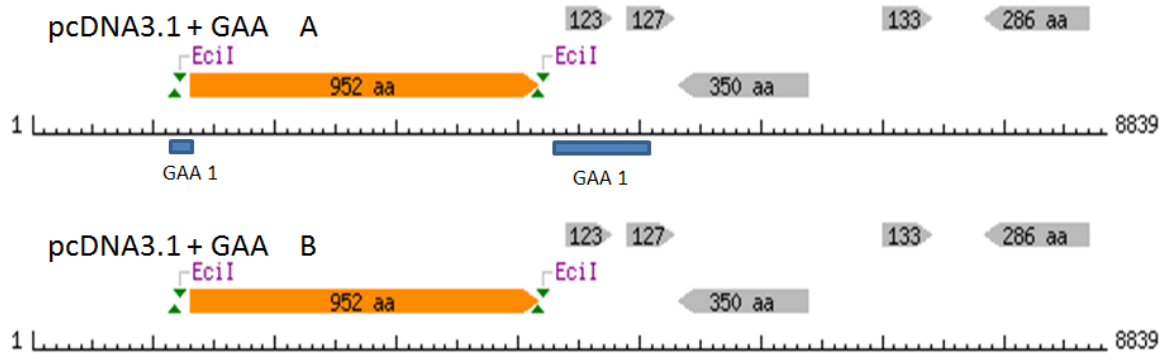


Figure 4-5. Graphical representation of sequencing results, blue bars represent regions coded by respective primers; orange sections represent GAA gene. Both pcDNA5/FRT + GAA A and B constructs received the complete gene in the correct direction. Both pcDNA3.1 + GAA A and B constructs did not receive the gene, or got it the wrong direction.

It is interesting to note how in construct pcDNA3.1 A, sequencing started on vector just before insertion point and continued after insertion point, proving that the gene is completely absent. However, in construct pcDNA3.1 B no alignment points were found, sign that either primers didn't find complementary sequences or that the genetic material is degraded. This explains why pcDNA3.1A size and direction of insertion agarose gels results were wrong.

For these reasons, only pcDNA5/FRT + GAA construct was used for subsequent experiments.

4.3 Transfection into mammalian CHO

An initial round of transient transfection was necessary in order to confirm that a CHO host could produce active GAA of the correct size. This process has the benefit of being a relatively quick way to generate material for research, compared to stable transfection. The downside is that as the genetic material is not integrated into the host genome, it is not replicated during cell division and is therefore lost after few generations. Western blot analysis confirmed that GAA was being produced in transient clones from day 2 for at least 5 days.

4.3.1 Transient transfection

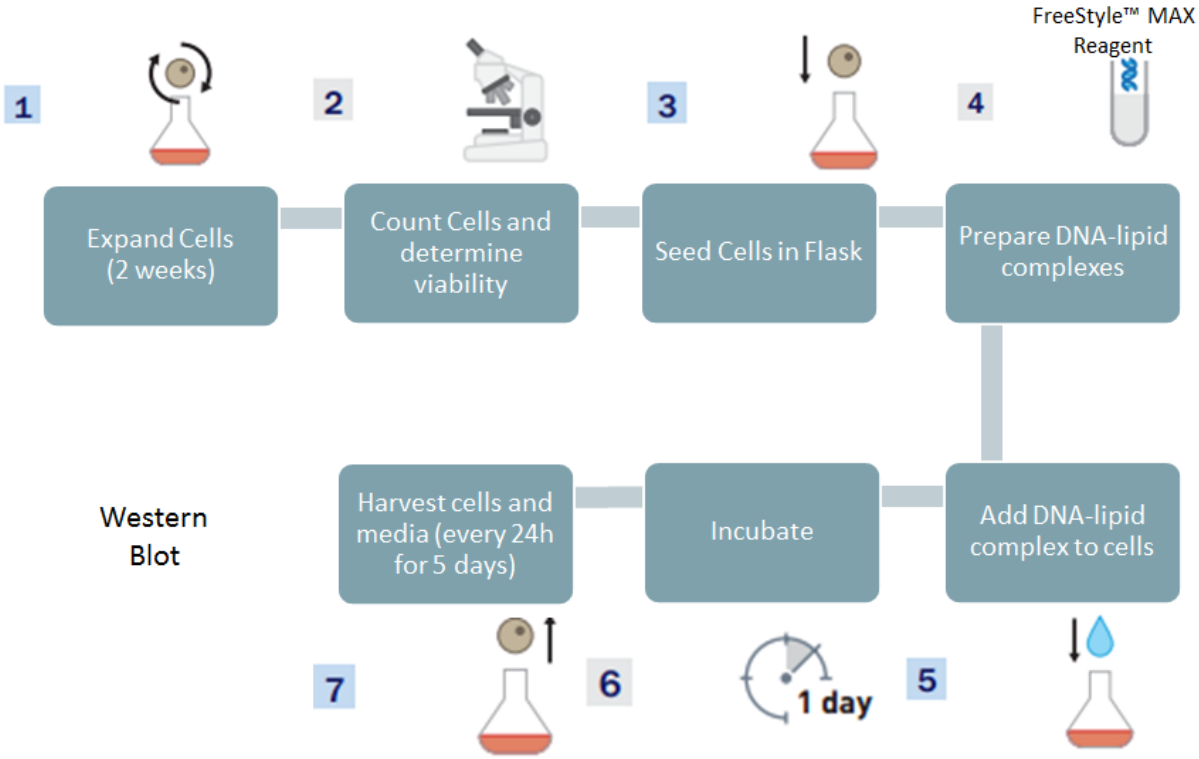


Figure 4-6. Schematic representation of a transient transfection cycle.

Transiently transfected cells were incubated for 5 days. Viability and cell count was assessed daily and 1 ml culture samples were taken to determine product production via western-blot. Both FRT constructs (A and B) showed similar doubling time (around 30 hours) and viability (>96%) as shown in Figure 4-7 below.

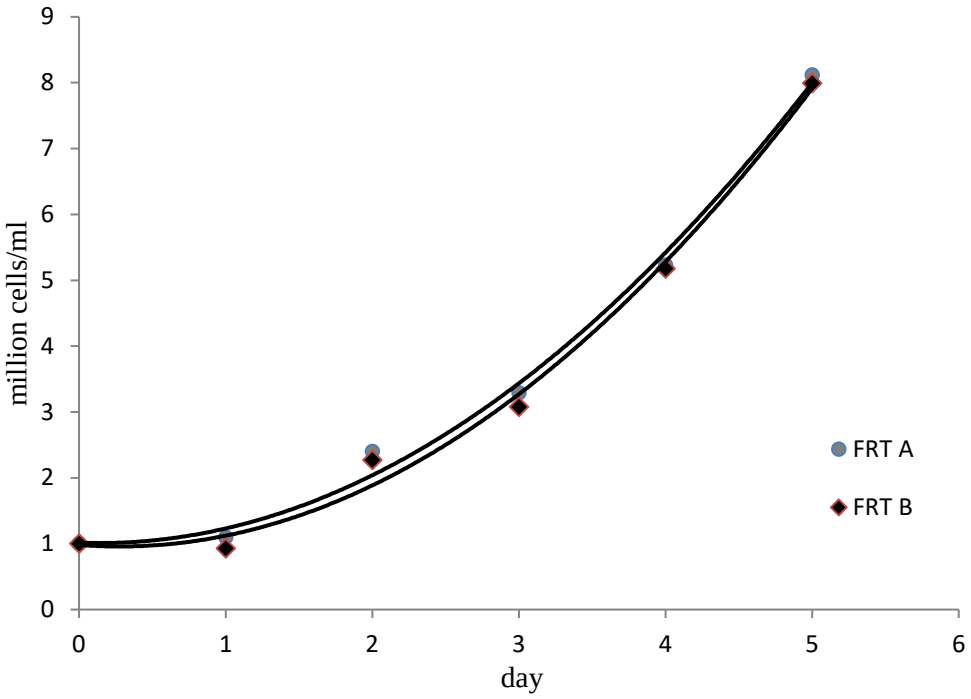


Figure 4-7. Transient transfection cultures growth curves. Cells were incubated for 5 days in 60ml CD-CHO media in 250ml cell flasks. Viability and cell count were assessed daily via ViCell (Beckman Coulter USA) and broth aliquots were taken to determine product production via western-blot. Both FRT constructs (A and B) showed similar doubling time (around 30 hours) and viability (>93%). Polynomial equation fitting. FRT A: $y = 0.295\ x^2 - 0.077x + 1.015$, $r^2 = 0.994$; FRT B: $y = 0.311x^2 - 0.171x + 0.981$, $r^2 = 0.994$

Product production was qualitatively assessed also via western blot. Both pcDNA/FRT A and B reach maximum titer at day 3.

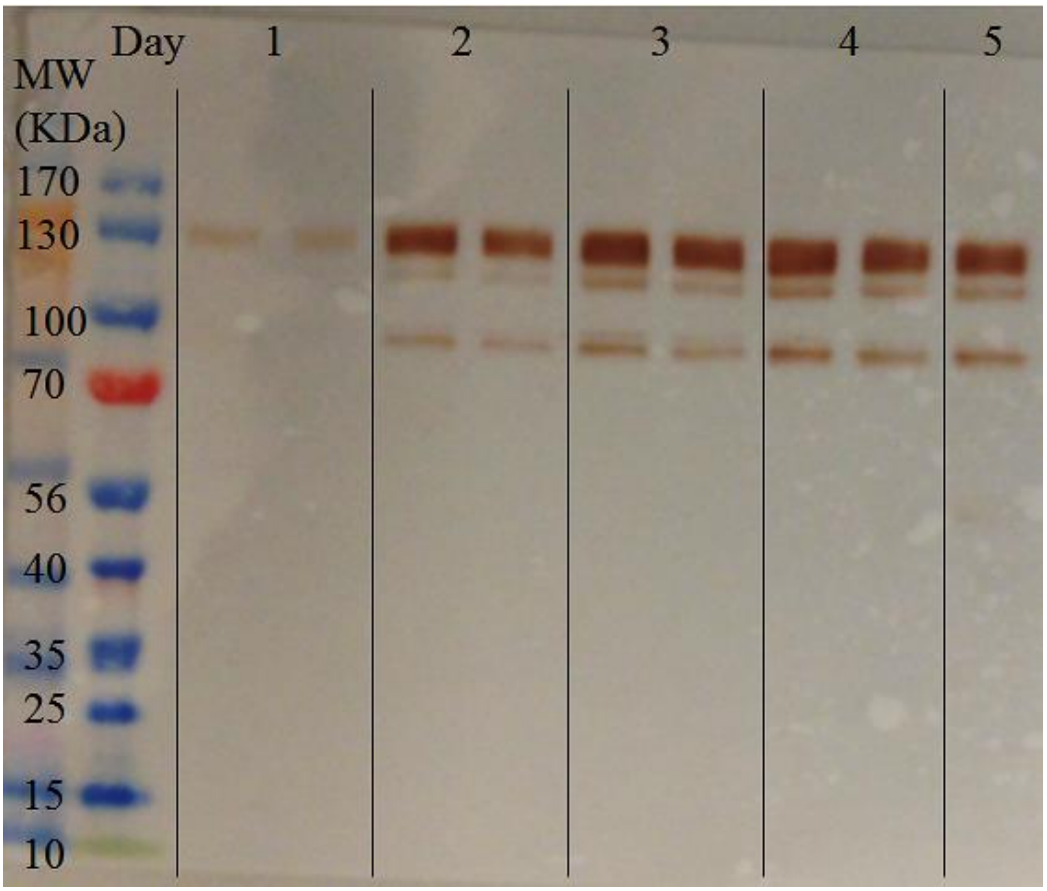


Figure 4-8. Western blot analysis of GAA production from day 1 to day 5 after transient transfection. Monoclonal rabbit anti-GAA 1:1000 working dilution was used (NewEnglandBiolabs, 2013). Both transfection reactions are represented per each day. Left pcDNA/FRT A, right pcDNA/FRT B. Three GAA isoforms are visible roughly at 95, and 120 and 130 kDa in both samples. These sizes are not common for GAA, they could perhaps be not fully processed isoforms of GAA. They are however recognized by anti-GAA monoclonal antibody.

Having confirmed that the transient transfection produced the expected results, stable transfection and clone selection followed.

4.3.2 Stable transfection

In order to stably express the target protein for an indefinite amount of time, a CHO FlpIn cell line was stably transfected with GAA recombinant gene. Stable transfection has the advantage

over transient transfection of producing recombinant cell lines that carry the transfection plasmid integrated into the genomes, ensuring a constant production of target protein.

Integration of recombinant DNA in host cell line genome normally happens randomly which means that expression level is unpredictable due to position effects (Huang et al., 2007). To overcome this, the site specific FRT/FLP system was used. This type of recombination offers an alternative strategy to develop high producing and stable clones in a reproducible and predictable manner. The key is to target the FRT sequence to chromosomal locations where there is a high rate of transcription and gene amplification, and the amplified genes can be maintained (Huang et al., 2007). The vector carrying the FRT gene also carries hygromycin gene resistance allowing for selection pressure using Hygromycin B antibiotic. The FRT/Flp-In system relies on availability of Flp recombinase, which greatly improves the recombination efficiency in mammalian cell lines by a process called site-specific recombination. The recombinase was produced by pOG44 plasmid, which was added to the reaction mix at the time of transfection.

In this experiment, two transfections were attempted using pcDNA5/FRT+GAA constructs (A and B) prepared. Transfection efficiency was evaluated using two controls: one control lacked of pOG44 (-P) which prevented the production of FLP recombinase needed for the first cell divisions; a second control lacked the transfection plasmid all together making it impossible for it to proliferate under selection pressure. The four samples are summarised in Table 4-4 below.

Sample	Cell line	Plasmid	pOG44	PEI
A	Flp-In	pcDNA/FRT A	✓	✓
B	Flp-In	pcDNA/FRT B	✓	✓
-P -pOG44 (first control)	Flp-In	pcDNA/FRT A	×	✓
U un-transfected (second control)	Flp-In	×	✓	✓

Table 4-4 Stable transfection reaction mixtures.

Cell viability of stably transfected cells was tracked over a period of 3 weeks.

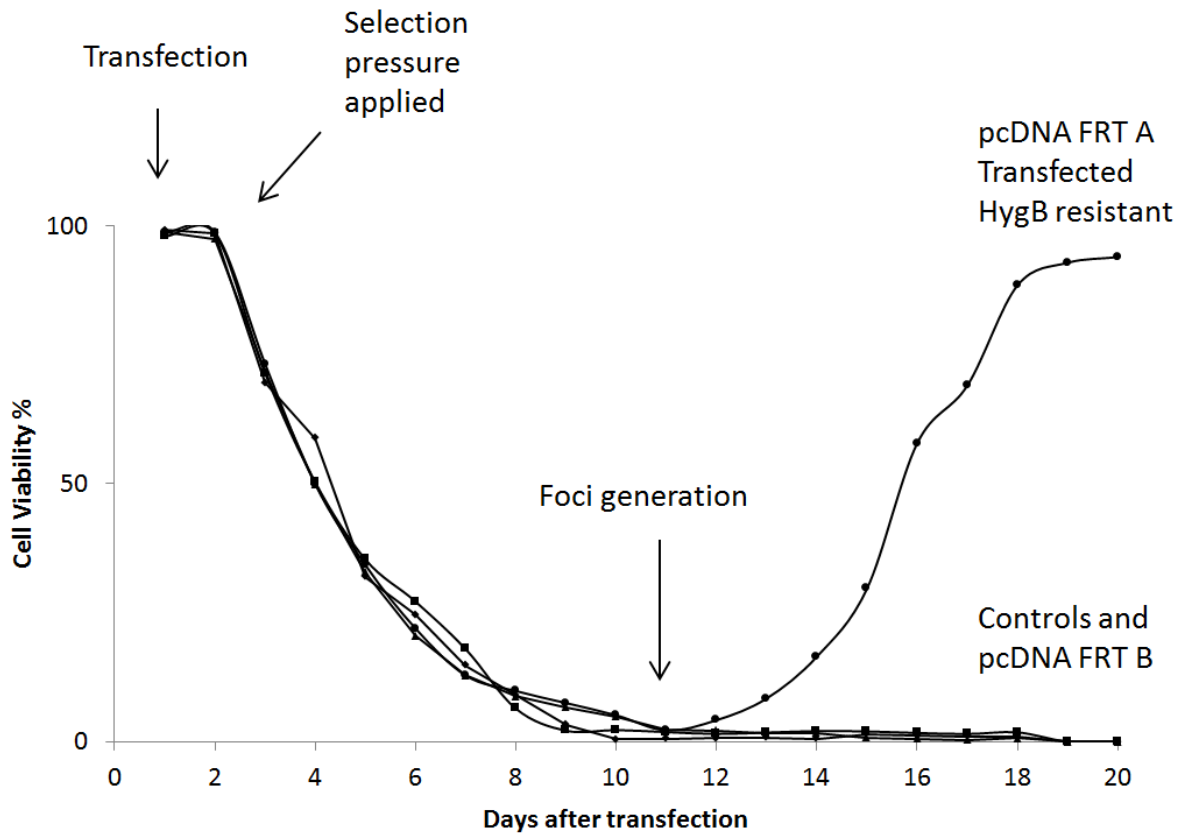


Figure 4-9 Stable transfection cell viability curve. For transfection 3µg of plasmid DNA was added to 27 µg of pOG44, incubated 5 min at RT, followed by 90 µg of linear PEI at a (1:9):3 ratio. Vial was vortexed, incubated 5 min at RT and added drop by drop to culture in flasks. Freshly transfected cultures were incubated at 37 ^{0}C 8% CO_2 125rpm spin tray incubator. The following day cells were spun at 300g 5 min and media changed with fresh media containing selection agent 500 µg/ml HygromycinB. Cell viability and number was assessed daily for three weeks, media was replaced by fresh HygB+ media every four days. Only FRT-A transfected CHO FlpIn viability improved after foci generation (day11). Data shows that as selection pressure is applied via HygB addition, cell viability for all reaction mixtures decreases sharply. Eventually only transfected cells from reaction FRT-A were able to pick up again and viability reached normal levels after day 20. Cells transfected with FRT B construct and both controls did not survive selection pressure.

During initial stages of stable transfection, the FRT plasmid carrying the target protein sequence and antibiotic hygromycin B resistance was added to the reaction mix and entered the host cells aided by PEI detergent. After 24h, hygromycin B selection agent was applied to all four reaction mixtures which caused cell viability to drop in all four samples. On day 11 viability of FRT A sample started increasing again reaching normal levels at day 20, while none of the other samples survived selection pressure. A possible explanation to this result is that at first, transfection supposedly happened only in sample A and B but not to either control, since they were lacking either the plasmid itself or the pOG44 plasmid. At this stage, in both transfected samples, the number of cells that acquired the plasmid, and therefore hygromycin resistance, was still incredibly smaller compared to the total number of cells present. For this reason the transfected cells samples, as a whole, seemed to perish as selection pressure was applied. After 11 days however, the number of cells with a copy of the plasmid, started to equal the number of cells lacking it as most of the latter ones perished. This is when the first foci were seen. In the following days, while all the remaining untransfected cells continued to die, the transfected ones increases in number and the viability raised again. By day 19 FRT A sample viability approached 100%.

At this stage the FTR A sample was composed by a number of cells that have inherited the FRT plasmid. While most of them should be high producers due to FRT presence, this was still a heterogeneous population. Cell line cloning and selection was therefore necessary to select for high producer clones lines.

4.4 Cell line selection

4.4.1 Clone generation

Limiting dilution cloning (LDC) was performed on transfected cell pool that survived selection pressure. This technique was used in order to create clonal lines from the heterogeneous

population of clones generated during transfection and evaluate clonal target molecule production. The technique involved diluting the cells in a way in which in each well of a 96 well plate there was a chance of having 1 or less cells. After dilution and plating, cells were monitored to assess cell division and growth. 6 x 96 well plates were used for this experiment. The outer perimeter of wells was filled with a solution of water and ethanol in order to prevent contamination from outside and evaporation from within the plate. This left 60 wells per plate usable and therefore 360 wells in total (100 µl per well, 36 ml total volume). On the day of the procedure, the cell count from FRT A sample (from stable transfection) was 0.7 10^6 c/ml (viability > 97%). Assuming a final concentration of 1 cell per well (100µl aliquot), the total number of cells needed was 360 in 36 ml of media. At the current concentration, half microliter of cells contains 360 cells, a volume too small to be prepared accurately. A series of serial dilution was therefore set up in order to get to the right dilution in a more precise way. Dilutions were setup as follow:

Dilution	**Neat**	**1:10**	**1:10**	**1:10**	**1:2**
Number of cells	700000	70000	7000	700	350
Volume (cells + media) ml	1	0.1 + 0.9	0.1 + 0.9	0.1 + 0.9	0.5 + 0.5

Table 4-5 serial dilution preparation for LDC in CD-CHO media.

After performing the serial dilution, the last aliquot containing approximately 350 cells was diluted in 36 ml of media containing 500 µg/ml HygromycinB, 100 µl aliquots were dispensed aseptically into the plate wells and sterile cover placed. Plates were incubated at 37 ^{0}C 5% CO_2. For the following 5 weeks the wells were monitored daily using an optical microscope to see whether clones started doubling. Evaporating media was added as needed.

In total only six clones proliferated and they were transferred into increasingly larger volumes of media in 24 well plate, then in 6 well plate and finally into 125ml flasks. A schematic representation of LDC is shown in *Figure 4-10* below.

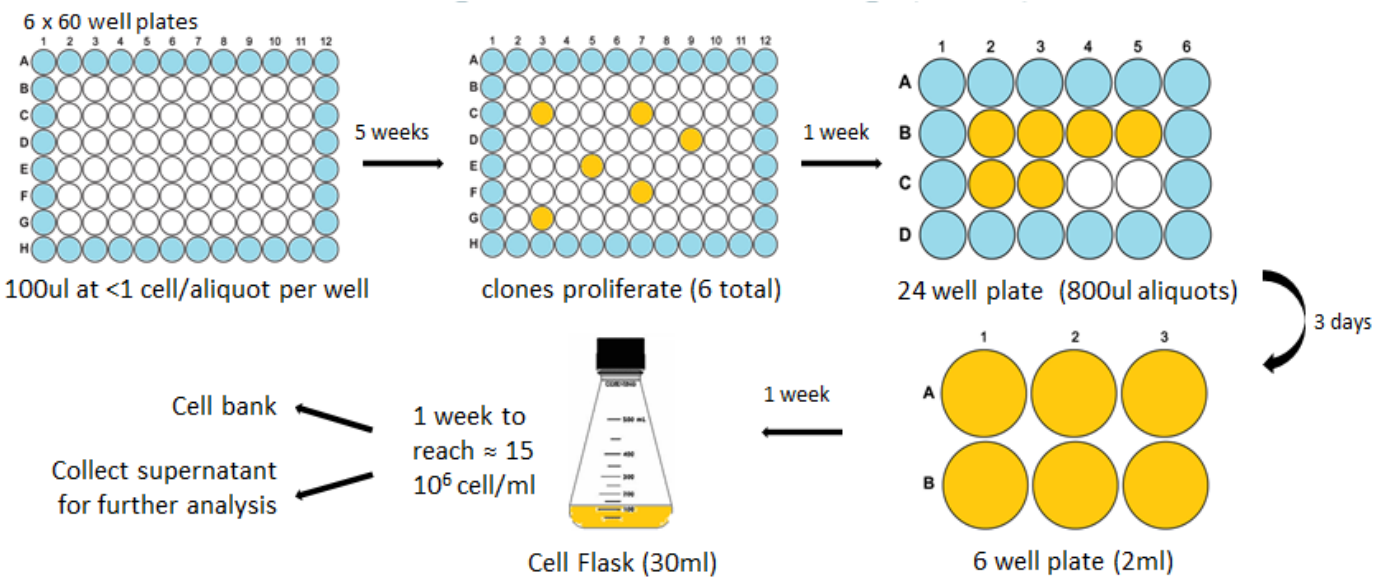
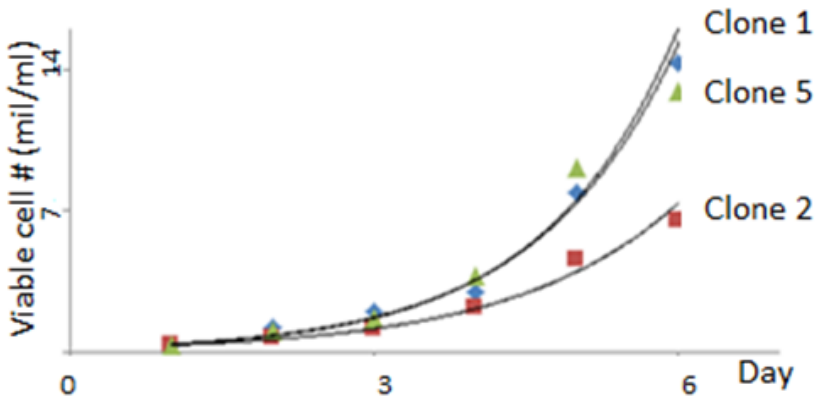


Figure 4-10 Schematic representation of foci generation via LDC. The heterogeneous FRT A population that survived selection pressure was put through one cycle of LDC to generate clonal lines. Out of 360 clones attempted only six showed doubling during the first five weeks of observation. Those six were transferred to gradually larger volumes up to 30ml flasks. Only three clones survived cell banking (clone 1, 2, 5).

4.4.2 Clone comparison

Clones were assessed for growth performance and target molecule production. Among the six viable clones from LDC, only three were ultimately viable after cell banking (freeze-thaw cycle). Viable clones (clone 1, 2 and 5) were revived from liquid nitrogen and expanded in 125ml flasks in CD-CHO media under moderate agitation at 37 ^{0}C 5% CO_2. Cells were passaged to 0.3 10^6 cell/ml when they reached above 3 10^6 c/ml for four times. These values were chosen based on expert advice from other laboratory users. On the last passage, cells were monitored daily for cell number and viability by ViCell until they reached 15 10^6 c/ml. This allowed the calculation of doubling times and growth performance (Figure 4-11 top). The clones were then passaged one last time and grown until the viable cell concentration reached 5 10^6 cells/ml. They were then harvested, spun to separate cells and the supernatant was micro filtered through 0.22um filter. Samples titer was assessed via the GAA diagnostic assay (Figure 4-11 bottom) and qualitatively via western blot (Figure 4-12). A suitable candidate was chosen based on good balance between growth performance and target molecule production levels.

Note. For assay development refer to 5.2 – Western blot and 5.4.2 – Quantification of Quantification by GAA diagnostic assay.



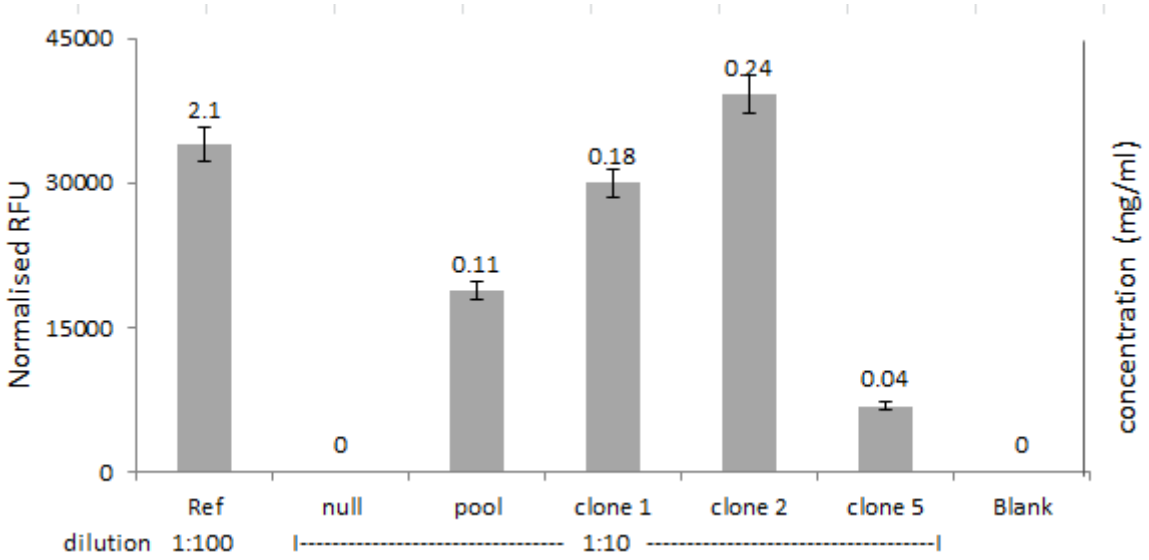


Figure 4-11. CHO GAA clone selection process. (top) doubling time and growth performance was assessed via Vi-cell™ XR Cell Viability Analyser; (bottom) GAA amount was assessed via diagnostic GAA assay (Okumiya et al., 2006). Although clone 2 had the best titer (0.24 g/l), clone 1 was chosen because of fairly good titer (0.18 g/l) and best growth performance (15 million cells 6 days after passage). Medium used ThermoFisher scientific CD-CHO with 250 µg/ml HygB, viability >93% at day 6. For assay development refer 5.4.2 Quantification by GAA diagnostic assay.

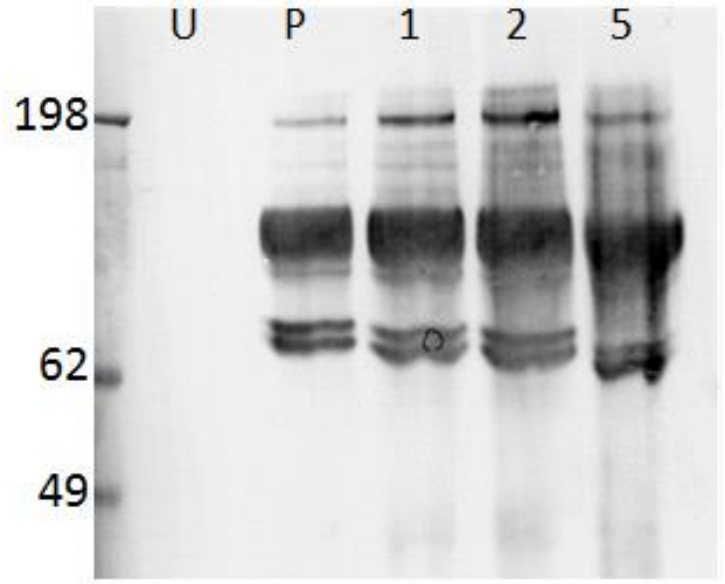


Figure 4-12 Western blot analysis of GAA production of heterogeneous pool (P) and clone 1, 2 and 5. All but untransfected cell sample (U) show similar amount of GAA produced. Three isoforms (76, 92, 106KDa) are visible. Monoclonal rabbit antiGAA antibody (1:1000 w.d.) was used (cat.# ab137068 (Abcam, 2013). For assay development refer to 5.2 – Western blot

4.5 Chapter conclusions

In this chapter the generation of a mammalian clonal cell line stably expressing human acid alpha glucosidase (GAA) is described. This was necessary in order to provide a well defined material throughout the duration of the project to develop and optimize the purification process for this therapeutic enzyme in accordance with the overall goals of this project.

The human GAA gene was purchased from Origene and cloned into a pcDNA/FRT plasmid, to be used in conjunction with site specific recombination Flp-In CHO expression system. A series of agarose gels and sequencing experiments confirmed correct integration of GAA into the vector and the proper formation of the vector. This was used to transiently transfect a CHO cell line, and GAA production was successfully confirmed via western blot over a five-day lab scale batch fermentation at 250ml scale. Although enzyme production was not quantified at this stage,

we visually confirmed GAA presence in transiently transfected cells. The generated plasmid was then utilised to stably transfect a suspension adapted CHO Flp-In mammalian cell line, targeting the FRT transcription site for site directed mutagenesis. This process greatly shortened cell clone selection process by eliminating location effect commonly associated with random gene integration. Antibiotic selection pressure was applied to transfected cells and eventually foci carrying gene and antibiotic resistance emerged.

The heterogeneous cell population deriving from the process was subjected to LDC to generate clonal cell lines. Generally, two rounds of LDC are suggested to generate true clonal lines, however in the interest of time only one cycle was performed. This process resulted in three viable clones out of 360 attempted. The three clones were expanded and GAA presence, titer and cell culture growth and performance determined. Titer determination was performed using a GAA diagnostic activity assay. Out of these three (clone 1, 2 and 5) clone 1 was chosen as candidate for future studies as it showed an acceptable titer (0.18g/l) and better growth performance when compared to the other two.

This is precisely the reason why Flp-In/FRT system was used: although the number of clones generated is quite low, they all produce good levels of target product.

5 Chapter two

Product characterization

5.1 Introduction

In this chapter, we describe the development of the analytical assays needed to characterize the targeted molecule produced by the stable GAA producer CHO clone 1 generated in the previous chapter.

Gel electrophoresis was used visualize the target molecule and other species present in clarified harvest. Western blot was utilized to identify GAA by anti-GAA mAb recognition and to gather information on its molecular weight. Identity of the target molecule was further confirmed via tandem mass spectrometry. Amylase activity assay (EnzCheck) was used to determine activity of target molecule and lastly immunoassay (AlphaLISA) and a fluorescence based assay (GAA diagnostic) were utilized to determine concentration of GAA in harvest and through the purification process. EnzCheck and AlphaLISA did not give successful results.

The features that have been evaluated are summarised in Table 5-1 below:

	SDS-PAGE	Western blot	MSMS	EnzCheck	AlphaLISA	GAA diagnostic
Size	✓	✓	✓			
Presence	✓	✓	✓		✓	
Activity				✓		✓
Identity			✓			
Amount					✓	✓

Table 5-1. List of analytical techniques described in this chapter and their purpose.

Size via SDS-PAGE

One method for measuring GAA size is via gel electrophoresis. This method is expected to show target molecule size based on the principle that linearized molecules travel through the mesh of a polyacrylamide gel while subjected to an electrostatic current with a speed proportional to their size. By running alongside a ladder of proteins of known MW it is possible to determine protein of interest size. This assay is however prone to misinterpretation especially in early clarified harvest samples due to the large number of impurities present.

Due to its simplicity this method was used also to determine presence of GAA at different stages of the downstream process.

Different GAA isoforms have been seen at different stages of production with the main one being at around 110KDa. Other GAA forms have also been reported in literature with sizes ranging from 70KDa to 105KDa (Reuser, 1985)

Presence and size via Western blot

Size and presence of the target molecule is more easily determined when the target protein can be seen clearly next to a molecular standard free from other impurities. Western blot's main advantage over SDS-PAGE is that the target protein is designed to be the only specie recognized by the pair of specific primary-secondary antibodies and as a result, it is the only band that appears on the PVDF membrane after incubation next to the molecular standard. The assay is however much more labor intensive than SDS-PAGE as it involves transferring proteins from gel onto a membrane, multiple washes and incubations with different antibody solutions. Assay optimization is also important as protein transfer time and antibodies working solutions must be determined.

Identity via MSMS

MALDI TOF MSMS is mostly used to acquire information on identity (and size) of the molecular species of interest via peptide ionization and time of flight analysis to create a peptide map fingerprint. This technique is labor intensive as it involves separating protein mixture via

SDS-PAGE, excision of single protein bands, peptide digestion using trypsin and spotting samples onto the MALDI plate for analysis. It also requires high technical know-how necessary to operate a mass spectrometer and data analysis. It is however the only technique that can provide information on identity of the peptides analysed by sequence comparison with online protein databases such as SwissProt.

Activity via EnzCheck

Activity of the produced target protein was assessed via EnzChek® Ultra Amylase Assay Kit (Thermo Fisher Scientific USA). This assay contains a proprietary starch derivative (DQ ™) labeled with BODIPY® FL dye to such a degree that fluorescence is quenched. In presence of amylase (GAA is an amylase), the substrate is degraded, quenching is relieved and as a result fluorescence is emitted. By measuring with fluorescence microplate reader the resulting signal, it is possible to determine the amount and activity of enzyme, as signal will be proportional to activity of amylase. One drawback of this assay is that the substrate is not specific to GAA only but also to beta and gamma amylases present.

Quantification via AlphaLISA.

Determination of the amount of target protein is of critical importance throughout the whole bioprocess. Developing an assay capable of determining target molecule concentration in a reliable and high throughput fashion will help determine titer levels and downstream process strategy.

The AlphaLISA immunoassay is a high throughput fluorescence based immunoassay in which signal is proportional to amount of analyte in solution. Primary and secondary antibodies are immobilized onto a donor and acceptor beads and the presence of analyte allows formation of a close association of the two beads which leads to emission of light at 615nm as explained in materials and methods (section 3.2.5).

Since the manufacturer does not provide already made detection kit for GAA, the assay had to be developed and optimised. This involved biotinylation of antibody, choosing the right antibody pair and calibrating analyte sensitivity and response. The AlphaLISA development guide was followed (PerkinElmer, 2012).

Quantification via GAA diagnostic assay.

Quantification of GAA can also be determined via the alternative activity method developed by Okumiya group (Okumiya et al., 2006). This assay is used in the clinics to detect presence of GAA in patients' blood and it is specific to GAA, which puts it at a clear advantage over other activity assays (such as EnzCheck) that detect generic amylase activity in samples. Activity is measured by fluorescence deriving from substrate degradation (glycogen and 4-methylumbelliferyl-α-D-glucopyranoside 4MU-αGlc). Specificity is achieved by incorporation of acarbose to eliminate the interference of unrelated alpha-glucosidases (predominantly maltase–glucoamylase).

This assay was used to determine GAA concentration in clarified harvest of different clones for cell line selection.

5.2 Target molecule size and presence.

The purpose of this experiment is to compare the clarified harvest material from the GAA producing CHO selected in section 4.4.2 (Clone 1) to the originating Null culture in order to show that the GAA CHO line over-expresses the target protein.

After growing both strains in 30ml of CD-CHO media in bottom baffled cell flasks for 8 days, cells were transferred to centrifuge tubes and spun. Supernatants were concentrated via Amicon 10000MWCO. 50µg of protein per sample was prepared according to protocol and run on the gel below (Figure 5-1).

This gel shows clearly the presence and size of the target protein at around 110KDa in the GAA CHO Clone 1. By running the sample alongside a control of the original cell line (Null CHO-S) we can observe that the 106KDa band corresponding to GAA is being strongly overexpressed in the cell supernatant during fermentation.

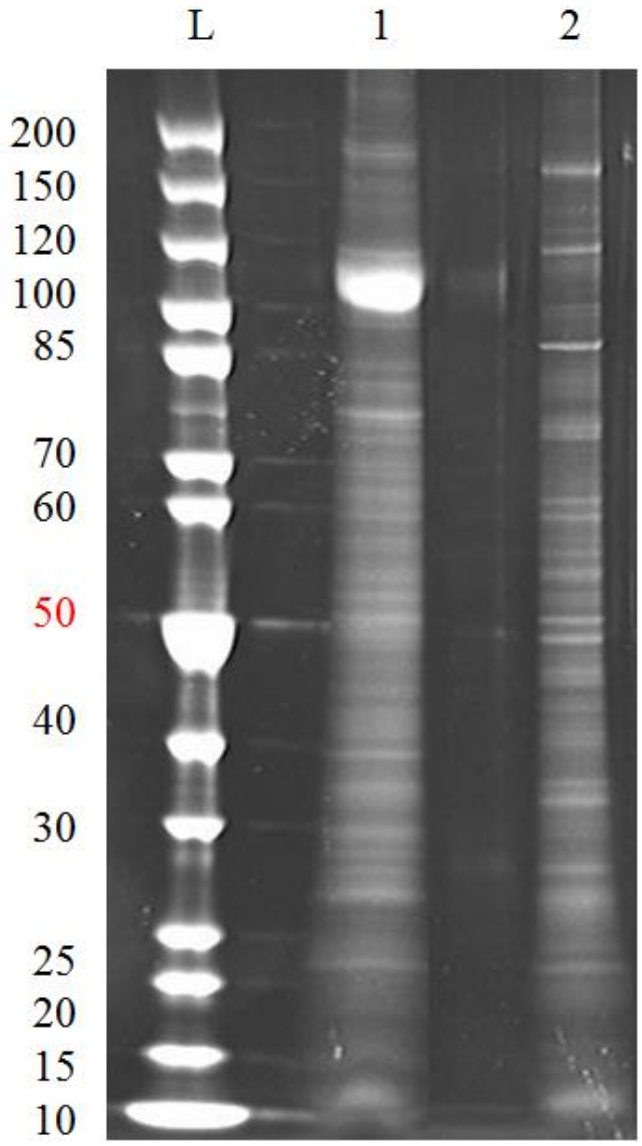


Figure 5-1 gel electrophoresis NuPAGE™ Novex™ 4-12% Bis-Tris (ladder cat# 26614 Thermo Fisher, USA) SYPRO ruby stained. Concentrated clarified cell culture supernatants. Sample 1 is GAA CHO Clone 1, sample 2 is the originating Null CHO-S. Data shows one clear band at around 110KDa in GAA CHO sample. This is believed to be over-expressed GAA. Null sample shows no such band.

GAA size and presence in the clarified harvest can also be assessed via western blot. This technique was also used to detect presence of target protein in chromatographic fractions as explained in the following chapter.

For this western blot experiment a clarified harvest sample from CHO Clone 1 was concentrated using a 10000MWCO and run on a SDS-PAGE gel alongside a sample of pure GAA drug substance provided by BioMarin pharmaceuticals Inc. Proteins were transferred to a PVDF membrane by applying a voltage of 30V for 1h.

After transfer, the membrane was washed in PBS-T 5% milk three times for 10 minutes each, then incubated overnight at 4^{o}C with the primary mAb in a solution of PBS-T 5% milk under gentle agitation. Next day the membrane was washed again in PBS-T 5% milk three times for 10 minutes each and the incubated for 1h in a solution of secondary Ab PBS-T and 5% milk under gentle agitation. SigmaFast Diaminobenzidine (DAB) tablets were used for detection.

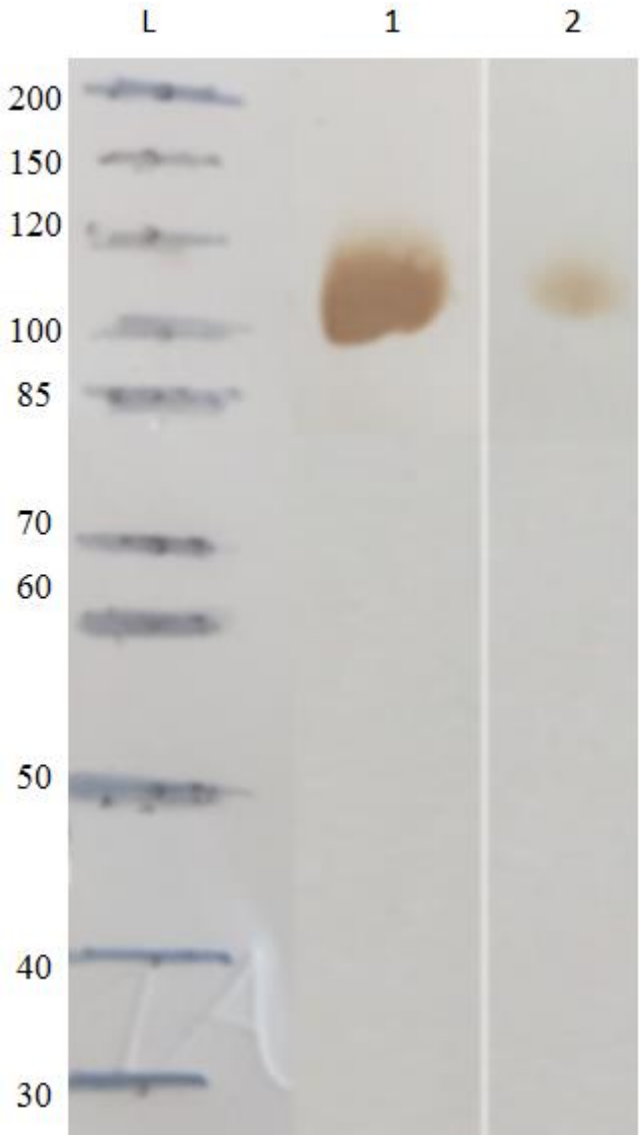


Figure 5-2 Western blot. Proteins transferred to PVDF membrane from Bis-Tris 4-12% Gel, ladder cat# 26614 Thermo Fisher, USA. Primary anti-GAA rabbit mAb (w.d.1:1000 Abcam cat.# ab137068), Secondary anti rabbit polyclonal Ab HRP-linked (w.d. 1:1000). Sample 1 is GAA drug substance reference (BioMarin, lot# 11132013), sample 2 is the GAA CHO clone 1 clarified harvest. Data shows one prominent band in CHO clone 1 harvest at just above 100 KDa. Accidental use of unstained ladder meant that bands had to be handwritten using gel as reference.

5.3 Target molecule identity

So far we have gained information on the target protein presence and size. The data seen so far tells us about the presence of a 100 to 110 KDa protein overexpressed in clone 1. While it is probable that this is recombinant human GAA, it must be proved by an orthogonal identity method. Mass spectrometry is usually the go-to method to determine protein identity by providing peptide fingerprint of the protein of interest and matching it to a peptide sequence database such as SwissProt.

For this experiment, the most visible band in sample 1 (at around 110 KDa) from SDS-PAGE gel Figure 5-1 was excised with a scalpel for further analysis. The gel fragment was processed via the protocol included in materials and methods (3.2.4 Identification of product – MALDI TOF MSMS) and digested with modified sequencing grade trypsin. After application of 0.5 µl of sample and standard to the MALDI plate followed by 1 µl of Super-DHB matrix, the plate was processed via mass fingerprint first and then the 10 most prominent peaks processed again via a second round of MS where the ion-trap first performed MS measurements on all intact peptide ions, and then in a second scan it performed a MS/MS experiment on the ten largest peptide signal peaks.

The additional sequence information provided by tandem MS can be extremely powerful, sometimes enabling definite protein identification to be made on the basis of a single peptide. However, tandem MS spectra of multiple peptides that arise from digestion of a given protein provide greater opportunity for obtaining definitive identification.

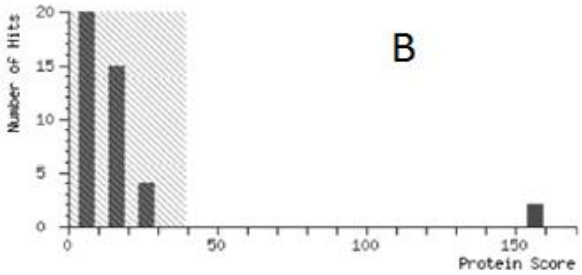
{MATRIX} {SCIENCE} Mascot Search Results

User	:	Damiano Migani
Email	:	d.migani.11@ucl.ac.uk
Search title	:	**Auto submitted by BioTools**
Database	:	**SwissProt 2015_10 (549646 sequences; 195983064 residues)**
Timestamp	:	**20 Oct 2015 at 11:39:03 GMT**
Protein hits	:	LYAG_HUMAN Lysosomal alpha-glucosidase OS=Homo sapiens GN=GAA PE=1 SV=4
		LYAG_PONAB Lysosomal alpha-glucosidase OS=Pongo abelii GN=GAA PE=2 SV=1
		CASA1_BOVIN Alpha-S1-casein OS=Bos taurus GN=CSN1S1 PE=1 SV=2
		CFA54_MOUSE Cilia- and flagella-associated protein 54 OS=Mus musculus GN=Cfap54 PE=2 SV=4
		CC88B_MOUSE Coiled-coil domain-containing protein 88B OS=Mus musculus GN=Ccdc88b PE=1 SV=2
		MURC_SYNFM UDP-N-acetylmuramate--L-alanine ligase OS=Syntrophobacter fumaroxidans (strain DSM 10017 / MPOB) GN=murC PE=3 SV=1
		YCEH_ECO81 UPF0502 protein YceH OS=Escherichia coli O81 (strain ED1a) GN=yceH PE=3 SV=1
		DDL_AERHH D-alanine--D-alanine ligase OS=Aeromonas hydrophila subsp. hydrophila (strain ATCC 7966 / DSM 30187 / JCM 1027 / KCTC 23
		CYSP3_SOLLC Cysteine proteinase 3 OS=Solanum lycopersicum GN=CYP-3 PE=2 SV=1
		SCP39_ARATH Serine carboxypeptidase-like 39 OS=Arabidopsis thaliana GN=SCPL39 PE=2 SV=1

A

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 39 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



B

Search Parameters

C

Type of search	:	MS/MS Ion Search
Enzyme	:	Trypsin
Fixed modifications	:	Carbamidomethyl (C)
Variable modifications	:	Oxidation (M)
Mass values	:	Monoisotopic
Protein Mass	:	Unrestricted
Peptide Mass Tolerance	:	± 50 ppm
Fragment Mass Tolerance	:	± 0.5 Da
Max Missed Cleavages	:	1
Instrument type	:	MALDI-TOF-TOF
Number of queries	:	100

1. LYAG_HUMAN **Mass:** 106112 **Score:** 156 **Matches:** 4(2) **Sequences:** 4(2) **emPAI:** 0.16
Lysosomal alpha-glucosidase OS=Homo sapiens GN=GAA PE=1 SV=4
☐ Check to include this hit in error tolerant search or archive report

D

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
28	1238.6892	1237.6819	1237.6819	0.02	0	49	0.0051	1	U	R.GAYTQVIFLAR.N
44	**1440.6875**	**1439.6802**	**1439.6827**	**-1.73**	**0**	**11**	**40**	**5**		**R.AVPTQCDVPPNSR.F**
70	1874.0005	1872.9932	1872.9945	-0.72	0	68	7.6e-005	1		R.AGYIIPLQGPGLTTTESR.Q
82	2223.1307	2222.1234	2222.1259	-1.15	0	28	0.6	1		R.APSPLYSVEFSEEPFGVIVR.R

Figure 5-3 MALDI-TOF-TOF on a UltrafleXtreme MALDI-TOF instrument (Bruker, Coventry, UK) in positive ion reflector mode and 50% laser power with MS/MS conducted on the ten most intense peaks for each target spot. Generated peptide masses with an ion score exceeding the threshold set for $p < 0.05$ were interrogated using the Mascot algorithm (matrix-science.com) to search all taxonomies in the SwissProt database. Fixed modifications: carbidomethyl (C); variable modifications: oxidation (M); mass values: monoisotopic; protein mass: unrestricted; peptide mass tolerance: +/− 50 ppm; fragment mass tolerance +/− 0.5 Da. Instrument MALDI-TOF-TOF

The Mascot results page shown in *Figure 5-3* was automatically generated after analysis and database search were completed. It is divided in different sections.

Section A at the top shows information on user identification, database searched (in this case the SwissProt database) and analysis timestamp are provided. Then follows a list of the species found in the sample in order of result confidence. LYAG_HUMAN (GAA) appears as first entry indicating that GAA was found in the sample mix in highest abundance.

Section B shows the histogram representation of the protein score. In this result LYAG_HUMAN (GAA) scores at 156, quite a bit away from the 'greyed out' (uncertain) threshold.

The protein score allows the software to judge whether a result is significant or not. From the Mascot website, the protein score is defined as 'probability that the observed match is a random event'. The higher the score, the lower is the probability that the result is a random event. A score is reported as $-10*LOG_{10}(P)$, where P is absolute probability. In this case (*Figure 5-3 B*) a probability of 10^-15.6 scores at 156.

An event is considered significant if it is expected to occur at random with a frequency of less than 5%. Events happening in the grey region are considered of no significance as their probability of being random identifications is higher than 5%. Events falling outside of this area,

such as the result shown in *Figure 5-3 B*, are non- random events. In this case GAA has a score which is well clear of the set threshold and as such is a clear result because it has a high score, leaving little room to doubt.

Section C shows the parameters used for database search. These are the filters that are used in the generation of the result in order to exclude impossible matches.

In this case sequencing grade modified trypsin (cat# V5111 Promega USA) was used as a peptide digestion enzyme. This enzyme works by hydrolysing peptide bonds at the carboxyl side of lysine and arginine residues. The enzyme is also modified by reductive methylation, rendering it resistant to proteolytic digestion (auto-proteolysis).

Under fixed modification Carbamidomethyl (C) was selected as a result of the deliberate post-translational modification introduced to cysteine residues by reacting with iodoacetamide (present in the digestion reaction mix). Under variable modifications Oxidation (M) was selected.

'Monoisotopic' mass values or the mass of the first peak of the isotope distribution is selected to differentiate it from an experimental average mass value.

Peptide mass tolerance of +- 50ppm was used to improve discrimination in peptide finger print and fragment mass tolerance of +- 0.5Da refers to MS/MS fragment ion matching.

Section D reports matched sequence molecular weight, in this case 106112 Da, the probability score and emPAI (Exponentially Modified Protein Abundance Index), the index used for approximate label free quantitation of proteins (Ishihama et al., 2005).

5.4 Target molecule quantification

5.4.1 Quantification by AlphaLISA

The AlphaLISA technology, also referred as 'no-wash ELISA alternative', allows the detection of molecules of interest in serum, plasma, cell culture supernatants or cell lysates in a very sensitive, quantitative, reproducible, and user-friendly way. Almost, any sandwich assay can be developed to detect an analyte of interest. As shown in Figure 5-4 below, an anti-analyte antibody which is biotinylated binds the streptavidin Donor bead while another anti-analyte antibody is conjugated to AlphaLISA Acceptor beads. In the presence of the analyte, the complex forms bringing the beads into close proximity (AlphaLISA development guide (PerkinElmer, 2012) and the excitation of the Donor beads with light at 680nm provokes the release of singlet oxygen molecules that triggers a cascade of energy transfer in the Acceptor beads, resulting in light emission at 615nm wavelength. The fluorescent light signal is measured using a Perkin Elmer EnVision reader and is directly proportional to amount of analyte in solution.

