The research described in this thesis was performed at the Laboratory of Protein science, Proteomics and Epigenetic Signaling of the University of Antwerp, Belgium.

Cover design: Ann Roelant, De Nieuwe Mediadienst University of Antwerp Printed by Provo, Gierle

Faculteit Farmaceutische, Biomedische en Diergeneeskundige Wetenschappen

Departement Biomedische Wetenschappen

Unraveling the molecular function of androglobin: a testis-specific globin

Het ontrafelen van de moleculaire functie van androglobin: een testis specifiek globine

An Bracke

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen: Biochemie en Biotechnologie
aan de Universiteit Antwerpen

Promotor

Prof. dr. Sylvia Dewilde

Antwerpen, 2018

Composition of the jury:

Promotor:	Prof. dr. Sylvia Dewilde (University of Antwerp)
Chairman:	Prof. dr. Dirk Snyders (University of Antwerp)
Internal member:	Prof. dr. Anne-Marie Lambeir (University of Antwerp)
External members:	Prof. dr. Jo Leroy (University of Antwerp)
	Prof. dr. Bart P. Braeckman (University of Gent)
	Prof. dr. Thomas Hankeln (University of Mainz)

Table of contents

List of abbreviations

A

ACTB	Beta-actin
ADGB	Androglobin
ART	Assisted reproductive technologies
AZF	Azoospermic factor

C

CAPN	Calpain
CB	Chromatoid body
CGH	Comparative genomic hybridization

D

DAG	Diacylglycerol
DTT	Dithiothreitol

E

ER	Endoplasmic reticulum
ESI	Electrospray ionization

F

FBS	Fetal bovine serum
FHB	Flavohemoglobin
FRET	Fluorescence resonance energy transfer
FS	Fibrous sheath

G

GCS	Globin coupled sensor
GO	Genome ontology
GPCR	G-protein-coupled receptor
GWAS	Genome wide association studies

H

Hb	Hemoglobin
HSP	Heat shock protein

I

ICSI	Intracytoplasmic sperm injection
IP	Immunoprecipitation

IP3	Inositol 1,4,5-triposphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVF	In vitro fertilization

L

LC-MS/MS	Liquid chromatography tandem mass spectrometry

M

Mb	Myoglobin
MIM	MIT-interacting motif
MIT	Microtubule-interacting and transport
MMAF	Multiple morphological anomalies of the flagella
MRM	Multiple reaction monitoring

N

NADH	Nicotinamide adenine dinucleotide (reduced)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NGB	Neuroglobin
NGS	Next generation sequencing
NMD	Nonsense-mediated mRNA decay
NOA	Non-obstructive azoospermia

O

OA	Obstructive azoospermia
OD	Optic density
ODF	Outer dense fiber

P

PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein databank
PIP2	Phosphatidylinositol 4,5-biphosphate
PIWIL1	Piwi-like protein 1
PKA	Protein kinase A
PTM	Post-translational modification
PTP	Proteotypic peptide

R

RNP	Ribonucleoprotein
RBP	RNA binding protein
ROS	Reactive oxygen species

RT-PCR	Reverse Transcriptase – polymerase chain reaction
RT-qPCR	Real time-quantitative polymerase chain reaction

S

SCO	Sertoli cell only
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gelelectrophorese
SNP	Single nucleotide polymorphism
SPATA20	Spermatogenesis-associated protein 20
SPE	Solid phase extraction

T

TDRD6	Tudor domain-containing protein 6

U

UPLC	Ultra performance liquid chromatography

W

WGS	Whole genome sequencing
WHO	World health organization

Summary

Globins are small heme-containing proteins which can reversibly bind oxygen and/or be involved in redox reactions. These proteins are ancient proteins since they occur in the three big domains of life: Bacteria, Archaea and Eukarya. Until the beginning of the 21^{the} century there were only two types of globins known in vertebrates: hemoglobin (Hb) and myoglobin (Mb). Since then, comparative genomic studies have led to the identification of novel globin genes. Recently, a novel globin lineage was identified, consisting of large chimeric proteins with an N-terminal calpain-like protease domain and a central globin domain. Because this protein is specifically expressed in testis tissue, it was named androglobin (Adgb). In mice, Adgb expression is associated with post-meiotic stages of spermatogenesis and male Adgb-deficient mice are infertile, displaying a maturation arrest in the late stages of spermatogenesis. Furthermore, an *in silico* expression analysis revealed that ADGB expression is decreased in sperm from infertile men compared with fertile men. These facts demonstrate that Adgb plays a crucial role in spermatogenesis. However, the molecular function of Adgb remains unknown.

The past decades, the incidence of male infertility has increased and semen quality has declined systematically. Therefore, the demand for diagnostic biomarkers and infertility treatments is rising. To fulfill this rising demand, we believe that a clear understanding of the fundamental molecular mechanisms underlying the processes of spermatogenesis and male infertility is necessary. Unraveling the molecular function of Adgb will help us to better understand the molecular fundamentals of spermatogenesis.

First, we reviewed the up-to-now characterized molecular defects underlying male idiopathic infertility (Chapter3). This formed an instrument for us to investigate the molecular function of Adgb in the context of spermatogenesis and male infertility.