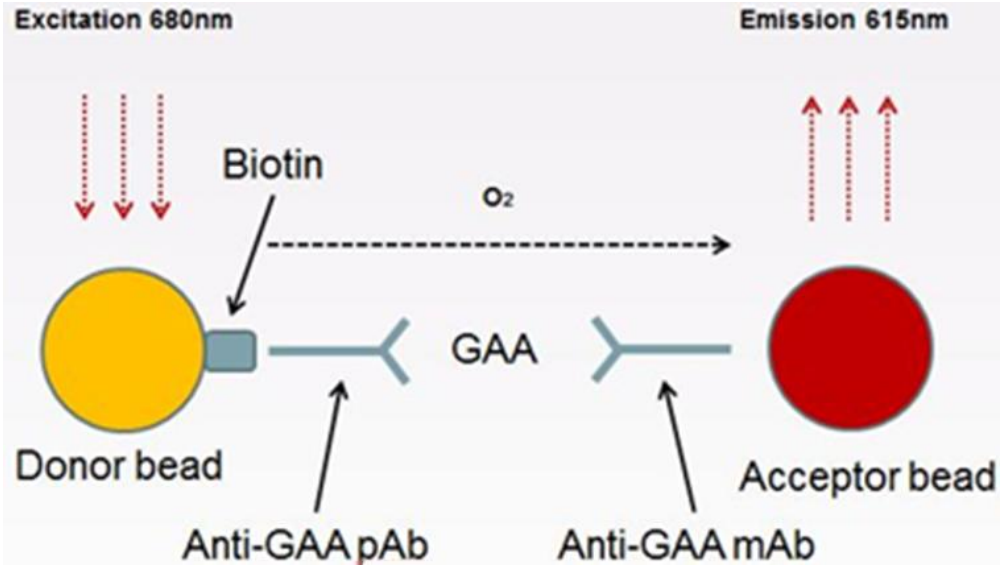


Figure 5-4 Perkin Elmer AlphaLISA assay. In presence of GAA in solution the polyclonal and monoclonal anti-GAA antibodies recognize different epitopes of the GAA molecule and bring

into close proximity the acceptor and the streptavidin coated donor bead. Once this complex is formed, excitation using light at 680nm causes transfer of an oxygen molecule to acceptor bead which triggers cascade of energy transfer which leads to emission of fluorescent light at 615nm wavelength. Signal is measured an EnVision AlphaLISA reader and signal is directly proportional to amount of GAA present in solution.

The vendor sells a number of off-the-shelf detection kits for several widely used assay such as for human IgG and Host Cell Protein (HCP) and they all differ in the antibodies that recognize the specific analyte. GAA, being a therapeutic enzyme for a rare genetic disease, had no available ready to use detection kit at the time this work was performed, which meant a new assay had to be developed.

Developing an AlphaLISA assay for a new analyte is a multi-step process which involves choosing the right pair of antibodies, biotinylation of one of the two antibodies, and conjugation of the other antibody onto the acceptor bead. Lastly, titration of biotinylated antibody is done to determine its working concentration.

The function of the antibody pair is to recognize the analyte of interest and bring the acceptor and donor beads in close proximity. Both antibodies need to be specific to the analyte and must recognize different epitopes of the molecule. One of the two antibodies will have to be biotinylated and the other one should be conjugated to AlphaLISA Acceptor beads.

5.4.1.1 Selection of antibody pairs

In this experiment the pair of monoclonal and polyclonal anti-GAA (specified in section 3.2.5 of materials and methods) was chosen. AlphaLISA development guide recommends testing the two possible combinations such as:

- biotinylated antibody A + antibody B-AlphaLISA Acceptor beads

- biotinylated antibody B + antibody A-AlphaLISA Acceptor beads

In this work however only one combination was tested:

Biotinylated polyclonal antibody + monoclonal antibody - AlphaLISA Acceptor beads.

5.4.1.2 Biotinylation of polyclonal antibody

Biotinylation is the process of attaching covalently biotin to a molecule. By attaching biotin to the polyclonal antibody, this will bind with the streptavidin found on the surface of the donor beads effectively securing the bead to the antibody, necessary step for the dual bead complex formation.

For this purpose we used a biotinylation kit ChromaLink™ Biotin Antibody Labeling Kit from Solulink, USA (cat. B-9007-105K).

Following manufacturer instructions we proceeded to buffer exchanging the polyclonal antibody to remove the 0.02% sodium azide it was stored in. This was achieved by using a 10000mwco dialysis cassette overnight against PBS buffer. The next day the recovered antibody was further desalted using Zeba™ Spin Desalting Columns (Thermofisher USA) and protein concentration was assessed by BCA assay at 3.62 mg/ml. This value was used with the Chromalink Biotin calculation tool in *Table 5-2* below.

ChromaLink Biotin (DMF-soluble) Protein Labeling Calculator	
Protein Name	rabbit IgG
Molecular weight (Daltons)	150000
Protein concentration (mg/mL)	3.62
Amount of Protein (µg)	362
Volume of Protein (µL)	100
Mass of solid ChromaLink Biotin (mg)	0.5
Volume Anyhdrous DMF (µL)	**100**
Concentration of ChromaLink Biotin Working Solution (mg/mL)	5
Equivalents of ChromaLink Biotin	10
Volume of ChromaLink Biotin working solution to add (µL)	**3.91**

Table 5-2 ChromaLink Biotin (DMF-soluble) Protein Labelling Calculator. By entering protein concentration and molecular weight of the protein to be biotinylated, the tool outputs in bold the volume Anhydrous DMF (Dimethylformamide – used to solubilize S-HyNic, part of Solulink's proprietary linker technology) and the volume of ChromaLink Biotin (3.91µl) to add for the biotinylation reaction to take place.

Mixing thoroughly 100 µl of antibody with 100 µl of anhydrous DMF with 3.91µl of Chromalink Biotin and incubating 2h at room temperature allows the antibody to bind covalently biotin.

After incubation is completed, a second buffer exchange allows for biotin in excess to be washed out of reaction solution. The last step in this process is to measure the Molar Substitution Rate (MSR) to assess how efficient the biotinylation reaction has been. This is done by measuring absorbance of the sample at 280 and 354 using a NanoDrop UV reader and using the SoluLink E1% ChromaLink Biotin MSR Calculator .

E1% ChromaLink Biotin MSR Calculator	
Protein name	Rabbit IgG
Protein molecular weight (daltons)	150000
Total volume of biotinylated protein recovered (µL)	90
A_{280} absorbance reading (1-cm pathlength)	0.37
A_{354} absorbance reading (1-cm pathlength)	0.48
Protein E1% (A_{280} of a 10 mg/ml solution)	13.50
Corrected A_{280} absorbance reading	0.260
Total biotinylated protein recovered (µg)	17.3
Corrected protein concentration (mg/mL)	0.19
MSR (biotins/protein)	**12.91**

Table 5-3 Solulink E1% ChromaLink Biotin MSR Calculator. By entering volume of biotinylated protein recovered, absorbance at 280 and 354nm, and standard extinction coefficient (E1%) for human antibody, this tool calculates the MSR. In this experiment per each mole of antibody, close to 13 moles of biotin were incorporated.

5.4.1.3 AlphaLISA acceptor bead conjugation with the antibody

Looking at Figure 5-4 showing the AlphaLISA principle, the analyte is recognized by one antibody biotinylated that will attach to the donor bead via biotin-streptavidin interaction, and by a second antibody conjugated to the acceptor bead. Here we discuss the process of conjugation of the monoclonal antibody to the acceptor bead.

The protocol provided by PerkinElmer discusses the conjugation of 5mg of acceptor beads with a coupling ratio of 50:1 to the antibody. Due to limited amount of acceptor beads available, the protocol was scaled down to 10 fold.

In a 0.5ml Eppendorf, 25µl of acceptor beads conjugated with 0.01mg (16.1 µl) of mAb (0.62 mg/ml) in presence of Sodium cyanoborohydride and 100mM HEPES according to protocol and incubated 20h at 37^0C in the Eppendorf shaker. Beads are washed the next day using a low salt buffer provided by the vendor and stored at 4^0C in dark vial until use. Final concentration 50 µg/ml.

5.4.1.4 Determination of optimal biotinylated antibody concentration

The last step in the AlphaLISA assay development is the determination of the optimal working concentration of biotinylated antibody for the reaction to take place as shown in *Figure 5-5*. This is done by adding increasing concentrations (0.1mM to 100mM) of biotinylated antibody to a fixed amount of sample (5ul of clarified cell culture material), in presence of acceptor beads coupled with anti-GAA mAb and streptavidin donor beads. This is done in a 384well AlphaPlate™ (cat. 6005350) using AlphaLISA Immunoassay Buffer (10X) (Cat No. AL000).

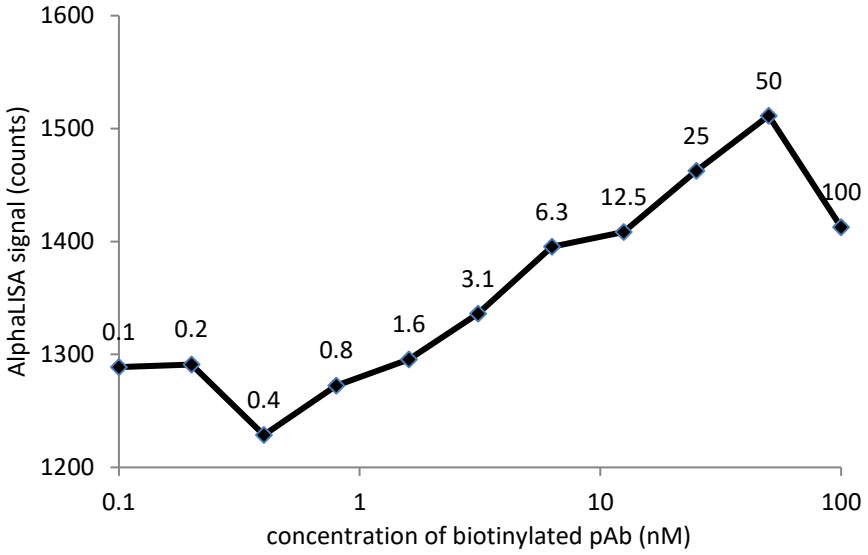


Figure 5-5. Titration curve for GAA detection assay prepared in 25mM HEPES pH 7.4. 5 µl of clarified cell broth were mixed with 10 µl of mAb coupled acceptor beads and 10µl of biotinylated polyclonal antibody serially diluted from 100 mM to 0.1mM and incubated 1h at 23 ^{0}C. At completion of incubation, 25 µl of streptavidin coated donor beads were mixed into each well in darkness and incubated 30 min at 23^{0}C. Plate was read in PerkinElmer EnVision AlphaLISA reader (excitation 680nm emission 615nm).

As described in the AlphaLISA development manual (PerkinElmer, 2012) a bell shape curve was obtained. The highest signal obtained (50 mM) indicates the hook point (highest signal before saturation of the bead binding capacity) of the biotinylated anti-GAA antibody. According to protocol, a sub-hooking concentration should be used as optimal working concentration. The working concentration is of the biotinylated antibody was therefore set as 25 mM.

5.4.1.5 Results

We used the prepared reagents to run an AlphLISA plate to measure amount of GAA in harvest of the three different clones generated in the clone comparison section (4.4.2) and compared them with data from the untransfected CHO cells.

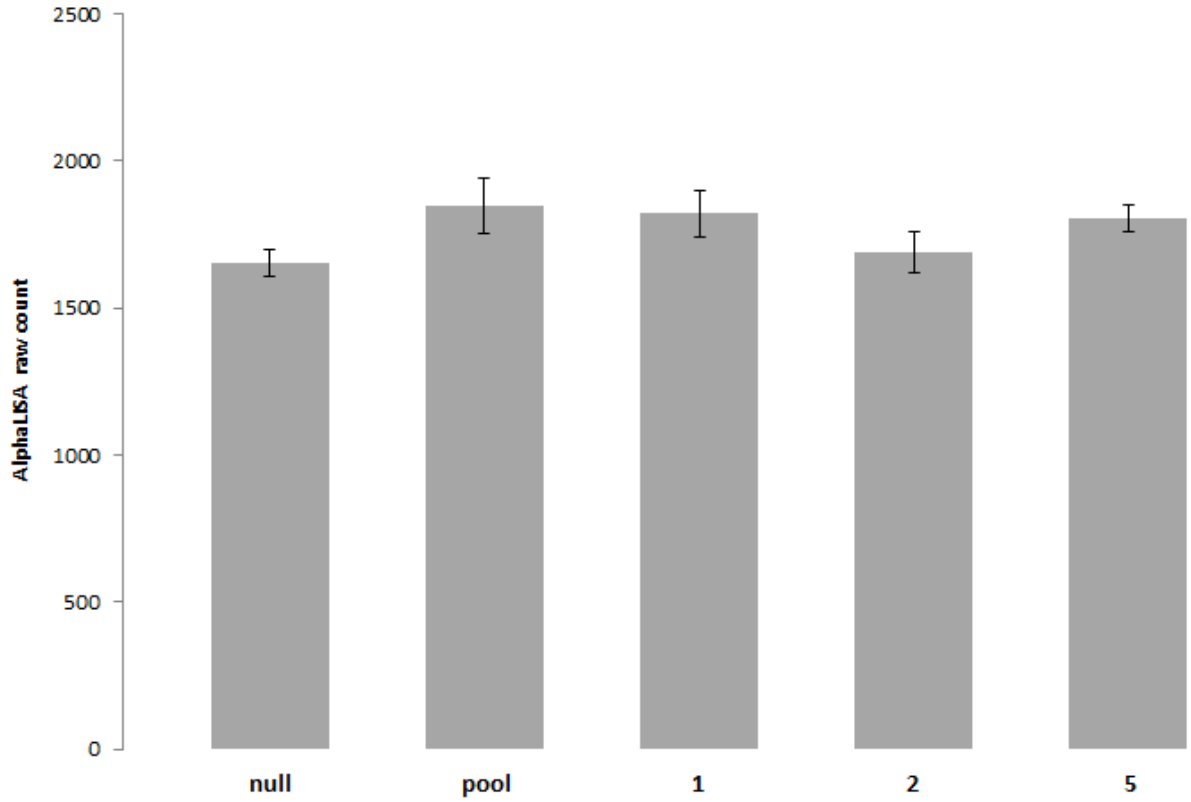


Figure 5-6. GAA amount comparison in CHO untransfected (Null), clonal pool and clones 1, 2 and 5. Samples prepared in triplicate on 96well half area microplate. 5 µl of clarified harvest from each culture mixed with 10 µl 25 mM biotinylated anti-GAA pAb mix and 10 µl acceptor beads coupled to anti-GAA mAb (50 µg/ml). After 1h incubation at r/t, 25 µl of streptavidin coated donor beads were added to each well in dark and seal-film applied. Plate was read after 30 min incubation at r/t using AlphaLISA EnVision reader. Data shows comparable levels of GAA in all samples.

Results shown in Figure 5-6 are in direct contrast with western blot and GAA diagnostic assay results presented in section 4.4.2. Data has previously shown the absence of GAA in Null culture, and its presence at different levels in pool and the three compared clones.

Despite the time invested in developing this assay it was not possible to obtain consistent GAA results from it. This can be attributed to a combination of factors.

Firstly, it is recommended that both antibody combinations are tested for biotinylation and conjugation to acceptor beads. That is biotinylation of antibody A and conjugation of antibody B and the vice-versa. Due to limited amount of antibody available and limited supplier stock at the time, only one combination was tested (biotinylation of pAb and conjugation of mAb). This was perhaps not the optimal combination.

Secondly, it would have been better to use pure GAA reference material instead of clarified cell culture material for developing the assay. As harvest is likely to contain large amount of HCP and other unknown substances that could interact with the AlphaLISA reagents and antibodies. BioMarin could not provide pure GAA reference material at the time the assay was being developed. Reference material would have also been useful for a spike study for quantification purposes.

Subsequently, we became aware of an alternative assay for quantification of GAA (described in detail in the following section 5.4.2). This assay, employed in clinics to determine amount of GAA in patients' blood, is based on a University Medical Center of Rotterdam study. We therefore decided that it would have made more sense to transfer this already working assay into UCL laboratories rather than continuing developing the AlphaLISA assay.

5.4.2 Quantification by GAA diagnostic assay

The search for an alternative method for the quantification of GAA lead to the adoption of a GAA diagnostic assay. This method is based on substrate degradation by GAA and unlike other amylase activity assay, it is specific to this enzyme thanks to the use of an alpha-glucosidase inhibitor. Currently this technique is used in clinics to test patients' blood for levels of GAA in the diagnosis of Pompe disease.

The 2006 paper by (Okumiya et al., 2006) discusses how this method employs glycogen and 4-methylumbelliferyl-α-D-glucopyranoside (4MU-αGlc) as substrates for measuring the lysosomal GAA activity, as previously also described in (Reuser et al., 1978) and (Van Hove, 1996) but incorporates acarbose to eliminate the interference of unrelated α-glucosidases (predominantly maltase-glucoamylase). In their study, the activity of GAA in mixed leucocytes from 25 Pompe patients and 30 healthy subjects was determined and they found that the addition of acarbose, an alpha-glucosidase inhibitor, created a clear separation between the patient and the control ranges.

Assay principle

At acidic pH, α-glucosidase hydrolyses the substrate 4MU-αGlc to 4-methylumbelliferone and glucose. Adding alkaline buffer stops the enzyme reaction and causes the 4-methylumbelliferone to fluoresce at a different wavelength from the unhydrolysed substrate, thereby allowing its measurement in the presence of a vast excess on unhydrolysed substrate.

The addition of acarbose eliminates interference by unrelated α-glucosidases, predominantly maltase-glucoamylase. It has been shown that 3.0 µmol/l acarbose completely inhibits the maltase-glucoamylase activity but the lysosomal acid α-glucosidase by only 5%.

Results:

The aim of this assay is to determine the concentration of GAA in clarified harvest material from clone 1,2 and 5 and compare it to the GAA concentration in heterogeneous clone population (pool) and untransfected culture (Null) (section 4.4.2)

First a serial dilution experiment using Biomarin pure GAA reference standard (2.09mg/ml lot# 11132013) was set up to determine the signal to dilution curve range (Figure 5-7). This experiment allowed us to understand what dilution range would be appropriate for in process sample analysis.

From Figure 5-7 below, we show how the dilution factor and fluorescence signal response are linked by a power curve relationship (resulting equation $y = 2E+06x^{-0.861}$). Reference material dilutions lower than 1:40 ratio resulted in saturation of the instrument absorbance limit (RFU 65000). Following this experiment, we empirically decided that a dilution of 1:100 (20 mg/L) for the GAA reference material would place it somewhere in the middle of the power response curve.

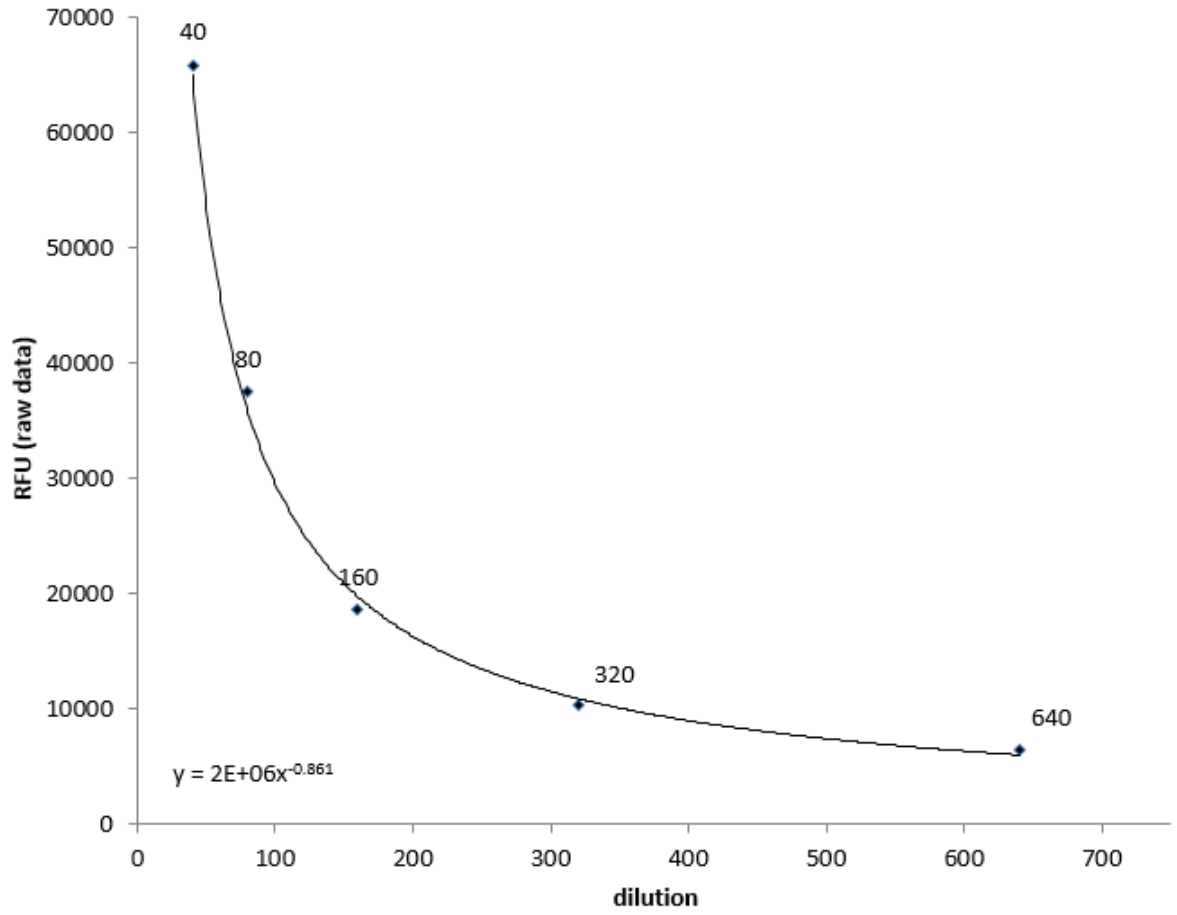


Figure 5-7. GAA Reference dilution response via diagnostic assay on BMG Fluorstar Galaxy fluorimeter. Reference material (2 mg/ml) was first diluted 1:10, then 1:2 dilutions 22 times in triplicate in assay buffer (range shown 1:40 to 1:640 or 50 to 3 mg/L). 5µl of each dilution was mixed with 100 µl substrate (4MU-αGlc pH4.0) and incubated at 37ºC for 60 min. At the end of incubation period, the reaction was stopped with 1M glycine buffer pH 10.4. Plate was read in scan mode at 360-480 nm range, with absorbance peak at 420 nm. Data range selected shows power curve relationship (with equation $y = 2E+06x^{-0.861}$) between fluorescent signal and dilution factor. Reference dilution factors lower than 1:40 saturated the instrument signal (instrument maximum signal is 65000 RFU)

Subsequently we tested clarified harvest from clones. As GAA amount in harvest is expected to be about 1/10th of the amount found in the reference (around 0.2mg/ml), the samples were further diluted 1:10 in order to allow GAA amount to be similar in concentration to the reference. After diluting the samples in assay buffer they were prepared in triplicate the same way as the reference.

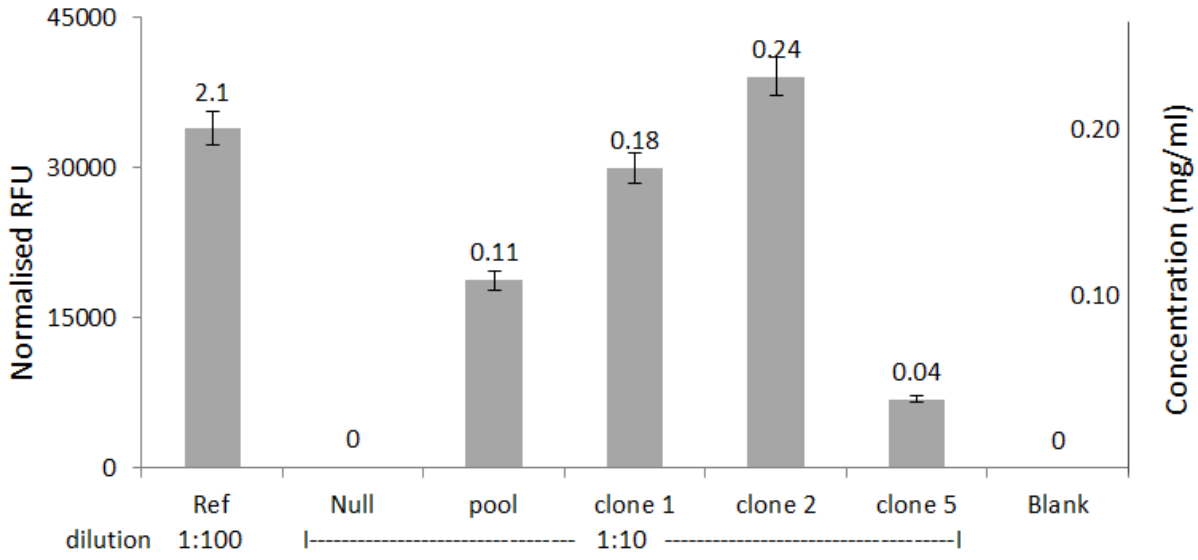


Figure 5-8. GAA diagnostic assay, determination of GAA amount in clarified harvest from Null, pool and clone 1, 2 and 5. Samples prepared in triplicate and read on a flat-bottom 96well polystyrene plate via BMG Fluorstar Galaxy fluorimeter. Reference material (2.09 mg/ml) was diluted 1:100 in assay buffer, all other samples were diluted 1:10. 5µl of each dilution was mixed with 100µl substrate (4MU-αGlc acarbose pH 4.0) and incubated at 37°C for 60 min. At the end of incubation period, the reaction was stopped with 1M glycine buffer pH 10.4. The plate was read in scan mode at 360-480 nm range, with absorbance peak at 420 nm. RFU was normalized by subtracting blank value and resulting RFU values were used to calculate GAA concentration based on reference value [i.e. clone1 concentration: (clone 1 RFU x 2.09 / reference RFU)/10 = (29997 x 2.09 / 34002) / 10 = 0.18 mg/ml]. Error bars are ±1 standard deviation on triplicate measurement.

Acknowledgment goes to Dr. Derralynn Hughes and her team at the haematology department of Royal Free Hospital in London for the training and allowing the use of their reagents and instruments.

5.5 Chapter conclusions

This chapter focused on the characterization of the target molecule GAA. The work here presented served as an important bridge between stable cell line generation (Result chapter one) and bioprocess design and characterization of impurities (Result Chapters three and four). Ensuring the stable cell line generated produced the expected molecule was critical before undergoing downstream development effort.

As GAA is known to exist in different isoforms of molecular weights ranging from 70 to 110KDa a series of analytical assays was put in place to characterize the target molecule and determine titer and molecular weight. While the main specie was found to be at 106KDa, other isoforms at different stages of maturity were also observed.

As the mature GAA was expected to have a molecular weight of 106KDa, we first ran clarified harvest material from stable clone 1 on a SDS-PAGE to visualise the target molecule and host cell protein (HCP) profile and compared this with a Null culture clarified harvest. This experiment showed two broadly similar HCP profiles, with only visible difference being Clone 1 having a clear band at around 110KDa mark. As shown later, most mammalian cell lines share a largely overlapping HCP profile it is therefore not surprising that the two samples looked so similar. The presence of the significant band at 110KDa was a clear indication that GAA is being over-expressed. This however needed to be confirmed with an identity or immuno based assay.

Western blot in this case was a good additional assay for the confirmation of size and presence of the target molecule as it relies on specific anti-GAA antibodies to recognize GAA. Clone 1

clarified harvest was run alongside GAA pure reference on a SDS-PAGE gel and transferred onto a PVDF membrane. We then proceeded with incubation with the specific antibody pair and exposure to diaminobenzidine (DAB) tablets which revealed that the overexpressed band at around 100KDa seen in SDS-PAGE was in fact GAA.

Identity confirmation was also provided by MS/MS analysis using MALDI-TOF-TOF on UltrafleXtreme MALDI-TOF instrument (Bruker, Coventry, UK). SDS-PAGE band of interest was excised, digested with sequencing grade porcine trypsin and analysed via tandem mass spectrometry, which confirmed the identity as GAA and molecular weight of 106KDa.

After having demonstrated the size and identity of target molecule we wanted an assay that could determine GAA concentration in samples in a fast and reliable way. Although western-blot can be used semi-quantitatively, due to its fairly long running time (2 days per gel) it was deemed not appropriate for this purpose. We investigated AlphaLISA, another immunoassay based technique for the determination of GAA concentration in samples. The development of this assay however did not produce satisfactory results mainly because it relied on antibodies and a pure GAA reference not available at the time of development.

Lastly we tried an enzymatic activity assay used in the diagnosis of Pompe disease from blood samples. This assay is capable of measuring GAA activity without interference from other amylases thanks to the addition of acarbose. This assay was transferred from the Royal Free Hospital Haematology Department and it was utilised to quantify GAA in clarified harvest of different clones for clone comparison purposes.

Overall, the work presented in this chapter showed how stable clone 1 CHO produces around 0.18 mg/ml of 106KDa GAA in 8 to 10 days of small scale (cell flask) fermentation in CD-CHO hygromycin+ media.

6 Chapter three

Downstream Purification process

6.1 Introduction

So far we have generated a stable mammalian cell line producing therapeutic enzyme GAA and developed some of the assays needed to characterize the target molecule during production. We have established that harvest supernatant of 8 day old clone 1 culture contains approximately 0.18 g/l of 106 KDa GAA. The theoretical pI was calculated as 5.4 and hydropathicity gradient (GRAVY) -0.129.

The aim of this chapter is to present the work that went into the development of the two-step chromatographic GAA purification process. This process is designed to mimic sponsor company BioMarin Pharmaceutical GILT-tag (Glycosylation Independent Lysosomal Targeting) GAA purification process.

The development of the purification process resulted in a two-step strategy based on exploitation of target molecule net charge first via anion exchange (IEX) using Capto-Q resin (GE) and then its hydrophobicity via hydrophobic interaction chromatography (HIC) using a butyl HIC resin. High throughput resin screening was used to help chose among different resins and binding conditions for the intermediate step.

Some of the analytical assays developed in the previous chapter such as western blot and the GAA diagnostic assay, were used to assess GAA presence and concentration in elution fractions. Impurity and protease analysis of process samples at different stages of the bioprocess are also reported in this chapter.

6.2 Fermentation and harvest

A cell bank vial (1 ml at approximately 10^7 cells) was thawed from liquid nitrogen in 37^0C water bath and quickly resuspended in 10 ml of fresh pre-warmed CD-CHO media. DMSO was removed by 2000 g, 5 min centrifugation and resuspension in 35 ml of pre-warmed media containing 250 µg/ml Hygromycin B in a 125 ml cell flask. They were then incubated for 3 days in 37^0C 5% CO_2 to allow recovery and passaged to 0.3 10^6 cell/ml in a 500ml shake flask containing 125ml CD-CHO (250µg/ml hygromycin B). Flasks were incubated under moderate agitation (125rpm on orbital shaker) at 37^0C 5% CO_2, until they reached at least 10 million cells/ml (6-8 days). Cell number and viability was assessed daily using a ViCell (Beckman Coulter USA). On harvest day, cell broth from each flask was transferred to multiple 50ml conical tubes and spun at 2000g for 10 minutes to pellet the cells. The supernatant was pooled and it was filtered through a 0.22µm Stericup® (Merck Millipore USA). An equal volume of buffer A-P (50 mM sodium acetate, 20 mM sodium phosphate, 1 M NaC1 adjusted to pH 5.2) was added to reduce the pH and inhibit phosphatase activity. This mixture was stored at -20^0 C until purification.

6.3 Capture step– anion exchange chromatography

The capture chromatographic step is designed to isolate, concentrate and stabilize the target protein.

IEX separates large proteins, small nucleotides and aminoacids on the basis of their charge. IEX media have charged functional groups that bind molecules with an opposite charge. Bound molecules are eluted from the medium by displacement, via the application of an increasing concentration of a similarly charged molecule. Proteins have numerous functional groups that can have either positive or negative charges. By adjusting the pH or the ionic concentration of the mobile phase, proteins can be separated. IEX is used for capture of the target protein from

bulk impurities from large-volumes, as an intermediate purification step, or as a final step for high resolution purification to remove impurities.

Proteins carry a charge based on the pH of the environment. The pH at which a molecule has no net charge is called its isoelectric point (pI). GAA theoretical pI based on its aminoacid sequence was calculated as 5.4 (ExPASy - SIB Swiss Institute of Bioinformatics) therefore at a pH above 5.4 the protein should carry a net negative charge and bind to an anion exchanger. Vice versa at pH below pI the protein should carry a net positive charge and bind to a cation exchanger. Although technically both exchangers could have been used by working at pH range 6-7 for the anion exchanger or between 4-5 pH for cation exchanger, keeping the target molecule in the cation exchanger low pH range, would expose it to proteolytic degradation threat. For this reason we decided to go for anion exchange separation.

Anion exchange resins come in two types: strong and weak. The main difference between the two is that in strong anion exchangers the number of charges remains constant regardless of the buffer pH. These types of resins retain their selectivity and capacity over a wide pH range. Quaternary ammonium (Q) is an example of strong anion exchanger and its functional group is $-N^+(CH_3)_3$. Weak anion exchangers in contrast, display pH-dependent behaviour and so deliver optimal performance over only a small pH range. When the pH of the buffer no longer matches the acid dissociation constant (pKa) of the resin functional group, these resins suffer significant capacity reduction. Diethylaminoethyl (DEAE) is an example of weak anion exchanger and its functional group is $-N^+(C_2H_5)_2$. Usually a strong ion exchanger is tried first due to its wider pH range of operation, weak exchangers are used when strong ones fail as selectivity of weak and strong ion exchangers often differ.

Bead size had to be taken into consideration. As this is early in the purification process (e.g., first step - the capture stage), speed of sample processing is more important than resolution as the

volume is large and the target protein needs to be removed from harvest and stabilized. For this reason particle size (dp) was selected to allow low backpressure at high flow rates. Some of the commonly used large particle size resins available are Sepharose (90nm) and Capto (75nm) (both GE Life Sciences).

Lastly, a small volume (1 ml) prepacked column (HiTrap) was used due to limited scale of the feed (lab research scale) and convenience.

After all these considerations a HiTrap Capto Q 1ml IEX column was chosen for this step.

A standard preloaded GE-developed linear gradient purification method was used as a starting method which provided good initial product separation with minor tweaking required. Resulting chromatogram is shown in *Figure 6-1* below.

6.3.1 Chromatogram

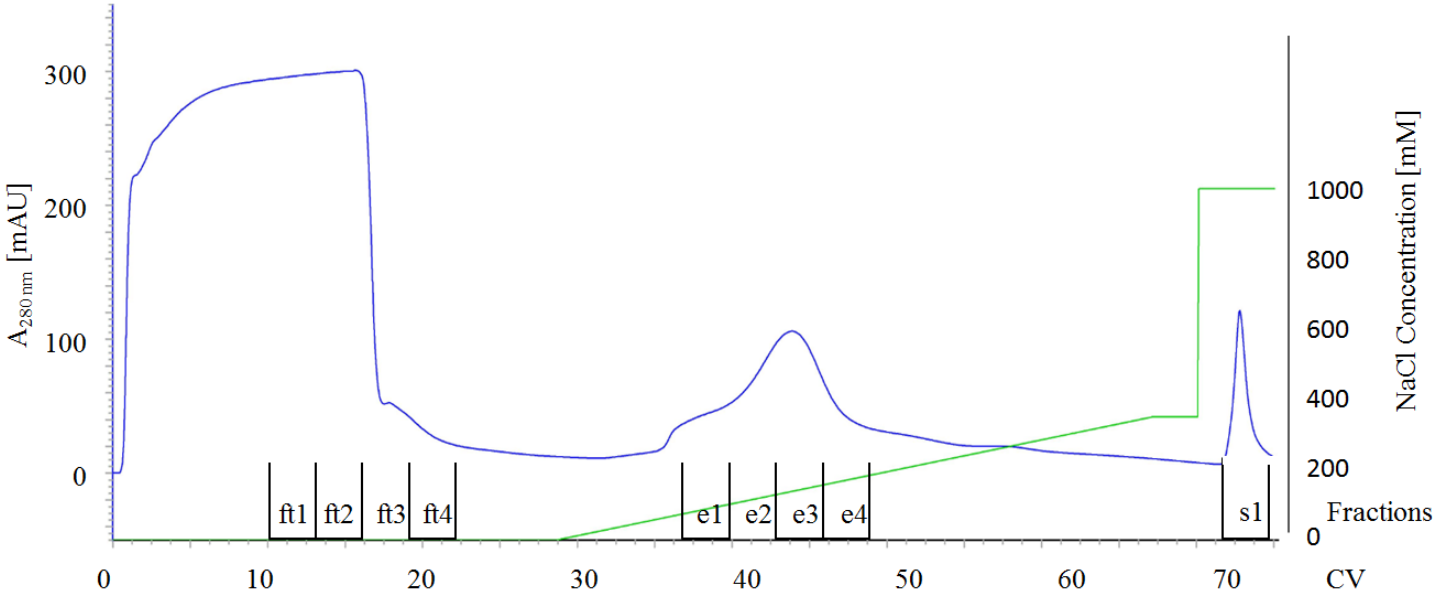


Figure 6-1 Clone 1 clarified harvest anion exchange chromatogram. Acquired on GE AKTA Avant 25 column GE CaptoQ HiTrap 1ml, 1 ml/min flow rate, equilibration buffer 50mM Sodium Acetate pH6.0 (BufferA), elution buffer 1M Sodium Chloride in 50mM Sodium Acetate pH6.0 (BufferB). 30ml of Clone 1 were diluted 1:4 into equilibration buffer to favour binding.

After equilibrating column with buffer A for 10CV, 1.8 mg of material was loaded at 1 ml/min via sampling port. Salt concentration was kept at 0 until completion of flow through (approximately 20CV) and then for wash cycle (15CV). The NaCl concentration was increased in a linear gradient fashion over 40CV up to 400mM to elute the GAA. Elution peak showed a pre-peak. 1M NaCl for 10CV allowed tightly bound material to elute as well. Column was regenerated after each injection with 20CV 500mM NaOH. 2ml fractions were taken as shown during flowthrough (ft), elution (e) and strip (s).

6.3.2 Fraction analysis

The purpose of the following experiments was to demonstrate whether all GAA elutes in elution fraction or some is lost in the flow through or strip fractions.

The 2ml fractions collected during the IEX chromatographic run were analysed via gel electrophoresis (Figure 6-2), western blot (Figure 6-3) and the GAA diagnostic assay (Figure 6-4) for presence and quantity of target molecule.

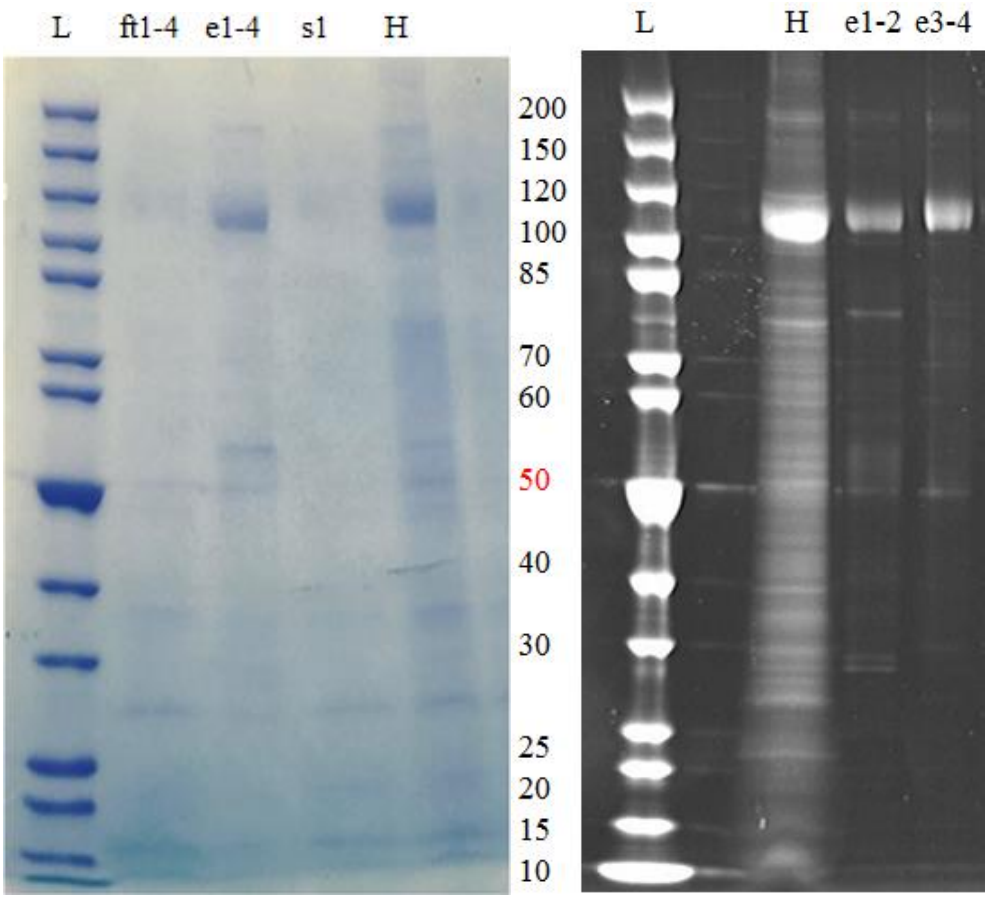


Figure 6-2 gel electrophoresis NuPAGE™ Novex™ 4-12% Bis-Tris (ladder cat# 26614 Thermo Fisher, USA) Comassie blue stained (left) SYPRO ruby stained (right). Samples: fractions from Figure 6-1. ft1-4 (pooled flow through fraction 1 to 4), e1-4 (pooled elution fraction 1 through 4), s1 (strip), H (clarified harvest), e1-2 (pooled elution fraction 1 and 2), e3-4 (pooled elution fraction 3 and 4). Data shows GAA over-expressed (~110KDa) in H together with impurities of sizes covering the whole gel size spectrum. e1-4 shows qualitatively the degree of purification of the capture step. A protein of similar size (presumably GAA) is present in elution together with other proteins at ~175, 55 and 30KDa. Same samples were rerun and gel was stained via SYPRO ruby to confirm results and increase sensitivity.

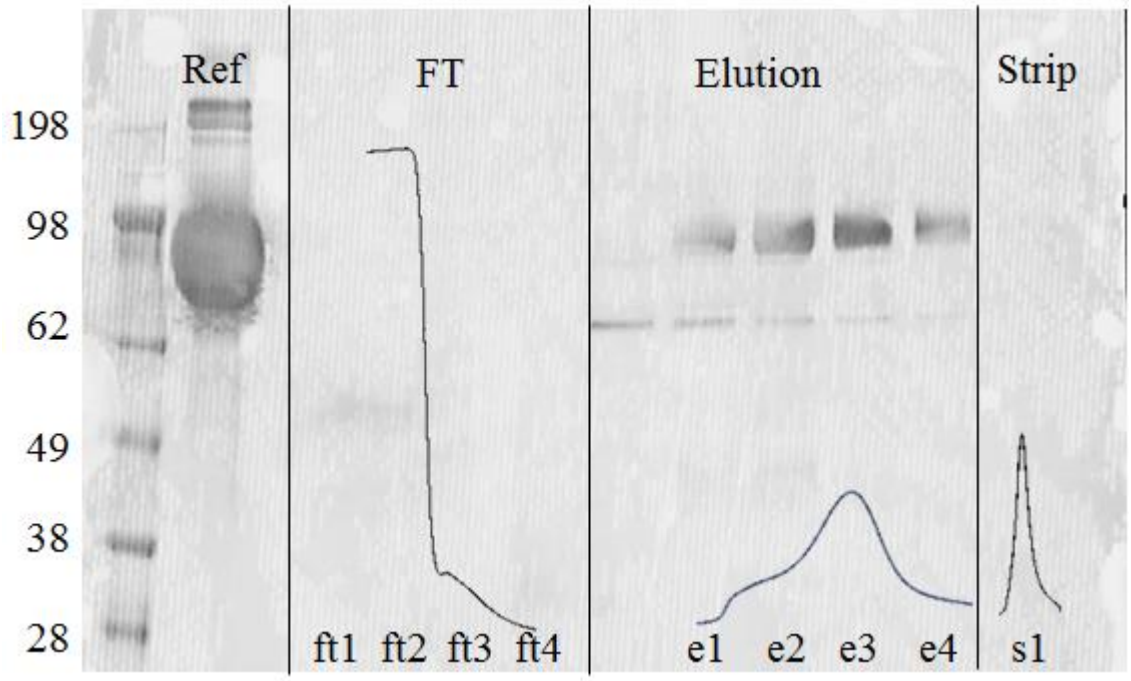


Figure 6-3. Western blot of anion exchange fractions (chromatogram shown in Figure 6-1). After collection, fraction samples were concentrated using 10000MWCO filters and 20µg of protein was loaded on 4-12% bis-tris SDS-PAGE gel and separated according to MW. Proteins where then transferred to PVDF membrane from Bis-Tris 4-12% Gel, (ladder cat# 26614 Thermo Fisher, USA). Primary anti-GAA rabbit mAb (w.d.1:1000 Abcam cat.# ab137068), Secondary anti rabbit polyclonal Ab HRP-linked (w.d. 1:1000). Ref sample is GAA drug substance reference sample 2.08 mg/ml (BioMarin, lot# 11132013). Chromatogram (from Figure 6-1) was copied on membrane to better show the fractions. Data shows how GAA can be seen present in varying concentration only in the four elution fractions and no GAA is observed in flowthrough and strip fractions. Elution fraction e3 seems to have the highest GAA amount, which corresponds to the peak in the chromatogram attached. The GAA bands were slightly below the expected molecular weight for GAA (106 KDa), however so was the reference standard. This can be attributed to the molecular weight standard being compromised or inaccurate.

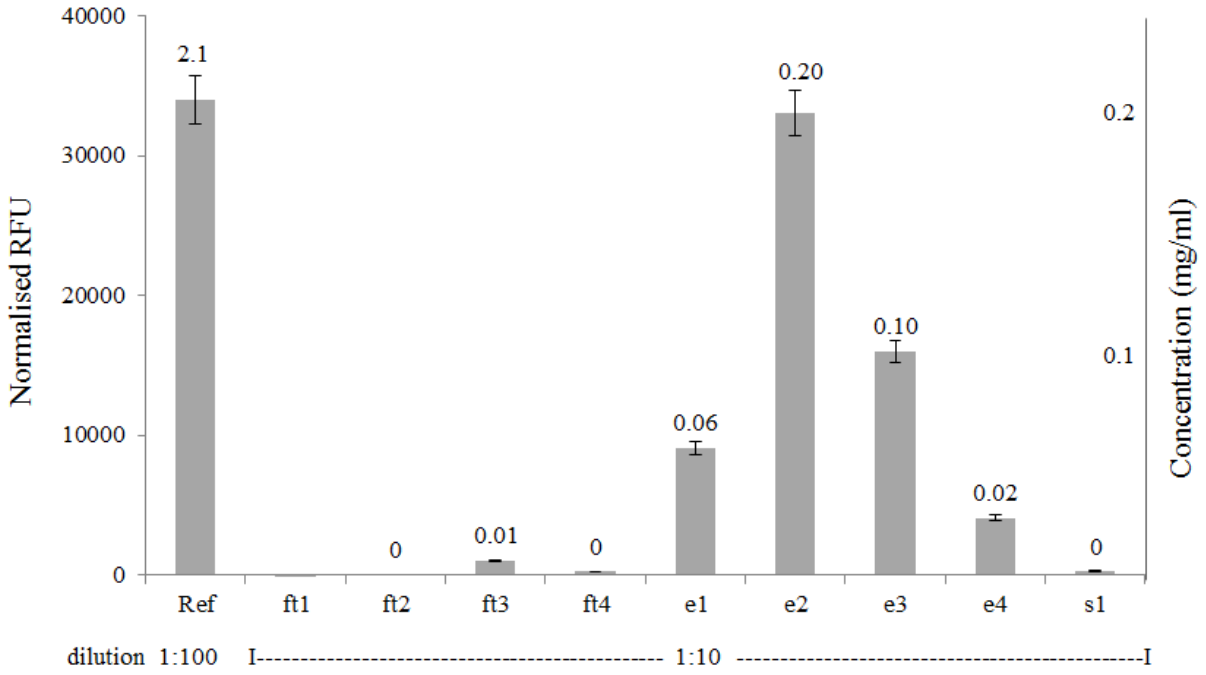


Figure 6-4. GAA diagnostic assay determination of GAA amount IEX fractions from figure 6.1. Samples prepared in triplicate and read on a flat-bottom 96 well polystyrene plate via BMG Fluorstar Galaxy fluorimeter. Reference material (2.09 mg/ml) was diluted 1:100 in assay buffer, all other samples 1:10. 5 µl of each dilution was mixed with 100µl substrate (4MU-αGlc acarbose pH 4.0) and incubated at 37 °C for 60min as per described assay method. At the end of incubation period, reaction was stopped with 50µl 1M glycine buffer pH 10.4. The plate was read in scan mode at 360-480 nm range, with absorbance peak at 420nm. RFU was normalized by subtracting blank value and resulting RFU values were used to calculate GAA concentration based on reference value.

As previously mentioned, the purpose of this step is to isolate, concentrate and stabilize the target protein. The figures shown in this section confirm the fact that the anion exchange step is able to capture the target molecule at low salt (pH 6.0). Elution occurs at a NaCl concentration between 200 and 400 mM.

From gel electrophoresis image (*Figure 6-2*) we can see a comparison between clarified harvest and post IEX fractions. In both gels (Comassie and SYPRO stained), the elution fraction is visibly purer than the harvest and it clearly contains a protein of the same molecular weight as

the one seen in harvest sample. Although we cannot be sure from this experiment that the marked band is GAA, we can assert that the MW is around 110KDa, the expected MW for GAA. The flow-through sample, strangely, shows very little protein present, perhaps due to wrong sample loading concentration. This is in clear discordance with chromatogram showing flowthrough absorbance maxing out the UV detector.

In order to identify the 110 KDa band in the elution fractions, the same samples were run again on a NuPage 4-12% BT gel and transferred onto a PVDF membrane for western blotting analysis (Figure 6-3). The processed membrane showed that the band seen in elution fractions from electrophoresis is in fact GAA as it is recognised by the specific monoclonal anti-GAA - secondary antibody couple linked to HRP. This also confirms that flow through and strip fractions are target molecule free.

The GAA diagnostic assay was used to quantify eluting GAA and calculate step yield. From Figure 6-4, up to 0.2 mg/ml of GAA is present in the four fractions. Each fraction volume is 2ml. As the load volume was 10ml and the concentration of GAA in the load was 0.18 mg/ml, we can calculate mass balance and purification yield.

Mass balance

Load (In): 10ml x 0.18mg/ml = 1.8 mg

Elution (Out): 4 x 2ml x 0.2mg/ml = 1.6 mg

Percent recovery

Mass out / mass in x 100 = 1.6 mg / 1.8 mg x 100 = 89%

6.4 Intermediate purification – hydrophobic interaction chromatography

The intermediate purification objective is to remove most of the bulk impurities such as other proteins, nucleic acids, HCPs and lipids. An effective separation strategy is commonly based on the principle of exploitation of a combination of techniques with independent selectivity. In this instance, the target molecule was first purified based on its charge via anion exchange, leaving it in a mixture of similarly charged molecules, and co-eluting species as later shown. In order to further purify GAA, the subsequent separation had to be based on a different selectivity. Hydrophobic interaction was picked as a good candidate for intermediate purification as able to separate molecules based on hydrophobicity. This was also decided to mimic the sponsor company downstream process.

HIC separation principle is complementary to ion exchange. Proteins are composed by amino acids, and amino acids can have either polar (hydrophilic) or non-polar (hydrophobic) side chains. The specific sequence of aminoacid determines protein folding structures. Depending on the type of side chains that are mostly exposed, the whole protein will acquire either polar on non-polar behaviour. When a sample containing a mixture of proteins is applied to a HIC column in high salt, hydrophobic and hydrophilic regions interact with the resin and bind to it. The salt interacts with water molecules to reduce solvation of the protein molecules in solution (Porath, 1986). As salt is gradually decreased, hydrophobic regions become exposed and are absorbed by the mobile media and therefore start eluting. The more hydrophobic the molecule, the less salt is needed to promote binding.

HIC solid phase is composed by resin beads with bound hydrophobic ligands. Bead size choice depends on application as it impacts resolving power needed and the pressure drop. For instance, larger beads are more suitable for capture to intermediate purification as they tend to have a lower pressure with impure and viscous samples, vice versa smaller beads are more suitable for

intermediate and polishing steps when the sample is mostly pure and more resolving power is required. In terms of ligands there are multiple types on the market with different levels of hydrophobicity and the two tested in this work are butyl and phenyl. A protein's adsorption behaviour is determined by the type of immobilized ligand. In general, straight chain alkyl ligands demonstrate hydrophobic character while aryl ligands show a mixed mode behaviour where both aromatic and hydrophobic interactions are possible (Hofstee and Otillio, 1978). The choice of ligand type is empirically determined following hydrophobic ligand order from GE.

Degree of ligand substitution must also be taken into consideration as protein binding capacity increases with an increased degree of substitution of the immobilized ligand. With a high level of ligand substitution, the binding capacity remains constant; however, the affinity of the interaction increases (Jennissen and Heilmeyer, 1975).

Lastly pH and temperature also have an impact on binding capacity of the resin as they can alter a protein conformation and therefore affect ionic interactions and hydrogen bonds between polypeptide chains. As seen in the next section a specific pH range was tested at room temperature.

6.4.1 HIC development strategy

The development of the HIC step was a particularly lengthy process as it did not produce satisfactory results due to the first principle approach strategy adopted. Most of the initial experiments were done using a ligand that turned out to be not the ideal one. Much time could have been saved had we used high throughput plates strategy from the beginning.

The starting point of the method development was a paper by Van Hove (Van Hove et al., 1997) in which purification of GAA was achieved by binding target molecule onto Concanavalin A Sepharose 4B column and then loading the eluate onto a phenylsepharose HP column in 0.75M

ammonium sulphate pH 5.3. This was the first experimental run tested: IEX eluate was buffer exchanged using 10000 mwco concentrator spinner filters in 0.75M ammonium sulphate pH 5.3 and loaded onto HiTrap Phenylsepharose low sub column. Elution was achieved by gradual decrease of conductivity, or in other words linear increase of the no-salt buffer (buffer B). The chromatogram expected elution region however showed little absorbance (data shown in next section). Due to the small amount of protein present in loaded sample, it was difficult to determine whether any binding occurred. To try to increase affinity we then switched the column with a higher degree of substitution Phenyl Sepharose HiTrap (high sub) (GE Healthcare, USA cat# 17-1355-01), with similar results.

At this point, a more systematic approach was adopted and a screening experiment was set up: GAA was loaded onto the Phenyl Sepharose HiTrap (high sub) column using different equilibration buffers at different salt concentrations and pH levels (details in next section). A total of 13 conditions were screened, this however did not lead to satisfactory results. The small amount of protein loaded onto the column would not bind to the beads and was observed in the flowthrough peak.

Different resins were then suggested for testing by BioMarin in a high throughput plate based format type of experiment. Binding of pure GAA material was tested using two types of resin (phenyl and butyl) in different salt buffers. This approach eventually led to successful result.

6.4.2 Screening of resins and binding conditions via AKTA

6.4.2.1 First principle approach

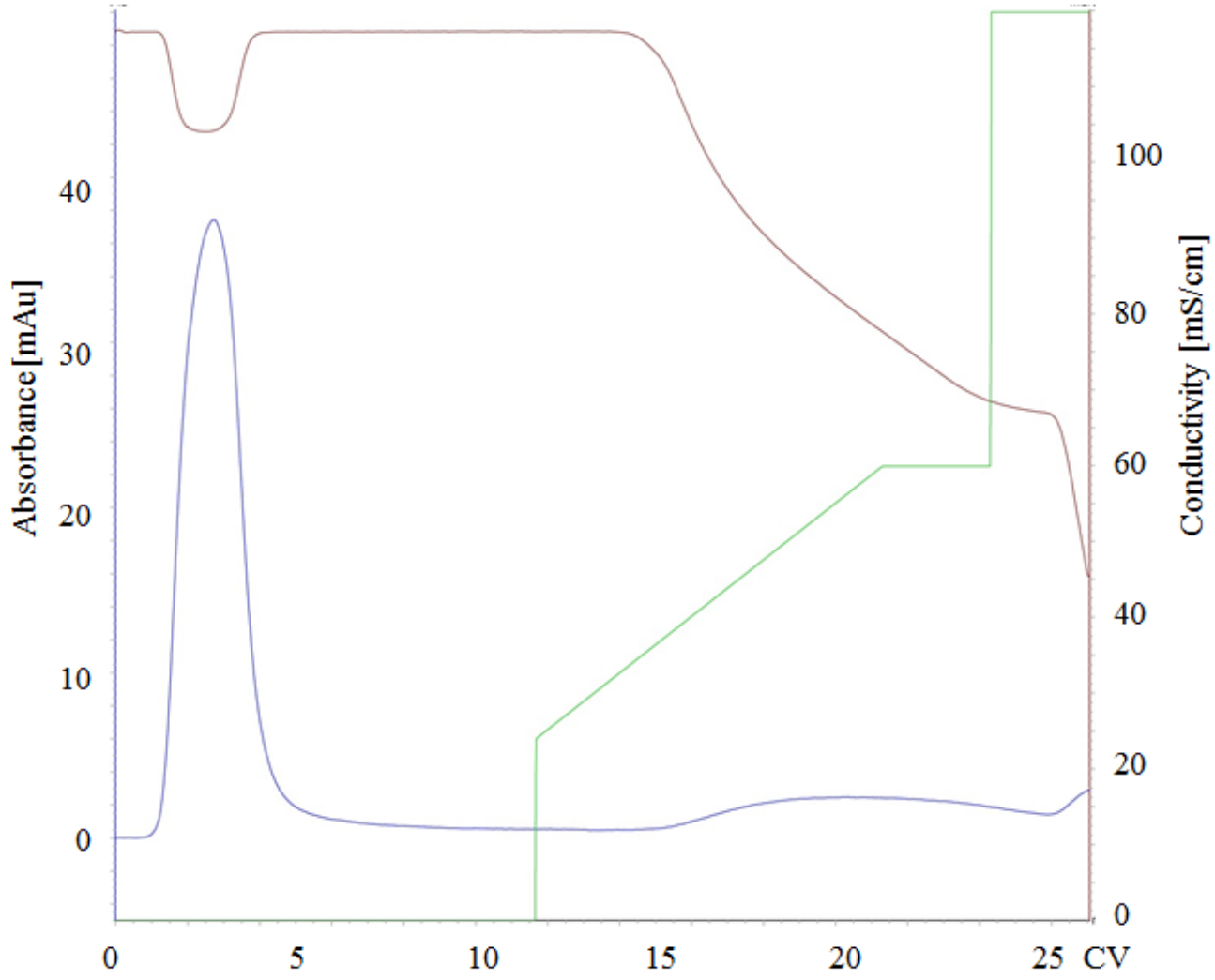


Figure 6-5. HIC chromatogram, Run 1: 0.75M ammonium sulphate pH 5.3. Blue: absorbance (mAu); green: % buffer B; brown: conductivity (mS/cm). IEX eluate fraction was buffer exchanged in Buffer A via multiple washes through 10000mwco Amicon filter keeping final volume constant. HIC method reagents: buffer A 0.75M ammonium sulphate pH 5.3, buffer B 20mM acetate pH 5.3, flow rate 0.5ml/min, column HiTrap phenyl FF low sub (cat# 17-1353-01 GE HealthCare) HIC method sequence: equilibration 5CV buffer A, injection 2ml via sample loop, wash 10CV buffer B, Elution 20% to 50% buffer B over 10CV, strip 100% buffer B 5CV, CIP 0.5M NaOH 5CV. Figure shows large portion of protein mixture does not bind. No elution peak visible.

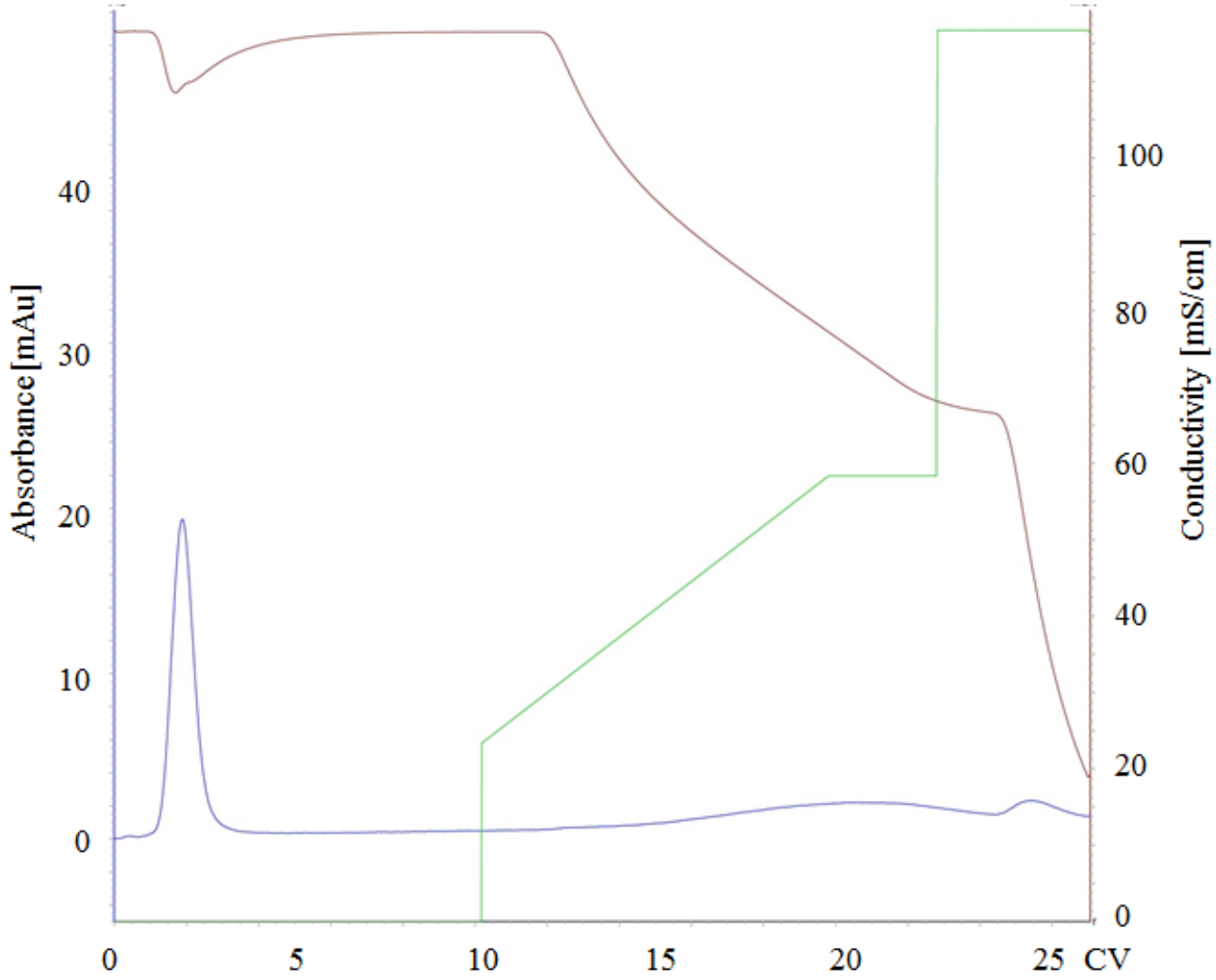


Figure 6-6. HIC chromatogram, Run 2: GAA reference in 0.75mM ammonium sulphate. Blue: absorbance (mAu); green: % buffer B; brown: conductivity (mS/cm). Injection of GAA reference material diluted in buffer A. HIC method reagents: Buffer A 0.75M ammonium sulphate pH 5.3 20 mM acetate, Buffer B 20mM acetate pH 5.3, flow rate 0.5ml/min, column HiTrap Phenyl FF low sub (cat# 17-1353-01 GE HealthCare) HIC method sequence: equilibration 5CV buffer A, injection 2ml via sample loop, wash 10CV buffer B, Elution 20% to 50% buffer B over 10CV, strip 100% buffer B 5CV, CIP 0.5M NaOH 5CV. Figure shows GAA reference does not bind under these conditions.

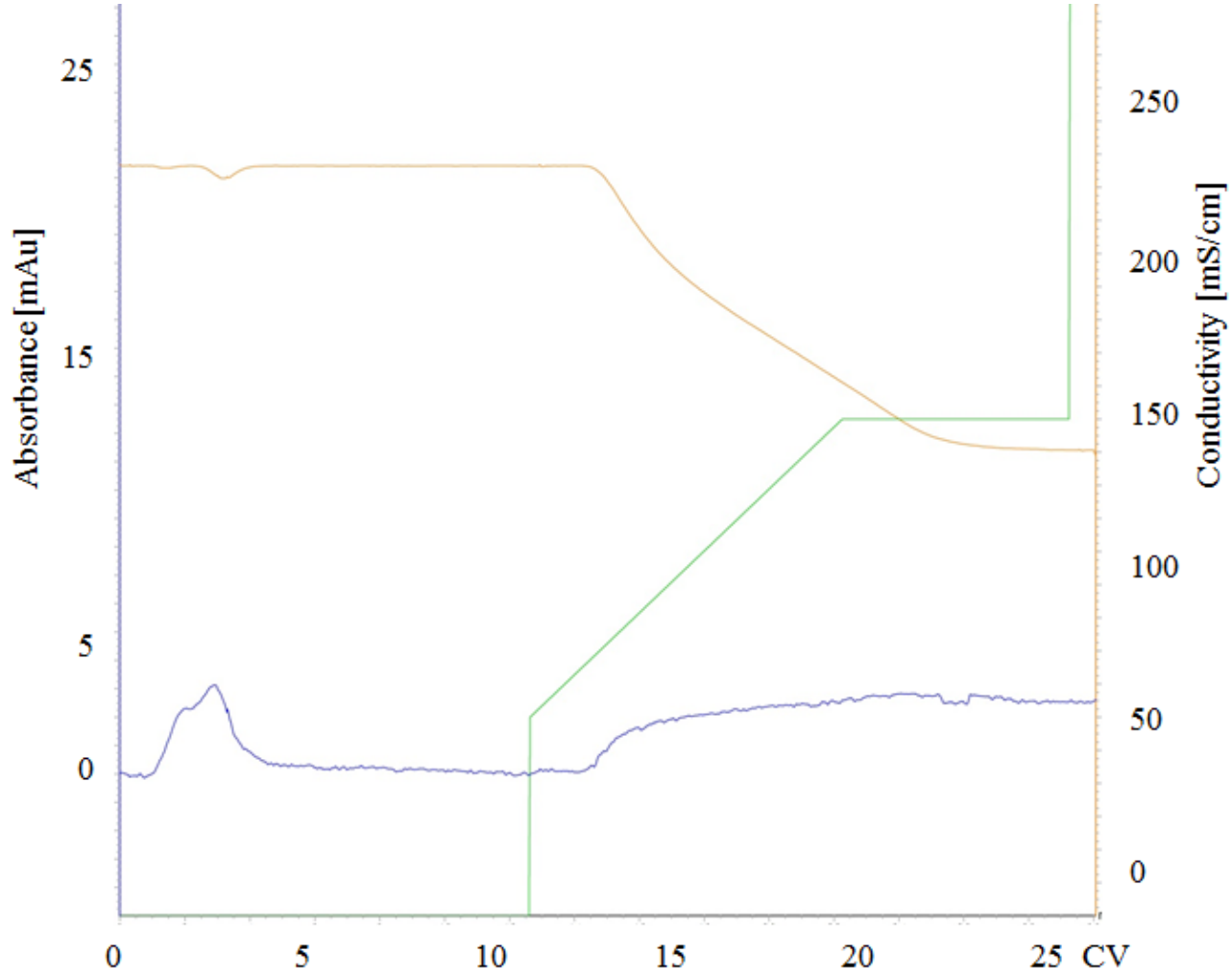


Figure 6-7. HIC chromatogram, Run 3: high-sub HiTrap Phenyl FF column used. Blue: absorbance (mAu); green: % buffer B; brown: conductivity (mS/cm). IEX eluate fraction was buffer exchanged in Buffer A via multiple washes through 10000mwco Amicon filter keeping final volume constant. HIC method reagents: Buffer A 2.0M ammonium sulphate pH 5.3, 20 mM acetate, Buffer B 20mM acetate pH 5.3, flow rate 0.5ml/min, column HiTrap Phenyl FF high sub (cat# 17-1355-01 GE HealthCare) HIC method sequence: equilibration 5CV buffer A, injection 2ml via sample loop, wash 10CV buffer B, Elution 20% to 50% buffer B over 10CV, strip 100% buffer B 5CV, CIP 0.5M NaOH 5CV. Figure shows large portion of protein mixture still does not bind. As conductivity starts lowering proteins start coming off the column however no peak can be identified.

6.4.2.2 Binding buffer screening

Buffer A	Buffer B	pH	Initial conductivity (mS/cm)
0.75 M AS	20 mM acetate	5.3	115
1.0 M AS	50 mM NaP	5	140
1.0 M AS	50 mM NaP	6	140
1.0 M AS	50 mM NaP	7	140
1.5 M AS	50 mM NaP	4	185
1.5 M AS	50 mM NaP	5	185
1.5 M AS	50 mM NaP	6	185
1.5 M AS	50 mM NaP	7	185
2.0 M AS	50 mM NaP	5	220
2.0 M AS	50 mM NaP	6	220
2.0 M AS	50 mM NaP	7	220

Table 6-1. HIC binding buffer screening study of IEX eluate fraction. AS: ammonium sulphate, NaP: sodium phosphate. Column HiTrap Phenyl FF high sub (cat# 17-1355-01 GE HealthCare).

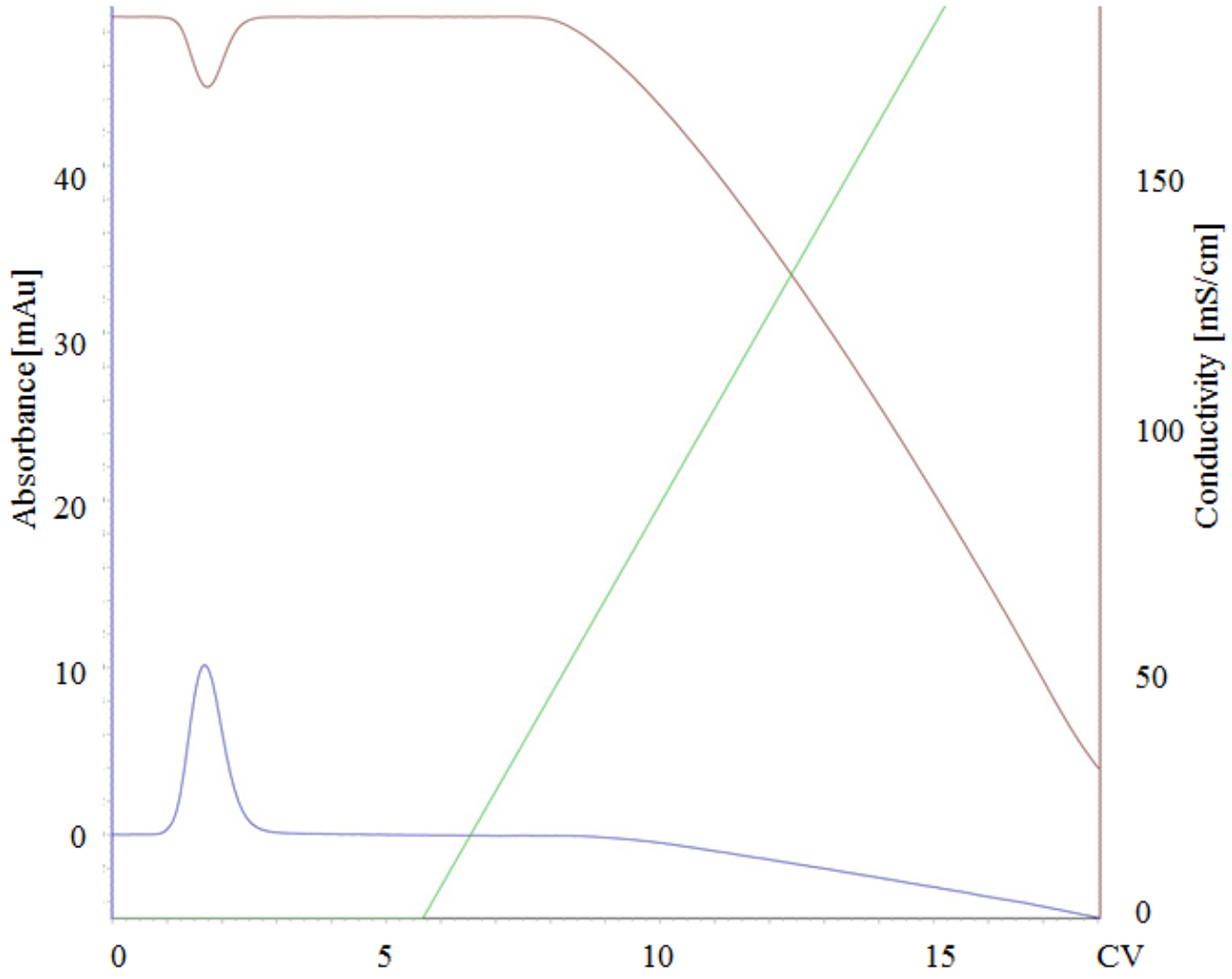


Figure 6-8 HIC binding screening study representative chromatogram (12 total chromatograms show similar profiles). Blue: absorbance (mAu); green: % buffer B; brown: conductivity (mS/cm). IEX eluate fraction was buffer exchanged in Buffer A (refer to table 6-1) via multiple washes through 10000mwco Amicon filter keeping final volume constant. HIC method reagents: Buffer A and B refer to table 6-1, flow rate 0.5ml/min, column HiTrap Phenyl FF high sub (cat# 17-1355-01 GE HealthCare) HIC method sequence: equilibration 5CV buffer A, injection 2ml via sample loop, wash 10CV buffer B, Elution 20% to 50% buffer B over 10CV, strip 100% buffer B 5CV, CIP 0.5M NaOH 5CV. Figure shows very little protein in breakthrough fraction and no elution occurring. IEX peak absorbance is at around 100 mAu, but here it is only 1/10th of that. Probable protein loss during buffer exchange could be the cause of this.

6.4.2.3 Resin binding screening

Having been unsuccessful in obtaining satisfactory separation using first principle approach we decided to screen alternative resins and buffers. This step was performed at BioMarin laboratories in Novato CA with Dr. John Henstrand's group assistance. This study was designed to determine GAA optimal binding condition on two different resins and multiple salts. This procedure allowed for quick screening of multiple conditions using 96 well plate setting.

(µg)			Load		W1	W2	W3	Wash		E1	E2	E3	Elution		S1	S2	S3	Strip		
Resin	Equilibration buffer (A)	Elution Buffer (B)	T=0	T=12	100% (A)	100% (A)	100% (A)	tot	% Bound	75% (A)	50% (A)	25% (A)	tot	% Recovery	100% (B)	100% (B)	100% (B)	tot	% Loss	% Total lost
Phenyl sepharose 6FF	1.0M NaCl	50mM sodium acetate	2000	2	6	2	4	12	99.3	15	83	483	581	29.1	769	327	128	1224	61.2	70.3
	1.5M NaCl		2000	1	4	12	0	16	99.2	5	127	481	613	30.7	769	365	128	1262	63.1	68.5
	2.0M NaCl		2000	1	3	0	0	3	99.8	2	327	353	682	34.1	769	378	128	1275	63.8	65.7
	1.0M AS		2000	123					93.9											
	1.5M AS		2000	70					96.5											
	2.0M AS		2000	35					98.3											
Buthyl sepharose 4FF	1.0M NaCl		2000	76					96.2											
	1.5M NaCl		2000	2	1	1	2	4	99.7	20	511	575	1106	55.3	480	75	12	567	28.4	44.4
	2.0M NaCl		2000	0	4	0	0	4	99.8	24	520	808	1352	67.6	321	32	4	357	17.9	32.2
	1.0M AS		2000	308					84.6											
	1.5M AS		2000	288					85.6											
	2.0M AS		2000	244					87.8											

Table 6-2 Resin binding screening experiment. Quantities refer to protein amount found in each 1 ml sample and are expressed in µg (green to red relative good to bad performance). 2000 µg of pure GAA reference material (BMN 103; 2.0mg/ml; ε = 1.56) was mixed with 200 µl of 80% slurry of resin in 1 ml of the different equilibration buffers as per table, mixed and incubated o/n under gentle agitation at 4°C to allow protein binding. The next day tubes were spun and supernatant concentration determined via absorbance at 280 nm (T=12h), beads were then resuspended in 1 ml of fresh buffer A, mixed, spun and concentration of supernatant determined and step repeated 3 times (W1, W2, W3). Elution was performed by resuspending the beads in 1 ml of elution buffers of decreasing salt concentrations E1 (75%A+25%B), E2 (50%A+50%B), E3 (25%A+75%B). For strip step beads were resuspended in 100% buffer B (no salt) 3 times as above (S1, S2, S3). The condition that yielded to the best protein recovery and least amount of loss was (A) 2M NaCl, (B) 50mM sodium acetate on Butyl Sepharose 4FF at pH 5.0. Conditions that did not show satisfactory results were not brought forward (blank spaces).

This experiment allowed the quick screening of 12 different conditions for binding of pure GAA onto HIC resin. After mixing 2mg of pure GAA reference material with the resin beads in specific equilibration buffer, o/n incubation and three wash cycles, only 5 conditions out of 12 show >99% binding (highlighted in green under wash in Table 6-2). This showed that binding generally occurred better under NaCl than ammonium sulphate. The conditions that did not achieve >99% protein binding after three cycles of buffer A washing, were not brought forward (amber and red). After 3 cycles of elution at decreasing salt concentrations, GAA showed better unbinding from butyl rather than phenyl resin, with one condition achieving over 67% recovery (highlighted in green under elution). Lastly the strip cycle showed how out of the 5 experimental conditions, more GAA came off at this stage from phenyl >60%) vs butyl resin (<30%) indicating butyl performing overall better in binding and eluting GAA minimizing protein loss.

Out of the 12 experimental conditions tested, the condition that yielded the best protein recovery and least amount of loss was binding in 2M NaCl and eluting in 50mM sodium acetate on Butyl Sepharose 4FF at pH 5.0.

6.4.3 Purification of IEX eluate on AKTA via HIC using resulting conditions from resin binding screening study.

6.4.3.1 GAA Reference

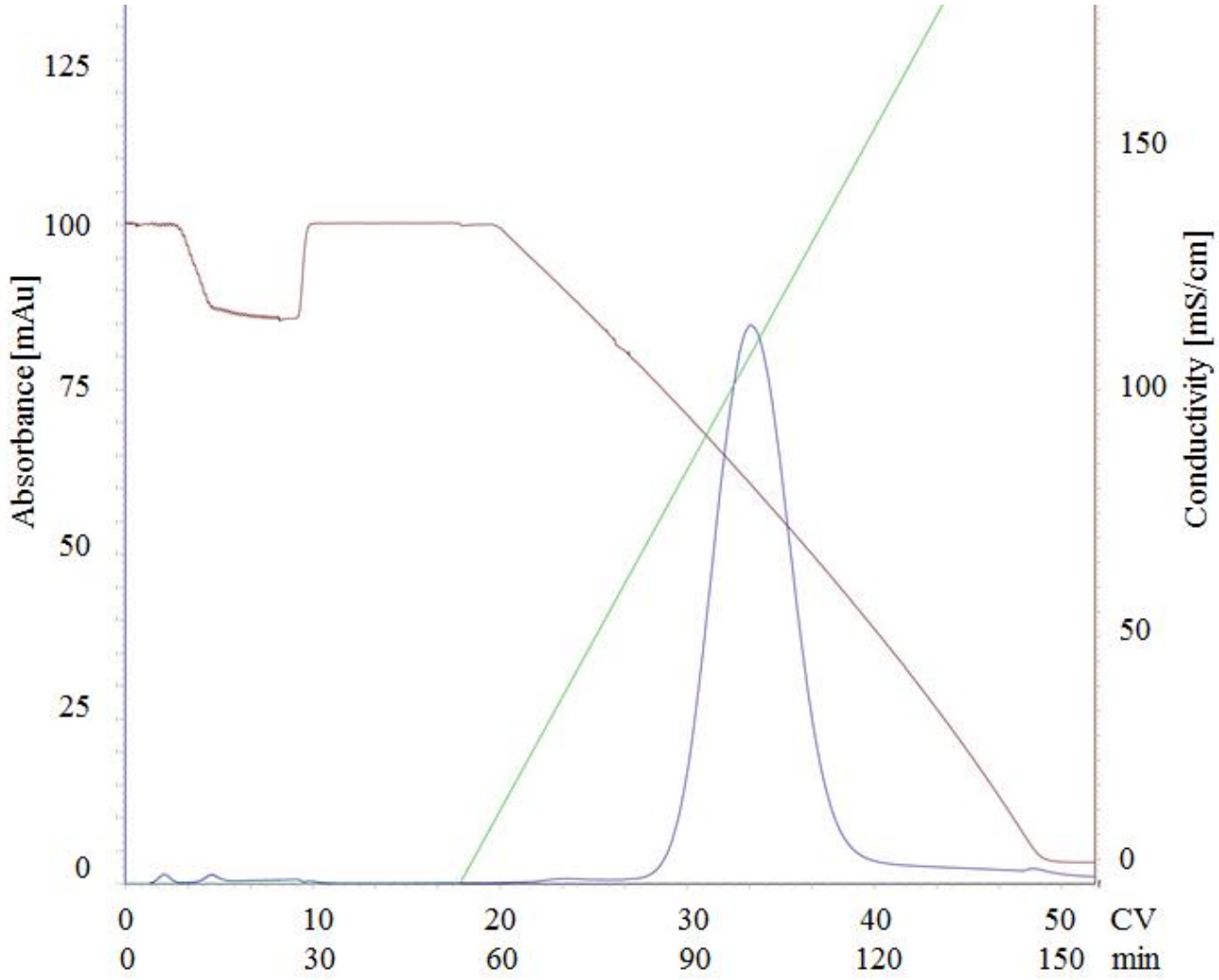


Figure 6-9. GAA reference material injection on Butyl Sepharose 4FF CV = 3 ml (GE Healthcare, USA cat# 17098001). Load: 5mg GAA diluted in 15 ml 2.0M NaCl, flow rate 1 ml/min, wash 3CV, elution 0-100% 50mM sodium acetate 10CV, 1ml fractions collected.

The separation conditions established via resin binding screening were tested on an AKTA Avant 25 using an in-house packed Butyl Sepharose 4FF (d = 0.66cm, bed height = 9 cm, CV = 3ml) column.

First we injected 5 mg of pure GAA diluted in 15 ml of 2.0 M NaCl at a flow rate of 1 ml/min onto the column. Elution was performed via linear gradient from 0 to 100% buffer B (50 mM sodium acetate) over 10v CV. The 33 1ml elution fractions collected were pooled together. Protein absorbance at 280nm was performed via NanoDrop and concentration determined to be 0.084 mg/ml ($\varepsilon = 1.56\ (mg/ml)^{-1}\ cm^{-1}$).

Protein recovery:

Load (IN): 15 ml at 0.33 mg/ml = 5 mg

Elution (OUT): 33 ml at 0.084 mg/ml = 2.77 mg

% protein recovery: 100 x 2.77 / 5 = 55.4%

The protein recovery calculated (55.4%) was slightly lower than resin binding screening experiment result (66.7%), This was probably due to experimental conditions not being comparable. Although high throughput screening was useful in selecting the right running conditions, results from the two systems cannot be fully comparable due to the different geometry and fluid dynamics of a plate microwell and a chromatographic column.

6.4.3.2 IEX Eluate sample

We then proceeded to injecting IEX eluate sample (from *Figure 6-1*). The four 2 ml eluate sample fractions were pooled together and buffer exchanged into 2.0 M NaCl pH 5.3 using an Amicon 10000mwco spinner concentrator maintaining the final volume constant (8 ml). The sample was then injected onto the Butyl Sepharose 4FF column (CV 3ml) at a flow rate of 1 ml/min. Column was washed with 3 CV of 2 M NaCl, followed by 10 CV of 0-100% linear grade elution with 50 mM sodium acetate to lower salt concentration (and buffer conductivity)

and allow protein to elute. Unlike previous attempts at purifying IEX eluate sample, this time a clear elution peak could be observed (*Figure 6-10*).

The HIC elution fractions were pooled together and the sample was run on SYPRO ruby SDS-PAGE gel to visually assess degree of purification compared to the capture step (*Figure 6-11*).

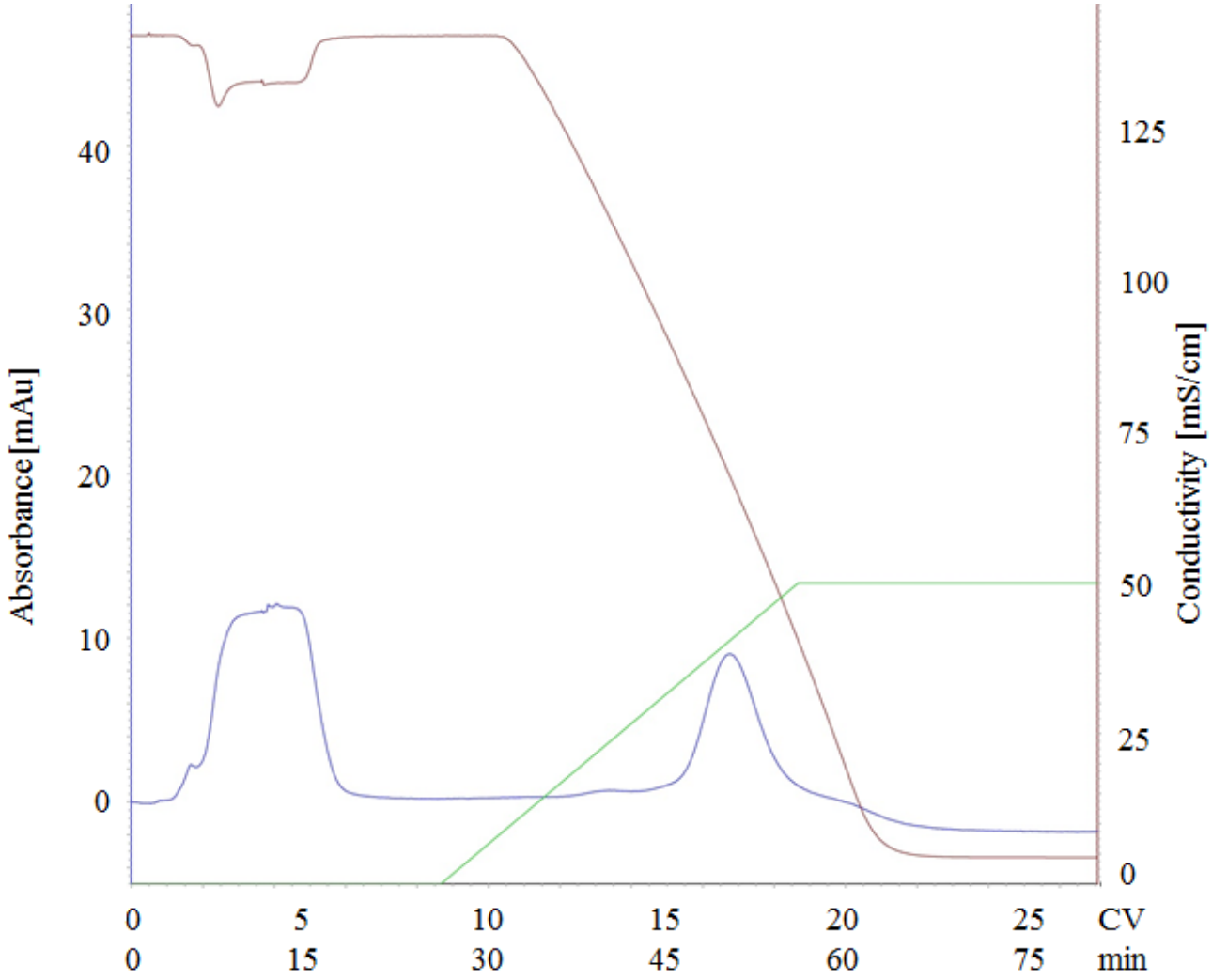


Figure 6-10. HIC separation of GAA from IEX elution peak. Blue: absorbance (mAu); green: % buffer B; brown: conductivity (mS/cm). Butyl Sepharose 4FF CV = 3 ml (GE Healthcare, USA cat# 17098001). Load: 0.4 mg in 2.0M NaCl, flow rate 1 ml/min, wash 3CV, elution 0-100% 50mM sodium acetate 10CV, 1ml fractions collected.

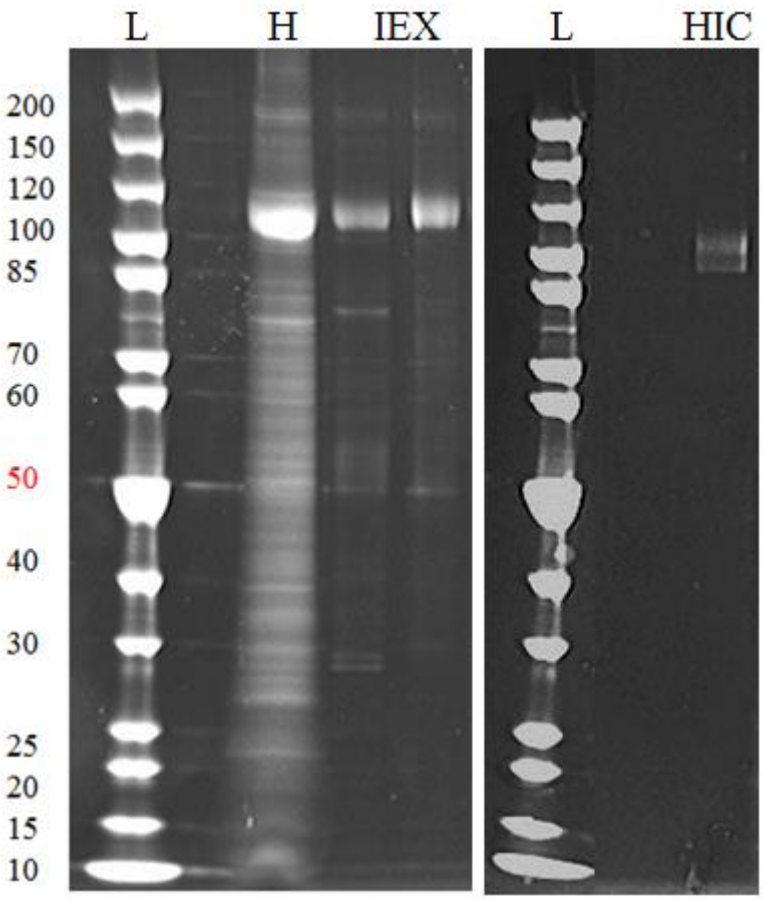


Figure 6-11. SYPRO ruby stained SDS PAGE comparison of H: HCCF sample, IEX elution pre-peak (left) and main peak (right) and HIC elution peak. GAA protein band at 106KDa in HIC sample shows a higher degree of purity than the IEX GAA band.

The chromatogram in *Figure 6-10* shows a good degree of separation occurring under the running conditions selected, in contrast with all the previous runs. The elution fraction was run on a SYPRO ruby SDS-PAGE (*Figure 6-11*) which confirmed presence of GAA at 106KDa and increased overall feed purity. The purity of GAA after HIC was determined via LC/MSMS to be above 80% as shown later in section 7.4.3.

6.5 Chapter conclusions

The development of a purification process capable of removing the biotherapeutic enzyme of interest (GAA) from a CHO cell culture process stream was an important milestone of the EngD

project. The main goal of this process was to investigate protease and HCP removal at the two chromatographic steps while achieving an acceptable level of GAA purity. Although important, achieving highest purity of the target molecule was never the end goal of this process mainly because of the purely research nature of the project. The GAA produced through this process was never meant for human use but rather should be seen as an academic exercise to process development and HCP identification and removal activities.

The harvest of a GAA producing clonal CHO cell culture, generated in results chapter one, was first clarified via centrifugation and filtration to remove whole cells and cell debris. The clarified Harvest Cell Culture Fluid (cHCCF) was then subjected to two rounds of chromatographic separation, first based on protein charge (capture IEX) then followed by protein hydrophobicity (intermediate HIC). The final target protein purity was above 80% as assessed by MS (section 7.4.3 Figure 7-10)

In the first chromatographic step, Anion Exchange, protein charge was used to capture GAA from the process stream using a 1ml HiTrap capto-Q column on an Akta Avant 25. GAA presence in elution fraction was confirmed by western blot. The step protein recovery was calculated as 89% and purification yield 46%.

For intermediate purification, Hydrophobic Interaction chromatography was investigated. After numerous failed attempts at separating GAA in IEX eluate on a Phenyl Sepharose 6FF HiTrap resin a resin binding study was set up at BioMarin laboratories. This experiment lead to the adoption of butyl resin instead of phenyl and determination of 2 M NaCl pH5.0 as the ideal equilibration buffer. IEX eluate was separated using the conditions determined in the screening experiment after being buffer exchanged into 2M NaCl. An SDS-PAGE gel of the HIC elution fraction was run next to the IEX elution fraction and HCCF sample and GAA was observed at 106KDa molecular weight. Protein purity was measured at 82% via MS analysis (next chapter).

7 Chapter four

HCP, proteases and cellular stress due to target protein overexpression.

7.1 Introduction

In the previous chapter we have discussed the development of the downstream process for the purification of GAA from cell line 1 harvest. The focus of this last chapter will be identification and quantification of HCP released by the chosen cell line, assessment of protease presence in the feed throughout the bioprocess and determination of whether GAA over-expression has an effect on general cell and lysosomal phenotype by acquiring high resolution images of affected organelles.

This chapter is divided in four result sections ordered by assay type: qualitative protease characterization via zymography; HCP screening via immunoassay (AlphaLISA); protease identification via mass spectrometry; and lastly lysosomal phenotype characterization.

Proteases present in in-process samples were at first qualitatively characterized via zymography. This technique was used early on in the project during the cloning to compare transfection strategies and their effect on protease release and later on to determine protease clarification performance of the capture step.

Subsequently we proceeded with assay development for total HCP quantification via AlphaLISA. This assay was used to quantify HCP before and after capture chromatography to determine step yield.

Once we had a good idea of the total amount of HCP present in the feed, we proceeded with identification of those HCP via mass spectrometry. We mainly focussed on HCP in GAA Clone 1 HCCF and capture step eluate and compared eluates from different cell lines as explained later. For these sections three tandem MS techniques were employed: (1) in-gel tryptic digestion

MALDI MS/MS, (2) MALDI TOF LC-MS/MS and (3) label free high resolution LC-MSMS. The first revealed to be labour intensive and error/contamination prone and best suited for product identification rather than HCPs, the second was more high throughput due to separation via LC but reliant on approximate quantification by emPAI index, while the latter was high throughput as well as truly quantitative via label free quantitation through direct comparison with a BSA standard reference. Other non-MS based strategies were considered for HCP identification, including Proteseeker by G-Biosciences, an assay based on exposure of protease mix to 12 different kinds of inhibitors together with colorimetric protease screening. This assay was attempted on IEX eluate fraction and however produced no satisfactory results.

In the last result section of this chapter we present TEM acquired images of cells lysosomes to assess whether GAA over expression has an impact on their morphology and overall cell stress, which could relate to overall levels of secreted proteases. Different cell lines lysosomes were compared for reference.

7.2 Qualitative protease characterization via zymography

The purpose of this assay was to qualitatively assess presence and approximate molecular weight of proteases present in in-process samples. Zymography combines two analytical methods: protein separation based on molecular weight via gel electrophoresis and protease digestion of specific substrate (casein) within the gel resulting in white bands showing on blue gel background.

Proteins in samples are linearized by the SDS in the sample buffer but not heat denatured in order to maintain their functionality. Linearization allows proteins to move through the gel mesh at a rate proportional to their molecular weight under the applied electrostatic current. At the end of the run, the proteins are renatured by incubating the gel in non-ionic detergents (Renaturing buffer, Thermo Fisher Scientific USA) to re-establish enzymatic activity. The gel is then

incubated overnight at RT in developing buffer (Thermo Fisher Scientific USA) which adds divalent metal cations required for enzymatic activity. The following day the gel is stained with Comassie blue and de-stained in 40%MeOH, 10%Acetone. Regions of protease activity appear as clear bands against a dark blue background where the protease has digested the substrate.

There are several variants commercially available and they differ by substrate used. The ones used here were 12%Novex® casein (0.05%) zymograms gels (cat# EC64052BOX Thermo Fisher Scientific, USA).

7.2.1 Protease distribution in transient and stable transfection cultures.

The first study was done to compare protease content in transient and stable transfection clones from 4.3.1 and 4.3.2

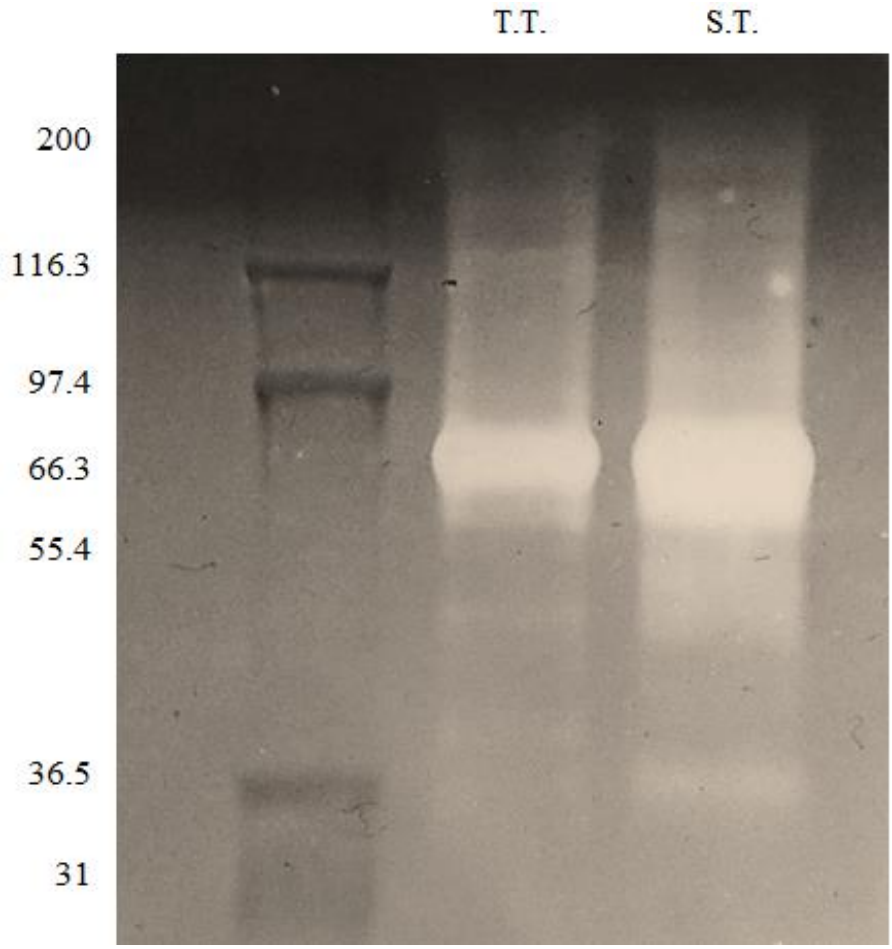


Figure 7-1. Comassie blue stained zymogram gel of centrifuged harvested cell culture fluid (HCCF) of transient transfection (T.T. section 4.3.1) and stable transfection (S.T. section 4.3.2) cultures. Incubation time 72h at 37°C pH 5.0 under moderate agitation. Molecular weight standard Mark12 unstained (cat# LC5677, Thermo Fisher Scientific, USA). Data shows the wide range of proteolytic activity in both harvest samples, with heavier presence at around 80KDa molecular weight. Also stable transfection GAA clone 1 sample seems to have heavier proteolytic presence then its transient transfection counterpart. Note. Absence of 66.3 and 55.4 KDa ladder marks makes data questionable. Note. GAA not being a protease is not visible in zymograms.

7.2.2 Protease distribution in Anion Exchange Capture chromatography

Following the capture step purification of HCCF of clone 1 (section 6.3.1), we wanted to determine protease presence in HCCF and ion exchange eluate fraction. Samples were buffer

exchanged into PBS and concentrated 4X in a 10000mwco spin concentrator. After determining sample total protein concentration via nano-drop, 5mg of protein was mixed with sample buffer as per protocol (material and methods section 3.4.1) and loaded onto the gel. At the end of the run the gel was incubated with renaturing buffer to allow proteins to renature and become active and then incubated for 72h in developing buffer pH 5.0 at 37^0C under moderate agitation. Gel was finally stained in Comassie blue and destained in 40% MeOH 10% Acetic Acid.

The original resulting data image (Figure 7-2 right) was enhanced (left) to better reveal the molecular weight standard. Image shows a general reduction in proteolytic activity from HCCF to eluate fraction, with the exception of a conspicuous amount left at 80KDa mark. This result is in accordance with later figures (*Figure 7-3*). According to MS identification data shown in result section 7.4.3 below (see also appendix 1 for complete HCP identification list), proteases of ~80KDa size present in GAA IEX eluates are Dipeptidyl peptidase 3, a metalloproteinase and Prolyl endopeptidase a serine protease. The implications of the presence of these two proteases coeluting with the product has not been evaluated due to lack of time and we think it would be valuable work to undertake in the future.

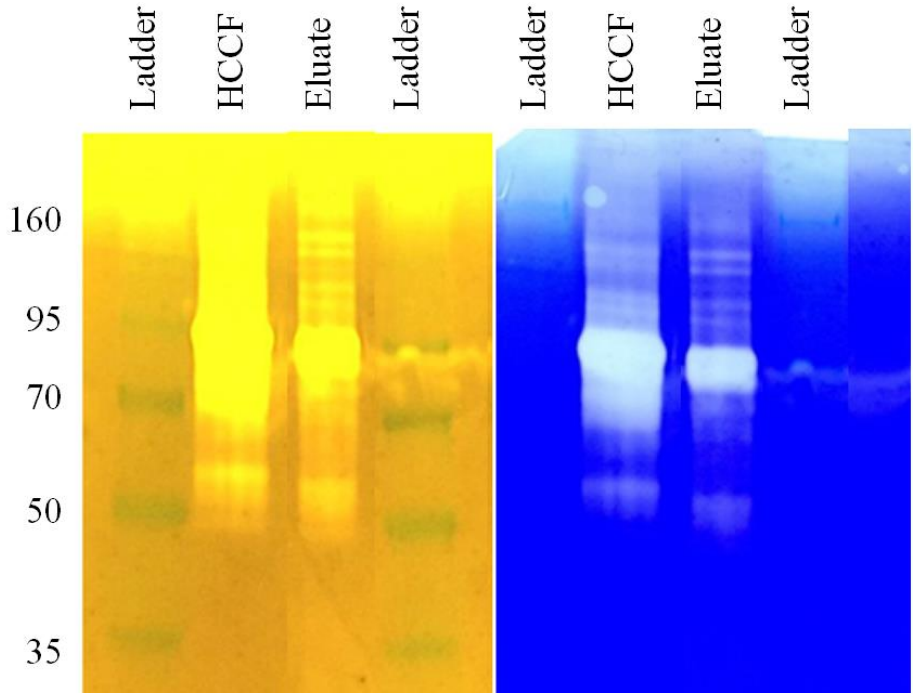


Figure 7-2 Comassie blue stained zymogram. Yellow tone added in image post processing to enhance ladder. Harvest (HCCF) from GAA Clone1 (300616) and IEX eluate were buffer exchanged into PBS and concentrated 4X, samples prepared as per manufacturer protocol. Incubation time in developing buffer was increased from over-night to 72h to increase sensitivity. Ladder EZ™-Unstained Protein Ladder Marker (Cat# PM001, Enzynomics, Korea). Data shows proteolytic activity in HCCF sample at wide range of molecular weights, with vast majority located at 80KD, later determined to be made up by Dipeptidyl peptidase 3, a metalloproteinase and Prolyl endopeptidase a serine protease. Similar pattern was observed to the HCCF harvest but with lower amounts.

7.2.3 IEX protease removal by comparing GAA Clonal cell line to originating Null

A further comparison of IEX protease removal was done by running HCCF and IEX eluate of the GAA producing clone selected and the originating Null cell line on a zymographic gel side by side. Both cultures were grown identically as previously discussed (section 6.2) in 50ml CD-CHO and harvested. Resulting gel was developed for 72h in pH 5.0 at 37^{O}C for increased sensitivity. Similarly to *Figure 7-2*, the gel in *Figure 7-3* shows qualitatively the amount of HCP with proteolytic activity that the capture anion exchange step is able to clear from HCCF samples.

Specifically, going from HCCF to IEX eluate while most proteases are cleared from the process stream, a significant amount of proteases at 80kDa coelutes. MS analysis putatively identified this band as dipeptidyl peptidase 3 and prolyl endopeptidase. The gel also shows how the protease profile of a Null and an overexpressing CHO cell line are broadly similar, as further demonstrated in section 7.4.3 below.

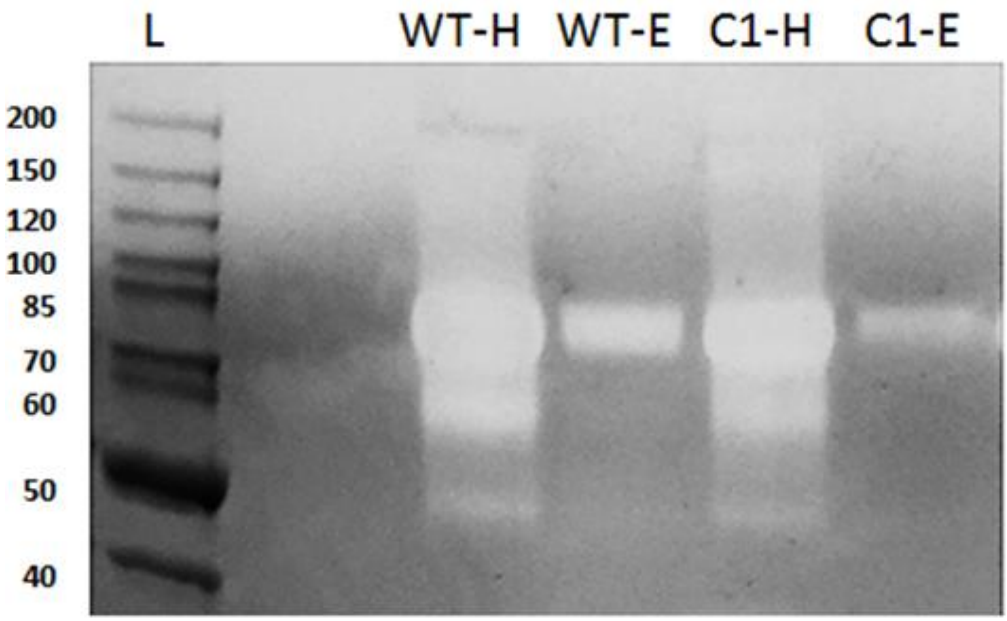


Figure 7-3 Comassie blue stained casein Zymogram. From left: Null HCCF (Null-H), Null IEX Eluate (Null-E), Cell line1 HCCF (C1-H), Cell line 1 IEX Eluate (C1-E). Protein samples are first separated according to molecular weight via gel electrophoresis, then renatured and incubated 72h in pH 5.0 at 37^{O}C in the gel containing casein substrate and finally stained and de-stained. Areas that show proteolytic activity appear as white bands over dark background. Data shows that IEX is able to clear out the majority of proteases from HCCF to eluate sample except for some co-eluting ones at around 80 KDa. These were later identified via MS as Dipeptidyl peptidase 3 and Prolyl endopeptidase. Furthermore, amount of protease in GAA cell line 1 eluate seems to be less than amount of proteases present in Null eluate. Note: GAA, not being a protease, is not visible in zymography.

7.3 CHO HCP screening via AlphaLISA

Host cell protein (HCP) release during production of therapeutic proteins is a critical quality attribute that must be monitored and reduced to acceptable levels during the production of a biotherapeutic protein (Bracewell et al., 2015). Being able to control HCP release during harvest and their removal via downstream processing is thus an important component of the manufacturing of such molecules and must be reported to the regulatory authorities. Typical purity targets include <100 ppm HCP, <10 ng/dose DNA and <5% product aggregates (Chon and Zarbis-Papastoitsis, 2011). Some HCPs are known to be co-purified with the recombinant protein target, whilst others, specifically proteases, are reported to lead to product degradation (Gao et al., 2011).

Historically, immunoassay based techniques such as ELISA have been used as high throughput methods to quantify HCP amounts during process development, manufacturing and in final product formulations. This approach has been important in the development of HCP assays and process development and product characterisation, however there are some inherent limitations with the use of ELISA. For example, the polyclonal antibodies used for detection are raised against the spectrum of HCPs found in the host cell (or in some cases mixed pools of cells), results are thus generally cumulative of all the HCPs present in a sample and cannot provide identification of single critical HCPs of interest, and only those HCPs which are immunogenic can be detected. These issues can be summarized as concerns as to whether all HCPs are being measured at all and with sufficient sensitivity (Bracewell et al., 2015, Zhu-Shimoni et al., 2014). For these reasons orthogonal analytical high throughput techniques are being developed based on peptide identity via mass spectrometry: the need to simultaneously quantify and identify HCPs in samples with greater sensitivity and precision, has resulted in the advancement of orthogonal

analytical techniques based on LC-MS/MS. High throughput monitoring of HCP via LC-MS/MS is becoming ever more common (Doneanu et al., 2012, Chiverton et al., 2016)

Nevertheless, immunoassays remain a valid and well established technique for HCP quantitation. Perkin Elmer AlphaLISA, already used (although unsuccessfully) for product quantitation (see section 5.4.1 for assay principle), was revisited for HCP quantitation. In this instance, the development work performed for product quantitation was not needed due to the fact that the HCP quantitation kits are commercially available through the manufacturer.

7.3.1 AlphaLISA standard curve

According to the manufacturer protocol the standard curve for this assay uses a Four Parameter Logistic Fit (4PL) with $1/y^2$ weighting. The applied weighting is used to offset heteroscedasticity by taking into account the change in variance which occurs with an increase in signal. Heteroscedasticity refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it. A four Parameter Logistic Fit (4PL) with $1/y^2$ weighting is represented by the equation

$$Y = d + \frac{a-d}{1+(\frac{x}{c})^b} \qquad \text{Equation 7-1}$$

After preparing the standards in triplicate as per section 3.4.2, the plate was inserted into the EnVision AlphaLISA reader, excitation set at 680 nm and signal emission recorded at 615 nm.

All standard readings were first corrected by the mean of the blank group measurements. The standard data points were then plotted (concentration vs. corrected measurement) and a Four Parameter Logistic Fit (4PL) with $1/y^2$ weighting was made through these points (source myassay.com)

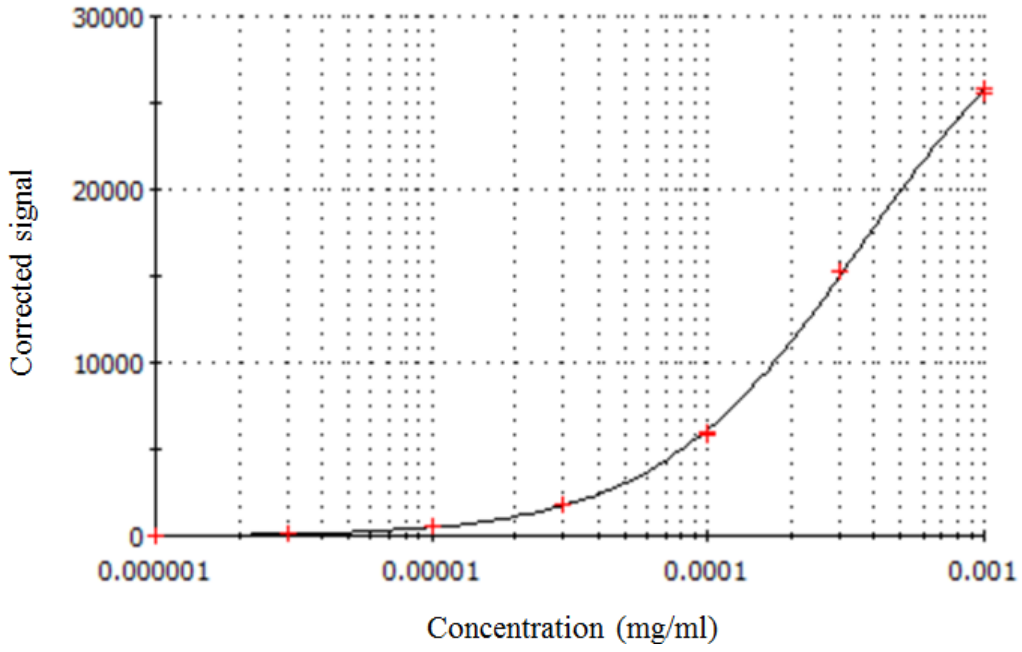


Figure 7-4 CHO HCP AlphaLISA assay standard curve. AlphaLISA CHO HCP 12 point standard dilution prepared in triplicate as per section 3.4.2 in materials and methods using CD-CHO (cell media) as diluent. 5 µl of each standard dilution was mixed with 20 µl Anti-CHO HCP acceptor beads + biotinylated Antibody mix in a 96 well ½ area plate and incubated 60 min at RT, then 25 µl of streptavidin coated donor beads were added to each well in subdued lighting. Plate sealed with black TopSeal and incubated at RT for a further 30 min. Plate read in an EnVision AlphaLISA reader (680/615 nm ex/em). Resulting signal was averaged and corrected subtracting blank reading and fitted to Eq 7.1 (Calculated parameters for this curve: a = -1.81693; b = 1.24996; c = 0.000286492; d = 15673; R^2 = 1

The CHO HCP analyte standard curve here presented showed acceptable intra replicate precision and broad dynamic range. Lower Detection Limit (LDL): 450 pg/mL; Lower Limit of Quantification (LLOQ): 1472 pg/mL (cat# AL301C Perkin Elmer).

7.3.2 HCP quantification in capture step via AlphaLISA

AlphaLISA was used to determine CHO HCP concentration in HCCF and IEX eluate fraction to determine purity. During this process we encountered difficulties in measuring accurately HCP in HCCF sample. Several plates containing HCCF and IEX eluate samples were run and

although HCP determination in IEX eluate was reliable, we couldn't get reproducible data for HCCF samples. We noticed that the quantification of HCP in HCCF was constantly giving inexplicably low results, often a fraction of what found in the respective purified sample. Surely HCP amount in HCCF should have been at least equal if not higher than amount found in the purified sample. We tried an array of sample diluents for HCCF samples from water, PBS, cell culture media. This seemed to have no effect on the result.

We then realised that this phenomenon had been described before in the literature and is known as 'the hook effect'. The hook effect is caused by excessively high concentrations of analyte in solution simultaneously saturating both capture and detector antibodies. The high-dose hook effect occurs mostly (but not exclusively) in one-step immunometric (sandwich) assays, giving a decrease in signal at very high concentration of analyte (Fernando and Wilson, 1992). In these conditions, reactions can go into antigen excess and result in falsely decreased results and potential misdiagnosis. In one step two-site immunoassays where capture and detection antibody are added simultaneously, free analyte and analyte bound to the labelled antibody compete for the limited number of antibody-binding sites of the detector and in the presence of very high analyte concentration will decrease instead of increase label bound to the solid phase, decreasing the resulting signal (Cole et al., 2001). As a solution to this problem the authors suggest decreasing amount of analyte in solution.

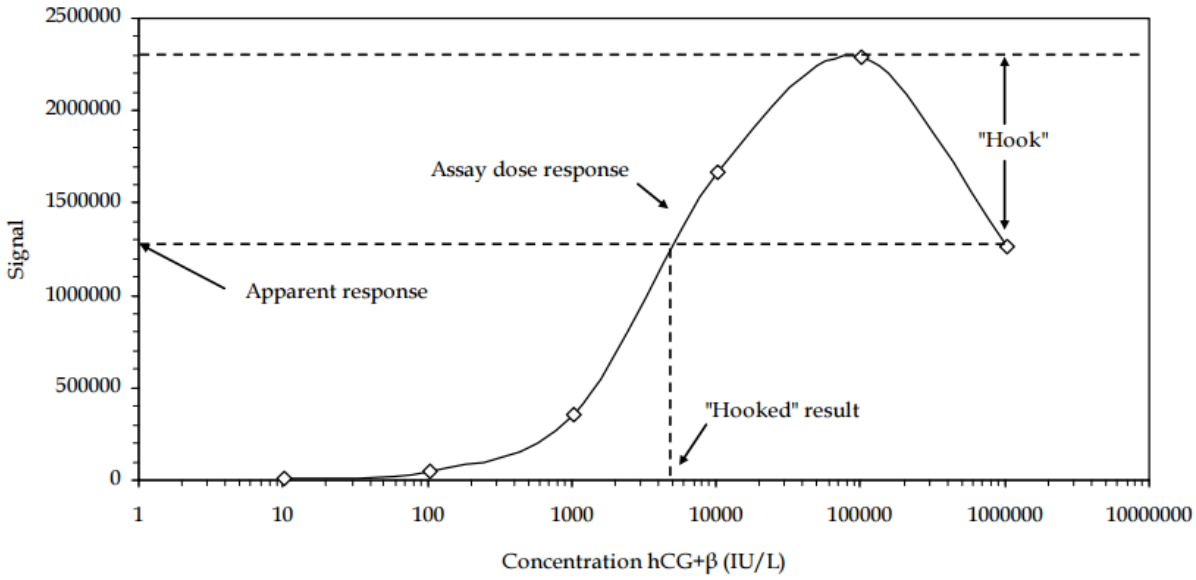


Figure 7-5. With permission from (InTech). High-dose hook effect in Elecsys hCG+β assay - an excessive amount of hCG overwhelms the binding capacity of the capture antibody. This results in an inappropriately low signal that causes an erroneous "hooked" result (6200 IU/L) for a patient with an excessively elevated serum hCG+ β concentration of 1 030 000 IU/L.

To overcome the hook effect a HCCF sample from Clone 1 culture was serially diluted in the 1:100 to 1:256000 range in CD-CHO media (12 dilutions in triplicate) and mixed with acceptor beads, biotinylated antibody and donor beads in the same exact way it was done for standard curve preparation explained above (section 7.3.1). The resulting signal was then averaged among the triplicates, corrected by subtracting the blank and converted to concentrations via the standard curve generated, taking into account the dilution factors. A recovery dilution factor between 80-120% was used to pick results close to the mean value. As seen in (Table 7-2) below only dilutions between 1:500 and 1:32000 were in the acceptable recovery range. This explains why previous attempts at determining HCP in HCCF samples failed, as sample dilutions were usually in the range 1:1 to 1:64, a range here shown to be outside acceptable recovery. This experiment allowed for the determination of HCP concentration in HCCF samples as

0.093mg/ml. The usable dilution range for future experiments was arbitrarily set between 1:1000 and 1:8000 dilutions.

Samples: HCCF (30/06)	Dilution factor	Repl1	Repl2	Repl3	Average signal	Corrected signal	Conc (pg/ml)	Conc (mg/ml)	Concentration x dil fact	% recovery (80-120%)	[HCP] conc (mg/ml)
1:100	100	12376	12753	11645	12258	12016	1.78E+06	0.0017796	0.1779580	174	
1:250	250	8544	8673	8226	8481	8239	575408	0.0005754	0.1438520	141	
1:500	500	4733	4919	4802	4818	4576	223459	0.0002235	0.1117295	109	
1:1000	1000	2582	2765	2500	2616	2374	102057	0.0001021	0.1020570	100	
1:2000	2000	1285	1312	1242	1280	1038	43400.4	0.0000434	0.0868008	85	
1:4000	4000	741	721	684	715	473	20441.4	0.0000204	0.0817656	80	0.093
1:8000	8000	547	487	504	513	271	12207.5	0.0000122	0.0976600	96	
1:16000	16000	310	376	358	348	106	5341.81	0.0000053	0.0854690	84	
1:32000	32000	308	287	261	285	43	2605.68	0.0000026	0.0833818	82	
1:64000	64000	264	266	316	282	40	2450.9	0.0000025	0.1568576	154	
1:128000	128000	243	229	234	235	-7	408.323	0.0000004	0.0522652	51	
1:256000	256000	294	255	261	270	28	1910.47	0.0000019	0.4890803	479	

Table 7-2. CHO HCP AlphaLISA HCCF serial dilution experiment. HCCF dilutions 1:100 to 1:256000 were prepared in triplicate in CD-CHO media and mixed in acceptor beads, biotinylated antibody and donor beads as explained above. Standard curve run in same plate was used to determine HCP concentration in samples. Only concentrations that satisfied the 80-120% recovery range (dilutions between 1:500 and 1:32000 in yellow) were used to calculate average HCP concentration. Total HCP concentration in HCCF sample prepared was measured at 0.093 mg/ml. The standard curve shown in section 7.3.1 is a representative one shown as an example. A unique standard curve is run with each plate and set of samples. The specific equation parameters are used with each set of samples for concentration calculations.

Having established HCP concentration in GAA Clone 1 HCCF and its appropriate dilution range for AlphaLISA assay, a new purification was set up on the AKTA as previously shown (section 6.3) and 2 ml fractions collected throughout the whole run (Figure 7-6). Fractions from each stage were pooled together and CHO HCP AlphaLISA was performed. A new plate 96 well ½ area plate was set up with standard curve dilutions and the 4 samples from one GAA clone 1 IEX purification: (H) HCCF load 7ml, (FT) Flowthrough (16 x 2ml fractions), (E) Eluate (7 x 2ml fractions), (St) Strip (1 x 2ml fraction). This data was used to measure IEX yield reported in section 6.3.

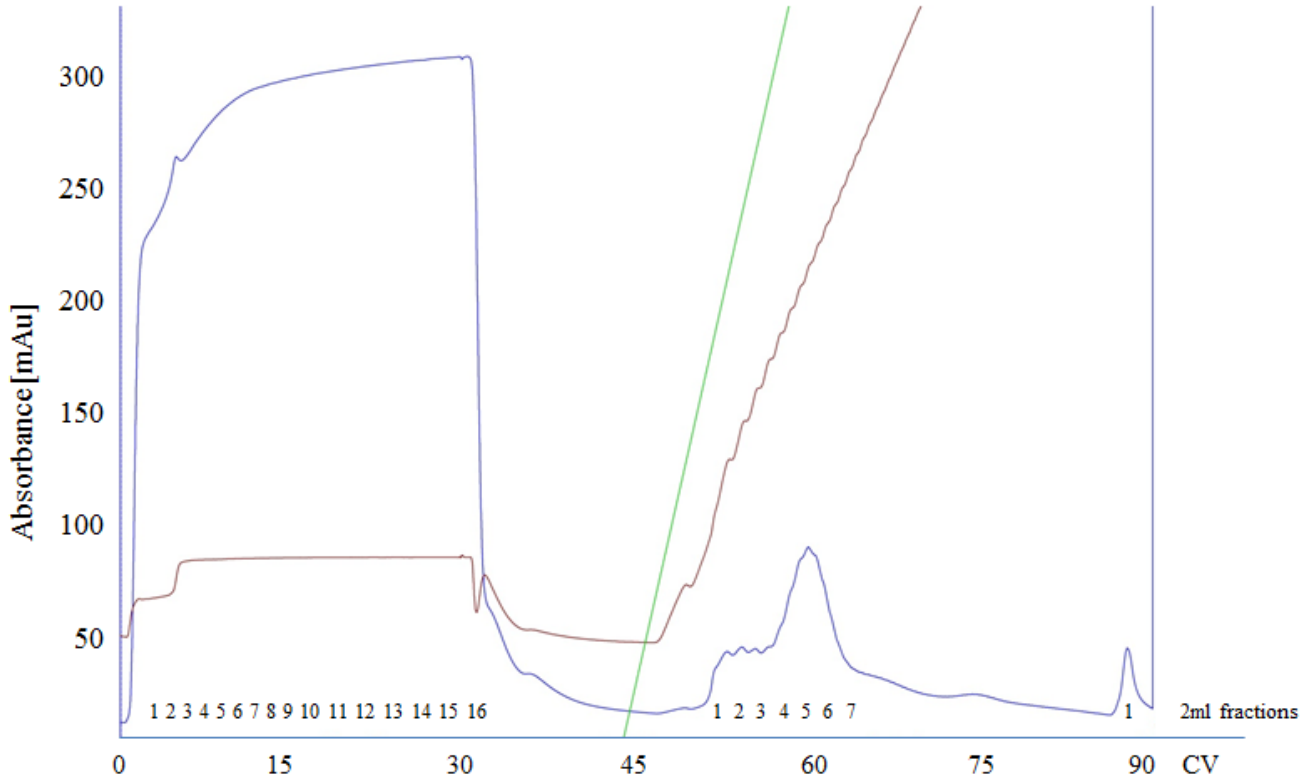


Figure 7-6. GAA Clone 1 HCCF IEX purification using AKTA avant. 2ml fractions collected as shown and used for CHO HCP quantitation via AlphaLISA assay. Separation details previously described in section 6.3.1.

	Sample	Sample Volume (ml)	HCP concentration (mg/ml)	HCP amount (mg)	Tot HCP (mg)	GAA concentration (mg/ml)
IN	HCCF	7	0.082	0.574	0.574	0.18
OUT	FT	32	0.005	0.160 (28%)	0.564 (98%)	0
	E	14	0.022	0.308 (54%)		0.2
	St	2	0.048	0.096 (17%)		0

Table 7-3. IEX HCP mass balance via AlphaLISA. HCP in samples was measured as described in table 7.2. GAA concentration determined in section 6.3.2

7.4 Protease identification via tandem mass spectrometry

In this section we present results of three tandem MS techniques aimed at identification and reliable quantitation of HCP present in GAA HCCF and capture eluate fraction: (1) in-gel tryptic digestion MALDI MS/MS, (2) MALDI TOF LC-MS/MS and (3) label free high resolution nano LC-MSMS.

The first MS technique was in-gel tryptic digestion MALDI MS/MS. This technique involved separation of proteins based on molecular weight on a SDS-PAGE gel, excision of visible bands, digestion of each band with trypsin, and spotting sample on a MALDI plate for MS/MS analysis. This methodology is labour intensive and inherently error prone for a few reasons: first of all, not all bands were found to contain only one protein but rather a complex mixture of proteins of similar molecular weights, this was especially true in an impure HCCF sample; also, low concentration HCPs could easily go undetected due to bands being too faint to see and excise; lastly, as this technique involved many steps, human contamination (mostly by keratin) became a serious issue.

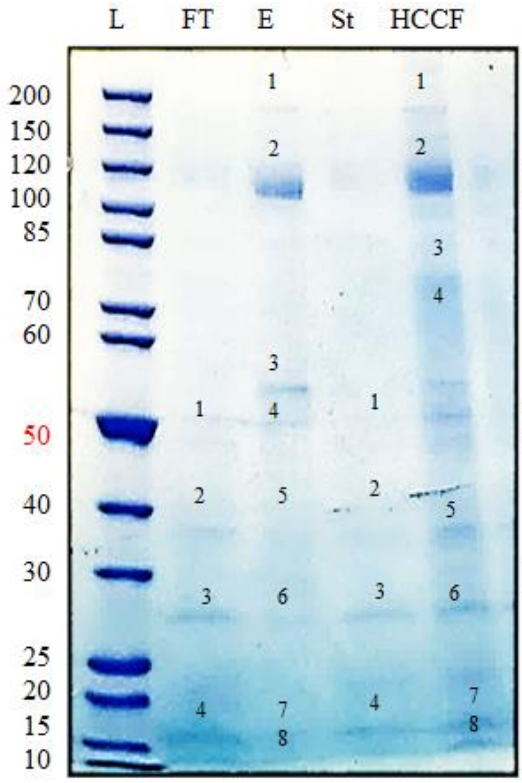
We then tried running the same samples on LC/MSMS using MALDI attached to fraction spotter. This technique relieved the analyst of the manually intensive gel digestion sample preparation. The benefit of this technique over the previous one was that the peptides digested by trypsin were separated by LC rather than the electrophoretic gel method. This resulted in a more reproducible and less labour intensive task. Elution fractions were spotted automatically onto a MALDI plate and run via tandem mass spectrometry. Quantification of species found was achieved via emPAI (Exponentially Modified Protein Abundance Index), the index used for approximate label free quantitation of proteins (Ishihama et al., 2005). Although peptide separation was sufficient resulting in identification of more of the HCPs present, quantification was still inherently approximate as based on protein coverage by the peptide matches in a

database search result (emPAI), rather than true positive quantitation via reference standard comparison.

Lastly, the same samples were tested via a high resolution label free quantitative nano LC-MSMS. This methodology still relies on robust LC peptide separation, however the relative quantification calculation of each protein identified through this type of analysis is achieved via direct comparison with a BSA reference standard spiked in the samples before sample injection. In order for quantification to be absolute there should be a reference standard per identified ion or protein species, a clearly unfeasible solution.

7.4.1 HCP identification via In-gel tryptic digestion and MALDI MS/MS

The aim of this experiment was to try to identify HCP present in HCCF and their distribution in IEX fractions flowthrough, eluate and strip, using the same technique utilized for product identification (section 3.2.4). This was important in order to determine which class of co-eluting proteases were responsible for product degradation. A GAA Clone 1 cell culture was expanded and harvested as previously shown (section 6.2). HCCF material was loaded onto CaptoQ anion exchange resin on AKTA Avant and separation was achieved as shown before (section 6.3). The four samples (HCCF, FT, E, St) were run on a NuPAGE® Novex® Bis-Tris Gels SDS-PAGE gel (Figure 7-7), excised with a scalpel, digested with sequencing grade trypsin and run on UltrafleXtreme MALDI-TOF instrument (Bruker, Coventry, UK) for MS/MS analysis. Generated peptide masses were interrogated using the Mascot algorithm (matrix-science.com) to search all taxonomies in the SwissProt database.



FT	2	DEF4_HUMAN	Neutrophil defensin
		FBLN3_HUMAN	EGF-containing fibulin-like extracellular matrix protein 1
	3	LACG_STAA1	6-phospho-beta-galactosidase
	4	MOAD_ECOLI	Molybdopterin synthase sulfur carrier subunit

E	1	K2C1_HUMAN	Keratin
	2	LYAG_HUMAN	Lysosomal alpha-glucosidase OS=Homo sapiens
	3	Mus musculus	Keratin
	4	LYAG_HUMAN	Lysosomal alpha-glucosidase
	5	K2C1_HUMAN	Keratin
	6	CATZ_HUMAN	Cathepsin Z OS=Homo sapiens
	7	LACG_STAA1	6-phospho-beta-galactosidase

St	1	K2C1_HUMAN	Keratin
	2	K2C1_HUMAN	Keratin
	3	K22E_HUMAN	Keratin
	4	-	-

HCCF	1	K1H1_HUMAN	Keratin
	2	LYAG_HUMAN	Lysosomal alpha-glucosidase
	4	-	
	5	PCOC1_MOUSE	Procollagen C-endopeptidase enhancer 1
	6	K2C1_HUMAN	Keratin
	7	LACG_STAA1	6-phospho-beta-galactosidase

Figure 7-7. MALDI-TOF-TOF on excised in-gel trypsin digested GAA fermentation ion exchange purified fractions. Comassie blue stained SDS-PAGE gel containing load (HCCF) and IEX separated fractions (FT, E, St) of a Clone 1 GAA cell culture harvest. Gel sharpness and contrast were enhanced to facilitate visualisation of protein bands. The gel fragments were processed via the protocol included in materials and methods (3.2.4) and digested with modified sequencing grade trypsin. 0.5 µl of digested sample followed by 1 µl of Super-DHB matrix was applied onto MALDI plate. UltrafleXtreme MALDI-TOF instrument (Bruker, Coventry, UK) in positive ion reflector mode and 50% laser power and MS/MS was conducted on the ten most intense peaks for each target spot. Generated peptide masses with an ion score exceeding the threshold set for $p < 0.05$ were interrogated using the Mascot algorithm (matrix-science.com) to search all taxonomies in the SwissProt database. Fixed modifications: carbidomethyl (C); variable modifications: oxidation (M); mass values: monoisotopic; protein mass: unrestricted; peptide mass tolerance: +/− 50 ppm; fragment mass tolerance +/− 0.5 Da. Instrument MALDI-TOF-TOF.

Results reconfirmed presence of GAA (Lysosomal alpha-glucosidase) at around 110KDa in HCCF and eluate fraction. Interestingly human Cathepsin Z, the serine protease suspected of

degrading GAA in the process stream, was also identified in eluate fraction at around 28KD. Despite identifying a few other HCPs (i.e. 6 phospho beta galactosidase in FT and E fraction), this methodology was revealed to be of too low a resolution and prone to keratin contamination. Also, the overall amount of HCP identified (especially in HCCF) was unexpectedly low. This was due to the fact that protein bands on gel were hard to see and therefore excise and digest. In order to improve data quality digested peptides should have been separated via reverse phase HPLC, whilst reducing analyst interaction (thus reducing contamination), as discussed in the next section.

7.4.2 HCP identification via MALDI TOF LC-MS/MS and quantitation via emPAI

The aim of this experiment was to identify and quantify HCPs present in GAA harvest HCCF and IEX eluate fractions. The in-gel digested MSMS analysis from the previous section showed that a better peptide separation before MSMS analysis was necessary in order to improve data quality. Also, quantification was only possible by based on band intensity. The purpose of adding the LC function to MSMS analysis was to allow the digested peptides to be separated under more robust conditions than the gel electrophoresis. A GAA harvest HCCF and IEX eluate fraction samples were digested with sequencing grade trypsin in the presence of DTT and IAM and applied to a C18 Acclaim PepMap100 column for on-line reverse phase HPLC. Eluates from each sample were divided in 120 fractions and spotted onto a MALDI plate for MS/MS analysis. A Mascot generated emPAI index was used for relative quantitation of species identified as shown in Figure 7-8.

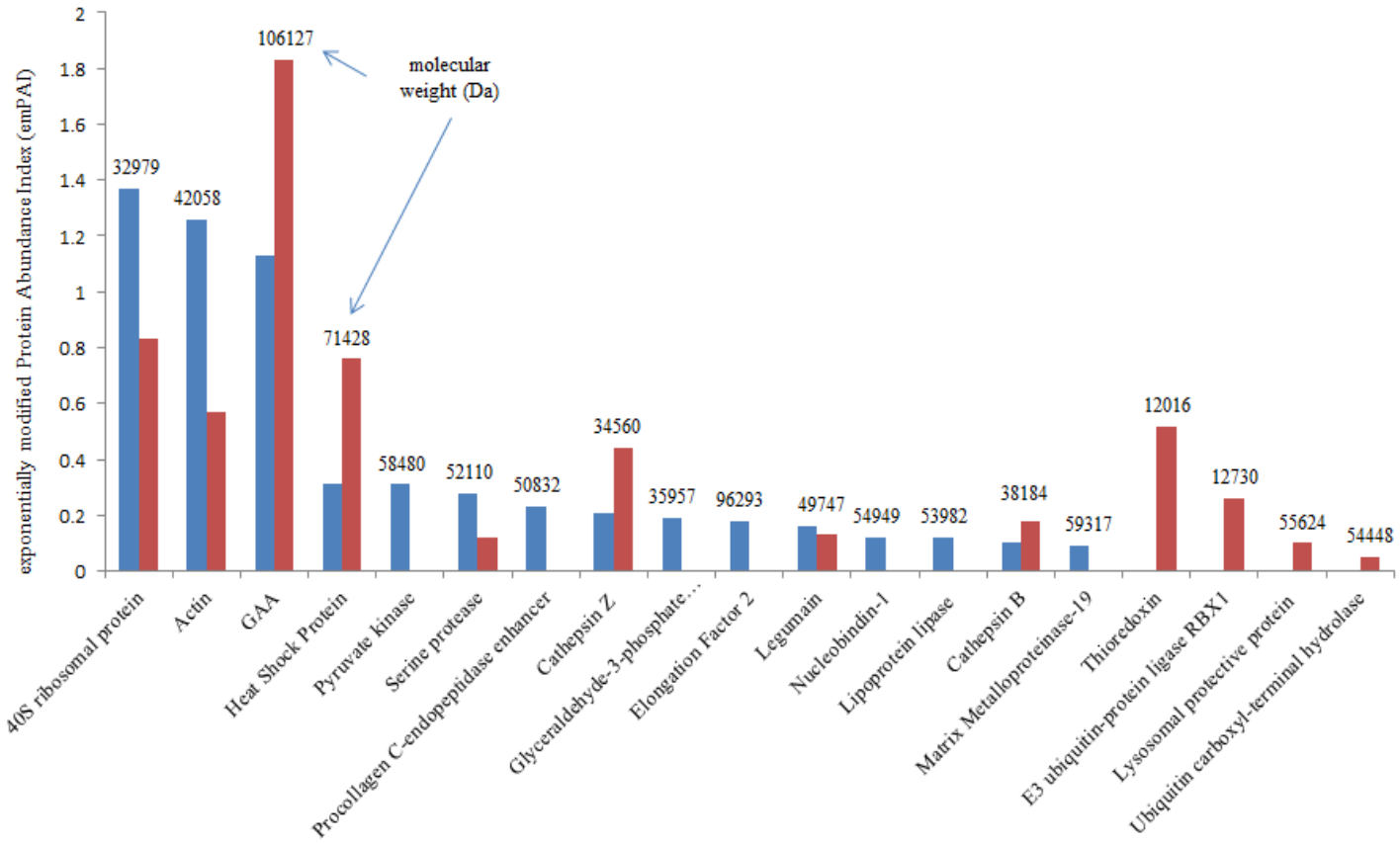


Figure 7-8. MALDI TOF LC-MS/MS HCP identification and quantitation via emPAI. GAA harvest HCCF (blue) and IEX eluate fraction (red) samples were digested with sequencing grade trypsin in presence of IAM and DTT, run on LC (NanoLC Ultimate3000) C18 Acclaim PepMap100 column (C18, 3 µm, 100A) (solvent A = 0.05% TFA, solvent B = 0.05% TFA 90% acetonitrile, 40 min cycle, flow rate of 300 nl/min) attached to fraction collector and spotted onto a MALDI plate for MSMS analysis on an UltrafleXtreme MALDI-TOF instrument in positive ion reflector mode and 50% laser power, MS-MS was conducted on the ten most intense peaks for each target spot. Generated peptide masses with an ion score exceeding the threshold set for $p < 0.05$ were interrogated using the Mascot algorithm (matrix-science.com) to search all taxonomies in the SwissProt database. Fixed modifications: carbidomethyl (C); variable modifications: oxidation (M); mass values: monoisotopic; protein mass: unrestricted; peptide mass tolerance: +/− 100 ppm; fragment mass tolerance +/− 0.5 Da. Instrument MALDI-TOF-TOF. In total 335 and 127 HCPs were respectively identified for HCCF and IEX eluate, here we only reported those amongst the better known or that showed proteolytic activity. Species were ordered by ascending relative amount in the HCCF sample.

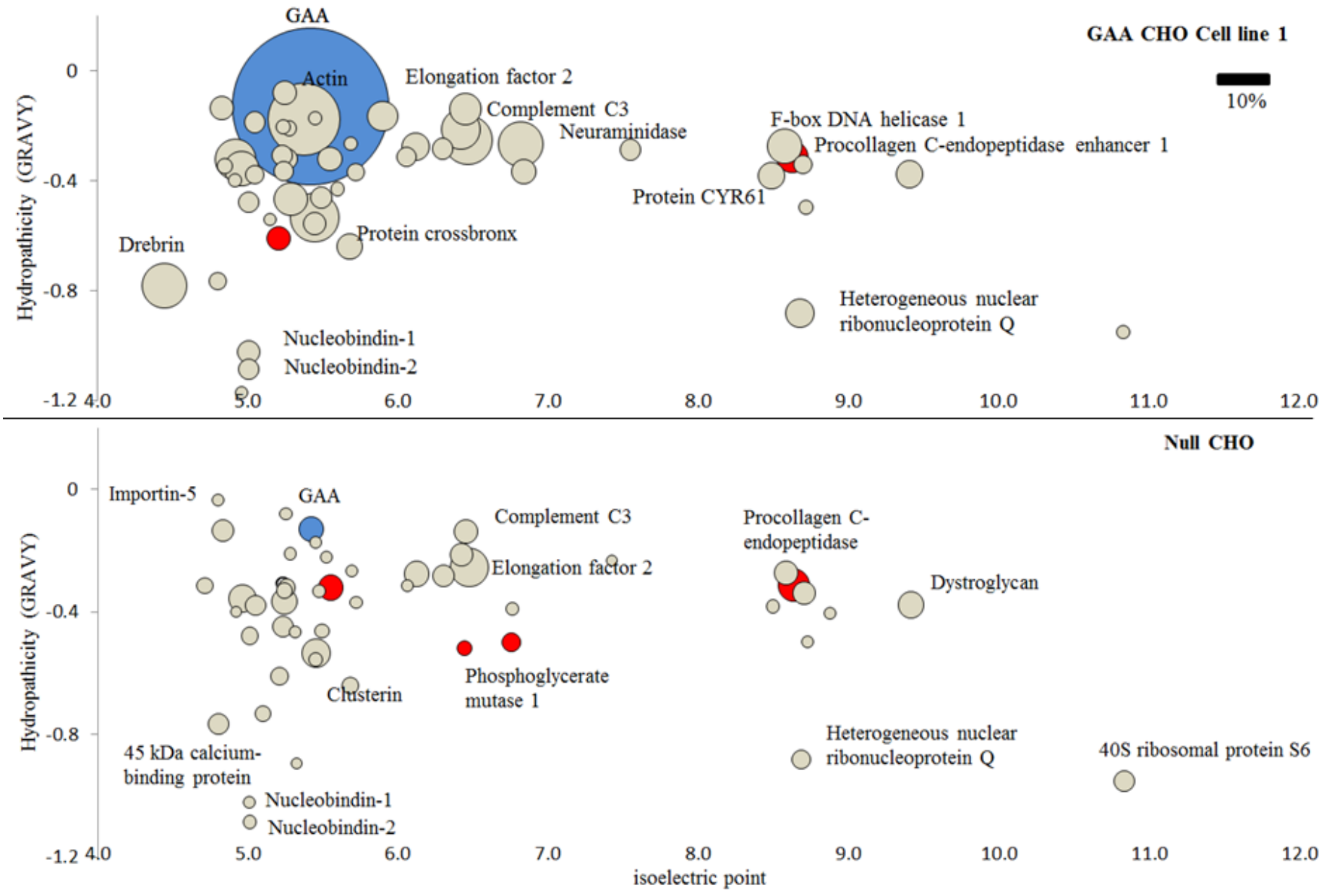
As shown in Figure 7-8 the emPAI index was used for quantification purposes of HCPs found in pre and post IEX purification of a GAA producing CHO culture harvest sample. emPAI (exponentially modified PAI) (Ishihama et al., 2005) is an identification-based algorithm for protein-abundance estimation, which is based on the number of observed peptides per protein normalized by the theoretical number of peptides, the so-called protein abundance index (PAI) (Rappsilber et al., 2002) and is generally used to get an idea of relative amounts of different species in a protein mixture. According to literature, although the emPAI is not as accurate as quantification using synthesized peptide standards, it is quite useful for obtaining a broad overview of proteome profiles (Shinoda et al., 2009).

According to the data, IEX was able to reduce amounts of most HCPs present in feed, while retaining the target molecule (GAA). Cathepsin B and Z were also retained during purification and coeluted with the product. Actin, 40S ribosomal protein, heat shock protein 71KDa, Legumain, nucleobinding-1, elongation factor 2 were found in moderate quantity in post capture step eluate in accordance with label free quantitative LC-MS/MS data reported in the next section.

7.4.3 HCP identification and quantification via high resolution nano LC-MS/MS

Scientific principle and novelty: the mass spectrometer used, a high resolution Waters SYNAPT G2-Si, was part of the Biomolecular Science Facility at University of Kent. The main difference between the Synapt G2-Si and other high resolution mass spectrometry systems is that it utilises the extra dimension of ion mobility separation (dt) on top of the common chromatography (rt) and mass separation (m/z) technologies. This maximizes selectivity and confidence in the results. By combining step-wave with quant-off and tri-wave technologies this system provides a step change in analytical performance for qualitative and quantitative applications.

In this experiment we compared HCP content found in capture step eluate fraction of three different CHO cell lines. The GAA producer, a Null and a mAb producing were expanded and purified under the same conditions and IEX eluate fractions were screened using high resolution nanoLC tandem mass spectrometry. The resulting data is presented in Figure 7-9 and Table 7-4. The data is shown in a bubble graph generated by plotting the 50 most abundant species found in each cell line. These were ordered by calculated pI (ExPASy Compute pI/Mw tool, SIB Swiss Institute of Bioinformatics) on the X axis and by hydrophobicity index GRAVY (ExPASy ProtoParam tool, SIB Swiss Institute of Bioinformatics) on the Y axis. Bubble size (diameter) represents percent of a given protein out of the total mass and was calculated by multiplying normalized abundance (averaged among triplicate LC injections) by the protein molecular weight. Quantification calculation of each species found through this type of analysis was done via direct comparison with the BSA reference standard spiked in the samples right before sample injection. As previously mentioned, in order for quantification to be truly representative there should be a reference standard per each identified species, a clearly unfeasible solution. For this reason, we are confident of HCP ID being correct, even though absolute amounts of each specie could perhaps be not fully accurate. This was however a clearly better solution than relative quantification via emPAI index in MALDI TOF TOF in the previous section.


GAA CHO Cell line 1
10%
GAA
Actin
Elongation factor 2
Complement C3
Neuraminidase
F-box DNA helicase 1
Procollagen C-endopeptidase enhancer 1
Protein CYR61
Protein crossbronx
Drebrin
Heterogeneous nuclear ribonucleoprotein Q
Nucleobindin-1
Nucleobindin-2
Hydropathicity (GRAVY)
0
-0.4
-0.8
-1.2
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
Null CHO
Importin-5
GAA
Complement C3
Elongation factor 2
Procollagen C-endopeptidase
Dystroglycan
Phosphoglycerate mutase 1
Clusterin
45 kDa calcium-binding protein
Heterogeneous nuclear ribonucleoprotein Q
40S ribosomal protein S6
Nucleobindin-1
Nucleobindin-2
Hydropathicity (GRAVY)
isoelectric point

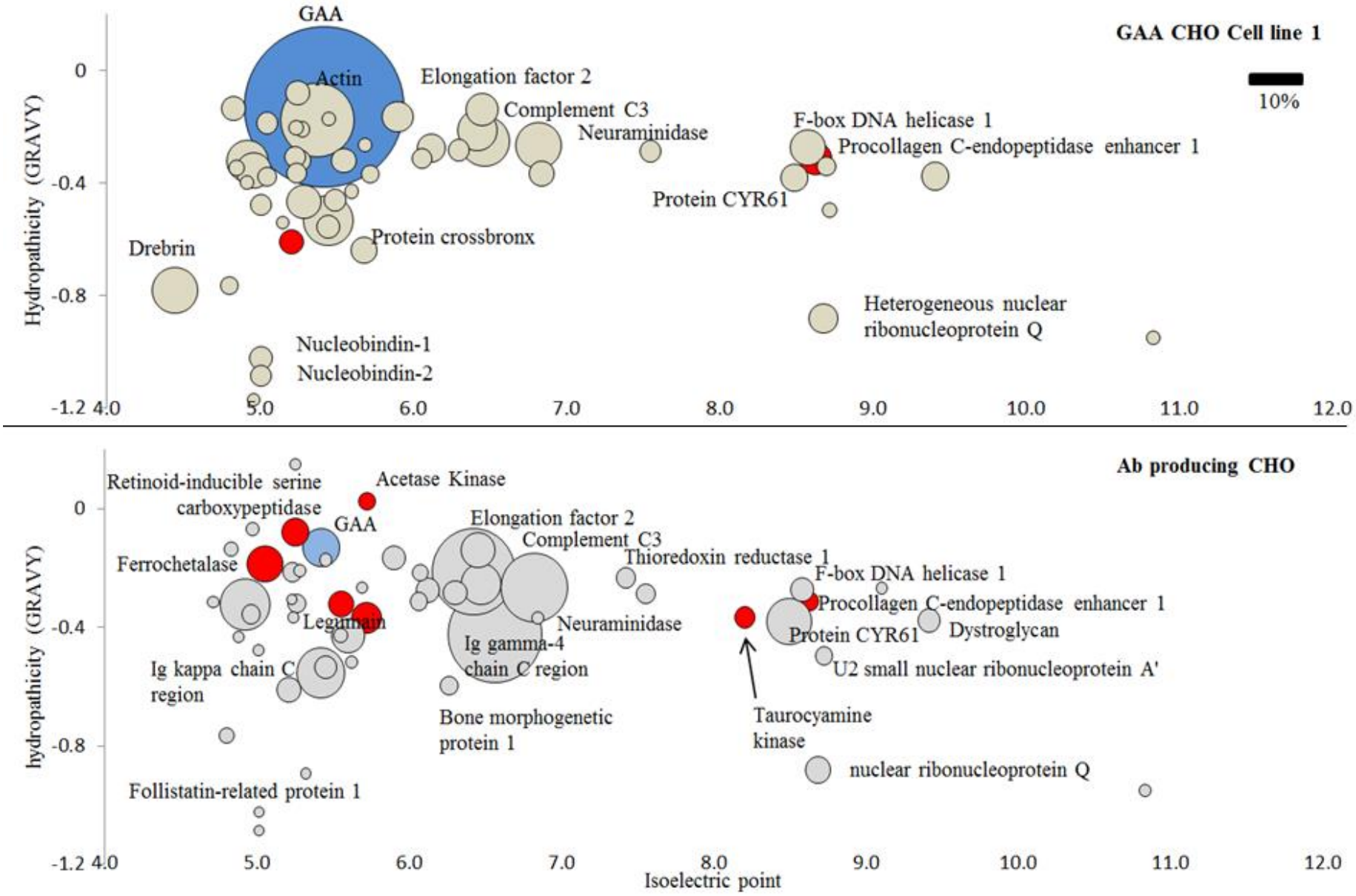

GAA CHO Cell line 1
10%
GAA
Actin
Elongation factor 2
Complement C3
Neuraminidase
F-box DNA helicase 1
Procollagen C-endopeptidase enhancer 1
Protein CYR61
Protein crossbronx
Drebrin
Heterogeneous nuclear ribonucleoprotein Q
Nucleobindin-1
Nucleobindin-2
Hydropathicity (GRAVY)
0
-0.4
-0.8
-1.2
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
Ab producing CHO
Retinoid-inducible serine carboxypeptidase
Acetase Kinase
GAA
Elongation factor 2
Complement C3
Thioredoxin reductase 1
F-box DNA helicase 1
Procollagen C-endopeptidase enhancer 1
Ferrochetalase
Legumain
Neuraminidase
Protein CYR61
Dystroglycan
Ig gamma-4 chain C region
Ig kappa chain C region
U2 small nuclear ribonucleoprotein A'
Bone morphogenetic protein 1
Taurocyamine kinase
nuclear ribonucleoprotein Q
Follistatin-related protein 1
hydropathicity (GRAVY)
0
-0.4
-0.8
-1.2
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
Isoelectric point

Figure 7-9 High resolution LC-MS/MS of IEX eluate. In blue target molecule GAA, in red HCPs with proteolytic activity, in grey all other HCPs. Data shows identification and quantification of HCP found in capture step (IEX) eluate fraction across the three cell lines: GAA cell line1 vs Null (top) and GAA cell line1 and Ab producing (bottom). Out of a total of 233 HCPs identified in the samples, 36 showed proteolytic activity. The first 50 more abundant species are reported per each cell line and ordered by calculated pI (ExPASy Compute pI/Mw tool, SIB Swiss Institute of Bioinformatics) on the X axis and by hydrophobicity index GRAVY (ExPASy ProtoParam tool, SIB Swiss Institute of Bioinformatics) on the Y axis. Bubble size represents percent of a given protein out of the total mass and was calculated by multiplying normalized abundance (averaged among triplicate LC injections) by the protein molecular weight. Post IEX column eluate fractions were concentrated and digested with sequencing grade modified trypsin, 50fmol of BSA was spiked in each digested sample before injection in LC to allow for label free quantitation. Separation via HSS T3 Acquity column (Waters, USA) 75 µm internal diameter x 15 cm (1.8 µm, 100A) on-line reverse phase UPLC (Acquity M-Class, Waters, USA). Elution by linear gradient from 3 to 40% B over 40 mins; (buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile), flow rate 300 nl/min. Separated peptides were injected via nanospray source to a Synapt G2-Si (Waters, USA), data collected in HDMSe mode. Data analysed using Progenesis QIP and searched against SwissProt using MSe Search and a false detection rate of 4%. Complete list of HCP identified in all samples is included in supporting data LC-MS/MS HCP.

Name (µg/ml)	GAA	null	Ab	Name (µg/ml)	GAA	null	Ab	Name (µg/ml)	GAA	null	Ab
Clusterin	3.5	7.6	0.2	Dipeptidyl peptidase 3	0.7	2.6	0.1	Bone morphogenetic protein 1	0.2	0.9	0.1
Complement C3	3.4	13.5	0.6	Nucleobindin-2	0.6	1.6	0.0	Annexin A5	0.2	0.3	0.0
Neuraminidase	2.8	0.1	1.6	Cathepsin B	0.6	1.8	0.0	Rho GDP-dissociation inhibitor 1	0.2	2.3	0.0
Elongation factor 2	2.2	4.8	2.5	78 kDa glucose-regulated protein	0.6	2.7	0.0	Receptor-type tyrosine-protein phosphatase S	0.2	1.5	0.0
F-box DNA helicase 1	1.7	4.9	0.2	Nidogen-1	0.5	5.7	0.0	Nuclear migration protein nudC	0.2	0.3	0.0
EGF-containing fibulin-like extracellular matrix protein 1	1.7	7.0	0.1	cAMP-dependent protein kinase catalytic subunit ga	0.5	4.7	0.0	Heat shock cognate 71 kDa protein	0.1	3.9	0.0
Pigment epithelium-derived factor	1.4	5.1	0.4	Legumain	0.4	1.4	0.3	30S ribosomal protein S6	0.1	0.2	0.0
Procollagen C-endopeptidase enhancer 1	1.4	9.4	0.1	45 kDa calcium-binding protein	0.4	3.9	0.1	Heat shock 70 kDa protein	0.1	0.9	0.0
Heterogeneous nuclear ribonucleoprotein Q	1.2	3.2	0.2	Actin-17	0.3	1.5	0.1	Small ubiquitin-related modifier 2	0.1	1.1	0.0
Dystroglycan	1.0	6.1	0.2	U2 small nuclear ribonucleoprotein A'	0.3	1.2	0.1	Chaperone protein dnaK2	0.1	2.4	0.0
Protein CYR61	1.0	1.6	0.7	Actin-10	0.3	0.4	0.0	Vitamin K-dependent protein S	0.1	0.6	0.0
Follistatin-related protein 1	0.8	2.9	0.2	Spermidine/putrescine import ATP-binding protein	0.3	1.0	0.0	POTE ankyrin domain family member E	0.1	0.0	0.0
Retinoid-inducible serine carboxypeptidase	0.8	1.4	0.2	Putative phospholipase B-like 2	0.2	1.2	0.0	Transmembrane protein 132A	0.1	0.8	0.0
Lysosomal protective protein	0.8	5.8	0.2	40S ribosomal protein S6	0.2	3.7	0.1	Matrix metalloproteinase-19	0.1	1.3	0.0
Importin-5	0.8	4.3	0.1	Actin, gamma	0.2	1.4	0.1	Protein disulfide-isomerase	0.1	1.0	0.0
Nucleobindin-1	0.7	1.3	0.0	Collagen alpha-1(VI) chain	0.2	0.9	0.0	Cathepsin Z	0.1	0.4	0.0

Name (µg/ml)	GAA	null	Ab
Actin	7.7	0.0	0.0
Drebrin	3.0	0.4	0.0
Probable dual-specificity RNA methyltransferase RlmN	0.9	0.0	0.3
EGF-containing fibulin-like extracellular matrix protein 1	0.4	0.0	0.1

Table 7-4 Overlapping and unique HCPs. List of the 50 most abundant overlapping HCP common to all three cell lines (top) and four unique to GAA HCP (bottom). Values (calculated µg/ml) are ordered by abundance in GAA sample. HCP are considered 'overlapping' when present in at least two out of three cell lines. Concentration (µg/ml) is calculated by multiplying normalized abundance value by molecular weight divided by sample volume (10µl). Normalized abundance value based upon BSA standard.

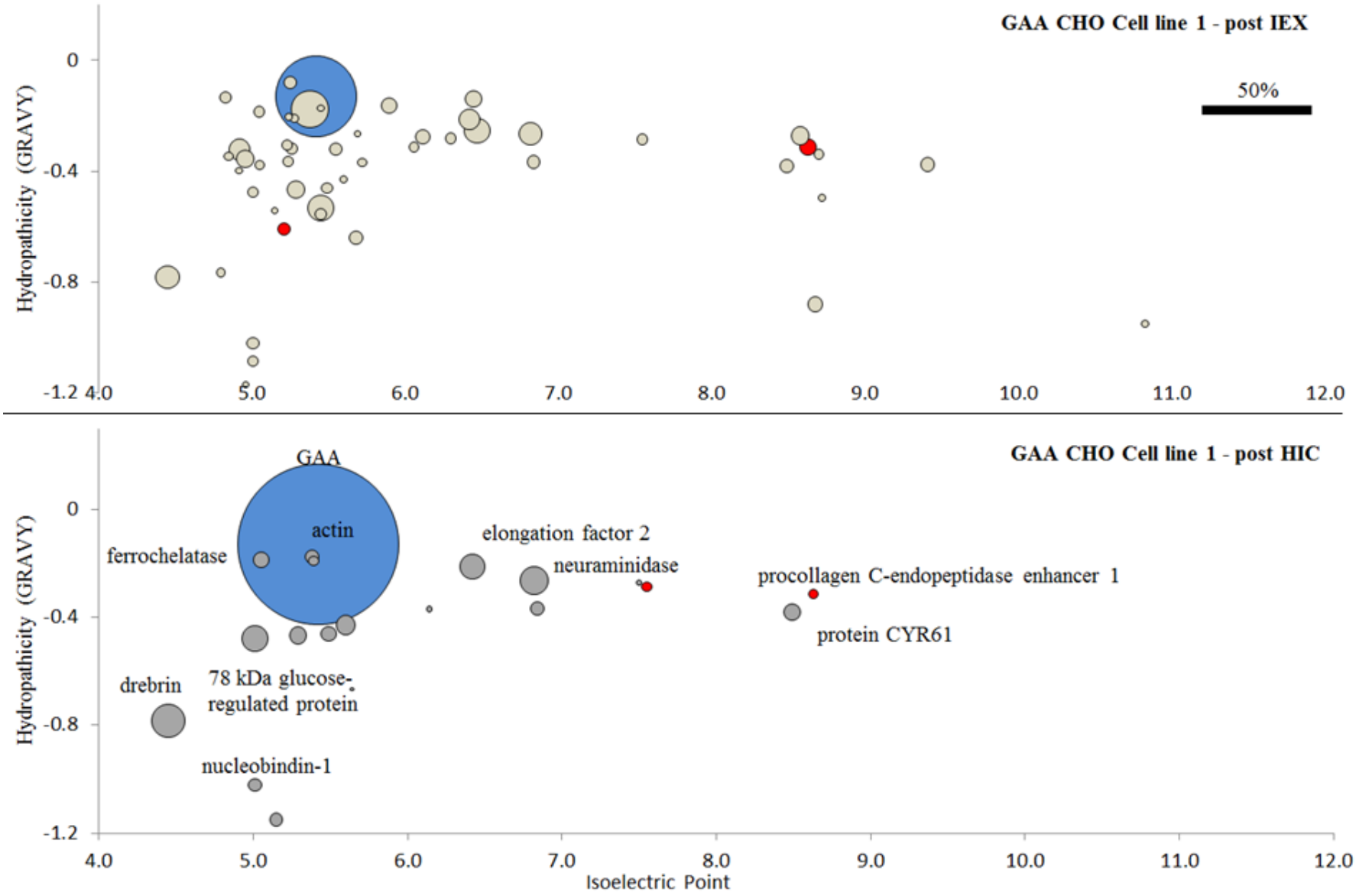

GAA CHO Cell line 1 - post IEX
50%
Hydropathicity (GRAVY)
0
-0.4
-0.8
-1.2
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
GAA CHO Cell line 1 - post HIC
GAA
actin
ferrochelatase
elongation factor 2
neuraminidase
procollagen C-endopeptidase enhancer 1
protein CYR61
drebrin
78 kDa glucose-regulated protein
nucleobindin-1
Isoelectric Point

Figure 7-10 High resolution LC-MS/MS of IEX vs HIC eluate. In blue target molecule GAA, in red HCPs with proteolytic activity, in grey all other HCPs. Data shows identification and quantification of HCP found in intermediate step (HIC) eluate fraction of the GAA producing cell line. Only 20 species had a non-negligible normalized abundance reading (above 10 counts) and were ordered by calculated pI (ExPASy Compute pI/Mw tool, SIB Swiss Institute of Bioinformatics) on the X axis and by hydrophobicity index GRAVY (ExPASy ProtoParam tool, SIB Swiss Institute of Bioinformatics) on the Y axis. Bubble size represents percent of a given protein out of the total mass calculated as explained in Figure 7-9. In post HIC eluate, 82% of total protein consists of product. Among the rest the most abundant are Drebrin (3.6%), Neuraminidase (2.5%), 78 kDa glucose-regulated protein (2.1%) and Elongation factor two (2%). Complete list of HCP identified in all samples is included in supporting data LC-MS/MS HCP.

As expected in the GAA producer, GAA was the dominating species in IEX eluate accounting for over 36% of the total by mass spectrometry. In the same GAA eluate sample, HCP quantified via AlphaLISA (PerkinElmer, MA USA) in section 7.3.2 accounted for approximately 54% of total protein. By comparison this enzyme accounts for approximately 2% by MS of the total species found in Null and mAb producing cell lines after IEX, further proof that basal GAA levels are found across many mammalian cells lines and that the GAA cell line is expressing recombinant material. Actin, clusterin, drebrin, elongation factor 2 and lysosomal protective protein (carboxypeptidase) were some of the more common HCP co-purifying due to their pI being within 1 unit of target molecule. HCPs with similar hydrophobicity to GAA but over 1 pI unit difference were thought to be passenger molecules that co-purified, some most likely by 'product-associated' interaction. This phenomenon of co-purification with the product due to interaction with the product itself (piggy backing), has been described previously for antibodies (Levy et al., 2014, Hogwood et al., 2016). In this category we found: neuraminidase (pI 6.82), the protease involved in cleaving glycosidic linkages in neuraminic acids; procollagen C-endopeptidase enhancer 1 (pI 8.63) involved in cartilage and bone formation; 40S ribosomal

protein S6 (pI 10.83); cAMP-dependent protein kinase catalytic subunit gamma (pI 8.7) a kinase; dystroglycan (pI 9.4) a complex involved in numerous processes such as cell survival and migration and membrane assembly (source Uniprot.org).

After intermediate purification HIC (*Figure 7-10*) GAA purity reached values of above 80%, while HCPs were greatly reduced. The top 5 HCP in order of abundance were Drebrin (3.6%), Neuraminidase (2.5%), 78 kDa glucose-regulated protein (2.1%) and Elongation factor 2 (2%) and Protein crossbronx (1%). Full HCP information can be found in supplemental information provided.

In the antibody producing cell line IEX eluate, some of the most abundant species found were antibody as expected, which account for about 25% of the total. In the pI range between 5 and 6, where direct adsorption to the IEX resin is anticipated, other proteins found include: legumain, a protease involved in lysosomal protein degradation; thioredoxin a protein responsible for various redox reactions through the reversible oxidation of its active center, and proteases lysosomal protective protein and ferrochelatase. Outside the typical pI range for direct interaction and therefore candidates for piggy-backing through the process are: CYR61 (pI 8.49), which promotes cell proliferation angiogenesis and cell adhesion; neuroaminidase (pI 6.8), dystroglycan (pI 9.4) and 40S ribosomal protein S6 (pI 10.8).

In the Null profile range of HCPs with pI between 4.5 and 6.5 we observed clusterin, lysosomal protective protein, Complement C3, thioredoxin (disulfide oxidoreductase), protocollagen C-endopeptidase enhancer (peptidase), phosphoglycerate mutase (kinase), lysosomal protective protein (serine carboxypeptidase), legumain (lysosomal hydrolysis) and the common heat shock cognate protein 70 KDa (Xiaohui Lu, 2015). HCPs with pI > 6.5 were identified as dystroglycan (pI 9.4), 40S ribosomal protein S6 (pI 10.8), matrix metalloproteinase-19 (pI 8.9). Interestingly most of the HCP identified in the pI > 6.5 range in this sample, were common with those found in

GAA and mAb producer, sign that their elution was not caused by product associated interaction, as product in the Null cell line is barely present. We believe that they might interact with another HCP that does interact with the resin.

Importantly, the cysteine protease cathepsin B and Z were ubiquitously present in moderate quantity in all three cell lines. More precisely cathepsin B was 0.60 μg/mL or 0.62 % of total protein content in GAA producer, 1.85 μg/mL or 0.83% of total protein in Null and 0.04 μg/mL or 0.20% of total protein in mAb producer. Cathepsin Z was 0.10 μg/mL or 0.10% of total protein in GAA producer, 0.41 μg/mL or 0.18% of total protein in Null and 0.01 μg/mL or 0.05% of total protein in mAb producer. This suggests these are not co-purifying with the products and are retained to some extent by the resin, or, that they also co-purify with another HCP that does interact with the resin.

The view of sponsor company BioMarin is that CatZ expression is linked to overexpression of lysosomal GAA. Although this could not be independently verified, this phenomenon could be due to the transport mechanism mannose–6–phosphate receptor being overloaded with GAA which induces CatZ and other lysosomal proteins being secreted as well. The commercial form of GAA, having the GILT tag is likely more sensitive to CatZ and other proteases due to potential clipping. Therefore overexpression of GAA/GILT results in CatZ expression which impacts stability of the target molecule.

With the exception of a small number of unique species found in excess in the GAA line and reported in the bottom part of Table 7-4, the HCP profile across the three cell lines was however found to be broadly similar and in agreement with other studies (Hogwood et al., 2016). A list of the 50 most abundant overlapping proteins common to all three cell lines is found in Table 7-4 (top). This overlap in HCPs across many different cell lines has been studied extensively in two recent studies. In the first one (Tait et al., 2012), the authors compared HCPs in harvest

supernatants of day 10, 12 and 14 cultures generated by a Null cell line and a recombinant IgG4-producing cell line and reported that whilst there were increases in HCPs, which they associated with viability decline later in culture, there were many similarities between the cell lines and days of harvest. In another study (Yuk et al., 2015), the HCP profile of three different Null cell lines was compared and it was reported that despite differences in CHO lineage, upstream process, and culture performance, the cell lines yielded similar HCP profiles. Furthermore, this study determined that about 80% of the most abundant 1000 proteins identified were common to all three lines.

While the HCP profile in this study is largely similar across the three cell lines, some are unique only to the GAA cell line. Table 7-4 (bottom) shows the relative distribution and calculated concentration of the four species that were found to be particularly abundant in the GAA line and almost absent in the other two. These are: (1) actin (cytoskeletal protein ubiquitously expressed in all eukaryotic cells) where a relatively low amount was found in Null and mAbs lines, a much greater amount was found in GAA line material. (2) drebrin (related to actin polymerization) which was also present predominantly in the GAA cell line material. (3) Probable dual-specificity RNA methyltransferase RlmN (involved in RNA formation) and (4) EGF-containing fibulin-like extracellular matrix protein 1 (involved in cell adhesion and migration). We suspect that the disparity in amounts observed for these particular species might be due to a combination of two factors: the large amount of GAA in GAA producer sample could possibly be responsible for mass spectrometer detector sensitivity saturation, which could explain quantification discrepancies with the other two lines. This could also account for differences in the mAb cell line where a high concentration of mAb is present. Also, the competitive behaviour of different species onto the anion exchange resin could have affected the amount of target bound coeluting species.

7.5 Lysosome phenotype characterization

Understanding the GAA production pathway in healthy mammalian cells is an important step in determining how GAA overexpression affects overall cell functionality. As discussed before in the GAA synthesis section (2.2.4.2), the newly formed polypeptide leaves the endoplasmic reticulum and goes through Golgi where M6P receptors bind to it and are responsible for its transfer into the preacidified compartment, before entering the lysosome. In this process only 10-20% of GAA fails to bind to the receptors and is secreted out of the cell through secretory granules (Kornfeld, 1987).

GAA overexpression in recombinant production CHO cells is likely to have an effect on cell physiology and as a result a potential impact on HCP release and product degradation. As most of the GAA produced by cells ends up accumulating in the lysosome, we hypothesize differences in phenotype between lysosomes from a Null and a GAA producer CHO maybe apparent. For this reason we compared images of lysosomes from two cell cultures using Transmission Electron Microscopy (TEM).

7.5.1 Investigating the lysosomal phenotype with TEM in GAA producing cells

For this experiment we wanted to investigate whether GAA over-expression affected lysosomal morphology by comparing images of lysosomes from the recombinant GAA expressing cell line and the originating Null CHO cells. The resulting data is presented in *Figure 7-11*, *Figure 7-12* and *Figure 7-13* and shows a clear difference in lysosomal phenotype between Null and GAA cells. While lysosomes are present in similar numbers (on average 18 per cell) and size (400 nm) in both cell types, they are generally full and dark in GAA overexpression cells and empty or mostly empty in Null cells. Specifically, over 72% of lysosomes are completely full (F) in the GAA cell line while only 8% present the same phenotype in the Null line. This trend was also observed when comparing lysosomes containing some (%HF – half full) to no (%E - empty) dark matter: only 17% of all lysosomes in the GAA overexpression cell line present this phenotype

versus 55% in Null. This data provides evidence that over expression of GAA leads to lysosomes becoming full and engorged. Potentially this could contribute to the overall cellular stress levels experienced by the cell.

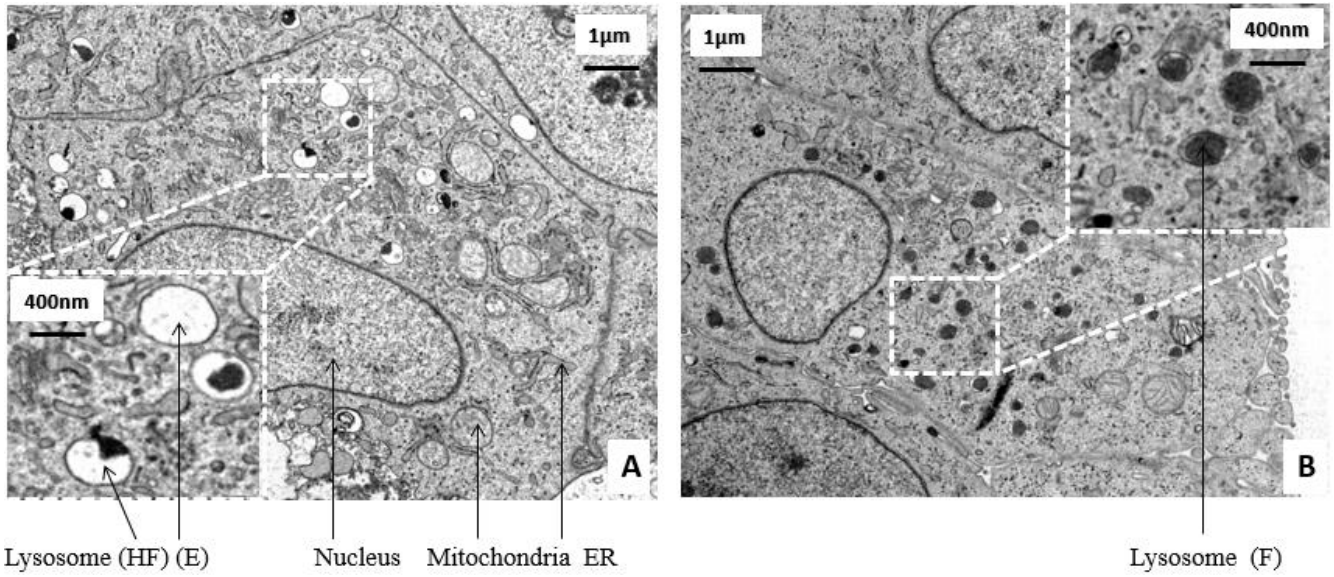


Figure 7-11 Transmission Electron Microscope images of Null (A) and GAA producer (B) cells. Cells were sampled at the same time from actively growing cultures, fixed and ultra-thin slices were prepared for TEM imaging (magnification 3000X). 33 images of single cells were taken per each sample. Lysosomes were counted and categorized based on content in Full (F), Half-Full (HF) and Empty (E). Images chosen for this figure are representative; all other images are included in supporting data "TEM images"

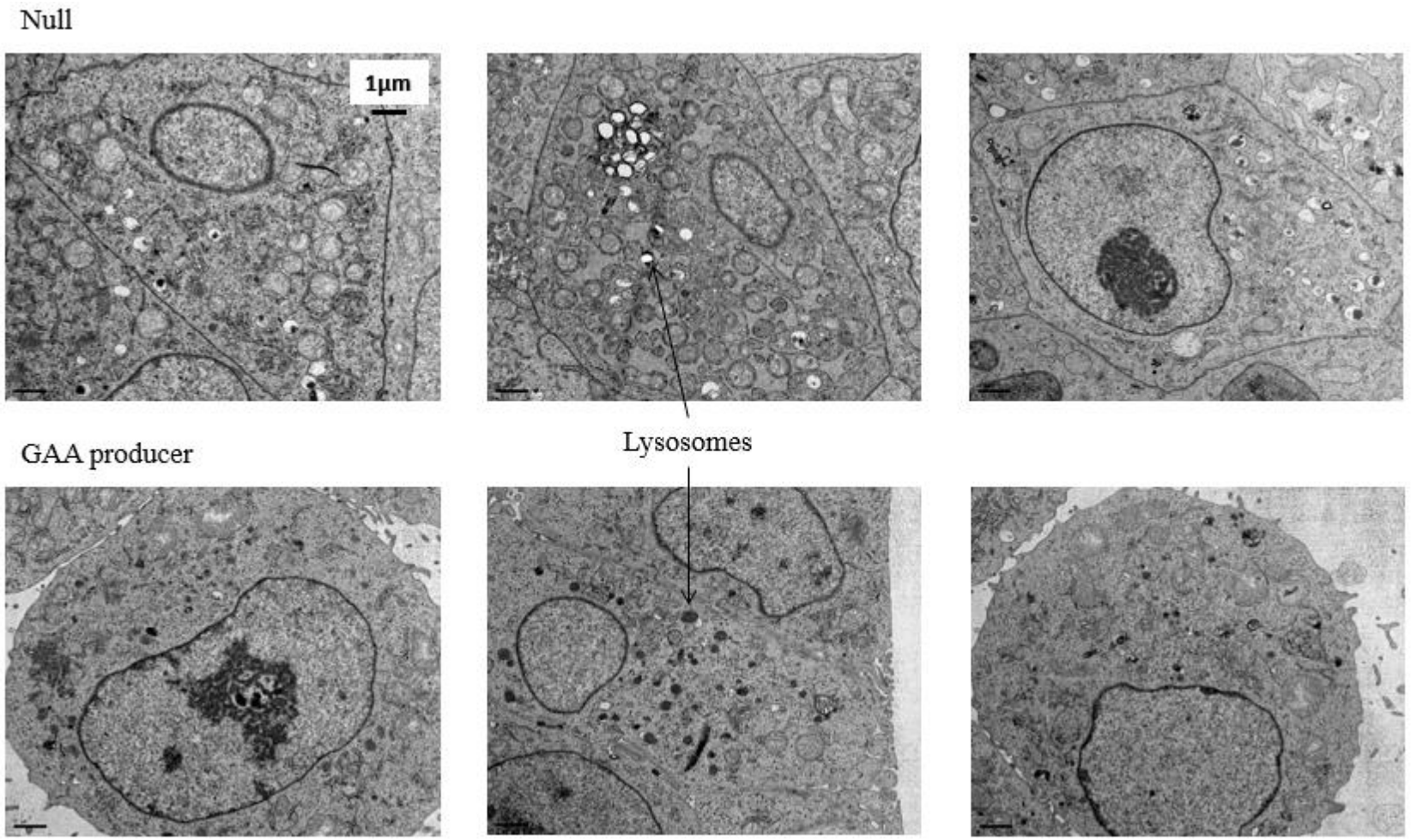


Figure 7-12. More TEM images of Null (top row) and GAA producer (bottom row) cells prepared as described in Figure 7-11 to highlight difference in lysosomal phenotypes. As previously shown lysosomes are dark and engorged in GAA producer and empty in Null.

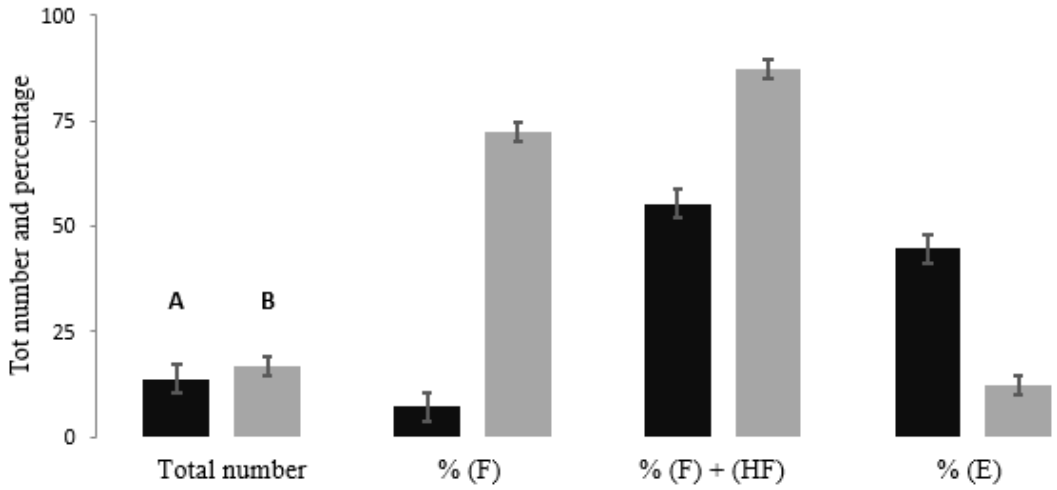


Figure 7-13 TEM Lysosomes data collated and analysed in excel (A – Null; B – GAA). Values derived from 33 images per each cell line.

7.6 Chapter conclusions

In this chapter we discussed the effects that overexpression of GAA from the CHO cell culture has on overall cell stress and how this relates to release of HCP and protease levels. This is important not only for potential product degradation caused by proteases but also for deciding the appropriate downstream processing strategy.

A first qualitative assessment of protease level in clarified harvest and capture step eluate via zymography (*Figure 7-3*) revealed that IEX is able to remove significant amounts of proteases from the process stream but with a notable coeluting bands at 80KDa. MS revealed this to be composed of Dipeptidyl peptidase 3, a metalloproteinase and Prolyl endopeptidase a serine protease.

Total HCP amount in the harvest and post IEX purification was quantified via immunoassay AlphaLISA, a non-wash higher throughput version of ELISA. This assay needed optimization to correctly quantify HCP in harvest due to the hook effect caused by high HCP concentration in harvest sample. HCP was determined to be around 0.09 mg/ml in harvest, 50% of which coeluting in IEX eluate fraction with GAA. By comparison HCP amount in post intermediate purification (HIC) accounted only for less than 20% of total protein as measured by MS. When comparing HCP amount in harvest determined via AlphaLISA to gel analysis images (i.e. *Figure 5-1*), it comes clear how AlphaLISA grossly underestimated the actual value, perhaps due to interference caused by cell media components. Limitations of as ELISA methods in the determination of HCP in harvest samples have been highlighted in (Wang et al., 2015, Zhu-Shimoni et al., 2014) and are thought to be due to a combination of underrepresentation of low abundance/immunogenicity species due to HCP heterogeneity and dilutional non-linearity caused by copurifying HCPs enriched in drug substance that may have limited antibody detection.

A proteomic approach was adopted in order to gather more information on the identity of the species found. Different tandem MS techniques were used to identify those HCPs eluting with GAA in the IEX eluate. After getting unsatisfactory results using in-gel tryptic digest MALDI MSMS and MALDI TOF LC/MSMS due to a series of issues, we used a high resolution nano-LC/MSMS to compare the HCP profile of the GAA IEX eluate sample with a Null and a monoclonal antibody producer CHO purified in the same way. This revealed that GAA is the dominating eluting species (as expected) in the GAA CHO, while this protein accounts for around 2% in the other two. Coeluting HCPs with a similar pI to that of the target protein were listed, together with other HCPs of pI value over 1 unit difference from GAA but eluting due to association to product or another coeluting HCP. Data was presented in bubble graph format in *Figure 7-9* and *Figure 7-10*. This analysis also revealed that the great majority of HCP profile in the three CHO cultures IEX overlaps, as also noted by similar studies (Yuk et al., 2015, Hogwood et al., 2016).

Cysteine proteases cathepsin Z and B were found in moderate quantity across the three cell lines IEX eluates suggesting that these do not co-purifying with the products and are retained to some extent by the resin, or, that they also co-purify with another HCP that does interact with the resin. Barely detectable amount of these two proteases were found in the subsequent HIC step eluate (less than 0.1% of the total protein), indicating that they were removed from the process stream in the HIC flowthrough.

Lastly to visually assess the effect that product overexpression has on producing cells, we captured images of lysosomes, the organelles that store GAA, and compared them with images of Null cells using Transmission Electron Microscope. This revealed that the majority of lysosomes (72%) in the GAA producer were of dark and engorged nature indicating that stress deriving from target

molecule overexpression had detectable effects on phenotype of the cells. The origin of the dark matter in lysosomes has not been evaluated, although we believe GAA to be part of it.

8 Conclusions

The work presented within this thesis has focused on a number of aspects of upstream and downstream processing. Much of this has been undertaken using methodologies for process development, allowing for product and HCP characterization at different steps of DSP. These investigations have benefited by the inclusion of analytical techniques for characterization purposes, with HCPs being pivotal process related impurities of interest. First, with reference to the research objectives defined in introduction, and as a means of summarizing the overall findings of this doctorate, a number of conclusions are drawn.

Objective one: Production of stable CHO cell line for material generation.

The generation of the mammalian cell line producing GAA was the first objective of the project and it was a necessary one in order for the rest of the work to take place. It started with the cloning of the GAA gene into a plasmid vector and the incorporation of the vector into a Flp-In CHO cell line. The Flp-In system was selected in order to shorten the clonal screening process due to its 'site-directed recombination', which facilitates integration of the gene of interest into a specific highly transcribed FRT site in the host genome, and allows all clones generated to produce considerable and comparable levels of recombinant product. Product production was first assessed via western blot of transiently transfected cultures. Stable transfection was then performed and clones were generated with limiting dilution cloning. Clones titer was lastly assessed using a GAA diagnostic assay and one clone was selected as candidate for material generation and downstream process development studies.

The chosen clone (GAA 1) was able to generate 0.18 g/l of human GAA after 6 days of batch culture in a 1 litre cell flask. The cell count at harvest was 15 million cells per millilitre.

Objective two: Development of assays for detection of product in the bioprocess stream

The aim of this chapter was to present the analytical techniques employed to characterize the product secreted by the clonal stable cell line generated in results chapter one. Before proceeding with downstream development and released HCPs characterization, it was important to ensure the correct isoform of GAA was produced and in what quantity. The clarified harvest of the selected clone underwent a series of analysis to determine GAA presence, amount and identity.

Comparison of GAA cell culture harvest with Null CHO line via gel electrophoresis highlighted the presence of a clear band at around 110KDa which was absent in the Null line. Analysis of the same sample by western blot, confirmed that the band at 110KDa was in fact recognised by anti-human GAA antibody. Product identity was confirmed by tryptic digestion of the protein band of interest and running it via tandem mass spectrometry. Peptide matching to SwissProt database positively identified it as 106112 Dalton LYAG_HUMAN with good result confidence.

GAA quantification was performed via a GAA diagnostic method used in clinics to diagnose presence of the enzyme in suspected Pompe patients. This method, unlike other enzyme activity assays, is specific to GAA thanks to the use of an alpha-glucosidase inhibitor. This permitted determination of clone 1 titer and was subsequently used to determine product content after purification steps.

Objective three: Bioprocess mimic

After having generated a stable cell line producing GAA and having developed the analytical assays to characterize the product, we proceeded with development of the purification process. Cell culture harvest was spun and filtered through a 0.22μm stericup filter to remove whole cells and cell debris. cHCCF (clarified Harvest Cell Culture Fluid) was then subjected to two rounds of chromatographic separation cycles, first based on protein charge (capture IEX) then followed

by protein hydrophobicity (intermediate HIC). The final target protein purity was above 80% as determined later on by MS.

Objective four: Characterization of cell stress and its impact on protease production

In the last results chapter we presented the data supporting the cell line and HCP characterization work. HCP presence in the process stream is a concern due to potentially harmful proteolytic product degradation effects and must be monitored and reduced to comply with strict regulatory requirements. A series of immunoassays, electrophoretic gels and mass spectrometry techniques were used to track specific HCPs throughout the bioprocess. Furthermore we showed how cell stress resulting from target product over-expression can have physical effects on cell internal structures and can possibly be related to increased HCP release.

HCP content throughout the bioprocess was assessed via electrophoresis and mass spectrometry. The degree of purification of the developed bioprocess can be appreciated by looking at Figure 6-11. The gel compares cHCCF to post capture step eluate to intermediate step eluate fractions. While the GAA band remained visible in all three samples at around 110KDa, impurities were greatly reduced.

LC-MS/MS analysis and emPAI quantification of cHCCF and post capture (IEX) eluate (Figure 7-8) showed how most identified HCPs were reduced by the capture step, while GAA relative amount increased due to product isolation and concentration. The HCP profile of the GAA producing cell line was compared to a Null and an antibody producing CHOs via high resolution nano LC-MS/MS which revealed that while the majority of HCPs overlap across the three species as seen in literature, some are unique to the GAA strain (Figure 7-9).

High resolution nano LC-MS/MS was also used to show the reduction of HCP in intermediate (HIC) step and to identify those species coeluting with the product (Figure 7-10). The number of

HCPs identified went from 233 in post IEX to less than 20 in post HIC eluate samples with the most abundant coeluting ones being drebrin (3.6%), neuraminidase (2.5%), 78 kDa glucose-regulated protein (2.1%) and elongation factor 2 (2%). Product purity went from 32% in post IEX eluate to 82% in post HIC.

General protease clarification of the capture IEX step was assessed via zymography and nano LC-MS/MS. Zymography revealed that most of the proteases present in cHCCF were cleared post capture purification except a significant amount at 80KDa. Mass spectrometry identified those as dipeptidyl peptidase 3 and prolyl endopeptidase respectively a serine and a metalloproteinase with hydrolytic activity. Effects of their existence in presence of GAA have not been evaluated.

In reference to effects of overexpression over cellular integrity, we presented images taken by a Joel 1010 transmission electron microscope of GAA producing cells and compared them with similar images of Null CHO. Images showed that while there is no significant difference in most cell organelles between the two cell lines, lysosomes of GAA producer are largely dark and engorged if compared with their counterparts in Null cell line. While we did not prove that GAA is in fact making up the dark matter observed, we know that the lysosome is where most GAA not secreted resides. Regardless of the nature of the dark matter observed, this experiment showed how product overexpression can subject the cell to physical stress.

9 Future work

Having identified the HCPs released by the generated CHO GAA producer cell line and how stress induced by overexpression impacts cellular behaviour has only in part fulfilled the original scope of this work. Given more time, additional areas could have been explored more in depth. In this section we propose some short and long term projects that could complement and enrich this research.

9.1 Short term

Recombinant GAA localization in GAA producer cell lines.

During the project, lysosomal cell stress was assessed by observation of the size and content of lysosomes via TEM imaging. We determined that GAA overexpressing cells lysosomes were generally darker than their counterparts in Null CHO cells and seemed to be full of matter which we putatively hypothesized as being GAA. While this was likely the case, we were not able to verify it.

Although we have a general idea of GAA cellular pathway from (Kornfeld, 1987), specific information on enzyme localization in overexpressing lines is still unavailable. In order to verify presence of GAA in lysosomes of GAA producer cell lines, a combination of confocal microscopy and immunostaining could be applied. By exposing the cells to a GAA primary mAb linked to gold particles (immunogold staining) it should be possible to highlight whether lysosomes do in fact contain GAA, or instead GAA is just present within cytoplasmic region. Because confocal microscope images are not as high resolution as TEM images, lysosomes are likely not going to be easily visible. By using a second staining directed to the acidic compartment themselves (such as lysosomes in this case) it should be possible to highlight presence of lysosomes and eventual coexistence of GAA enzyme within them. LysoTracker®

Red DND-99 (Thermo Fisher cat# L7528) provides labeling and tracking of acidic organelles in live cells.

Polish step development

The bioprocess designed in this work consists in a capture IEX step followed by intermediate HIC step. The final purity of the GAA product was measured via LC-MS/MS at 82%. While we demonstrated the process to be suitable for the project scope, the final purity of the product is still too low to meet to any regulatory requirements for commercial use. We propose the use of an additional chromatographic step to increase final product purity. At this stage, as separations by charge, hydrophobicity have already been used, a high resolution gel filtration (GF) step would be ideal for polishing. GF is the slowest of all chromatography techniques and the size of the column determines the volume of sample that can be applied, it is therefore most logical to use GF after techniques which reduce sample volume so that smaller columns can be used. We suggest the use of Superdex media, a high resolving gel filtration media for short run times and good recovery. Superdex is the first choice at laboratory scale and Superdex prep grade for large scale applications. Typical flow velocity is up to 75 cm/h.

9.2 Medium Term

Protease inhibitors screening to determine which protease in process feed is actually of concern.

From the proteomic approach to process development we have learned how certain classes of proteases behave during DSP. We now know that the two potentially harmful cysteine and serine proteases (Cathepsin B and Z) that were thought to damage GAA, are present in the capture step IEX elution, but eventually are cleared out in intermediate purification by HIC.

This information is useful in determining the appropriate conditions for capture step elution material storage in order to avoid proteolytic activity and degradation of the product. The information generated by this approach however, could potentially be used in another useful way.

Having identified all the HCPs present throughout the process stream of a GAA producer, allows for additional screening to take place to determine which ones are actually affecting GAA structural integrity. While not all HCPs have proteolytic activity, those that do could be individually tested in presence of GAA reference standard and eventual GAA fragmentation could be assessed via gel electrophoretic methodology. While incubating GAA with every identified protease is unrealistic, we propose an approach based on the use of protease inhibitors. GAA reference could be incubated in cell harvest supernatant and different classes of protease inhibitors could be individually screened in a high throughput mode. The commercially available protease screening kit ProteSEEKER (VWR cat# 786-325) comes supplied with 11 concentrated protease inhibitors designed to inhibit all classes of proteases known. By comparing GAA degradation occurring with each individual inhibitor, it should be possible to determine which class of protease degrades GAA.

This type of approach would be beneficial not just in GAA process development but could potentially be repeated for other therapeutic proteins production where protease degradation is of concern.

9.3 Long term

Due to enormous financial consequences of late stage failure of a drug candidate, one of pharmaceutical companies' major concern is to make the drug development process as quick, reliable and cost effective as possible. For this reason, considerable resources are invested in cell and process development. Recent advances in the proteomics analytical capabilities such

as the novel high resolution mass spectrometry features shown in this work, are allowing ever simpler access to information not only on identification of molecular species present in in-process samples, but also on biomarker and target discovery, protein mechanisms of action, prediction of toxicity. As these techniques become ever more powerful and fast, the amount of generated data can quickly become overwhelming and unfortunately, a substantial number of these efforts result only in long lists of proteins but fall short in demonstrating the effects of the controlled expression on the desired phenotype. We believe the industry would benefit from a new approach at data interpretation, where specialized software would constantly monitor process stream and adapt in real time a dynamic DSP based on specific proteases found in the feed. As omics technologies advance (genomics, proteomics and metabolomics), producing cells could also be specifically engineered to silence genes known of expressing harmful proteases, and diminish the amount of HCP present to the bare minimum necessary for cell survival and at the same time diminish cellular stress deriving from overexpression.

List of publications by author

Migani, D., Smales, C. M. and Bracewell, D. G. (2017), Effects of lysosomal biotherapeutic recombinant protein expression on cell stress and protease and general host cell protein release in Chinese hamster ovary cells. Biotechnol Progress. doi:10.1002/btpr.2455

Acknowledgments

Prof Daniel G. Bracewell, University College London for his endless patience, support, advice and availability.

Prof C. Mark Smales, University of Kent, for his cell culture and mass spec invaluable support.

Dr John Henstrand, BioMarin Pharmaceuticals, for providing reference materials and assistance with high throughput binding screening experiment at BioMarin laboratories, and also for hosting me during my visit to BioMarin.

Dr Erik Fouts, BioMarin Pharmaceuticals, for guidance and financial support.

Dr Naichi Liu, BioMarin Pharmaceuticals and her lab, for assistance with high throughput binding screening experiment at BioMarin laboratories.

Dr Alex Kinna, University College London for assistance in the transfection experiments.

Mark Turmaine, University College London for assistance with Transmission Electron Microscopy.

Dr Kevin Howland, University of Kent, for assistance in setting up and analyzing/interpreting mass spec experiments.

Dr Derralynn Hughes, haematology department at Royal Free Hospital and her lab for assisting in setting up the GAA diagnostic assay.

Dr Matthew Davis for his absolute dedication to researchers' safety.

List of Abbreviations

4MU-αGlc	4-methylumbelliferyl-α-D-glucopyranoside
ACE	Alglucosidase Alfa Control and Education
BANA	Benzoylarginine naphthylamide
CatZ	Cathepsin Z
CBE	Conduritol B epoxide
CHCA	α-cyano-4-hydroxycinnamic acid
CHO	Chinese hamster ovary
CIP	Clean in Place
CV	Column Volume
Da	Dalton
DAB	Diaminobenzidine
DHB	2,5-dihydroxybenzoic acid
dhfr	Dihydrofolate reductase
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSP	Downstream Processing
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
EURORDIS	European Organization for Rare Diseases
ExPASy	Expert Protein Analysis System
FDA	Food and Drug Administration
FF	Fast Flow
FRT	Flippase recognition target
FT	Flow through
GAA	Acid alpha-glucosidase
GFP	Green Fluorescent protein
GILT	Glycosylation independent lysosomal targeting
GRAVY	Grand average of hydropathicity
HCCF	Harvest Cell Culture Fluid
HCP	Host cell protein
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC	Hydrophobic interaction chromatography
HRP	Horseradish Peroxidase
IAA	Iodoacetic acid
IAM	Iodoacetamide
IGF2	Insulin-like growth factor
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LC	Liquid chromatography
LDC	Limiting dilution cloning
LEAP	Laser-enabled analysis and processing
LSD	lysosomal glycogen storage disease
M6P	Mannose 6 phosphate
mAbs	Monolconal antibody
MALDI	Matrix-assisted laser desorption/ionization
MBP	Maltose binding protein

MCA	Methyl Cumaryl Amide
MGAM	maltase-glucoamylase
MMP	matrix metalloproteases
MOPS	3-(N-morpholino)propanesulfonic acid
MPS	Mucopolysaccharidoses
MS	mass spectrometry
MSR	Molar Substitution Rate
MTX	Methotrexate
MW	Molecular Weight
MWCO	Molecular Weight Cut Off
NaCl	Sodium Chloride
NEM	n-ethylmaleimide
NMEs	New Molecular Entities
NS0	murine myeloma cell
OD	optical density
pAb	polyclonal Antibody
PBS-T	Phosphate Buffered Saline Tween
PCMB	Chloromercuribenzoic acid
PEI	Polyethylenimine
PER.C6	human retina derived
pH	potential of hydrogen
PMSF	phenylmethane-sulphonylfluoride
PVDF	Polyvinylidene fluoride
R&D	Research and Development
RE	Restriction Enzyme
REMS	Risk evaluation and mitigation strategy
RER	Rough endoplasmic reticulum
RFU	Relative Fluorescent Units
RNA	Ribonucleic acid
RT	Retention Time
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TEM	Transmission Electron Microscope
TIMPs	Tissue Inhibitor MetalloProteases
TOF	Time of Flight
TSAP	Thermosensitive Alkaline Phosphatase
UCL	University College London
UPLC	Ultra Performance Liquid Chromatography
w.d.	working dilution
WHO	World Health Organization

Bibliography

ABCAM. 2012. *WESTERN BLOTTING* [Online]. Available: http://www.abcam.com/ps/pdf/protocols/WB-beginner.pdf [Accessed].

ABCAM. 2013. *Anti-GAA antibody [EPR4716(2)] (ab137068)* [Online]. Available: http://www.abcam.com/gaa-antibody-epr47162-ab137068.html [Accessed].

AGARWAL, S. K. 1990. Proteases Cathepsins - a View. *Biochemical Education,* 18**,** 67-72.

BAXTER. 2011. *Recombinate* [Online]. Available: http://www.recombinate.com/ [Accessed 25-02 2013].

BEN-NAIM, A. 1980. Hydrophobic interactions. *In:* LONDON, P. P. N. Y. A. (ed.) *Biochemistry and Molecular Biology Education.*

BERTI, P. J. & STORER, A. C. 1995. Alignment Phylogeny of the Papain Superfamily of Cysteine Proteases. *Journal of Molecular Biology,* 246**,** 273-283.

BIOGENIDEC/ELAN. 2012. *Tysabri* [Online]. Available: http://www.tysabri.com/ [Accessed 22-01 2013].

BIOGENIDEC/GENENTECH. 2012. *Non-Hodgkin lymphoma (NHL)* [Online]. Company website. Available: http://www.rituxan.com/index.html [Accessed 22-01 2013].

BIOMARIN 2016. BIOMARIN UPDATE FOR THE POMPE COMMUNITY: 22 JUNE 2016. *In:* INC., B. P. (ed.).

BIOMARIN_PHARMACEUTICAL_INC. 2013. *BioMarin Pipeline* [Online]. Available: http://www.bmrn.com/products/index.php [Accessed 09-01 2013].

BIRCH, J. R. & RACHER, A. J. 2006. Antibody production. *Advanced Drug Delivery Reviews,* 58**,** 671-685.

BRACEWELL, D. G., FRANCIS, R. & SMALES, C. M. 2015. The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnology and Bioengineering,* 112**,** 1727-37.

BROACH, J. R., GUARASCIO, V. R. & JAYARAM, M. 1982. Recombination within the yeast plasmid 2mu circle is site-specific. *Cell,* 29**,** 227-34.

BROACH, J. R. & HICKS, J. B. 1980. Replication and recombination functions associated with the yeast plasmid, 2 mu circle. *Cell,* 21**,** 501-8.

BROWNE, S. M. & AL-RUBEAI, M. 2007. Selection methods for high-producing mammalian cell lines. *Trends in Biotechnology,* 25**,** 425-432.

BUCHHOLZ, F., RINGROSE, L., ANGRAND, P. O., ROSSI, F. & STEWART, A. F. 1996. Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. *Nucleic Acids Res,* 24**,** 4256-62.

CHIBA, S. 1997. Molecular mechanism in alpha-glucosidase and glucoamylase. *Bioscience Biotechnology and Biochemistry,* 61**,** 1233-1239.

CHIVERTON, L. M., EVANS, C., PANDHAL, J., LANDELS, A. R., REES, B. J., LEVISON, P. R., WRIGHT, P. C. & SMALES, C. M. 2016. Quantitative definition and monitoring of the host cell protein proteome using iTRAQ - a study of an industrial mAb producing CHO-S cell line. *Biotechnol J,* 11**,** 1014-24.

CHON, J. H. & ZARBIS-PAPASTOITSIS, G. 2011. Advances in the production and downstream processing of antibodies. *N Biotechnol,* 28**,** 458-63.

CLINCKE, M. F., GUEDON, E., YEN, F. T., OGIER, V. & GOERGEN, J. L. 2011. Characterization of metalloprotease and serine protease activities in batch CHO cell cultures: control of human recombinant IFN-gamma proteolysis by addition of iron citrate. *BMC Proc,* 5 Suppl 8**,** P115.

CLINICALTRIALS.GOV. 2013a. *A Long-Term Study for Extended BMN 701 Treatment of Patients With Pompe Disease Who Have Completed BMN 701 Studies* [Online]. Available: http://www.clinicaltrials.gov/ct2/show/NCT01435772 [Accessed 24-01 2013].

CLINICALTRIALS.GOV. 2013b. *Safety/Tolerability/Pharmacokinetic (PK)/Pharmacodynamics (PD) Study of BMN701 in Patients With Late-Onset Pompe Disease* [Online]. Available: https://www.clinicaltrials.gov/ct2/show/results/NCT01230801 [Accessed 23-02 2016].

COLE, L. A., SHAHABI, S., BUTLER, S. A., MITCHELL, H., NEWLANDS, E. S., BEHRMAN, H. R. & VERRILL, H. L. 2001. Utility of commonly used commercial human chorionic gonadotropin immunoassays in the diagnosis and management of trophoblastic diseases. *Clinical Chemistry,* 47**,** 308-15.

CRAIG, N. L. 1988. The mechanism of conservative site-specific recombination. *Annu Rev Genet,* 22**,** 77-105.

DICK, L. W., MAHON, D., QIU, D. F. & CHENG, K. C. 2009. Peptide mapping of therapeutic monoclonal antibodies: Improvements for increased speed and fewer artifacts. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 877**,** 230-236.

DONEANU, C. E., XENOPOULOS, A., FADGEN, K., MURPHY, J., SKILTON, S. J., PRENTICE, H., STAPELS, M. & CHEN, W. 2012. Analysis of host-cell proteins in biotherapeutic proteins by comprehensive online two-dimensional liquid chromatography/mass spectrometry. *MAbs,* 4**,** 24-44.

ELFERINK, R. P. J. O., BROUWERKELDER, E. M., SURYA, I., STRIJLAND, A., KROOS, M., REUSER, A. J. J. & TAGER, J. M. 1984. Isolation and Characterization of a Precursor Form of Lysosomal Alpha-Glucosidase from Human-Urine. *European Journal of Biochemistry,* 139**,** 489-495.

ELLIOTT, P., HOHMANN, A. & SPANOS, J. 2003. Protease expression in the supernatant of Chinese Hamster Ovary cells grown in serum-free culture. *Biotechnology Letters,* 25**,** 1949-1952.

EUROPEANCOMMISSION. 2004. *Useful information on rare diseases from an EU perspective* [Online]. Available: http://ec.europa.eu/health/ph_information/documents/ev20040705_rd05_en.pdf [Accessed 24-01 2013].

FDA. 2010. *FDA Approves New Treatment for Late-Onset Pompe Disease* [Online]. FDA webpage. Available: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm213282.htm [Accessed].

FDA. 2013. *Developing Products for Rare Diseases & Conditions* [Online]. FDA webpage. Available: http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/default.htm [Accessed 24-01 2013].

FERNANDO, S. A. & WILSON, G. S. 1992. Studies of the 'hook' effect in the one-step sandwich immunoassay. *Journal of Immunological Methods,* 151**,** 47-66.

FISHER, L. M. 1986. Biotechnology spotlight now shines on Chiron. *New York Times Magazine*.

GABEL, C. A., GOLDBERG, D. E. & KORNFELD, S. 1982. Lysosomal-Enzyme Oligosaccharide Phosphorylation in Mouse Lymphoma-Cells - Specificity and Kinetics of Binding to the Mannose 6-Phosphate Receptor Invivo. *Journal of Cell Biology,* 95**,** 536-542.

GAO, S. X., ZHANG, Y., STANSBERRY-PERKINS, K., BUKO, A., BAI, S., NGUYEN, V. & BRADER, M. L. 2011. Fragmentation of a highly purified monoclonal antibody attributed to residual CHO cell protease activity. *Biotechnology and Bioengineering,* 108**,** 977-82.

GENENTECH. 1985. *FDA Approves Genentech's Drug to Treat Children's Growth Disorder* [Online]. Available: http://www.gene.com/media/press-releases/4235/1985-10-18/fda-approves-genentechs-drug-to-treat-ch/ [Accessed 22 2013].

GORMAN, C. & BULLOCK, C. 2000. Site-specific gene targeting for gene expression in eukaryotes. *Current Opinion in Biotechnology,* 11**,** 455-460.

HALL, T. 2013. BioEdit.

HERMANS, M. M. P., KROOS, M. A., VANBEEUMEN, J., OOSTRA, B. A. & REUSER, A. J. J. 1991. Human Lysosomal Alpha-Glucosidase - Characterization of the Catalytic Site. *Journal of Biological Chemistry,* 266**,** 13507-13512.

HO, S. L. 2011. Structural Bioinformatics Analysis of Acid Alpha-Glucosidase Mutants with Pharmacological Chaperones. *Bioinformatics - Trends and Methodologies*.

HOEFSLOOT, L. H., HOOGEVEENWESTERVELD, M., KROOS, M. A., VANBEEUMEN, J., REUSER, A. J. J. & OOSTRA, B. A. 1988. Primary Structure and Processing of Lysosomal Alpha-Glucosidase - Homology with the Intestinal Sucrase Isomaltase Complex. *EMBO Journal,* 7**,** 1697-1704.

HOFSTEE, B. H. & OTILLIO, N. F. 1978. Non-ionic adsorption chromatography of proteins. *Journal of Chromatography,* 159**,** 57-69.

HOGWOOD, C. E., AHMAD, S. S., TARRANT, R. D., BRACEWELL, D. G. & SMALES, C. M. 2016. An ultra scale-down approach identifies host cell protein differences across a panel of mAb producing CHO cell line variants. *Biotechnol J,* 11**,** 415-24.

HOOKS, M. A., WADE, C. S. & MILLIKAN, W. J. 1991. Muromonab Cd-3 - a Review of Its Pharmacology, Pharmacokinetics, and Clinical Use in Transplantation. *Pharmacotherapy,* 11**,** 26-37.

HUANG, Y., LI, Y., WANG, Y. G., GU, X., WANG, Y. & SHEN, B. F. 2007. An efficient and targeted gene integration system for high-level antibody expression. *Journal of Immunological Methods,* 322**,** 28-39.

HUANG, Y. M., HU, W., RUSTANDI, E., CHANG, K., YUSUF-MAKAGIANSAR, H. & RYLL, T. 2010. Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnol Prog,* 26**,** 1400-10.

INVITROGEN. 2013a. *pcDNA3.1/V5-His-TOPO* [Online]. Available: http://www.lifetechnologies.com/order/catalog/product/V600520 [Accessed].

INVITROGEN. 2013b. *pcDNA5/FRT* [Online]. Available: http://products.invitrogen.com/ivgn/product/V601020 [Accessed].

ISHIHAMA, Y., ODA, Y., TABATA, T., SATO, T., NAGASU, T., RAPPSILBER, J. & MANN, M. 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics,* 4**,** 1265-72.

JEFFREY, P. L., BROWN, D. H. & BROWN, B. I. 1970. Studies of Lysosomal Alpha-Glucosidase .1. Purification and Properties of Rat Liver Enzyme. *Biochemistry,* 9**,** 1403-&.

JENNISSEN, H. P. & HEILMEYER, L. M., JR. 1975. General aspects of hydrophobic chromatography. Adsorption and elution characteristics of some skeletal muscle enzymes. *Biochemistry,* 14**,** 754-60.

KORNFELD, S. 1987. Trafficking of Lysosomal-Enzymes. *FASEB Journal,* 1**,** 462-468.

KORNFELD, S. & MELLMAN, I. 1989. The Biogenesis of Lysosomes. *Annual Review of Cell Biology,* 5**,** 483-525.

KOSTER, J. F., SLEE, R. G., VANDERKLEIVANMOORSEL, J. M., RIETRA, P. J. G. M. & LUCAS, C. J. 1976. Physicochemical and Immunological Properties of Acid Alpha-Glucosidase from Various Human Tissues in Relation to Glycogenosis Type-2 (Pompes Disease). *Clinica Chimica Acta,* 68**,** 49-58.

LANTZ, M. S. & CIBOROWSKI, P. 1994. Zymographic Techniques for Detection and Characterization of Microbial Proteases. *Bacterial Pathogenesis, Pt A,* 235**,** 563-594.

LEUTY, R. 2010. BioMarin to buy ZyStor for up to $115M. *San Francisco Business Time*.

LEVY, N. E., VALENTE, K. N., CHOE, L. H., LEE, K. H. & LENHOFF, A. M. 2014. Identification and Characterization of Host Cell Protein Product-Associated Impurities in Monoclonal Antibody Bioprocessing. *Biotechnology and Bioengineering,* 111**,** 904-912.

LI, F., VIJAYASANKARAN, N., SHEN, A. Y., KISS, R. & AMANULLAH, A. 2010. Cell culture processes for monoclonal antibody production. *MAbs,* 2**,** 466-79.

MADALA, P. K., TYNDALL, J. D. A., NALL, T. & FAIRLIE, D. P. 2010. Update 1 of: Proteases Universally Recognize Beta Strands In Their Active Sites. *Chemical Reviews,* 110**,** Pr1-Pr31.

MAGA, J. A., ZHOU, J., KAMBAMPATI, R., PENG, S., WANG, X., BOHNSACK, R. N., THOMM, A., GOLATA, S., TOM, P., DAHMS, N. M., BYRNE, B. J. & LEBOWITZ, J. H. 2013. Glycosylation-independent lysosomal targeting of acid alpha-glucosidase enhances muscle glycogen clearance in pompe mice. *J Biol Chem,* 288**,** 1428-38.

MARTINIUK, F., ELLENBOGEN, A., HIRSCHHORN, K. & HIRSCHHORN, R. 1985. Further Regional Localization of the Genes for Human Acid Alpha-Glucosidase (Gaa), Peptidase-D (Pepd), and Alpha-Mannosidase-B (Manb) by Somatic-Cell Hybridization. *Human Genetics,* 69**,** 109-111.

MARTINIUK, F., TZALL, S. & CHEN, A. 1992. Recombinant Human Acid Alpha-Glucosidase Generated in Bacteria - Antigenic, but Enzymatically Inactive. *DNA and Cell Biology,* 11**,** 701-706.

MATHEWS, C. K., VAN HOLDE, K. E. & AHERN, K. G. 2003. *Biochemistry,* San Francisco, Calif., Benjamin Cummings.

MIFIT. 2010. *mifit* [Online]. google. Available: http://code.google.com/p/mifit/ [Accessed 01-02 2013].

MOE. 2012. *Molecular Operating Environment* [Online]. Available: http://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm [Accessed 01-02 2013].

MOORE, W. M. & SPILBURG, C. A. 1986. Purification of Human Collagenases with a Hydroxamic Acid Affinity Column. *Biochemistry,* 25**,** 5189-5195.

NAGASE, H. & WOESSNER, J. F. 1999. Matrix metalloproteinases. *Journal of Biological Chemistry,* 274**,** 21491-21494.

NEWENGLANDBIOLABS. 2013. *Anti-rabbit IgG, HRP-linked Antibody #7074* [Online]. Available: http://www.cellsignal.com/products/7074.html [Accessed].

OKUMIYA, T., KEULEMANS, J. L., KROOS, M. A., VAN DER BEEK, N. M., BOER, M. A., TAKEUCHI, H., VAN DIGGELEN, O. P. & REUSER, A. J. 2006. A new diagnostic assay for glycogen storage disease type II in mixed leukocytes. *Molecular Genetics and Metabolism,* 88**,** 22-8.

ONDR, J. K. & PHAM, C. T. N. 2004. Characterization of murine cathepsin W and its role in cell-mediated cytotoxicity. *Journal of Biological Chemistry,* 279**,** 27525-27533.

ORIGENE. 2013. *GAA (NM_000152) Human cDNA Clone* [Online]. Available: http://www.origene.com/human_cdna/NM_000152/SC125512/GAA.aspx [Accessed 28/01 2013].

PARKS, W. C. 1999. Matrix metalloproteinases in repair. *Wound Repair and Regeneration,* 7**,** 423-432.

PERKINELMER 2012. AlphaLISA® Assay Development Guide

PERKINELMER. 2014. *Alpha Lisa Immunoassay* [Online]. Available: http://www.perkinelmer.co.uk/resources/technicalresources/applicationsupportknowledgebase/alphalisa-alphascreen-no-washassays/immunoassay-kits.xhtml [Accessed 15/03/2014 2014].

PORATH, J. 1986. Salt-promoted adsorption: recent developments. *Journal of Chromatography,* 376**,** 331-41.

PRIMERBIOSOFT. 2013. *PCR Primer Design Guidelines* [Online]. Available: http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html [Accessed].

PROMEGA. 2013a. *Tm Calculations for Oligos* [Online]. Available: http://www.promega.com/techserv/tools/biomath/calc11.htm [Accessed].

PROMEGA. 2013b. *TSAP Thermosensitive Alkaline Phosphatase* [Online]. Available: http://www.promega.co.uk/products/cloning-and-dna-markers/cloning-tools-and-competent-cells/tsap-thermosensitive-alkaline-phosphatase/ [Accessed 03-06 2013].

QIAGEN. 2013. *QIAquick Gel Extraction Kit* [Online]. Available: http://www.qiagen.com/resources/Download.aspx?id={F4BA2D24-8218-452C-AD6F-1B6F43194425}&lang=en&ver=1 [Accessed 18 April 2013].

RAPPSILBER, J., RYDER, U., LAMOND, A. I. & MANN, M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Research,* 12**,** 1231-45.

RAVINDRANATH, B. 1989. *Principles and Practice of Chromatography*, Ellis Horwood Limited.

REUSER, A. J. J., KOSTER, J. F., KOSTER, J. F., HOOGEVEEN, A. & GALJAARD, H. 1978. Biochemical, Immunological, and Cell Genetic Studies in Glycogenosis Type-Ii. *American Journal of Human Genetics,* 30**,** 132-143.

REUSER, A. J. J., KROOS, M., OUDE ELFERINK, R.P.J., TAGER, J.M. 1985. Defects in synthesis, phosphorylation, and maturation of acid α-glucosidase in glycogenosis type II. *Journal of Biological Chemistry,* 260**,** 8336-8341.

ROBERT, F., BIERAU, H., ROSSI, M., AGUGIARO, D., SORANZO, T., BROLY, H. & MITCHELL-LOGEAN, C. 2009. Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of a CHO cathepsin D protease. *Biotechnology and Bioengineering,* 104**,** 1132-41.

ROSENFELD, E. L. 1975. Alpha-Glucosidases (Gamma-Amylases) in Human and Animal Organisms. *Pathologie Biologie,* 23**,** 71-84.

SANTAMARIA, I., VELASCO, G., PENDAS, A. M., FUEYO, A. & LOPEZ-OTIN, C. 1998. Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. *Journal of Biological Chemistry,* 273**,** 16816-16823.

SATOH, M., HOSOI, S. & SATO, S. 1990. Chinese-Hamster Ovary Cells Continuously Secrete a Cysteine Endopeptidase. *In Vitro Cellular and Developmental Biology,* 26**,** 1101-1104.

SAUER, B. 1994. Site-specific recombination: developments and applications. *Curr Opin Biotechnol,* 5**,** 521-7.

SHINODA, K., TOMITA, M. & ISHIHAMA, Y. 2009. emPAI Calc—for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. *Bioinformatics,* 26**,** 576-577.

SIGMA-ALDRICH. 2014. *SIGMAFAST™ 3,3´-Diaminobenzidine tablets* [Online]. Available: http://www.sigmaaldrich.com/catalog/product/sigma/d4293?lang=en&region=GB [Accessed].

STACEY, G. N. & MASTERS, J. R. 2008. Cryopreservation and banking of mammalian cell lines. *Nature Protocols,* 3**,** 1981-9.

STATISTICS, C. N. B. O. 2012. Biopharmaceutical Sales in China.

TAIT, A. S., HOGWOOD, C. E., SMALES, C. M. & BRACEWELL, D. G. 2012. Host cell protein dynamics in the supernatant of a mAb producing CHO cell line. *Biotechnology and Bioengineering,* 109**,** 971-82.

UNIPROT. 2012. *Lysosomal alpha-glucosidase* [Online]. Available: http://www.uniprot.org/uniprot/P10253 [Accessed 30-01 2013].

VAN HOVE, J. L. K., YANG, H. W., OLIVER, L. M., PENNYBACKER, M. F. & CHEN, Y. T. 1997. Purification of recombinant human precursor acid alpha-glucosidase. *Biochemistry and Molecular Biology International,* 43**,** 613-623.

VAN HOVE, J. L. K., YANG, H.W., WU, J.-., BRADY, R.O. & CHEN, Y.- 1996. High-level production of recombinant human lysosomal acid α-glucosidase in Chinese hamster ovary cells which targets to heart muscle and corrects glycogen accumulation in fibroblasts from patients with Pompe disease. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 65-70.

VANDERHORST, G. T. J., HOEFSLOOT, E. H., KROOS, M. A. & REUSER, A. J. J. 1987. Cell-Free Translation of Human Lysosomal Alpha-Glucosidase - Evidence for Reduced Precursor Synthesis in an Adult Patient with Glycogenosis Type-Ii. *Biochimica et Biophysica Acta,* 910**,** 123-129.

VINCZE, T., POSFAI, J. & ROBERTS, R. J. 2003. NEBcutter: a program to cleave DNA with restriction enzymes *Nucleic Acids Res,* 31**,** 3688-3691.

WANG, F., RICHARDSON, D. & SHAMEEM, M. 2015. Host-Cell Protein Measurement and Control. *BioPharm International* 28**,** 32-38.

WHO. 2012. Coming together to combat rare diseases. *Bulletin of the World Health Organization* [Online], 90. Available: http://www.who.int/bulletin/volumes/90/6/12-020612/en/index.html [Accessed 22-01-2012].

WOLFFE, A. P. 1998. When more is less. *Nature Genetics,* 18**,** 5-6.

WU, J. Y., VANHOVE, J. L. K., HUANG, Y. S. & CHEN, Y. T. 1996. Expression of catalytically active human multifunctional glycogen-debranching enzyme and lysosomal acid alpha-glucosidase in insect cells. *Biochemistry and Molecular Biology International,* 39**,** 755-764.

XIAOHUI LU, C.-F. X., SULI LIU, EMILY MENESALE, LI ZANG, DONG XU, ZORAN SOSIC AND SVETLANA BERGELSON 2015. Identification and Quantification of Heat-Shock Protein 70: A Major Host-Cell Protein Contaminant from HEK Host Cells. *BioProcess Technical***,** 46-58.

YOSHIMOTO, N., KIDA, A., JIE, X., KUROKAWA, M., IIJIMA, M., NIIMI, T., MATURANA, A. D., NIKAIDO, I., UEDA, H. R., TATEMATSU, K., TANIZAWA, K., KONDO, A., FUJII, I. & KURODA, S. 2013. An automated system for high-throughput single cell-based breeding. *Sci Rep,* 3**,** 1191.

YUK, I. H., NISHIHARA, J., WALKER, D., HUANG, E., GUNAWAN, F., SUBRAMANIAN, J., PYNN, A. F. J., YU, X. C., ZHU-SHIMONI, J., VANDERLAAN, M. & KRAWITZ, D. C. 2015. More similar than different: Host cell protein production using three null CHO cell lines. *Biotechnology and Bioengineering,* 112**,** 2068-2083.

ZHOU, H., LIU, Z. G., SUN, Z. W., HUANG, Y. & YU, W. Y. 2010. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. *Journal of Biotechnology,* 147**,** 122-129.

ZHU-SHIMONI, J., YU, C., NISHIHARA, J., WONG, R. M., GUNAWAN, F., LIN, M., KRAWITZ, D., LIU, P., SANDOVAL, W. & VANDERLAAN, M. 2014. Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnology and Bioengineering,* 111**,** 2367-79.

APPENDICES

Description	Accession ID	Confidence score	Anova (p)	pI	MW	GRAVY	Concentration (mg/L)				Percent total (%)			
							IEX			HIC	IEX			HIC
							GAA	Null	mAb	GAA	GAA	Null	mAb	GAA
14-3-3 protein theta	Q3SZI4	9.3	4.655E-08	4.68	27764.3	-0.512	0.006	0.059	0.000		0.01	0.03	0.00	
28 kDa heat- and acid-stable phosphoprotein	Q13442	17.2	4.482E-09	8.84	20630.0	-1.612	0.005	0.111	0.003		0.00	0.05	0.02	
28 kDa heat- and acid-stable phosphoprotein	Q3UHX2	17.7	6.999E-09	6.8	20604.9	-1.606	0.000	0.013	0.000		0.00	0.01	0.00	
30S ribosomal protein S6	Q89MW3	9.8	4.197E-06	5.43	18627.6	-0.987	0.141	0.162	0.020		0.15	0.07	0.11	
40S ribosomal protein S6	P47838	11.3	8.44E-08	10.83	28654.5	-0.951	0.240	3.692	0.052		0.25	1.66	0.29	
45 kDa calcium-binding protein	Q61112	44.5	1.04E-09	4.8	38308.5	-0.765	0.422	3.854	0.073		0.44	1.73	0.42	
4-hydroxy-tetrahydrodipicolinate synthase	A5D2Q5	7.8	1.259E-08	5.53	31588.3	0.084	0.002	0.128	0.001		0.00	0.06	0.00	
78 kDa glucose-regulated protein	Q0VCX2	89.0	9.29E-08	5.01	70464.6	-0.477	0.587	2.715	0.045	0.549	0.61	1.22	0.25	2.13
Acetate kinase	Q6LLD4	4.8	0.0470184	5.72	42875.9	0.026	0.191	0.004	0.104		0.20	0.00	0.59	
Acidic phospholipase	P81243	14.8	0.0824282	4.58	14304.0	-0.652	0.000	0.001	0.000		0.00	0.00	0.00	
Actin	P53689	25.0	1.591E-09	5.38	41622.6	-0.177	7.386	0.099	0.000	0.150	7.66	0.04	0.00	0.58
Actin, gamma	Q9UVW9	47.7	2.649E-07	5.45	41607.5	-0.173	0.236	1.373	0.059		0.24	0.62	0.33	
Actin-10	Q54GX7	44.4	7.769E-05	5.24	41746.7	-0.204	0.273	0.353	0.005		0.28	0.16	0.03	
Actin-17	Q554S6	20.8	2.88E-09	5.28	41571.4	-0.209	0.311	1.475	0.056		0.32	0.66	0.31	
Actin-2	Q9Y707	34.1	1.712E-08	5.31	41663.4	-0.225	0.063	0.469	0.019		0.07	0.21	0.11	
Adapter molecule crk	P10365	42.4	5.051E-10	5.39	41635.4	-0.206	0.034	0.674	0.000		0.04	0.30	0.00	
Adapter molecule crk	Q04929	15.4	1.305E-08	5.36	33805.6	-0.658	0.005	0.074	0.000		0.01	0.03	0.00	
Angiopoietin-related protein 4	Q9Z1P8	24.0	4.522E-09	8.21	43423.8	-0.733	0.093	0.785	0.000		0.10	0.35	0.00	
Annexin A5	P08758	44.9	1.806E-06	4.93	35805.6	-0.33	0.181	0.314	0.027		0.19	0.14	0.15	
Astrocytic phosphoprotein PEA-15	Q9Z297	6.2	1.073E-08	4.93	15040.1	-0.644	0.003	0.132	0.000		0.00	0.06	0.00	
B box-binding protein	Q00436	11.4	0.0044479	9.35	34484.6	-1.514	0.000	0.005	0.000		0.00	0.00	0.00	
Beta-1,4-glucuronyltransferase 1	Q8BWP8	30.6	1.078E-07	6.84	47384.3	-0.266	0.057	0.918	0.021		0.06	0.41	0.12	
Beta-1,4-glucuronyltransferase 1	O43505	30.0	1.223E-07	6.77	47119.1	-0.224	0.011	0.174	0.004	0.078	0.01	0.08	0.02	0.30
Beta-actin-like protein 2	Q562R1	11.4	3.103E-05	5.39	42003.2	-0.19	0.022	0.004	0.000		0.02	0.00	0.00	

Bifunctional protease/dUTPase	P31625	5.0	9.515E-08	8.9	31311.1	-0.094	0.001	0.098	0.000		0.00	0.04	0.00	
Bifunctional purine biosynthesis protein PurH	Q87VR9	21.6	1.286E-06	5.81	57463.4	-0.049	0.008	0.359	0.001		0.01	0.16	0.01	
Bleomycin hydrolase	P70645	5.5	4.055E-09	6.04	52322.9	-0.389	0.032	0.319	0.001		0.03	0.14	0.01	
Bone morphogenetic protein 1	P13497	50.3	2.308E-07	6.26	98049.8	-0.597	0.192	0.884	0.109		0.20	0.40	0.62	
cAMP-dependent protein kinase catalytic subunit gamma	P22612	5.5	5.278E-08	8.7	40303.3	-0.339	0.475	4.676	0.016		0.49	2.10	0.09	
Cathepsin B	P10605	41.4	3.205E-08	5.23	27577.7	-0.306	0.596	1.849	0.035	0.025	0.62	0.83	0.20	0.09
Cathepsin Z	Q9R1T3	18.8	3.976E-07	5.51	27106.2	-0.511	0.099	0.410	0.009	0.015	0.10	0.18	0.05	0.06
Cell division control protein 42 homolog	Q2KJ93	15.2	4.434E-06	6.16	20933.2	-0.222	0.001	0.107	0.000		0.00	0.05	0.00	
Chaperone protein DnaJ	Q9RUG2	10.4	0.0007799	5.62	40238.4	-0.517	0.099	0.796	0.046		0.10	0.36	0.26	
Chaperone protein DnaK	P71331	9.8	1.309E-08	4.88	68262.9	-0.432	0.062	0.272	0.043		0.06	0.12	0.25	
Chaperone protein DnaK	Q7UM31	19.2	1.103E-06	4.62	69422.7	-0.414	0.047	0.843	0.017		0.05	0.38	0.09	
Chaperone protein dnaK2	Q7U3C4	33.0	2.103E-07	4.71	67888.3	-0.314	0.125	2.433	0.048		0.13	1.09	0.27	
Chemotaxis response regulator protein-glutamate methylesterase 1	Q221I1	5.2	0.0003197	6.62	39828.4	0.182	0.033	0.675	0.011		0.03	0.30	0.06	
Chromosomal replication initiator protein DnaA	Q1I2G4	23.6	2.809E-09	6.73	57096.2	-0.307	0.008	0.283	0.000		0.01	0.13	0.00	
Clusterin	P05371	96.5	2.741E-06	5.45	49076.4	-0.534	3.451	7.611	0.180		3.58	3.42	1.02	
Clusterin	Q06890	94.4	7.644E-06	5.45	49344.6	-0.554	0.707	1.736	0.032		0.73	0.78	0.18	
Clusterin (Fragment)	P14683	27.8	2.671E-05	5.68	22098.9	-0.639	0.924	2.439	0.026		0.96	1.10	0.15	
Coiled-coil domain-containing protein 6	Q16204	12.8	5.89E-10	6.95	53159.7	-0.981	0.006	0.925	0.002		0.01	0.42	0.01	
Collagen alpha-1(VI) chain	Q04857	24.5	2.854E-06	5.15	106386.8	-0.541	0.193	0.874	0.020		0.20	0.39	0.12	
Complement C3	P01026	55.3	9.63E-08	6.12	184110.5	-0.275	1.120	5.623	0.199		1.16	2.53	1.13	
Complement C3	P01025	39.7	3.187E-07	6.06	184601.1	-0.313	0.511	1.395	0.110		0.53	0.63	0.62	
Complement C3	Q2UVX4	44.6	2.495E-11	6.37	185047.4	-0.369	0.014	0.281	0.002		0.01	0.13	0.01	
Complement C3	P12387	38.2	1.812E-07	6.47	184441.5	-0.253	3.404	13.518	0.562		3.53	6.07	3.18	
Complement C3	P01027	62.5	5.104E-07	6.3	184180.1	-0.282	0.646	4.077	0.196		0.67	1.83	1.11	
Cytochrome b5	P00171	13.7	9.103E-09	4.94	15197.9	-0.516	0.015	0.470	0.001		0.02	0.21	0.01	
Cytochrome b6-f complex subunit 8	Q06H04	5.8	1.692E-06	5.59	3197.9	1.628	0.000	0.000	0.000		0.00	0.00	0.00	

Cytochrome c2	P86323	5.6	2.112E-09	9.3	11587.1	-0.642	0.033	0.266	0.012		0.03	0.12	0.07	
D-alanine aminotransferase	Q5HF24	16.4	7.153E-05	4.95	31893.9	-0.318	0.002	0.086	0.003		0.00	0.04	0.02	
Dermcidin	P81605	11.5	2.806E-05	5.64	9259.3	-0.667	0.021	0.219	0.022	0.012	0.02	0.10	0.13	0.04
Dihydropyrimidinase-related protein 2	O02675	18.8	5.392E-05	5.95	62277.6	-0.267	0.030	0.093	0.002		0.03	0.04	0.01	
Dipeptidyl peptidase 3	Q99KK7	83.4	4.301E-06	5.26	82766.5	-0.32	0.668	2.568	0.122		0.69	1.15	0.69	
Dolichyl-diphosphooligosaccharide--protein glycotransferas	P12244	15.7	1.604E-08	4.76	55052.9	-0.505	0.019	0.138	0.001		0.02	0.06	0.00	
Doublesex- and mab-3-related transcription factor	Q3LH63	7.9	0.0021919	8.06	36999.2	-0.765	0.017	0.000	0.000		0.02	0.00	0.00	
Drebrin	Q07266	9.3	7.445E-10	4.45	77340.8	-0.782	2.896	0.814	0.002	0.929	3.00	0.37	0.01	3.56
Dystroglycan	Q62165	107.6	5.09E-08	9.41	67629.1	-0.376	1.021	6.132	0.181		1.06	2.75	1.03	
Dystroglycan	Q29243	66.1	5.891E-08	9.11	65832.4	-0.345	0.035	0.296	0.010		0.04	0.13	0.05	
Dystroglycan	Q9TSZ6	65.2	4.8E-05	9.36	67762.3	-0.404	0.005	0.074	0.002		0.00	0.03	0.01	
Dystroglycan	O18738	83.9	2.052E-07	9.33	67773.2	-0.427	0.007	0.178	0.003		0.01	0.08	0.02	
Echinoderm microtubule-associated protein-like 2	Q7TNG5	28.7	1.462E-08	5.83	70733.7	-0.111	0.033	0.603	0.001		0.03	0.27	0.01	
EGF-containing fibulin-like extracellular matrix protein 1	Q8BPB5	142.6	4.185E-10	4.96	53047.4	-0.355	1.651	6.973	0.128		1.71	3.13	0.72	
EGF-containing fibulin-like extracellular matrix protein 1	Q12805	128.4	1.629E-09	4.85	52765.1	-0.346	0.363	0.038	0.025		0.38	0.02	0.14	
Elongation factor 1-alpha 1	Q90835	22.7	1.251E-05	9.1	50156.9	-0.267	0.041	0.165	0.046		0.04	0.07	0.26	
Elongation factor 2	Q3SYU2	68.9	6.237E-05	6.42	95237.0	-0.212	2.157	4.832	2.468	0.542	2.24	2.17	13.97	2.08
Endoplasmic reticulum vesicle protein 25	Q5B5L5	4.4	3.908E-07	6.14	21702.7	-0.369	0.003	0.000	0.000	0.028	0.00	0.00	0.00	0.11
Endoribonuclease YbeY	Q161G7	4.7	8.222E-09	4.26	17798.3	0.284	0.021	0.000	0.000		0.02	0.00	0.00	
Enolase (Fragment)	P42897	21.6	4.349E-10	5.43	42884.0	-0.192	0.006	0.192	0.000		0.01	0.09	0.00	
Exosome complex component Rrp41	A0RXU1	5.6	8.174E-05	5.04	26319.0	-0.17	0.036	0.413	0.014		0.04	0.19	0.08	
F-actin-capping protein subunit alpha-1	P47753	9.1	8.968E-08	5.34	32808.6	-0.657	0.013	0.088	0.001		0.01	0.04	0.01	
F-box DNA helicase 1	Q8NFZ0	27.9	0.0002112	8.58	117686.3	-0.272	1.694	4.889	0.190		1.76	2.20	1.08	
Ferrochelatase	A9BEE9	14.6	7.699E-06	5.05	44606.0	-0.186	0.615	0.038	0.457	0.188	0.64	0.02	2.59	0.72
Follistatin-related protein 1	Q58D84	74.9	3.903E-09	5.21	32593.5	-0.609	0.807	2.899	0.202		0.84	1.30	1.14	

G kinase-anchoring protein 1	Q5VSY0	5.1	8.707E-12	8.89	42078.1	-1.201	0.000	0.600	0.000		0.00	0.27	0.00	
Gametocyte-specific factor 1-like	Q9CWD0	5.3	4.423E-05	8.12	17281.7	-0.767	0.089	0.912	0.023		0.09	0.41	0.13	
Glia maturation factor beta	P60984	5.4	1.309E-07	5.19	16582.0	-0.546	0.011	0.080	0.000		0.01	0.04	0.00	
Glyceraldehyde-3-phosphate dehydrogenase	P17244	54.7	7.054E-08	8.5	35616.7	-0.084	0.071	0.335	0.001		0.07	0.15	0.01	
Glyceraldehyde-3-phosphate dehydrogenase 1	O16027	12.1	1.467E-08	6.71	34579.6	-0.055	0.046	0.381	0.023		0.05	0.17	0.13	
GTP-binding protein RHO1	Q9HF54	16.0	1.502E-07	7.5	22734.2	-0.271	0.004	0.028	0.001	0.024	0.00	0.01	0.00	0.09
HD domain-containing protein 2	Q0P565	10.1	2.828E-07	5.08	23192.2	-0.573	0.004	0.110	0.000		0.00	0.05	0.00	
Heat shock 70 kDa	Q2TBX4	7.6	8.88E-09	5.23	49729.8	-0.213	0.084	0.669	0.130		0.09	0.30	0.74	
Heat shock 70 kDa protein	Q05944	43.7	0.0008125	5.65	71467.8	-0.406	0.134	0.923	0.003		0.14	0.41	0.02	
Heat shock 70 kDa protein	P16019	27.9	4.848E-08	5.31	71007.1	-0.464	0.020	1.225	0.003		0.02	0.55	0.02	
Heat shock 70 kDa protein	P26791	27.8	9.798E-06	5.14	72051.8	-0.394	0.001	0.041	0.027		0.00	0.02	0.16	
Heat shock 70 kDa protein	Q91233	45.8	0.0971841	5.41	70976.1	-0.485	0.000	0.013	0.000		0.00	0.01	0.00	
Heat shock 70 kDa protein 13	Q5R8D9	23.9	3.153E-06	5.52	49651.6	-0.221	0.078	1.311	0.002		0.08	0.59	0.01	
Heat shock 70 kDa protein 1A	Q27975	30.3	8.365E-07	5.68	70127.3	-0.402	0.049	0.107	0.000		0.05	0.05	0.00	
Heat shock 70 kDa protein 1A	Q61696	54.0	7.128E-07	5.52	69948.0	-0.379	0.024	0.229	0.001		0.02	0.10	0.00	
Heat shock 70 kDa protein 4	P11145	31.5	2.2E-10	5.24	71436.5	-0.331	0.152	2.300	0.025		0.16	1.03	0.14	
Heat shock 70 kDa protein 6	Q9N1U2	21.5	0.2963867	6.06	71156.3	-0.449	0.007	0.084	0.014		0.01	0.04	0.08	
Heat shock 70 kDa protein cognate 1	P29843	53.0	1.535E-07	5.33	70686.4	-0.462	0.006	0.052	0.000		0.01	0.02	0.00	
Heat shock 70 kDa protein IV	Q06248	39.5	0.0003235	5.55	69749.8	-0.428	0.040	0.562	0.070		0.04	0.25	0.40	
Heat shock cognate 70 kDa protein	Q24789	60.4	2.721E-09	5.66	72554.1	-0.451	0.007	0.076	0.001		0.01	0.03	0.00	
Heat shock cognate 70 kDa protein 1	P36415	43.2	1.576E-09	5.34	70414.6	-0.489	0.032	0.140	0.002		0.03	0.06	0.01	
Heat shock cognate 70 kDa protein 3	Q54BE0	54.5	1.163E-05	5.49	69947.0	-0.461	0.654	1.857	0.018	0.186	0.68	0.83	0.10	0.71
Heat shock cognate 71 kDa protein	P19378	215.5	9.552E-07	5.23	70673.7	-0.448	0.146	3.922	0.021		0.15	1.76	0.12	
Heat shock cognate 71 kDa protein	P19120	229.4	1.941E-07	5.37	71109.3	-0.454	0.076	0.782	0.011		0.08	0.35	0.06	
Heat shock cognate 71 kDa protein	O73885	165.6	8.95E-09	5.46	70827.0	-0.449	0.015	0.398	0.001		0.02	0.18	0.01	
Heterogeneous nuclear ribonucleoprotein A/B	Q99020	23.7	7.245E-07	7.69	30831.3	-0.893	0.001	0.297	0.000		0.00	0.13	0.00	
Heterogeneous nuclear ribonucleoprotein D0	Q14103	5.8	1.427E-09	7.6	38303.0	-0.932	0.007	0.309	0.000		0.01	0.14	0.00	

Heterogeneous nuclear ribonucleoprotein K	Q3T0D0	28.0	1.299E-08	5.14	51019.2	-0.702	0.021	0.450	0.000		0.02	0.20	0.00	
Heterogeneous nuclear ribonucleoprotein Q	O60506	77.2	4.05E-08	8.68	69471.4	-0.881	1.156	3.176	0.226		1.20	1.43	1.28	
Hsp90 co-chaperone Cdc37	Q5EAC6	28.6	5.911E-07	5.09	44580.2	-0.987	0.006	0.334	0.000		0.01	0.15	0.00	
Hsp90 co-chaperone Cdc37	Q63692	39.5	7.95E-07	5.24	44510.4	-0.969	0.002	0.195	0.001		0.00	0.09	0.00	
Ig gamma-4 chain C region	P01861	80.4	0.0001372	6.56	35940.6	-0.423	0.001	0.032	3.104		0.00	0.01	17.58	
Ig kappa chain C region	P01834	32.9	1.686E-10	5.42	11608.9	-0.553	0.001	0.000	0.823		0.00	0.00	4.66	
Importin-5	O00410	94.0	1.47E-09	4.83	123498.7	-0.134	0.754	4.259	0.070		0.78	1.91	0.39	
Insulin-like growth factor-binding protein 4	P47879	35.4	8.08E-09	6.62	25760.3	-0.535	0.037	0.919	0.017		0.04	0.41	0.10	
Isopentenyl-diphosphate delta-isomerase	A8AUV1	7.7	2.478E-10	5.39	37621.8	-0.328	0.001	0.165	0.000		0.00	0.07	0.00	
Kin of IRRE-like protein 1	Q96J84	15.6	1.146E-10	5.55	81714.6	-0.404	0.011	0.139	0.000		0.01	0.06	0.00	
Lamin-B1	P20700	13.6	2.487E-08	5.11	65961.7	-0.811	0.009	0.180	0.000		0.01	0.08	0.00	
Legumain	O89017	31.6	3.451E-05	5.72	34818.2	-0.368	0.433	1.425	0.318		0.45	0.64	1.80	
Lipoprotein lipase	Q06000	21.7	0.0027366	8.22	50298.0	-0.422	0.068	0.282	0.002		0.07	0.13	0.01	
Lysine--tRNA ligase	P37879	51.6	3.691E-09	5.84	67859.0	-0.418	0.038	0.906	0.001		0.04	0.41	0.01	
Lysosomal alpha-glucosidase	P10253	553.1	0.0155703	5.42	105323.7	-0.129	34.791	5.319	0.491	21.317	36.07	2.39	2.78	81.83
Lysosomal protective protein	P16675	50.6	9.753E-07	5.55	51395.2	-0.321	0.800	5.823	0.223		0.83	2.62	1.26	
Lysyl oxidase homolog 1	P97873	21.9	1.963E-08	6.17	56452.1	-0.765	0.003	0.597	0.002		0.00	0.27	0.01	
Matrix metalloproteinase-19	Q9JHI0	11.8	4.96E-08	8.87	46555.8	-0.404	0.108	1.255	0.007		0.11	0.56	0.04	
Metalloproteinase inhibitor	P30120	5.3	1.384E-07	8.27	21536.8	-0.109	0.051	0.185	0.011		0.05	0.08	0.06	
Methionine aminopeptidase 2	P50579	10.5	7.263E-10	5.57	52760.4	-0.687	0.001	0.161	0.000		0.00	0.07	0.00	
Multiple inositol polyphosphate phosphatase 1	Q9Z2L6	5.6	2.332E-07	6.5	51546.9	-0.389	0.006	0.365	0.000		0.01	0.16	0.00	
Myosin light chain 6B	Q8CI43	5.5	7.276E-09	5.41	22617.7	-0.559	0.006	0.213	0.000		0.01	0.10	0.00	
Myotrophin	P62775	20.9	1.934E-07	5.28	12729.6	-0.224	0.012	0.557	0.003		0.01	0.25	0.01	
Na(+)/H(+) exchange regulatory cofactor NHE-RF1	P70441	22.7	8.103E-10	5.63	38468.9	-0.75	0.011	0.247	0.000		0.01	0.11	0.00	
N-acetylglucosamine-6-sulfatase	Q8BFR4	15.2	1.816E-09	7.63	57632.5	-0.411	0.075	0.208	0.001		0.08	0.09	0.01	
NAD-dependent protein deacylase	Q607X6	9.3	3.017E-09	5.98	27494.1	-0.179	0.048	0.422	0.021		0.05	0.19	0.12	
Nascent polypeptide-associated complex subunit	P70670	34.8	1.914E-09	9.39	220499.4	-0.329	0.026	0.543	0.001		0.03	0.24	0.01	

alpha, muscle-specific form														
Neudesin	Q9CQ45	19.3	1.953E-07	4.79	15612.4	-0.674	0.008	0.222	0.000		0.01	0.10	0.00	
Neudesin	Q9UMX5	18.6	4.539E-08	4.81	15684.5	-0.69	0.003	0.077	0.000		0.00	0.03	0.00	
Neudesin	Q6IUR5	18.5	8.449E-10	4.79	15684.5	-0.62	0.002	0.071	0.000		0.00	0.03	0.00	
Neuraminidase	P67923	9.9	1.473E-07	6.82	51381.6	-0.265	2.844	0.123	1.590	0.648	2.95	0.06	9.00	2.49
Nidogen-1	P10493	91.2	8.645E-08	5.24	133562.1	-0.365	0.547	5.698	0.045		0.57	2.56	0.25	
Nidogen-1	P14543	40.3	5.759E-09	5.05	133457.5	-0.378	0.463	3.698	0.013		0.48	1.66	0.08	
Nuclear migration protein nudC	O35685	33.9	0.0016435	5.17	38358.0	-1.061	0.007	0.073	0.000		0.01	0.03	0.00	
Nuclear migration protein nudC	Q63525	35.4	2.075E-05	5.27	38412.1	-1.05	0.151	0.321	0.001		0.16	0.14	0.00	
Nuclear migration protein nudC	Q17QG2	24.2	6.283E-08	5.22	38242.9	-1.028	0.024	0.426	0.000		0.02	0.19	0.00	
Nuclear transport factor 2	Q32KP9	5.1	4.986E-07	5.1	14478.5	-0.201	0.002	0.074	0.000		0.00	0.03	0.00	
Nuclease-sensitive element-binding protein 1	P67808	40.7	1.79E-07	9.87	35793.0	-1.495	0.033	0.727	0.029		0.03	0.33	0.16	
Nucleobindin-1	Q63083	110.6	0.0021546	5.01	50919.5	-1.022	0.727	1.260	0.036	0.136	0.75	0.57	0.20	0.52
Nucleobindin-1	Q02819	85.1	1.475E-09	4.96	50821.4	-1.025	0.099	0.991	0.008		0.10	0.45	0.05	
Nucleobindin-1	Q0P569	94.3	0.00041	5.05	52483.3	-1.007	0.028	0.249	0.000		0.03	0.11	0.00	
Nucleobindin-2	P81117	68.8	0.0001451	5.01	47423.9	-1.085	0.615	1.571	0.036		0.64	0.71	0.21	
Nucleobindin-2	Q9JI85	18.8	2.654E-07	4.96	47253.8	-1.171	0.210	0.901	0.003		0.22	0.40	0.02	
Olfactomedin-like protein 3	Q9NRN5	22.8	8.036E-11	6.18	43800.6	-0.577	0.017	0.385	0.000		0.02	0.17	0.00	
Out at first protein homolog	Q8QZR4	5.2	4.213E-08	6.49	28462.4	-0.257	0.073	0.836	0.000		0.08	0.38	0.00	
Peptide methionine sulfoxide reductase MsrA	A6WYH5	4.6	0.1352595	5.01	24158.1	-0.416	0.073	0.027	0.027		0.08	0.01	0.15	
Peptidyl-prolyl cis-trans isomerase B	P80311	11.4	2.541E-07	9.13	20200.2	-0.454	0.018	0.265	0.002		0.02	0.12	0.01	
Peroxiredoxin-1	Q9JKY1	22.1	3.391E-10	8.21	22131.4	-0.226	0.009	0.081	0.000		0.01	0.04	0.00	
Phosphoglycerate mutase 1	P18669	90.6	7.735E-08	6.75	28672.7	-0.498	0.094	3.083	0.027		0.10	1.38	0.15	
Phosphoglycerate mutase 1	Q3SZ62	81.4	1.174E-07	6.44	28720.8	-0.518	0.057	1.859	0.016		0.06	0.83	0.09	
Phosphoserine aminotransferase	B4SP45	4.4	0.0519085	5.74	38907.2	0.006	0.000	0.005	0.000		0.00	0.00	0.00	
Phosphoserine phosphatase	Q99LS3	17.0	6.447E-08	5.8	25095.9	-0.116	0.003	0.155	0.000		0.00	0.07	0.00	
Pigment epithelium-derived factor	P97298	48.8	5.516E-08	6.45	44284.7	-0.138	1.435	5.130	0.415		1.49	2.30	2.35	

Pigment epithelium-derived factor	P36955	35.6	9.811E-07	5.9	44387.8	-0.164	1.297	0.966	0.184		1.34	0.43	1.04	
Plastin-2	Q6P698	42.8	1.705E-09	5.26	69900.1	-0.294	0.027	0.252	0.000		0.03	0.11	0.00	
Plastin-3	O88818	72.7	7.127E-09	5.47	70795.0	-0.331	0.070	1.525	0.001		0.07	0.68	0.01	
POTE ankyrin domain family member E	Q6S8J3	57.7	1.961E-05	5.83	121363.4	-0.683	0.114	0.025	0.006		0.12	0.01	0.04	
Probable aquaporin PIP-type pTOM75	Q6B411	5.0	8.2E-08	8.31	14962.2	-0.647	0.023	0.071	0.009		0.02	0.03	0.05	
Probable dual-specificity RNA methyltransferase RlmN	A1A1I4	19.9	1.96E-08	6.84	42346.1	-0.367	0.890	0.005	0.051	0.157	0.92	0.00	0.29	0.60
Probable pectinesterase 30	Q3EAY9	14.0	0.0008701	6.71	53060.4	-0.24	0.007	0.269	0.000		0.01	0.12	0.00	
Procollagen C-endopeptidase enhancer 1	Q61398	88.7	1.085E-07	8.63	47630.9	-0.313	1.386	9.429	0.109	0.072	1.44	4.23	0.62	0.28
Procollagen C-endopeptidase enhancer 1	Q15113	25.2	3.573E-05	7.55	45549.5	-0.286	0.620	0.575	0.129	0.080	0.64	0.26	0.73	0.31
Procollagen-lysine,2-oxoglutarate 5-dioxygenase	Q63321	23.9	0.0005039	6.27	81646.8	-0.4	0.045	0.477	0.020		0.05	0.21	0.11	
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	O77588	20.1	2.078E-08	6.17	81468.6	-0.391	0.023	0.488	0.000		0.02	0.22	0.00	
Prolyl endopeptidase	Q9XTA2	53.4	6.122E-08	5.56	80641.4	-0.353	0.096	0.332	0.002		0.10	0.15	0.01	
Protein crossbronx	B4HT57	13.4	0.0517061	5.29	28180.0	-0.466	1.552	0.748	0.024	0.251	1.61	0.34	0.14	0.96
Protein CYR61	O00622	45.5	0.0017795	8.49	39438.4	-0.381	0.975	1.586	0.715	0.211	1.01	0.71	4.05	0.81
Protein disulfide-isomerase	Q8R4U2	74.8	1.078E-09	4.73	54997.9	-0.453	0.104	0.982	0.008		0.11	0.44	0.05	
Protein disulfide-isomerase A3	P38657	46.7	6.409E-08	5.78	54398.4	-0.637	0.099	0.304	0.001		0.10	0.14	0.01	
Protein disulfide-isomerase A3	P86235	124.9	2.691E-09	4.64	23387.9	-0.55	0.070	0.754	0.010		0.07	0.34	0.06	
Protein disulfide-isomerase A3	Q8JG64	67.9	1.585E-10	5.57	53782.7	-0.557	0.021	0.357	0.013		0.02	0.16	0.07	
Protein S100-A13	P79342	6.0	0.0145453	5.5	11198.9	-0.434	0.002	0.029	0.000		0.00	0.01	0.00	
Proteinase K	P06873	12.9	1.012E-08	8.25	28906.8	-0.218	0.014	0.187	0.001		0.01	0.08	0.00	
Purine nucleoside phosphorylase	P23492	23.1	1.864E-08	5.78	32277.1	-0.132	0.009	0.645	0.001		0.01	0.29	0.01	
Putative gene 49 protein	O48403	5.1	0.0597136	4.79	13332.3	-0.043	0.000	0.017	0.000		0.00	0.01	0.00	
Putative NADH dehydrogenase/NAD(P)H nitroreductase Reut_A1586	Q471I1	9.1	3.522E-05	6.73	21531.5	-0.239	0.056	0.003	0.001		0.06	0.00	0.00	
Putative phospholipase B-like 2	Q4QQW8	15.8	1.061E-07	5.69	61943.4	-0.265	0.244	1.197	0.041		0.25	0.54	0.23	
Pyruvate kinase PKM	P11980	32.8	7.584E-06	6.69	57686.6	-0.096	0.048	0.714	0.022		0.05	0.32	0.12	
Pyruvate kinase PKM	P14618	38.4	3.493E-09	7.95	57805.7	-0.132	0.046	0.379	0.007		0.05	0.17	0.04	

Ran-specific GTPase-activating protein	P34022	10.3	3.301E-10	5.15	23465.2	-1.15	0.018	0.096	0.000	0.139	0.02	0.04	0.00	0.54
Ras-like GTP-binding protein O-RHO	P22122	19.1	1.685E-07	5.82	21140.4	-0.359	0.002	0.274	0.001		0.00	0.12	0.00	
Receptor-type tyrosine-protein phosphatase S	B0V2N1	52.0	8.29E-06	6.76	208981.4	-0.39	0.163	1.467	0.002		0.17	0.66	0.01	
Recombination protein RecR	B3QYJ4	5.3	7.166E-07	5.61	22842.4	-0.147	0.057	0.861	0.000		0.06	0.39	0.00	
Recombination protein RecR	Q2RZH8	4.8	1.929E-07	4.19	25489.1	-0.38	0.003	0.227	0.001		0.00	0.10	0.00	
Retinoid-inducible serine carboxypeptidase	Q920A5	43.5	0.012753	5.25	47816.8	-0.079	0.801	1.413	0.247		0.83	0.63	1.40	
Rho GDP-dissociation inhibitor 1	Q5XI73	45.6	6.446E-07	5.1	23276.2	-0.73	0.171	2.293	0.014		0.18	1.03	0.08	
Ribonuclease T2	Q9CQ01	34.1	5.189E-09	5.92	26755.4	-0.589	0.006	0.253	0.009		0.01	0.11	0.05	
Ribonucleoside-diphosphate reductase subunit M2	Q60561	10.4	2.588E-10	5.25	44482.1	-0.179	0.001	0.058	0.000		0.00	0.03	0.00	
Ribose-5-phosphate isomerase	P47968	41.3	6.096E-10	7.81	32450.9	-0.233	0.001	0.137	0.001		0.00	0.06	0.00	
R-phycocyanin-1 beta chain	P37208	4.9	3.131E-08	4.92	18170.6	0.084	0.014	0.075	0.000		0.01	0.03	0.00	
Septin-2	Q2NKY7	20.1	6.628E-09	6.15	41571.5	-0.529	0.004	0.123	0.000		0.00	0.06	0.00	
Serine protease HTRA1	Q92743	16.4	0.0001173	7.89	49048.1	-0.202	0.020	0.051	0.000		0.02	0.02	0.00	
Serine/threonine-protein kinase PAK 1	Q08E52	9.5	1.112E-10	5.63	60414.6	-0.605	0.003	0.097	0.000		0.00	0.04	0.00	
Serum albumin	P02769	131.3	n/a	5.6	66433.0	-0.429	0.276	0.358	0.387	0.308	0.29	0.16	2.19	1.18
SH3 domain-binding glutamic acid-rich-like protein	O75368	10.7	7.369E-07	5.22	12774.3	-0.691	0.002	0.027	0.001		0.00	0.01	0.00	
Small cysteine-rich outer membrane protein	B0B816	5.4	1.452E-08	5.13	7474.4	-0.671	0.006	0.029	0.000		0.01	0.01	0.00	
Small ubiquitin-related modifier 2	Q6LDZ8	7.4	1.241E-06	5.32	10608.9	-0.893	0.126	1.137	0.041		0.13	0.51	0.23	
Soluble calcium-activated nucleotidase 1	Q8VCF1	30.4	1.62E-08	6.28	45652.5	-0.446	0.003	0.085	0.001		0.00	0.04	0.01	
Soluble calcium-activated nucleotidase 1	Q8WVQ1	43.7	3.765E-08	5.72	44839.6	-0.339	0.029	0.944	0.028		0.03	0.42	0.16	
Spermidine/putrescine import ATP-binding protein	A1TXH7	5.3	1.124E-07	4.92	42106.8	-0.398	0.255	1.027	0.018		0.26	0.46	0.10	
Superoxide dismutase [Cu-Zn] 1	O42724	4.6	0.00086	5.82	15757.2	-0.444	0.000	0.008	0.000		0.00	0.00	0.00	
Suprabasin	A6QQF6	10.1	1.397E-07	6.74	54505.4	-0.745	0.059	0.365	0.034		0.06	0.16	0.19	
Synaptic vesicle membrane protein VAT-1	Q99536	5.8	2.518E-09	5.88	41789.1	-0.043	0.051	0.231	0.000		0.05	0.10	0.00	
Taurocyamine kinase	P16641	17.5	6.933E-08	8.2	83891.0	-0.366	0.017	0.178	0.146		0.02	0.08	0.83	
Tetranectin	P05452	28.5	2.312E-09	5.8	20138.9	-0.484	0.034	0.426	0.000		0.04	0.19	0.00	

Thioredoxin	P10639	12.7	1.935E-08	4.8	11544.3	-0.034	0.072	1.438	0.015		0.07	0.65	0.08	
Thioredoxin reductase 1, cytoplasmic	O62768	17.5	4.709E-06	6.07	54770.5	-0.216	0.028	0.891	0.095		0.03	0.40	0.54	
Thioredoxin reductase 1, cytoplasmic	Q9JMH6	44.3	0.0003475	7.42	67083.8	-0.233	0.042	1.087	0.138		0.04	0.49	0.78	
TIP41-like protein	O75663	14.6	1.011E-08	5.59	31444.0	-0.401	0.001	0.040	0.000		0.00	0.02	0.00	
Tolloid-like protein 2	Q9Y6L7	34.2	4.124E-11	5.8	97638.1	-0.55	0.075	0.584	0.000		0.08	0.26	0.00	
Transcription initiation factor TFIID subunit 3	Q9P6P0	19.0	1.134E-06	4.7	17681.0	-0.32	0.073	0.394	0.004		0.08	0.18	0.02	
Transmembrane protein 132A	Q922P8	15.6	1.144E-07	5.37	106794.2	-0.262	0.112	0.808	0.007		0.12	0.36	0.04	
Tripeptidyl-peptidase 1	O89023	11.9	0.0003276	5.93	39877.6	-0.173	0.082	0.550	0.016		0.09	0.25	0.09	
tRNA 2-thiocytidine biosynthesis protein TtcA	A3PNA9	15.3	0.0002483	6.26	33150.2	-0.298	0.035	0.003	0.019		0.04	0.00	0.11	
Trypsin-1	P16049	9.4	6.737E-08	6.36	23835.9	-0.073	0.002	0.067	0.000		0.00	0.03	0.00	
Tryptophan--tRNA ligase, cytoplasmic	Q5R4J1	30.9	9E-10	5.99	53241.6	-0.374	0.010	0.312	0.000		0.01	0.14	0.00	
Tryptophan--tRNA ligase, cytoplasmic	P32921	31.5	6.973E-09	6.44	54357.9	-0.37	0.000	0.098	0.000		0.00	0.04	0.00	
Tubulin alpha-1 chain	O22347	6.0	5.552E-08	4.89	49731.1	-0.191	0.002	0.183	0.001		0.00	0.08	0.00	
Twisted gastrulation protein homolog 1	Q98T89	24.7	4.393E-06	4.97	22179.3	-0.067	0.089	0.707	0.056		0.09	0.32	0.32	
U2 small nuclear ribonucleoprotein A'	P09661	20.7	2.356E-07	8.72	28284.4	-0.496	0.310	1.203	0.103		0.32	0.54	0.58	
U2 small nuclear ribonucleoprotein A'	P57784	21.7	4.623E-07	8.72	28226.3	-0.482	0.000	0.022	0.000		0.00	0.01	0.00	
Ubiquitin-conjugating enzyme E2 K	P61085	17.0	2.253E-09	5.33	22275.5	-0.28	0.003	0.288	0.000		0.00	0.13	0.00	
Uncharacterized protein in bps2 5'region (Fragment)	P55031	27.3	0.2039077	4.92	51088.1	-0.322	2.342	0.872	0.873		2.43	0.39	4.94	
UPF0134 protein MPN_139	P75259	5.4	5.315E-09	9.15	19115.8	-0.855	0.042	0.361	0.000		0.04	0.16	0.00	
UV excision repair protein RAD23 homolog B	P54727	18.2	5.843E-07	4.77	43171.2	-0.388	0.010	0.185	0.002		0.01	0.08	0.01	
Vitamin K-dependent protein S	P53813	5.5	3.934E-11	4.99	69874.2	-0.285	0.117	0.623	0.000		0.12	0.28	0.00	
V-type proton ATPase subunit S1	Q9R1Q9	22.1	0.0003401	5.25	47737.6	0.149	0.050	0.429	0.043		0.05	0.19	0.24	
TOTAL (ng per 10µl sample)							96.5	222.7	17.7	26.1	100	100	100	100

Supporting data: HCP Full list comparison between GAA, Null and mAb CHO producer. *Progenesis QIP software provided resulting data in the form of normalised abundance (molar equivalent) in triplicate readings of each species matched in the SwissProt database and the average of each triplicate was converted into concentration by dividing it by the specie molecular weight. Respective specie percentages were calculated based on relative amount of each specie out of the sum of*

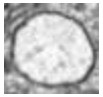
amounts of all species in each 10 µl sample and were used as bubble graph data set (figure 6). Accession ID, confidence score and Anova (p) were provided by Progenesis QIP software. Isoelectric point was calculated using ExPASy pI calculator online tool (ExPASy. Compute pI/Mw tool). Hydropathicity index (GRAVY) was also calculated based on amino acidic sequence.

Supporting data. TEM images of GAA producer CHO Clone 1

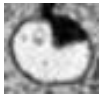
Sample name	Empty lysosomes (E)	Half lysosomes (H)	Full lysosomes (F)	Total lysosomes (T)	%F	%H+F
GAA 1	9	9	12	30	40	70
GAA 2	2	0	12	14	86	86
GAA 3	1	0	27	28	96	96
GAA 4	1	1	0	2	0	50
GAA 5	3	3	4	10	40	70
GAA 6	2	2	8	12	67	83
GAA 7	0	1	23	24	96	100
GAA 8	4	4	8	16	50	75
GAA 9	2	1	24	27	89	93
GAA 10	4	2	20	26	77	85
GAA 11	0	0	6	6	100	100
GAA 12	2	4	12	18	67	89
GAA 13	1	0	23	24	96	96
GAA 14	5	1	4	10	40	50
GAA 15	2	1	8	11	73	82
GAA 16	1	0	26	27	96	96
GAA 17	0	0	3	3	100	100
GAA 18	0	3	0	3	0	100
GAA 19	1	3	8	12	67	92
GAA 20	0	0	11	11	100	100
GAA 21	5	8	12	25	48	80
GAA 22	1	1	11	13	85	92
GAA 23	0	0	16	16	100	100
GAA 24	0	0	11	11	100	100
GAA 25	1	0	18	19	95	95
GAA 26	1	4	2	7	29	86
GAA 27	2	1	7	10	70	80
GAA 28	2	3	36	41	88	95
GAA 29	5	5	9	19	47	74
GAA 30	1	1	19	21	90	95
GAA 31	0	0	28	28	100	100
GAA 32	2	0	9	11	82	82

	Total lysosomes (T)	%F	%H+F	%E
Average	17	72	87	12
StDev	9.1	28.5	13.3	13.3
StError	1.61	5.04	2.36	2.36

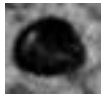
Representative lysosomes



Empty

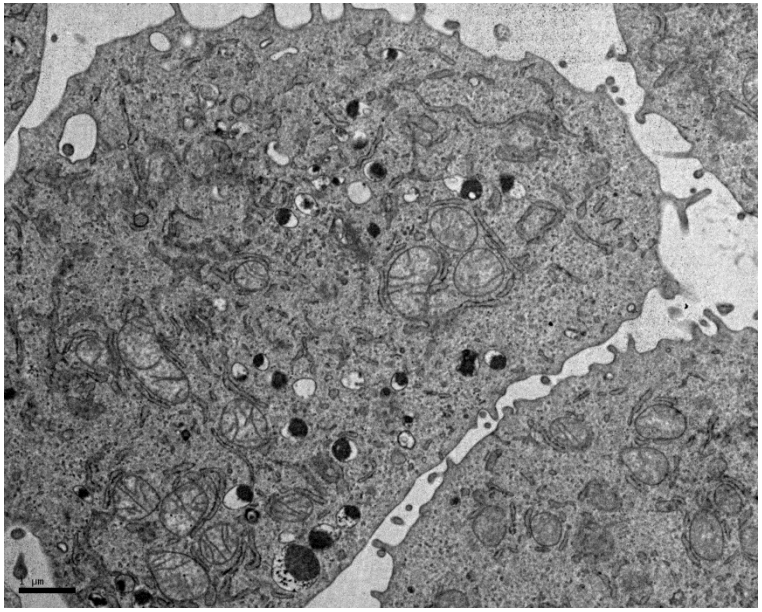


Half

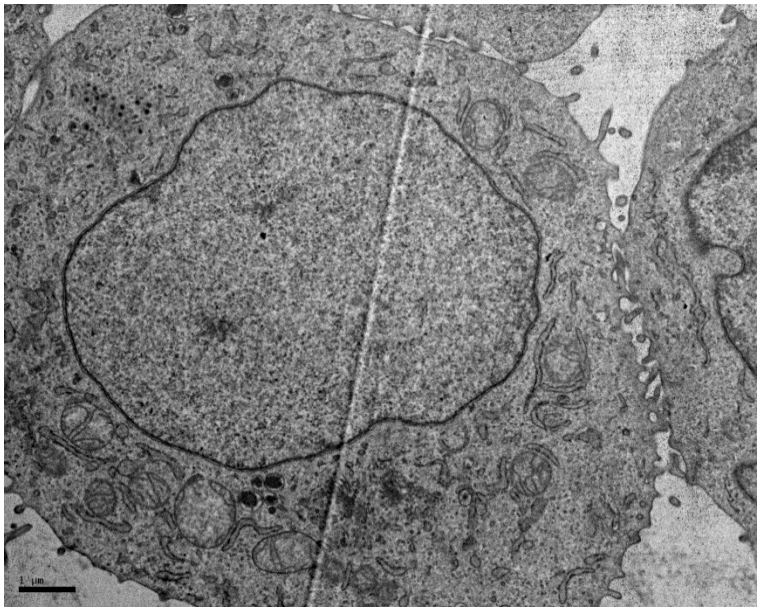


Full

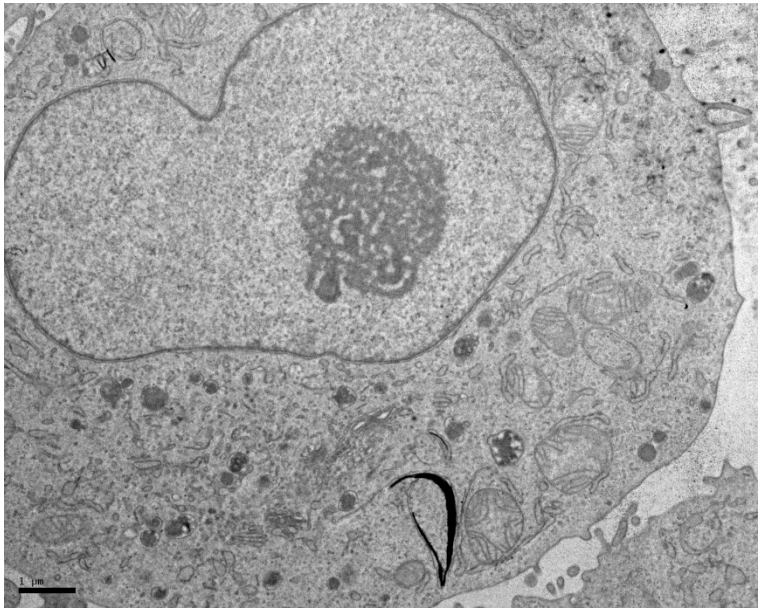
GAA 1



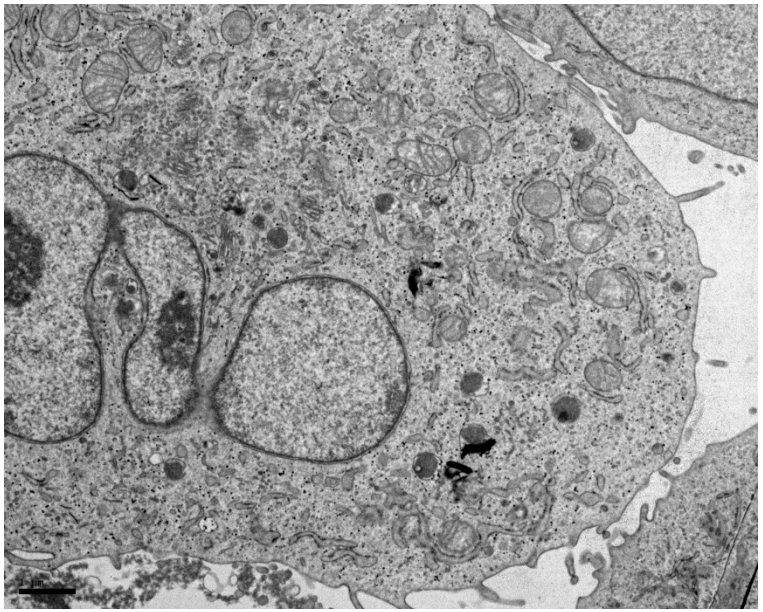
GAA2



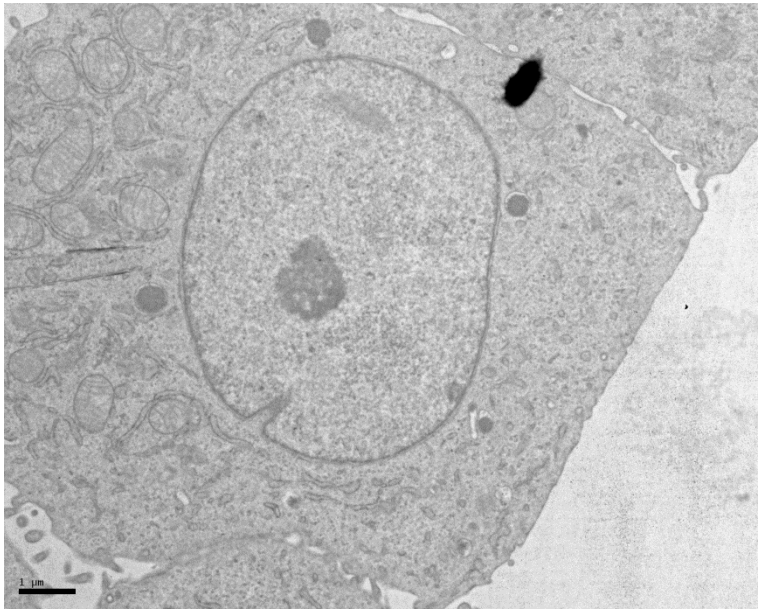
GAA 3



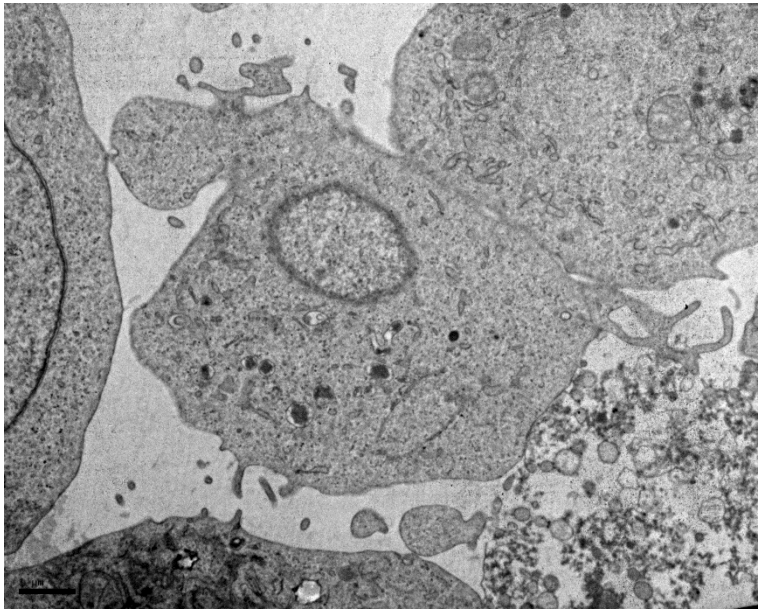
GAA 4



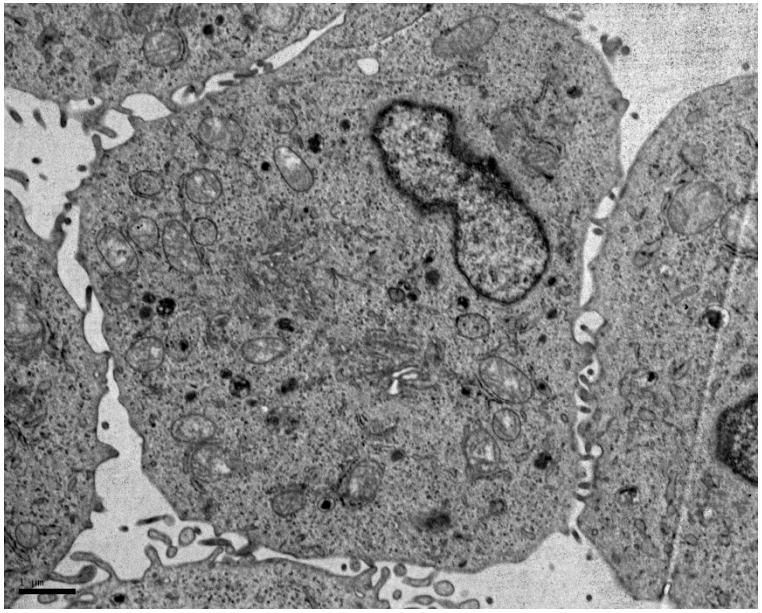
GAA 5



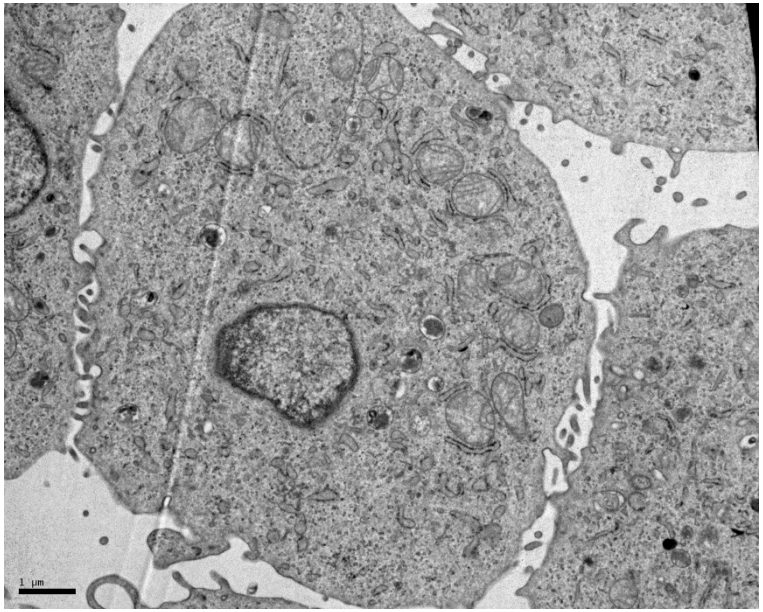
GAA 6



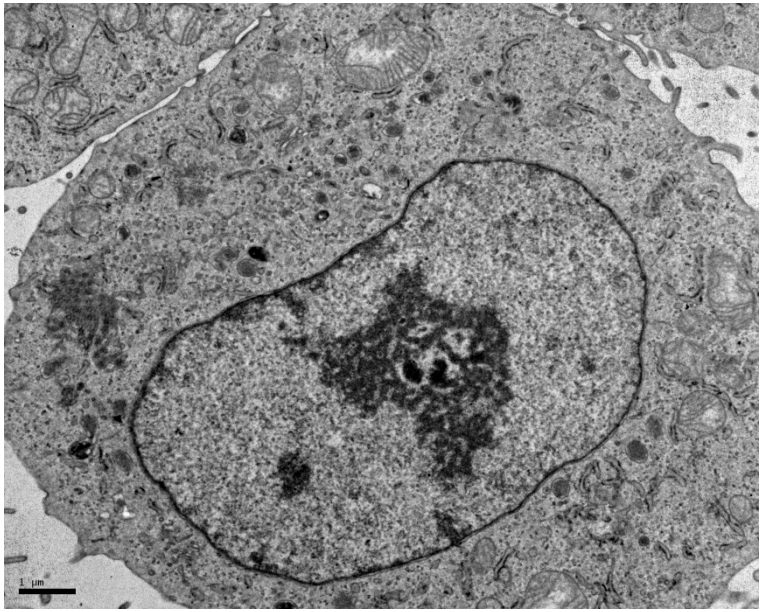
GAA 7



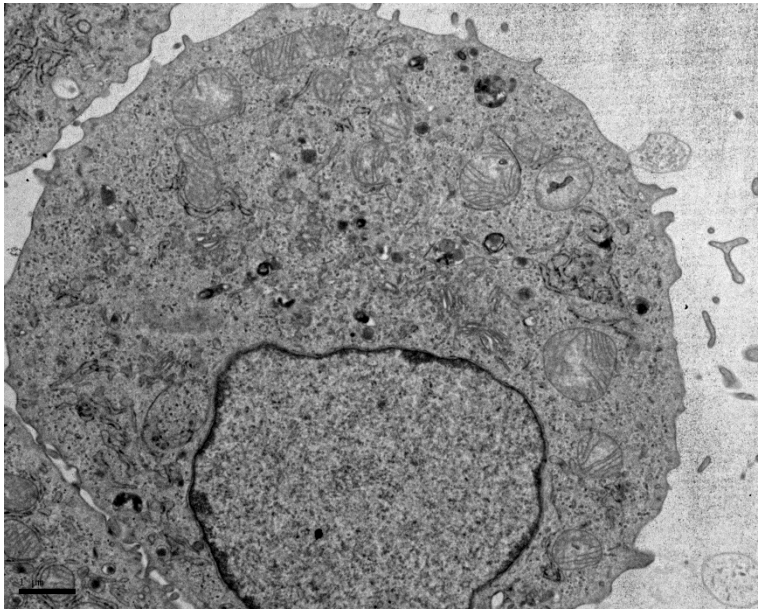
GAA 8



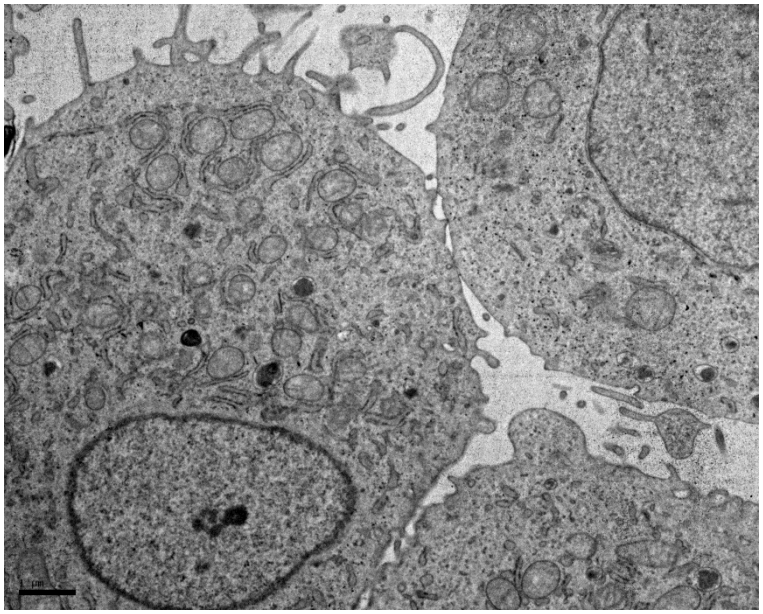
GAA 9



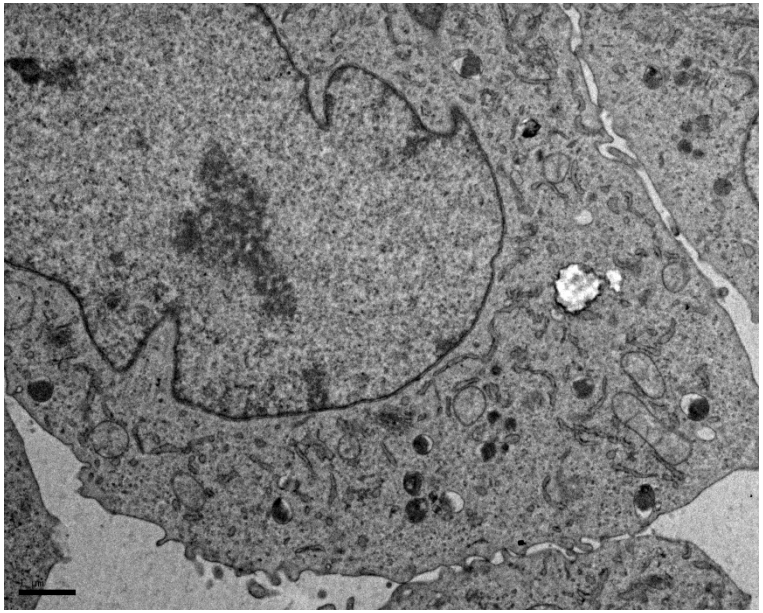
GAA 10



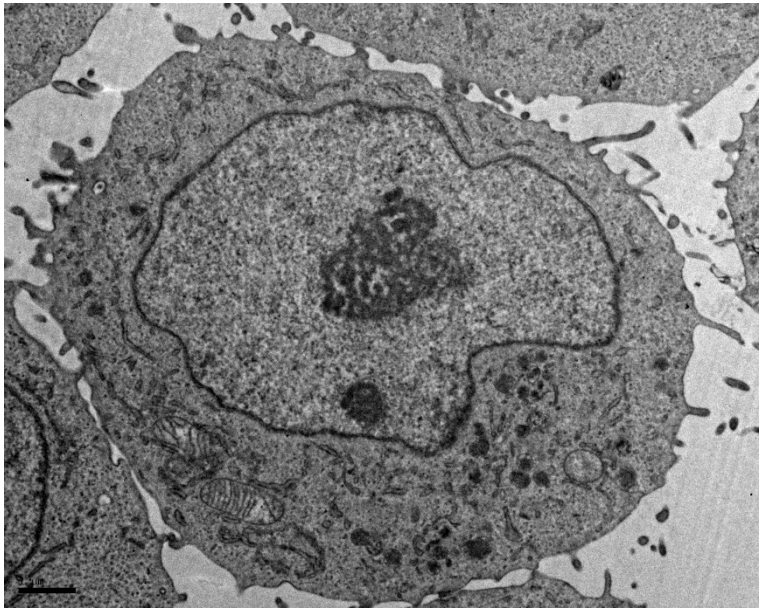
GAA 11



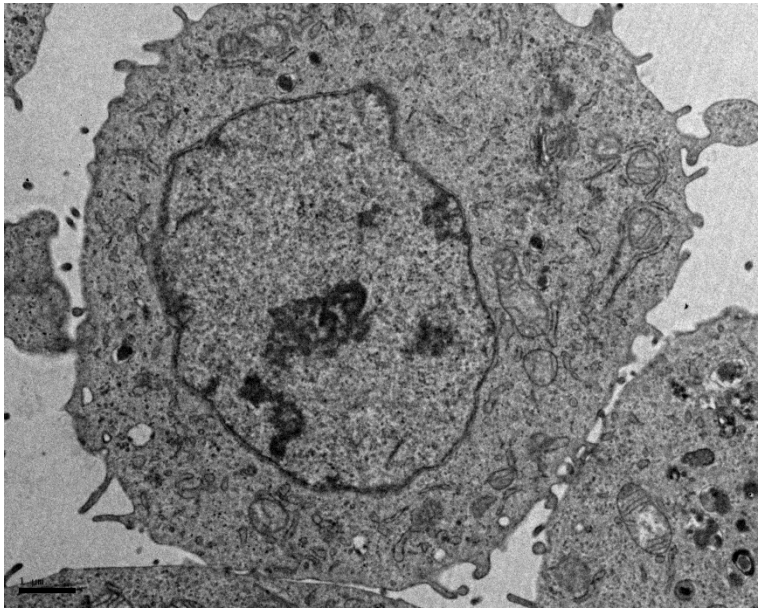
GAA 12



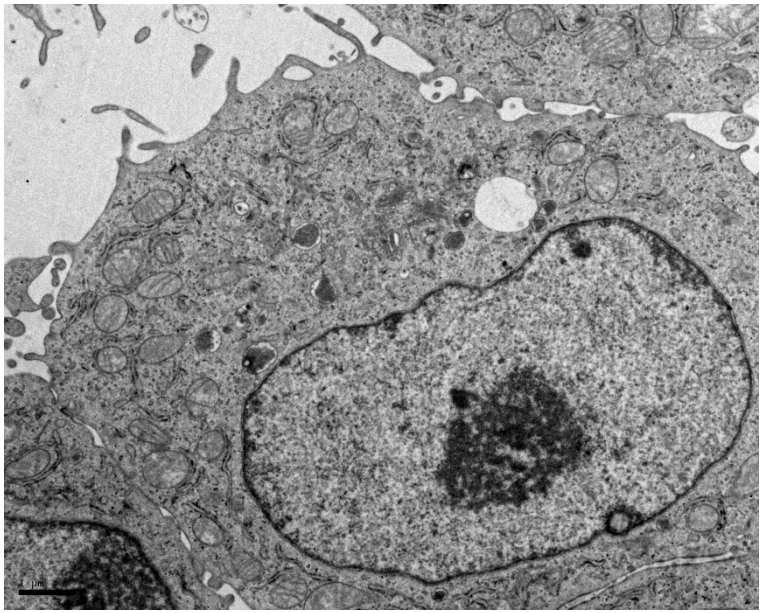
GAA 13



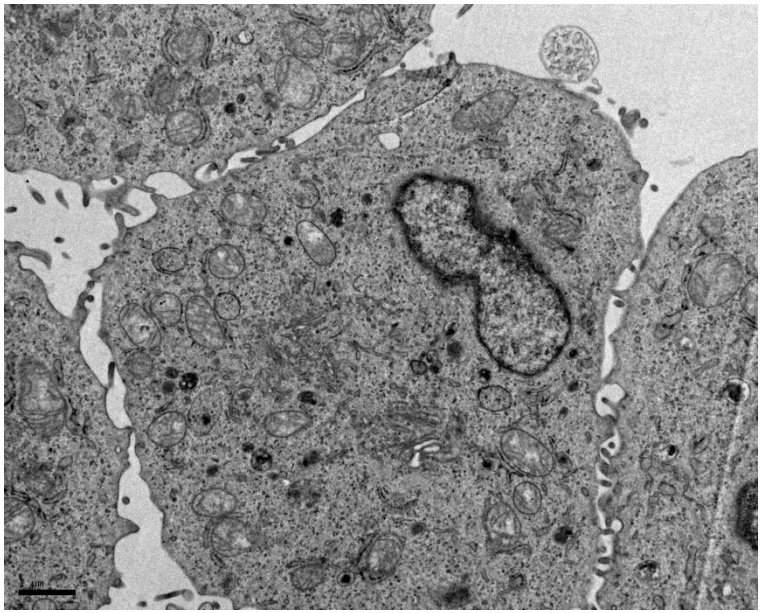
GAA 14



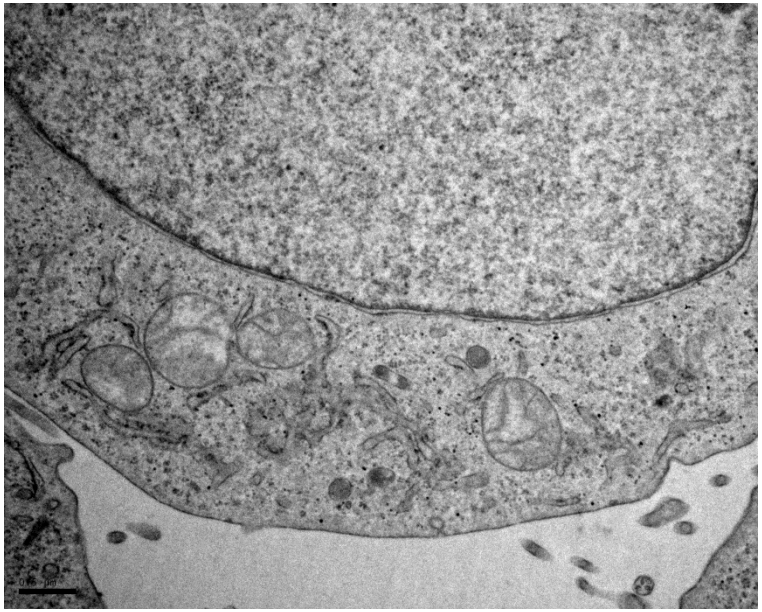
GAA 15



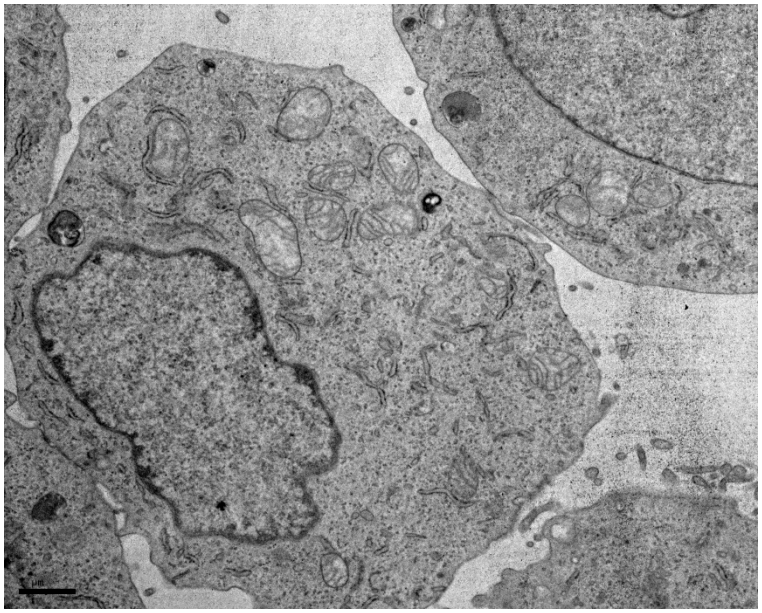
GAA 16



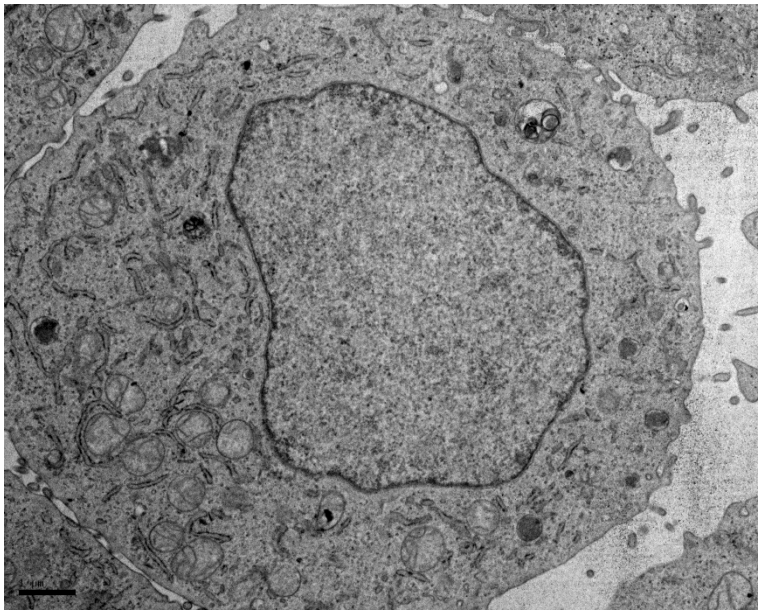
GAA 17



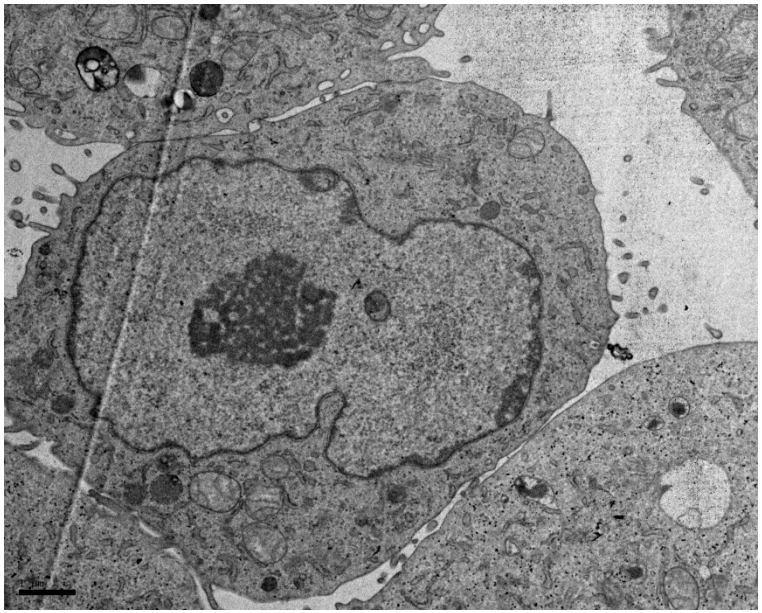
GAA 18



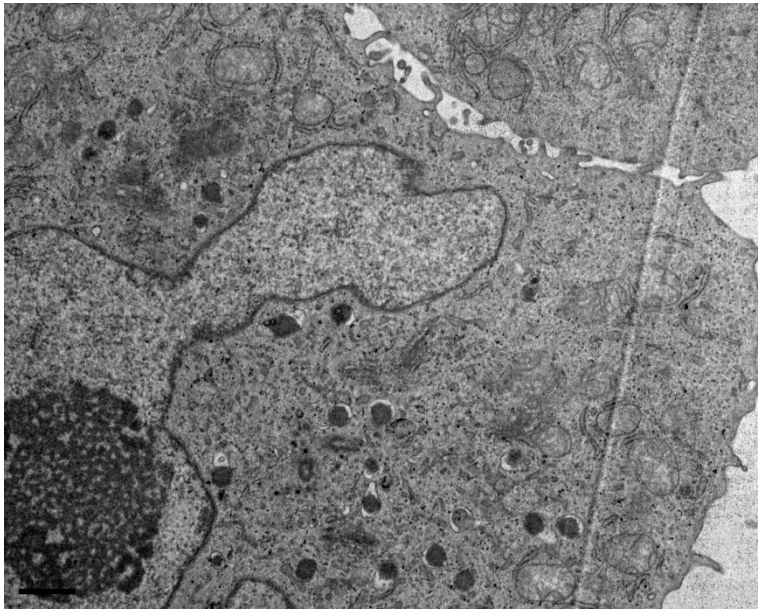
GAA 19



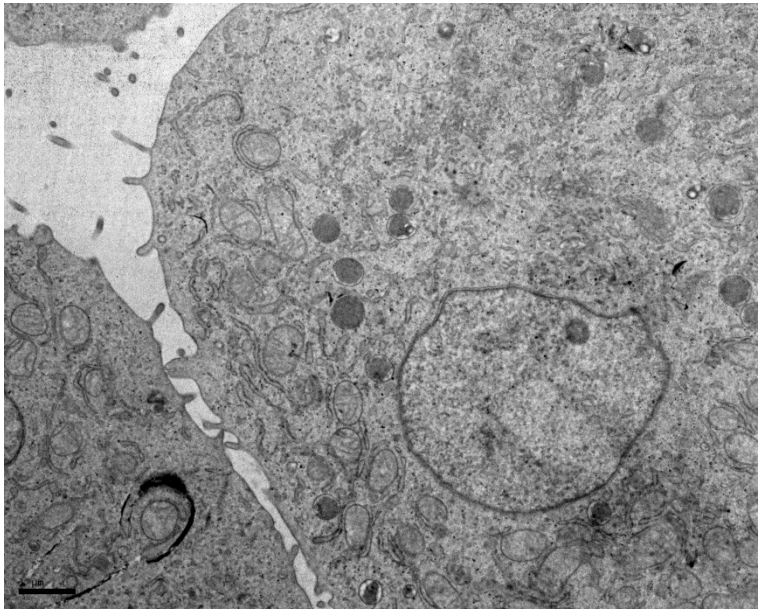
GAA 20



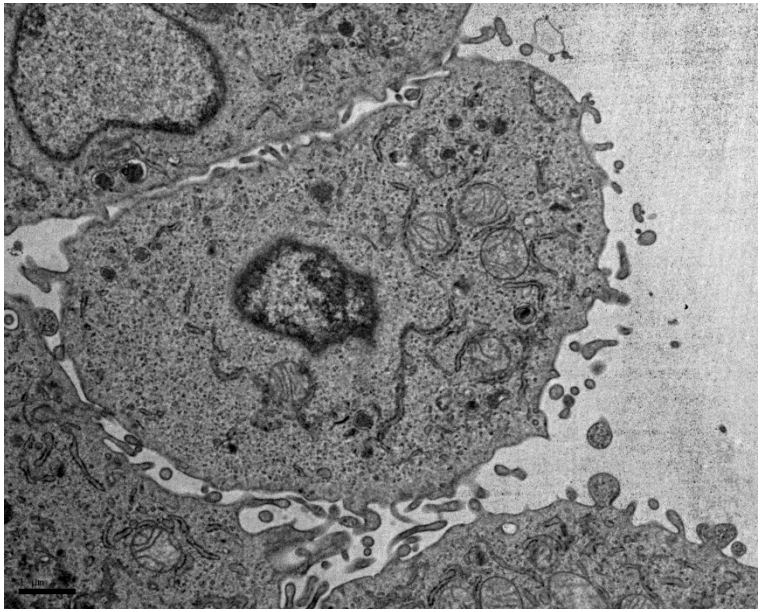
GAA 21



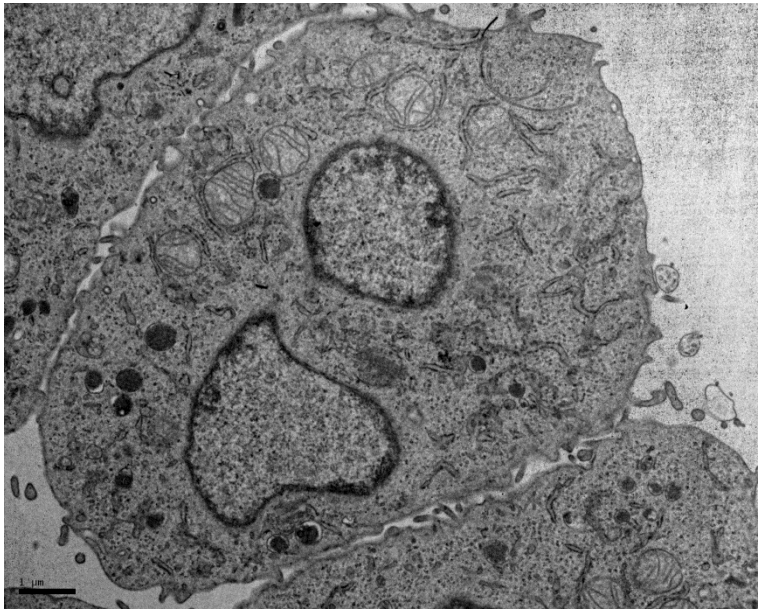
GAA 22



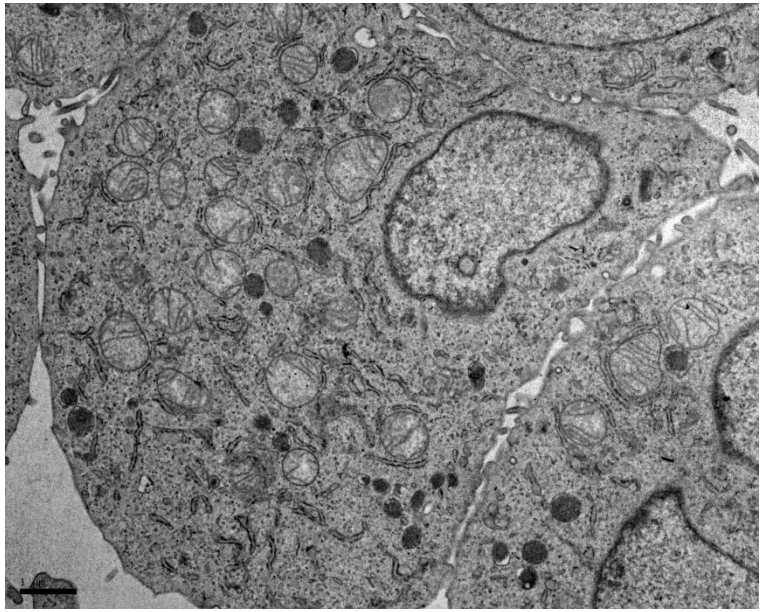
GAA 23



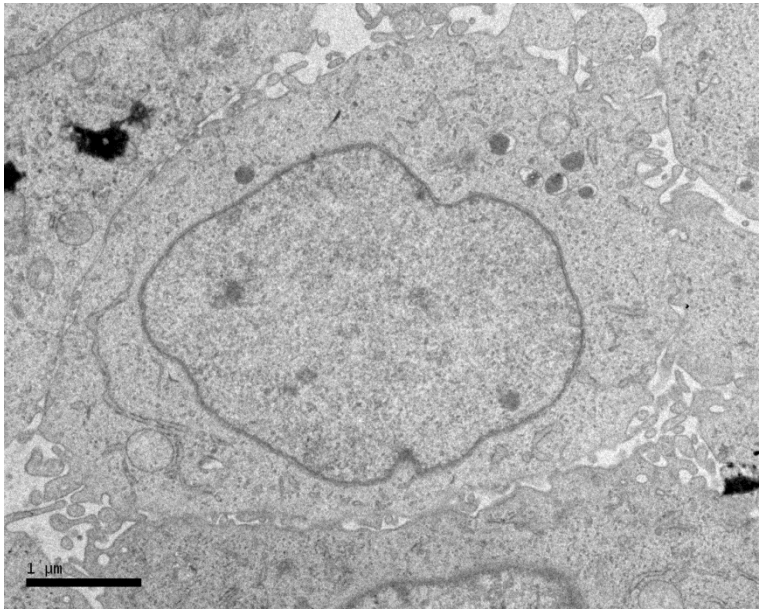
GAA 24



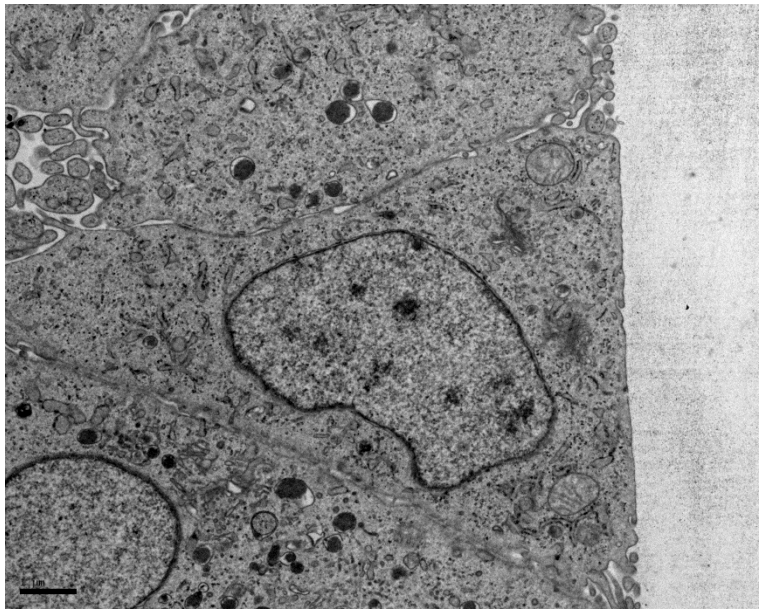
GAA 25



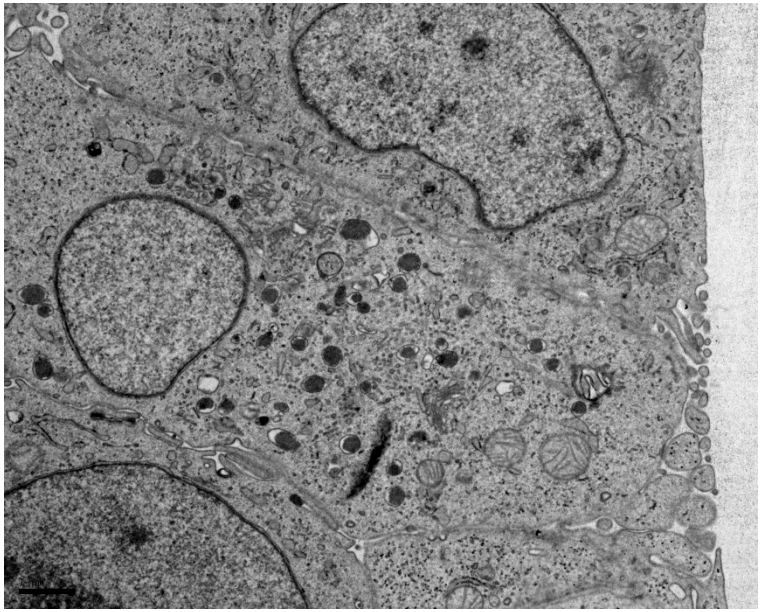
GAA 26



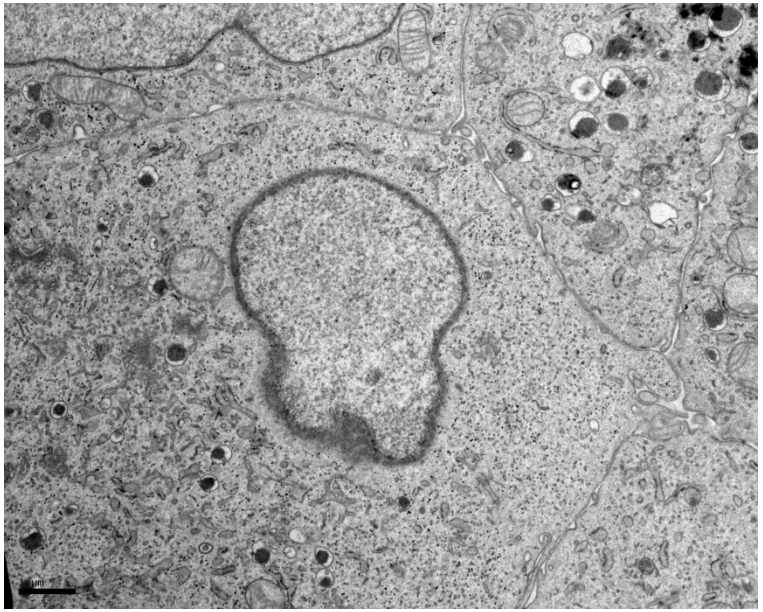
GAA 27



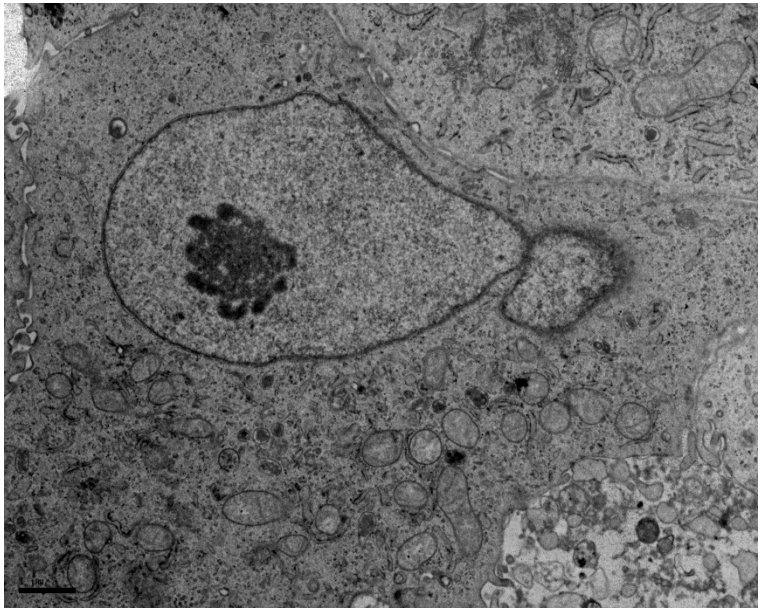
GAA 28



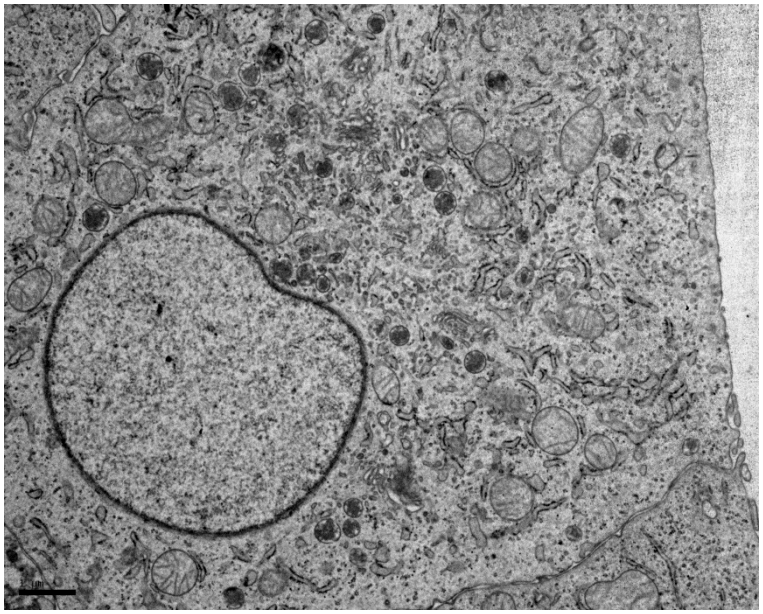
GAA 29



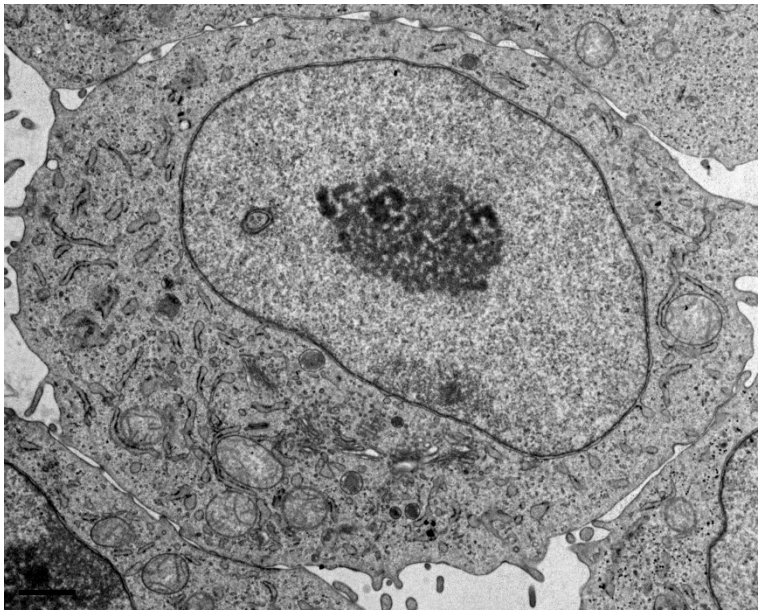
GAA 30



GAA 31



GAA 32

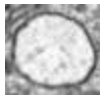


Supporting data. TEM images of Null producer CHO Clone 1

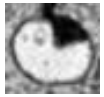
Sample name	Empty lysosomes E	Half lysosomes (H)	Full lysosomes (F)	Total lysosomes (T)	%F	%H+F	%E
Null 1	3	6	1	10	10	70.0	30.0
Null 2	0	5	0	5	0	100.0	0.0
Null 3	5	5	1	11	9	54.5	45.5
Null 4	17	6	1	24	4	29.2	70.8
Null 5	6	3	1	10	10	40.0	60.0
Null 6	0	8	0	8	0	100.0	0.0
Null 7	7	9	1	17	6	58.8	41.2
Null 8	15	6	0	21	0	28.6	71.4
Null 9	9	11	1	21	5	57.1	42.9
Null 10	6	6	7	19	37	68.4	31.6
Null 11	11	21	1	33	3	66.7	33.3
Null 12	6	3	0	9	0	33.3	66.7
Null 13	4	5	2	11	18	63.6	36.4
Null 14	10	5	0	15	0	33.3	66.7
Null 15	4	3	4	11	36	63.6	36.4
Null 16	6	10	2	18	11	66.7	33.3
Null 17	6	4	0	10	0	40.0	60.0
Null 18	5	6	0	11	0	54.5	45.5
Null 19	3	3	0	6	0	50.0	50.0
Null 20	9	6	2	17	12	47.1	52.9
Null 21	20	4	0	24	0	16.7	83.3
Null 22	1	3	0	4	0	75.0	25.0
Null 23	9	7	0	16	0	43.8	56.3
Null 24	3	2	0	5	0	40.0	60.0
Null 25	3	4	4	11	36	72.7	27.3
Null 26	3	8	0	11	0	72.7	27.3
Null 27	5	5	2	12	17	58.3	41.7
Null 28	3	18	0	21	0	85.7	14.3
Null 29	10	5	0	15	0	33.3	66.7
Null 30	1	2	0	3	0	66.7	33.3
Null 31	6	3	1	10	10	40.0	60.0
Null 32	4	14	0	18	0	77.8	22.2
Null 33	12	4	2	18	11	33.3	66.7

	Total lysosomes (T)	%F	%H+F
Average	14	7	55
StDev	6.6	10.7	20.1
St Error	1.15	1.86	3.50

Representative lysosomes



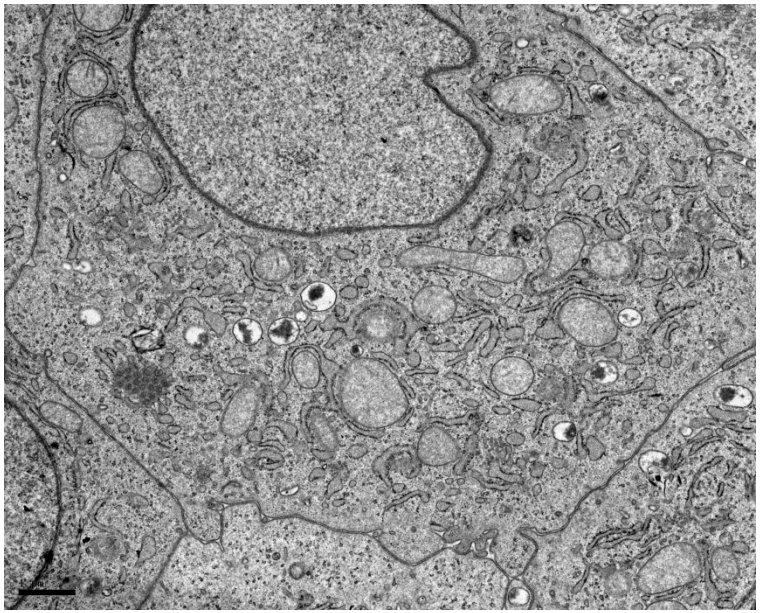
Empty



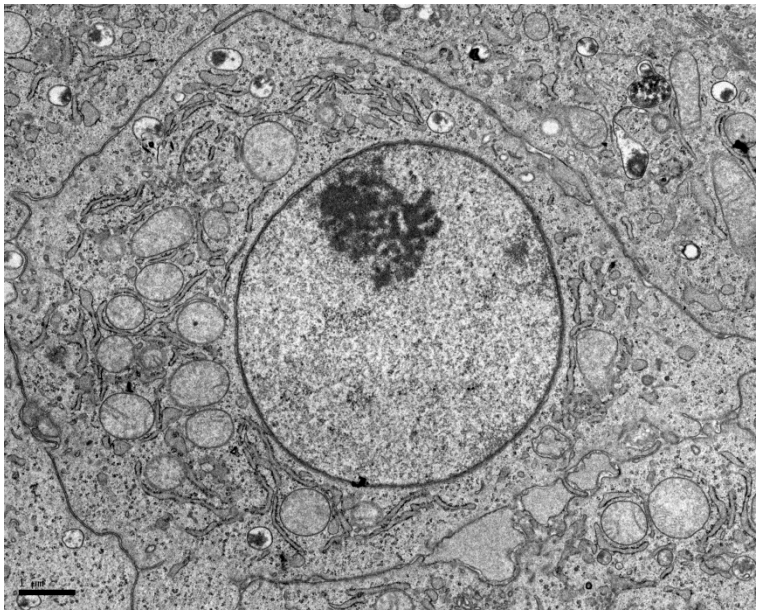
Half

Full

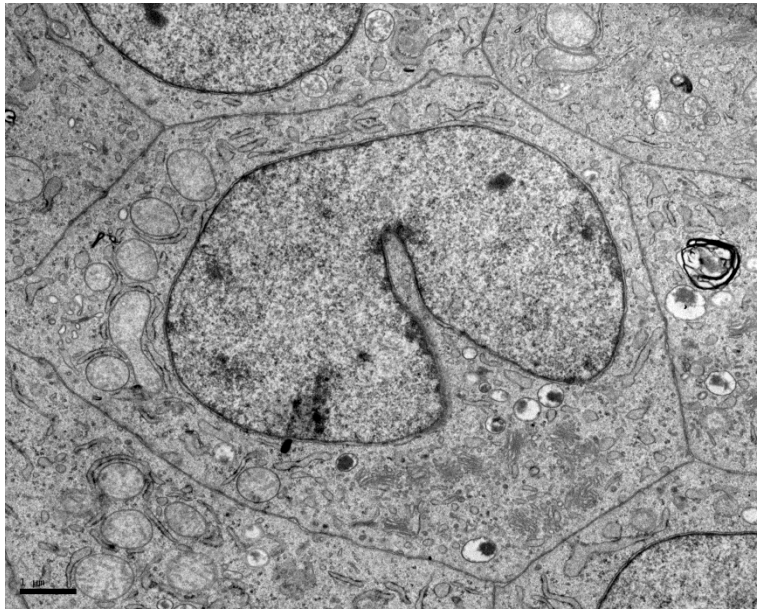
Null 1



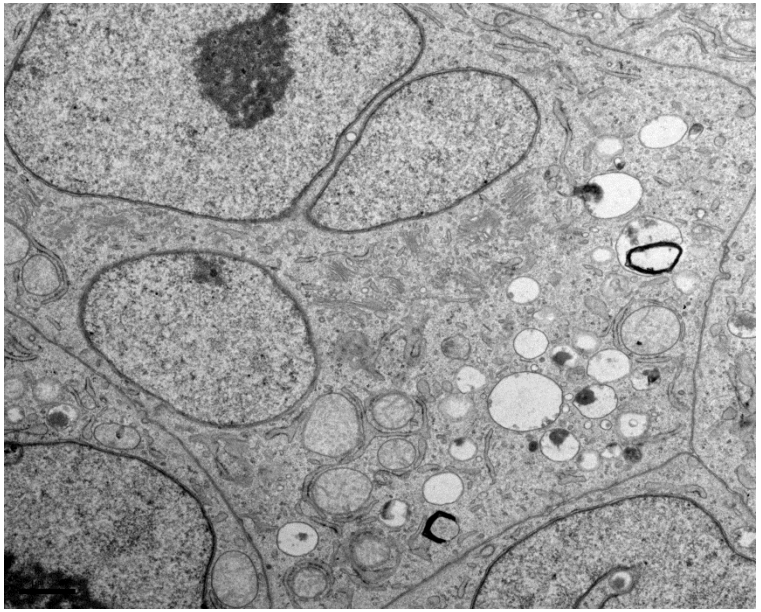
Null2



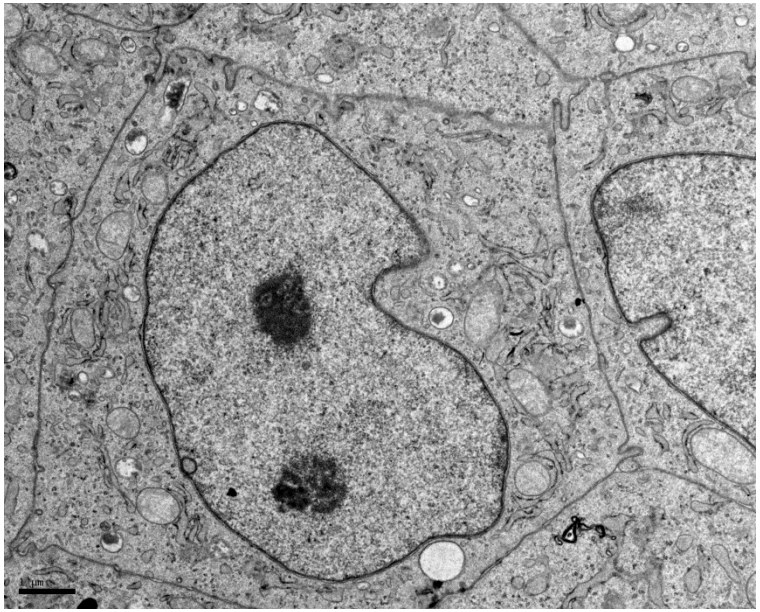
Null 3



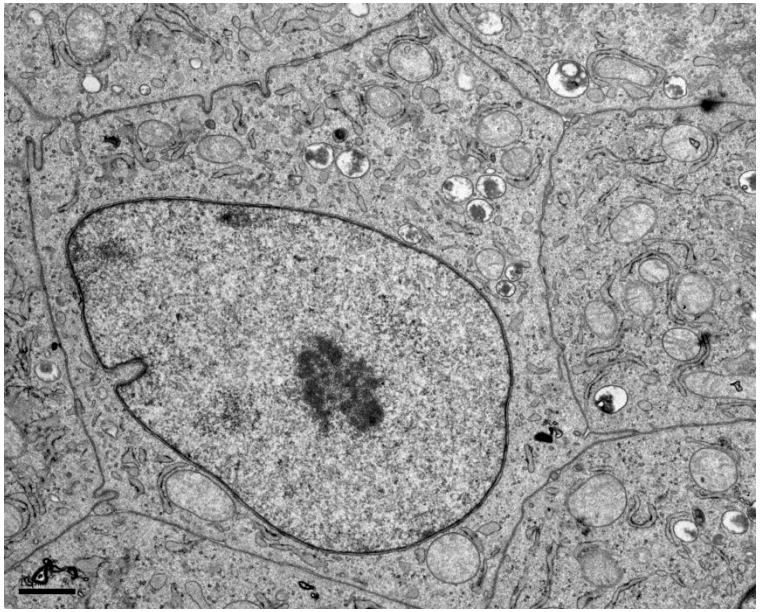
Null 4



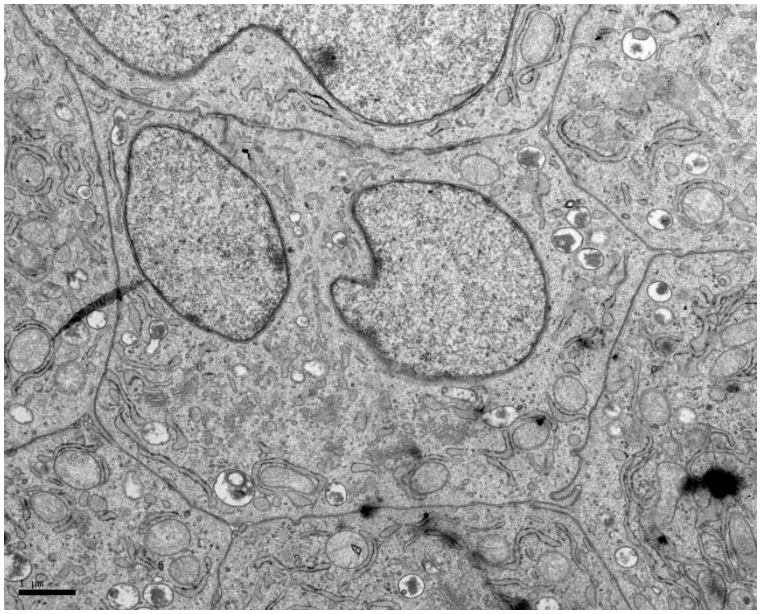
Null 5



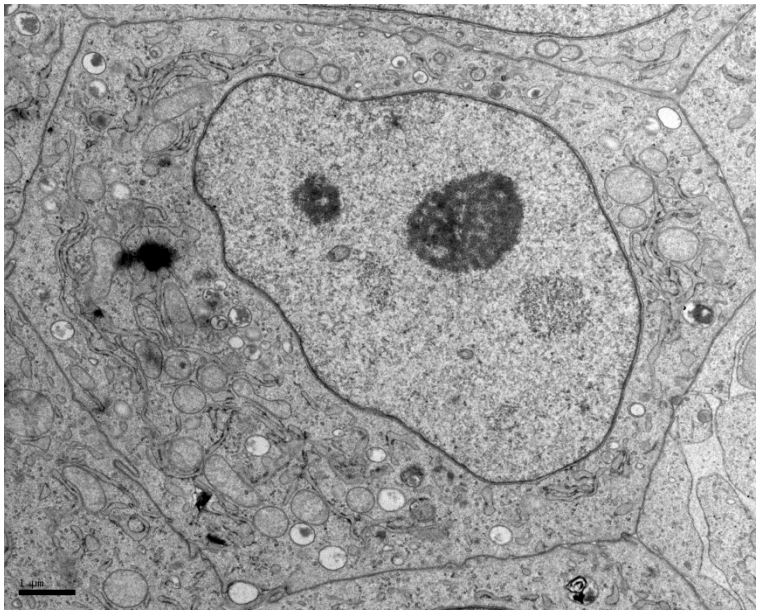
Null 6



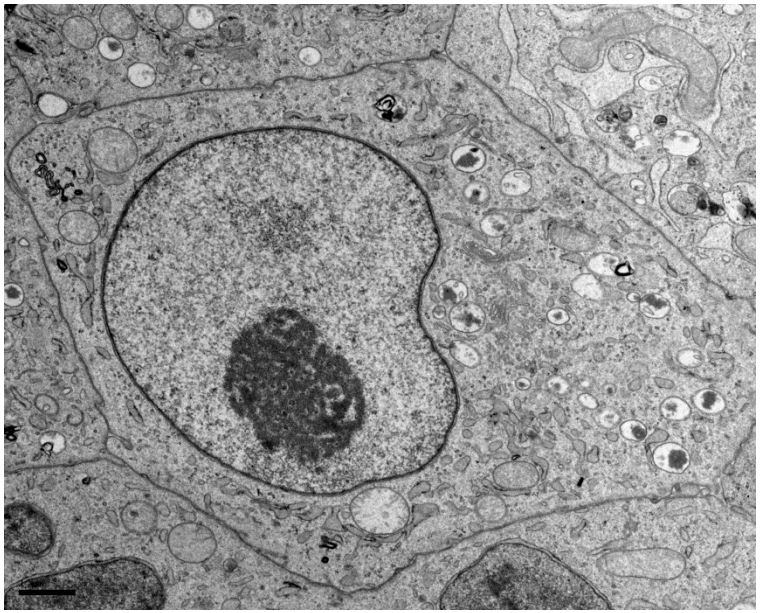
Null 7



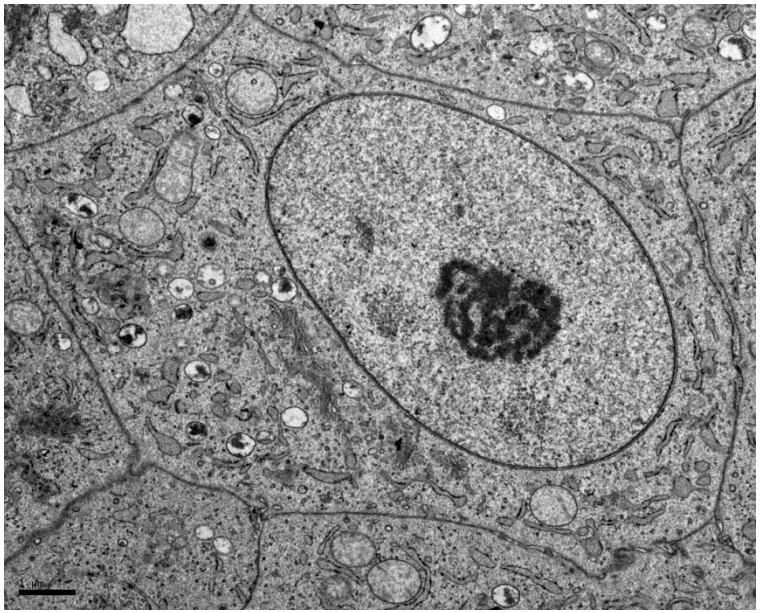
Null 8



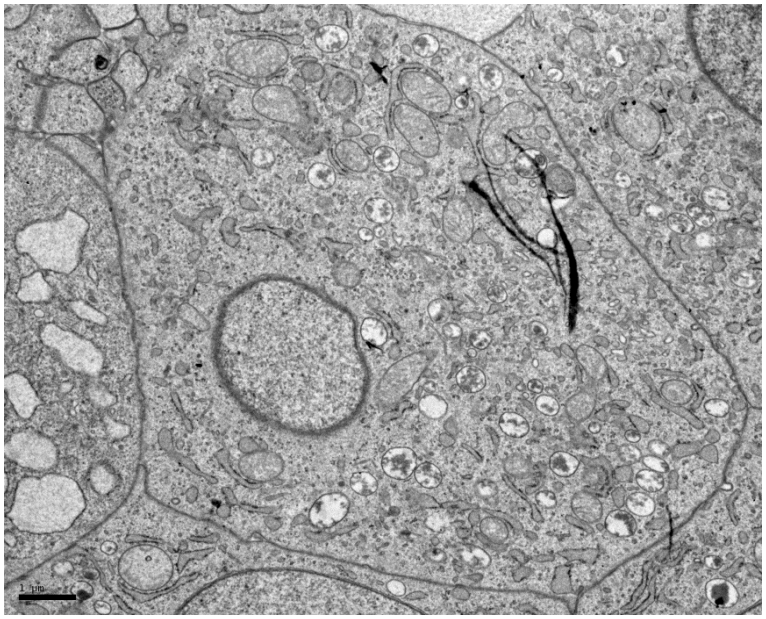
Null 9



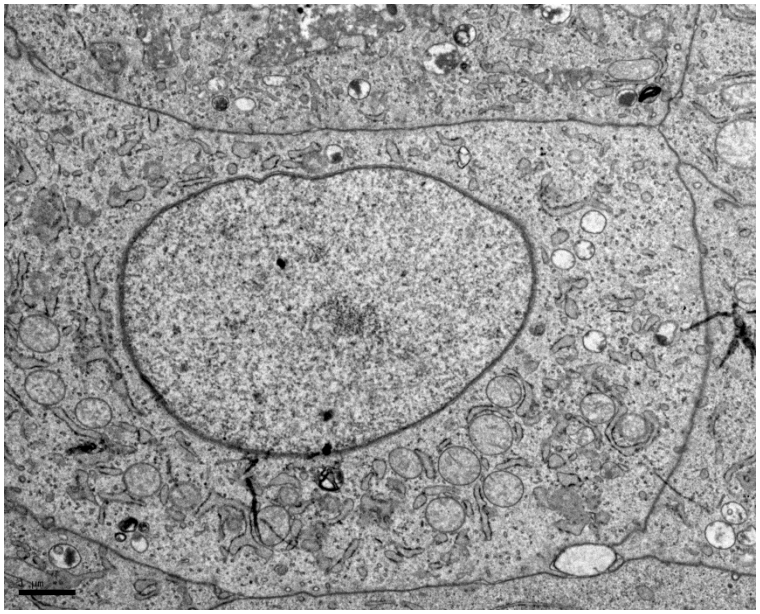
Null 10



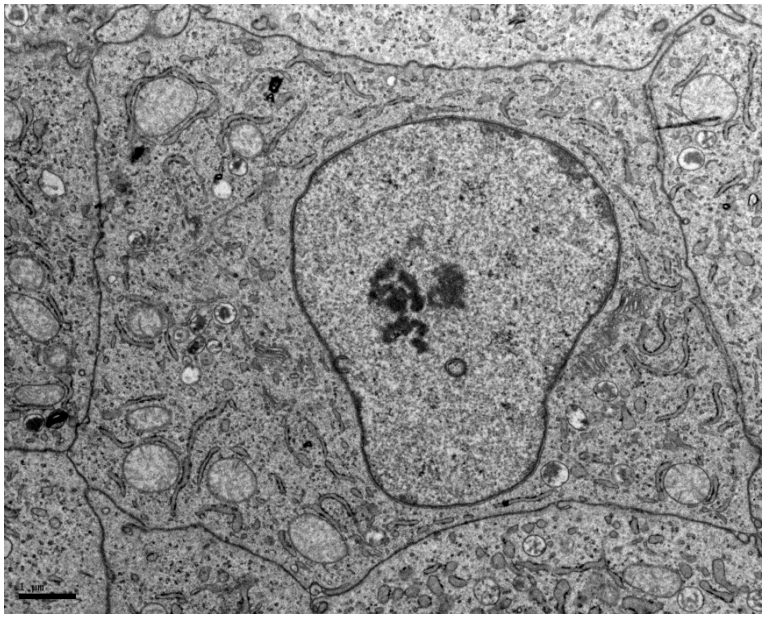
Null 11



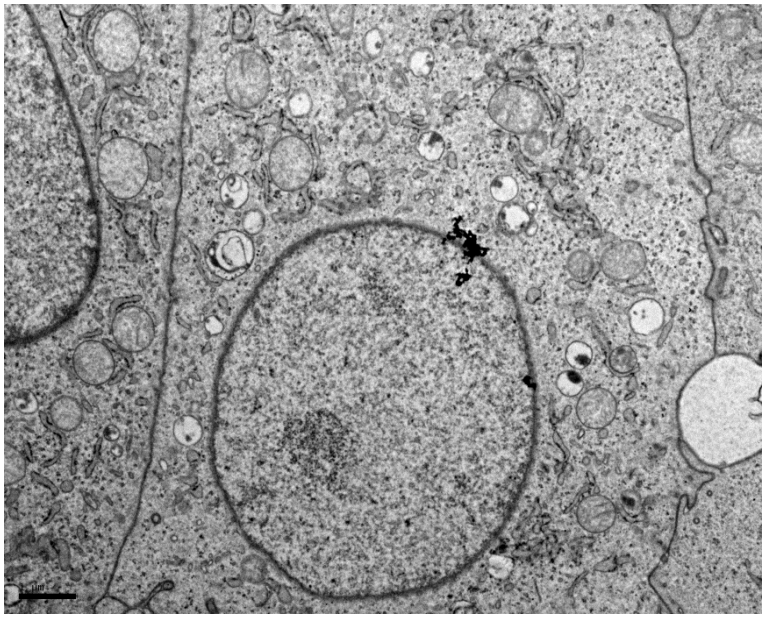
Null 12



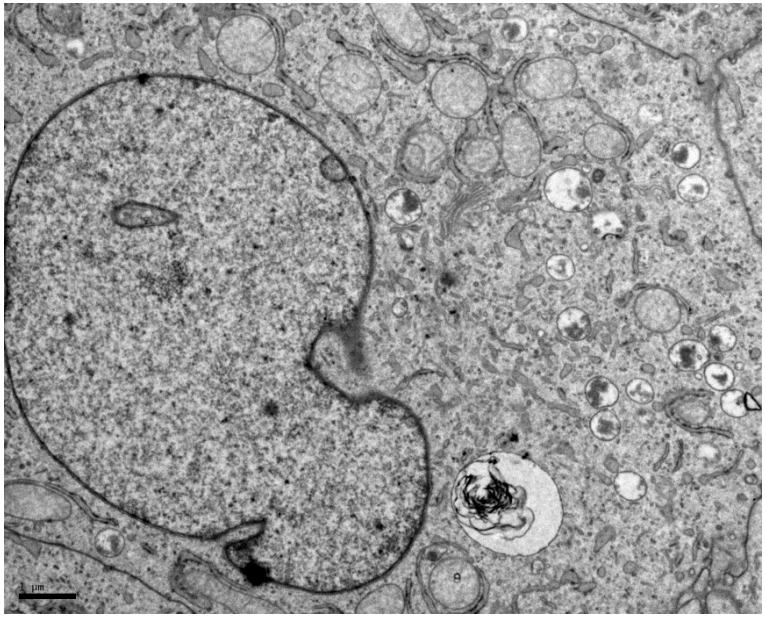
Null 13



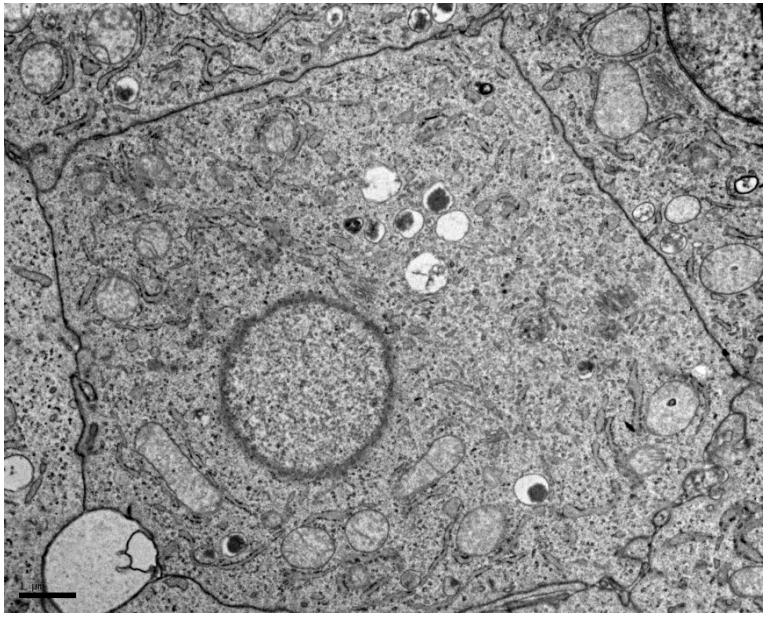
Null 14



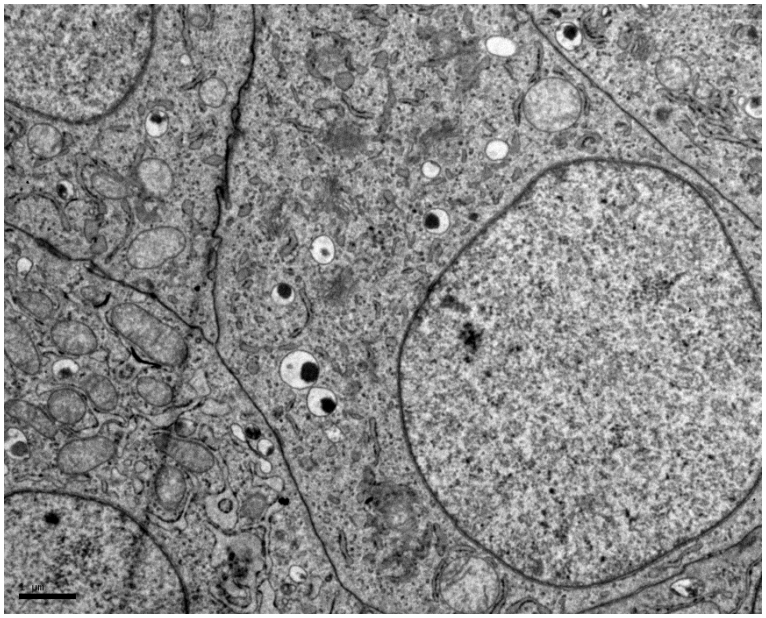
Null 15



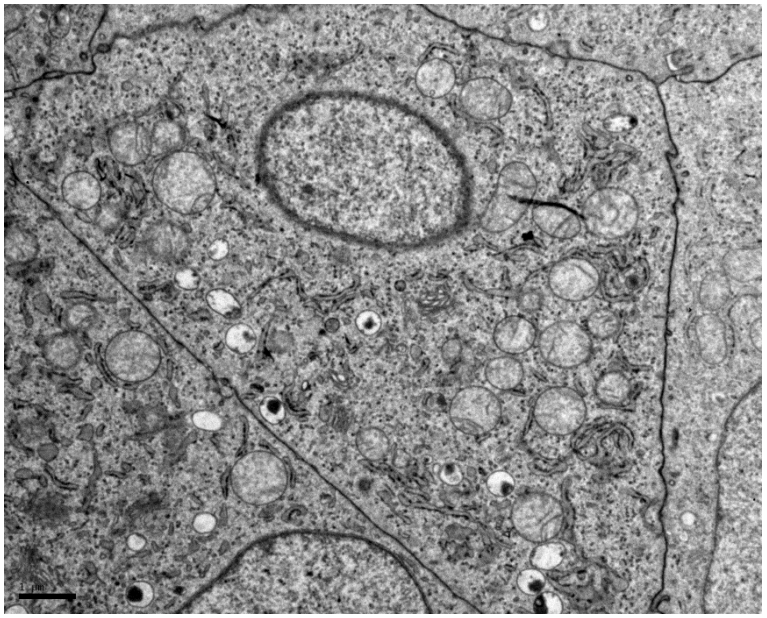
Null 16



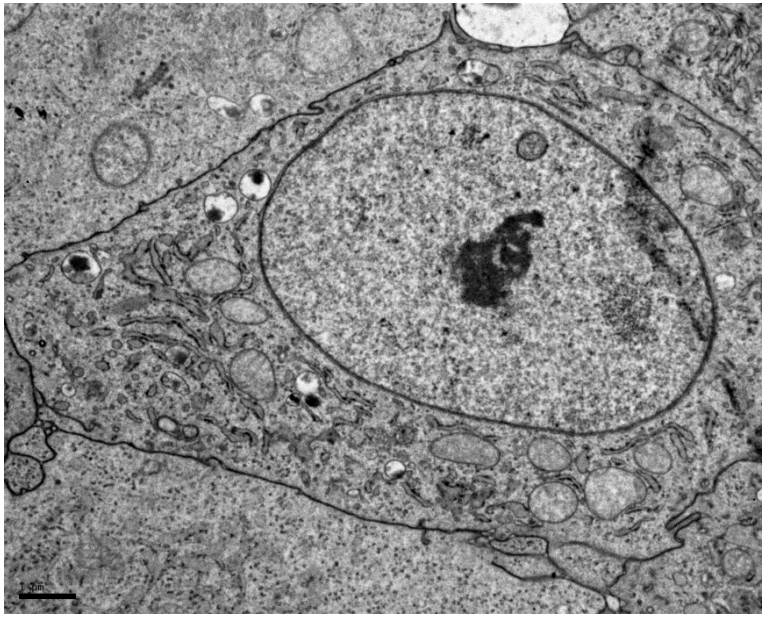
Null 17



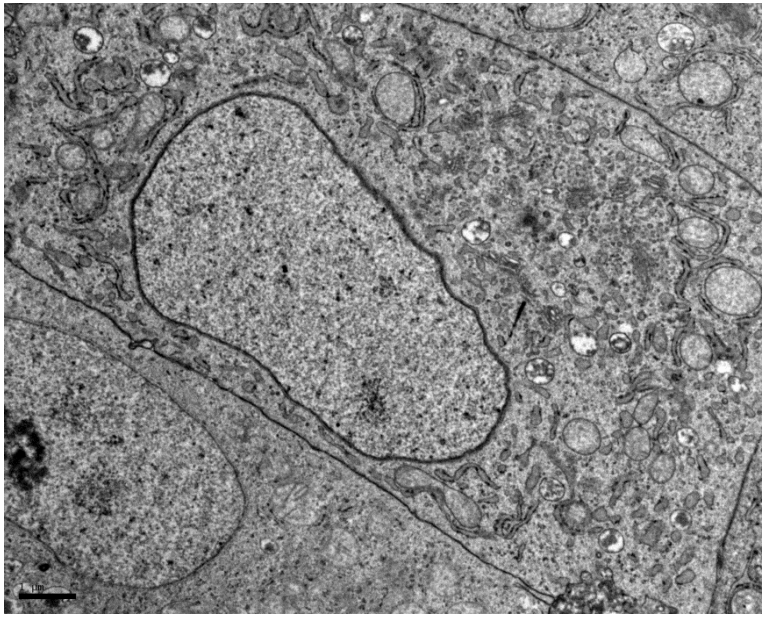
Null 18



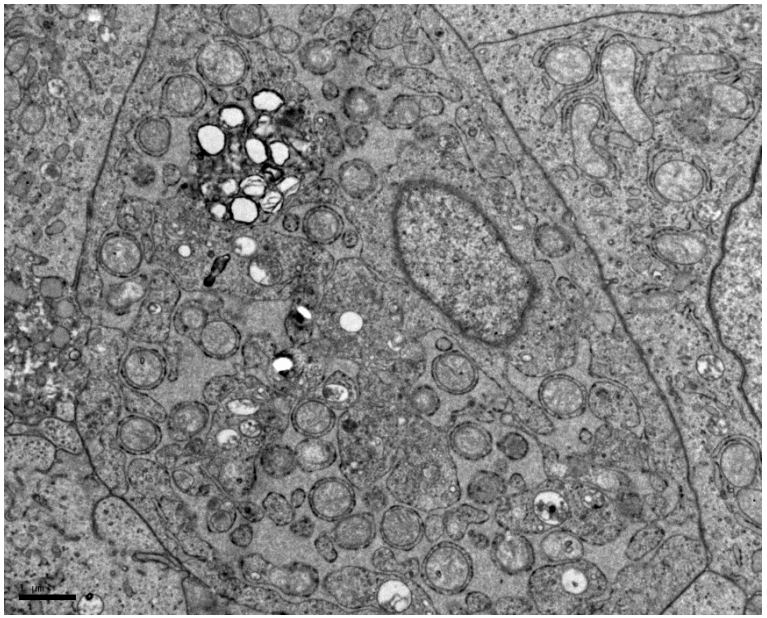
Null 19



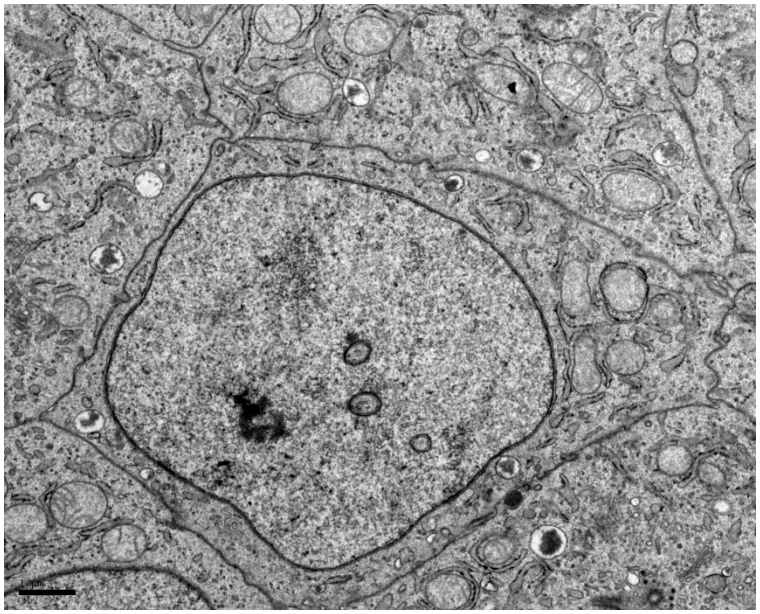
Null 20



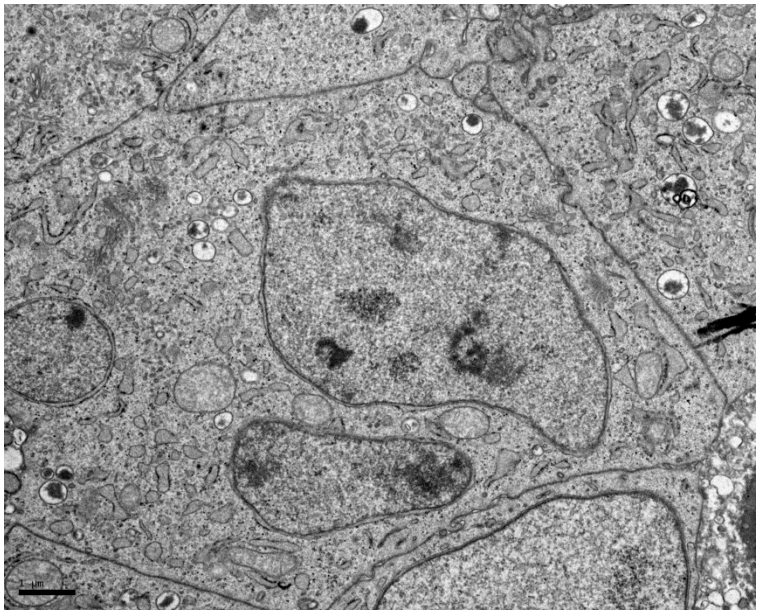
Null 21



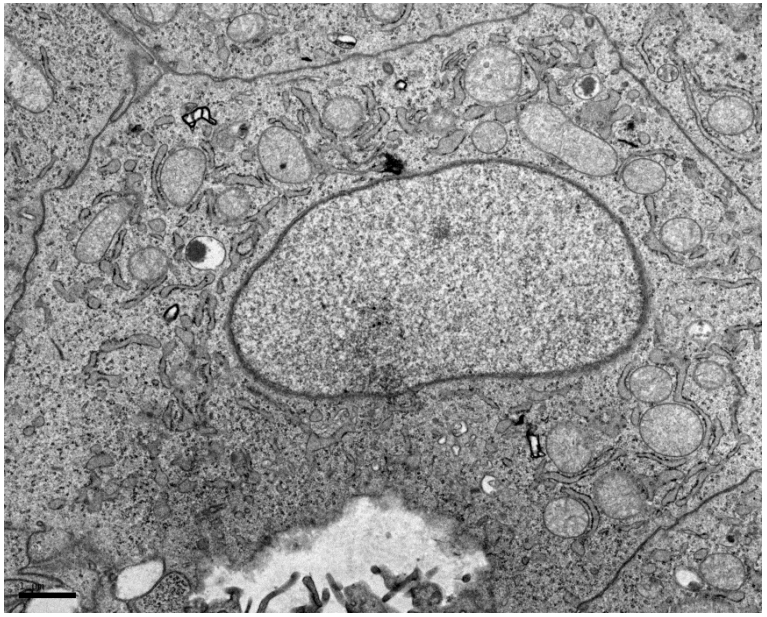
Null 22



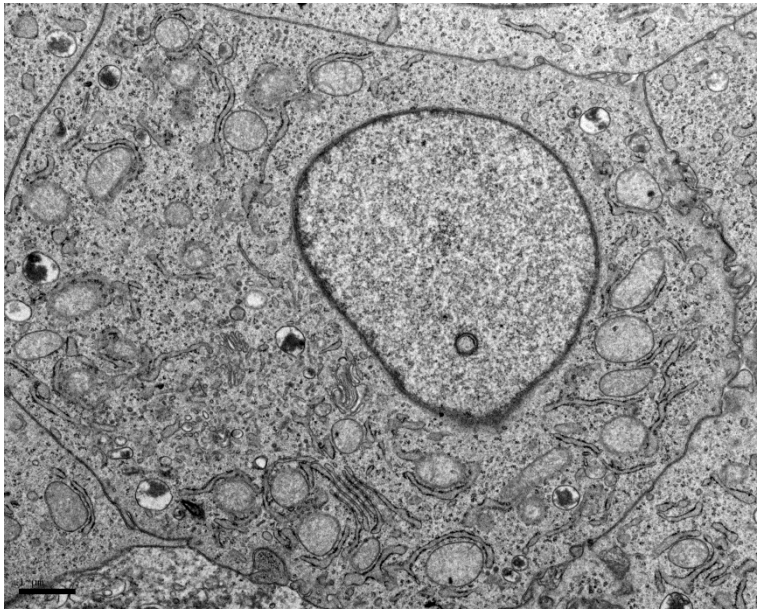
Null 23



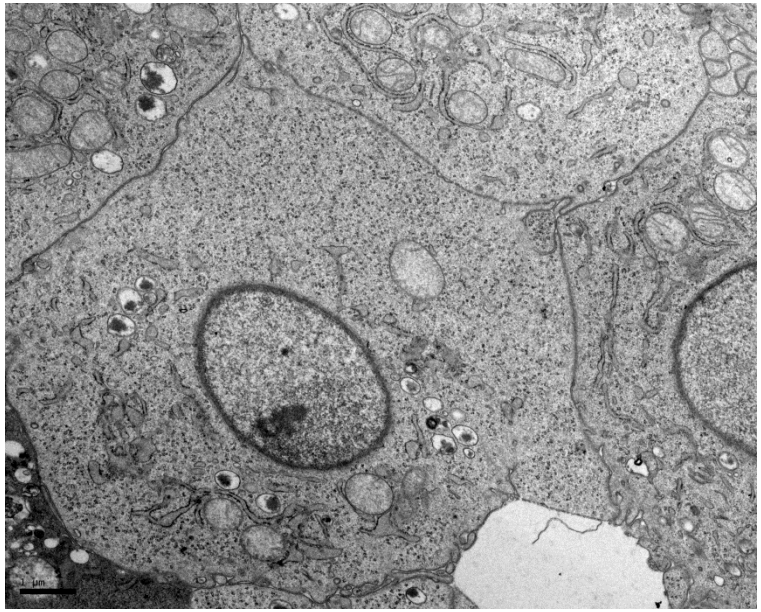
Null 24



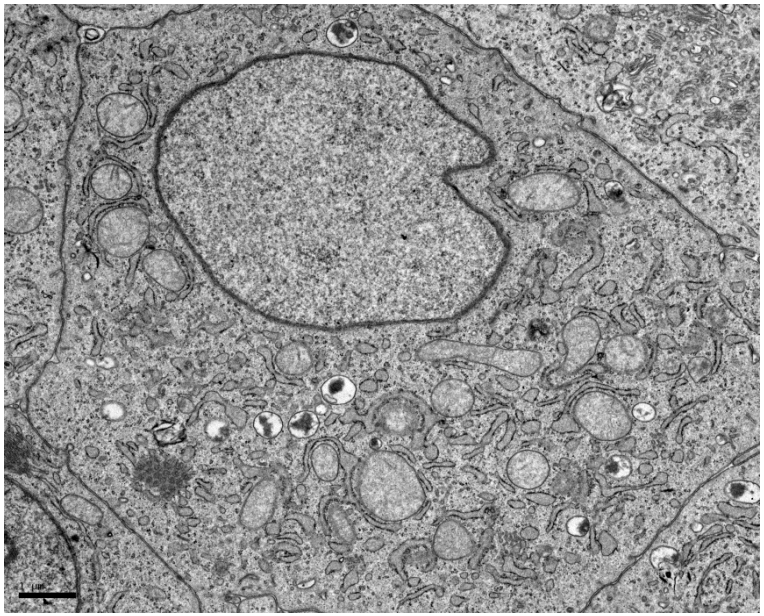
Null 25



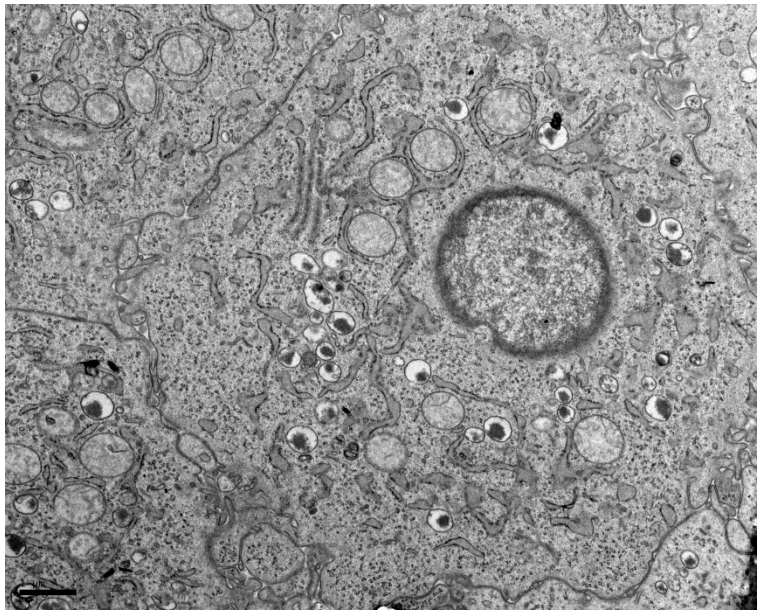
Null 26



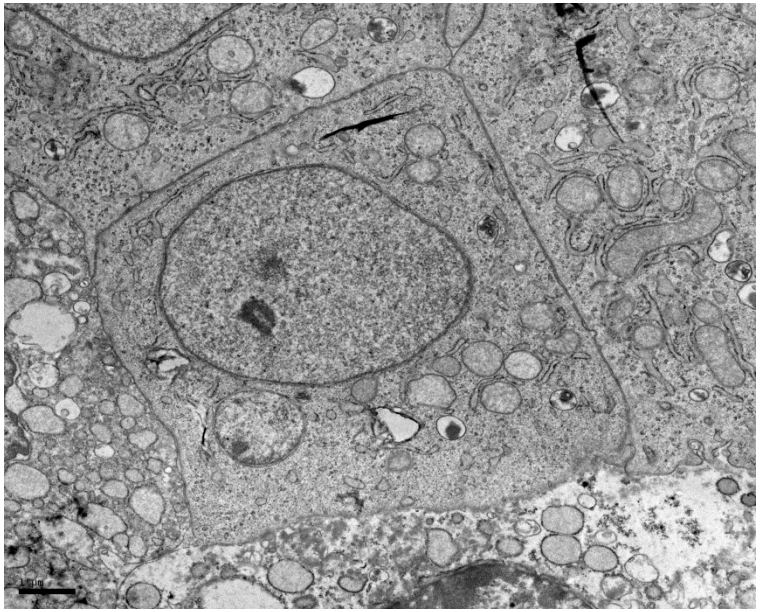
Null 27



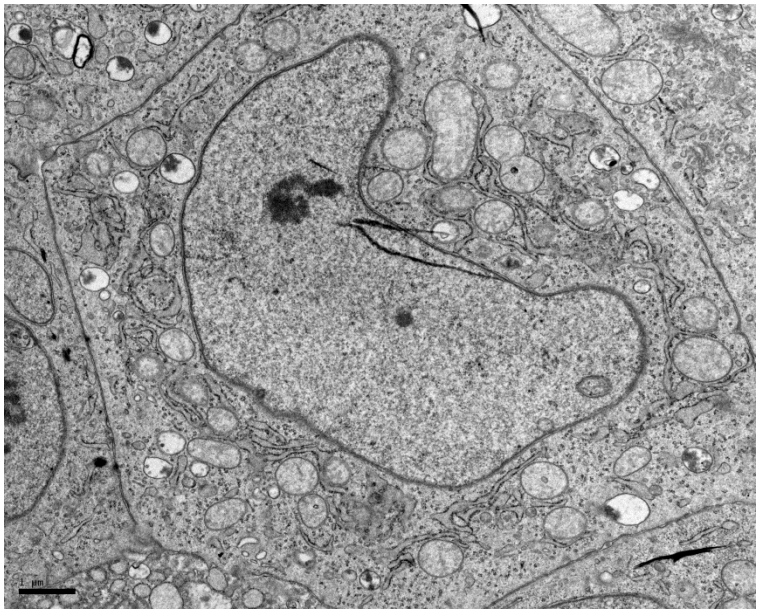
Null 28



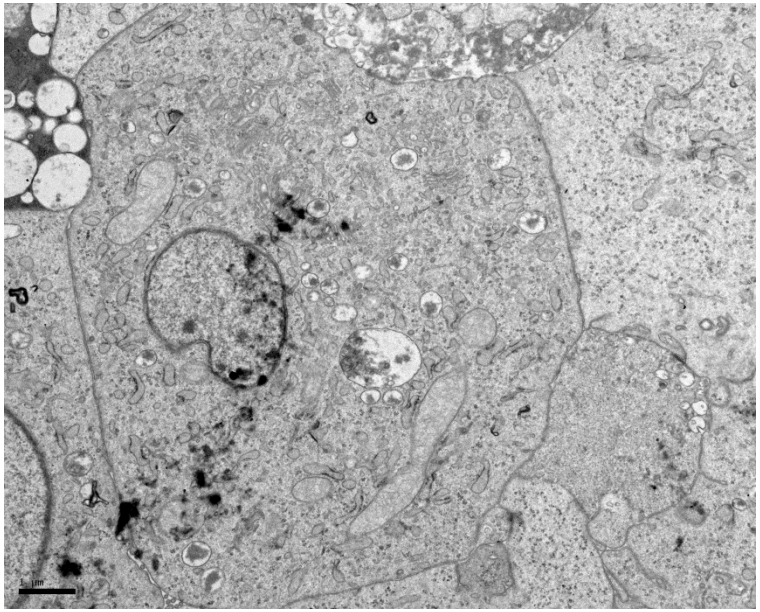
Null 29



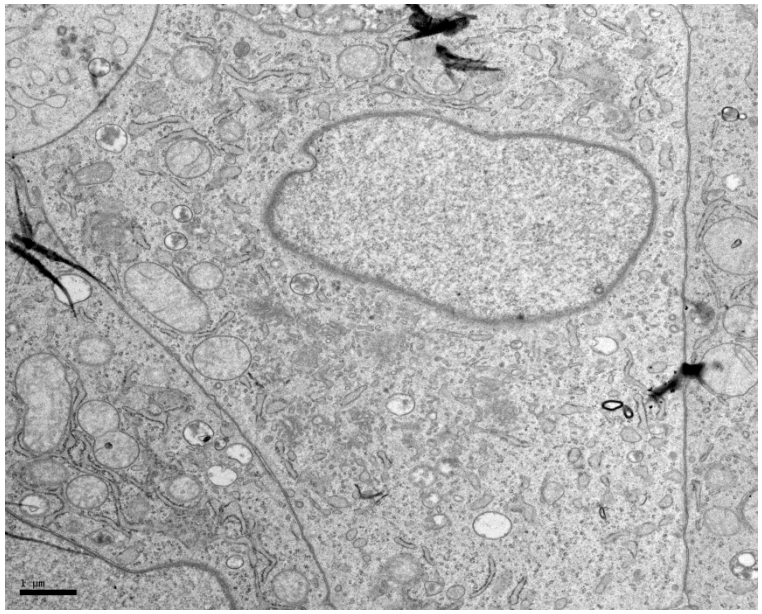
Null 30



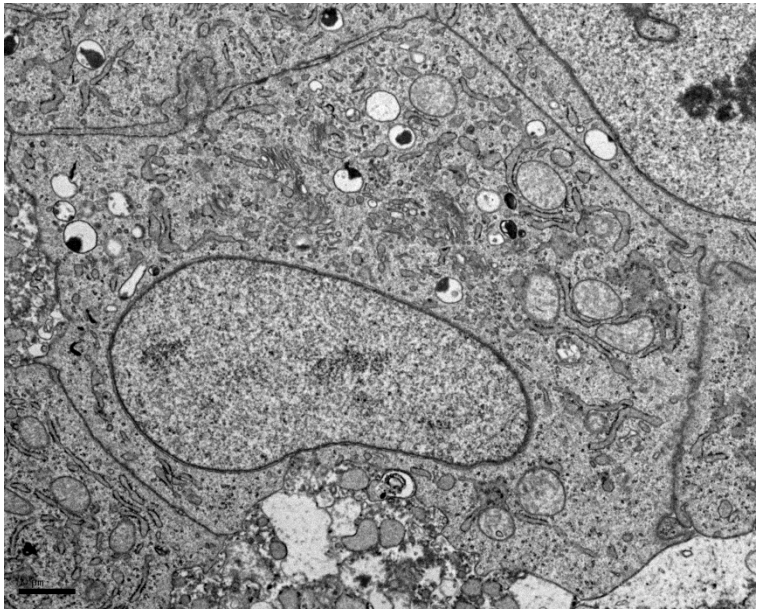
Null 31



Null 32



Null 33



BENGGP15 Bioprocess Validation Coursework

Describe briefly objectives of your research.

Proteases play a critical role in the functioning of the metabolism of all cells but, unfortunately, lead to challenges in the purification of proteins for use as pharmaceuticals in the biotechnology industry. During the purification of a target protein from a CHO line, both cysteine and serine proteases are thought to present particular problems. The understanding of the interaction of these proteases with the target protein and their behaviour during chromatography is the topic of this work.

Here is a list of objectives this research intends to achieve:

- Create a stable CHO cell line capable of producing the target protein, acid alpha glucosidase (GAA) to supply material for purification studies.
- Develop analytical assays able to characterize the cell line, its growth performance, product production, product identification, and protease analysis in the feed stream.
- Design a two column chromatographic process consisting of AIEX and HIC aimed at purifying GAA and eliminating proteases from the process stream.

2. Describe the significance of your research; especially its (potential) contribution to the healthcare sector (1000 characters).

The CHO host expression cell line generated in house is used to provide material used to mimic and improve on the industrial bioprocess in place at BioMarin Pharmaceuticals for the production of recombinant human Acid Alpha Glucosidase (GAA). This therapeutic enzyme, currently undergoing phase 3 clinical trials, is essential in the treatment of Pompe disease, a rare genetic disorder part of the lysosomal storage disorder (LDS) group. This disorder has an incidence of one in 40000 live births (Martiniuk et al., 1998) and it causes glycogen accumulation in the muscle cells. Adult Pompe patients present with progressive skeletal muscle weakness and show difficulties moving and breathing and depend on a lifelong supply of replacement GAA enzyme.

At the moment only one drug is approved for late onset Pompe by the FDA (FDA, 2014), but it lacks the ability to bind to key cell receptors, giving Bio Marin's molecule a clear advantage. Developing a purification method capable of ridding specific co-eluting proteases that are thought of affect final product quality would constitute a clear contribution to the healthcare sector.

3. Summarize the key validation issues faced by your sponsoring company (i.e. as for equipment supplier, or contract manufacturer, or biotech early phase, or big pharma early or late phase or consultant etc.) (4000 characters)

More companies in the pharmaceutical industry today are adopting the principles of Quality by Design (QbD) for pharmaceutical development and manufacturing. Described in ICH Q8, Q9 and Q10 guidance documents, QbD enables enhanced process understanding, and a more systematic and scientific approach to development, so that better controls may be implemented. The end goal is more robust manufacturing processes than those that typically result from traditional approaches to drug development (Nethercote, 2010).

The bioprocess currently in place at BioMarin has been validated for the manufacture of the therapeutic enzyme GAA to supply material for phase 3 clinical trials.

It was recently assessed that the first precolumn filter (capture step) in manufacturing facility, needed changing more often than previously thought due to higher than usual amount of cell debris in the process stream. An internal investigation has attributed this issue to high amount of shear that the cells experience as they are accelerated into the spinning bowl of the centrifuge. It has been decided to replace the current aging disk stack centrifuges with the more modern low shear hydrohermetic design which supposedly decreases shear stress that cells experience at harvest by creating a sealed area filled with liquid for the cells to accelerate to separation velocity. This change reduces considerably the amount of debris and as a consequence allows the reduction in filter size requirement in the pre chromatography filtration step with substantial economic advantage over the overall downstream processing running costs. However, the reduction of filter size has the negative impact of reducing flow rate, increasing the possibility of proliferation of viruses and bacteria.

It is a requirement of GMP that a bioprocess change in a pharmaceutical company is validated according to regulatory agency standards.

The first study that was done was a risk analysis assessment, which aim was to identify the critical quality attributes (CQAs) that might be affected by the change of filter size and centrifuge type. From this analysis it resulted that only process related attributes were going to be affected and specifically viruses and LPS levels due to decreased flow rate (longer residence time), and leachables levels (due to new filter potentially leaching in the process stream). From the analysis it also emerged that product related attributes were cause for lesser concern as although these changes have a great effect on product quality functionality and patient safety, the bioprocess change had little to no effect on their occurrence, therefore they were assigned a low class of risk.

Summarised in the table below are the steps required to validate the process change.

Step	Aim
Scale down studies	• Calculate amount of debris generated by new centrifuge using USD device and sigma concept. • Determine chromatography resin lifespan and efficiency.

	• Assess Virus and sterile final filtration. • Leachable release.
Identify vendors	• Identify multiple vendors to face 'out of stock' situations. • Let vendor do validation of filter based on specific process stream physical characteristics and operation parameters.
IQ/OQ	• Pipes, supports, and all new instruments should be installed by qualified personnel. • Equipment should be tested by installer and SOP provided to customer. • Release form should be obtained.
PQ and Process Validation	• Test instrument for robustness and repeatability. • Determine performance limits. • Run three consecutive batches successfully at operating parameters.
Cleaning Validation	• Ensure chemicals used for cleaning don't leave residue. • Ensure cleaning is done properly.

Following this validation plan ensures that the change in the bioprocess has no ultimate detrimental effect on the final product quality.

NOTE: The scenario here presented is a hypothetical scenario and does not reflect the sponsoring company procedure.

4. Discuss the regulatory (validation) issues which may affect the take up of your research (4000 characters)

As previously explained this research project is designed to mimic the complete bioprocess in place at BioMarin to try to improve on certain aspects such as a two-column chromatographic purification. Trying to address all the regulatory issues that a therapeutic enzyme produced in house might face on its way to market seems unrealistic, particularly due to the fact that the production facility (ACBE) does not abide GMP regulations.

There are two major potential regulatory issues that I feel need to be addressed:

<u>Product Quality: Isoforms</u>

At this stage of the research project it has been established that the molecular weight of the human recombinant GAA enzyme active form is 110 KDa, and that other isoforms are present due to post translational modifications that happen in Golgi and RER before secretion. These species have been identified as 100, 95, 91, 76 and 70kDa (Reuser, 1985) and some of them are often visible in Western Blot analysis of cell culture supernatant. From a regulatory point of view, allowing any isoform other than the 110KDa in the final drug product would not be acceptable as it could cause unknown immunogenic response to patients. The current purification method is based on a two column chromatographic process, namely an anion exchange column as capture and hydrophobic

interaction as intermediate. Although these two steps achieve close to 85% purification, they cannot guarantee the complete removal of isoforms as some of them have similar hydrophobicity and isoelectric point values of the target molecule. In order to clear out isoforms and potential aggregates a final polishing gel filtration step should be added to the process to allow separation by molecular weight of the final mixtures of isoforms.

Process: HCP and Proteases

By measuring the concentration of Host Cell Protein in initial feed and in anion exchange eluate using HCP kit AlphaLISA it was determined that the captoQ HiTrap column removes close to 85% of HCP present in the harvest.

The second intermediate HIC step utilises Phenyl FF Hitrap column to purify the target enzyme. This step will be developed to reduce HCP content even further.

It has been shown however that certain proteases, namely cathepsin D and Z co-purify with the target molecule and can affect product activity and purity.

Before this process development work is to be transferred to a commercial platform, more studies have to be done in the protease identification and separation areas, specifically casein zymograms, protease screening assay (such as GBiosciences proteSEEKER) and mass spec analysis on feed and column eluates.

A Design of experiment (DOE) study could also come in handy here. Multiple HIC resins and conditions could be screened to determine a design space for efficient removal of extra proteases while delivering expected purification levels.

Bibliography

FDA. 2014. *FDA expands approval of drug to treat Pompe disease to patients of all ages; removes risk mitigation strategy requirements* [Online]. Available: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm407563.htm [Accessed 13/may 2014].

MARTINIUK, F., CHEN, A., MACK, A., ARVANITOPOULOS, E., CHEN, Y., ROM, W. N., CODD, W. J., HANNA, B., ALCABES, P., RABEN, N. & PLOTZ, P. 1998. Letter to the editor - Carrier frequency for glycogen storage disease type II in New York and estimates of affected individuals born with the disease. *American Journal of Medical Genetics,* 79**,** 69-72.

NETHERCOTE. 2010. *QbD for Better Method Validation and Transfer* [Online]. pharmaceutical manufacturing. Available: http://www.pharmamanufacturing.com/articles/2010/060/.

REUSER, A. J. J., KROOS, M., OUDE ELFERINK, R.P.J., TAGER, J.M. 1985. Defects in synthesis, phosphorylation, and maturation of acid α-glucosidase in glycogenosis type II. *Journal of Biological Chemistry,* 260**,** 8336-8341.