Second, we wanted to perform an *in vitro* biochemical analysis of the globin domain of Adgb. (Chapter4) For this purpose, the globin domain was recombinantly expressed in *Escherichia coli*. We noticed, however, that the globin domain was not able to be expressed as a soluble form. In the end, the protein was purified in its denatured form

and was *in vitro* refolded. An absorption spectrum of this refolded protein was recorded, which resembled that of hexa-coordinated hemeproteins, such as neuroglobin or cytoglobin. In general, it is assumed that penta-coordinated globin, such as Hb and Mb, are associated with functions in oxygen transport or storage and that hexa-coordinated globin are associated with other functions, such as oxygen sensing, redox signaling and scavenging of reactive oxygen and nitrogen species. Further *in vitro* biochemical characterization of the globin domain was not possible due to its instability. In order to improve soluble expression, we chose to express the protein in two higher eukaryotic expression systems (*Pichia pastoris* and baculovirus infected insect cells), that are able to carry out more complex post-translational modifications and have a better folding machinery for mammalian proteins. Despite these efforts, the globin domain remained unstable and we concluded that it requires the context of the full length Adgb for a stable folding. The full length Adgb was successfully expressed in baculovirus-infected insect cells, but was cleaved over time, possibly by intracellular proteases of the insect cells or by auto-cleavage by the internal calpain-like protease domain of Adgb. The truncation of Adgb precluded further biochemical characterization of the full length protein. In the future, we will try to avoid this truncation by site-directed mutagenesis of the putative catalytic sites of the calpain-like domain.

Third, the cellular protein-interaction partners of Adgb were determined with a co-immunoprecipitation (co-IP) pull down in mice testis tissue, followed by mass spectrometer analysis (Chapter5). Three candidate interaction partners (Tdrd6, Piwil1 and Spata20) were subsequently validated as Adgb-interaction partner with reciprocal co-IP and fluorescence resonance energy transfer (FRET). Tdrd6 and Piwil1 are both main components of the chromatoid body, a specialized organelle of the male germ cell, of which it is assumed to serve as a center for mRNA storage and processing in the late, transcriptional-silent phases of spermatogenesis. We do not known yet which function Adgb might play in this specialized organelle, therefore a further characterization of the observed protein-protein interactions is required.

Finally, in order to investigate the role of ADGB in male infertility and its potential function as diagnostic biomarker, we performed an expression analysis in semen and testis tissue from fertile and infertile men (Chapter6). RT-qPCR exposed that Adgb

expression is down-regulated or even absent in testis tissue from men with disturbed or absent spermatogenesis. On protein-level we were not able to detect ADGB in mature spermatozoa or testis tissue. This may indicate that ADGB plays a very transient role during spermatogenesis and that the protein is not present anymore in the mature sperm cell. Considering this low cellular concentrations for ADGB in testis tissue and spermatozoa, we had to conclude that ADGB is unsuitable as candidate diagnostic protein-biomarker for male infertility.

In conclusion, we can say that several pieces of the puzzle have now been laid out and that the knowledge obtained in this doctoral thesis forms an excellent starting point for further unraveling the molecular function of Adgb. In the end, this research will contribute to a better comprehension of spermatogenesis and might eventually help to improve infertility treatments.

Samenvatting

Globines zijn kleine heem-bevattende eiwitten die in staat zijn om reversibel zuurstof te binden en/of redox reacties uit te voeren. Deze eiwitten zijn evolutionair gezien oud, aangezien ze voorkomen in zowel Archaea, Bacteria als Eukaryota, de drie grote domeinen van het leven. Tot aan het begin van de 21ste eeuw waren er slechts twee soorten globines gekend: hemoglobine (Hb) en myoglobine (Mb). Sinds dan, hebben verschillende vergelijkende genoom studies geleid tot de identificatie van talloze nieuwe globine genen. In 2012 werd er een nieuw soort globine ontdekt dat specifiek wordt aangemaakt in testis weefsel, hierdoor kreeg het de naam "androglobine" (Adgb). Adgb is een groot chimeer eiwit met meerdere domeinen, waaronder een N-terminaal calpaïne-achtig protease domein en een centraal gelegen globine domein. In muizen werd er aangetoond dat Adgb wordt aangemaakt vanaf de post-meiotische fasen van de spermatogenese. Bovendien zijn mannelijke muizen waarin het Adgb gen is uitgeschakeld, onvruchtbaar met een maturatie arrest in de late fasen van de spermatogenese. Daarnaast werd er een *in silico* expressie analyse uitgevoerd in menselijk weefsel, waaruit bleek dat Adgb minder wordt aangemaakt in sperma van onvruchtbare mannen versus dat van vruchtbare mannen. Uit bovenstaande feiten blijkt dat Adgb een cruciale rol speelt in de spermatogenese. De moleculaire functie van Adgb in de spermatogenese is echter nog niet gekend.

Het is gebleken dat de kwaliteit van menselijk sperma systematisch gedaald is de laatste decennia en dat mannelijke onvruchtbaarheid steeds vaker voorkomt. Als gevolg hiervan stijgt de vraag naar goede diagnostische biomerkers en fertiliteitsbehandelingen. Om aan deze vraag te voldoen, is het noodzakelijk om een fundamentele kennis te verwerven over de moleculaire mechanismen die nodig zijn voor een goed werkende spermatogenese en mannelijke fertiliteit. Het ontrafelen van de moleculaire functie van Adgb is daarom belangrijk.

Als eerste hebben we een overzicht gemaakt van de tot hiertoe gekarakteriseerde moleculaire defecten onderliggend aan mannelijke infertiliteit (Hoofdstuk3). Dit

overzicht helpt ons om het verdere onderzoek naar de moleculaire functie van Adgb te kaderen in de context van spermatogenese en mannelijke onvruchtbaarheid.

Ten tweede wilden we een *in vitro* biochemische karakterisering uitvoeren van het globine domain van Adgb, om zo de driedimensionale structuur, de ligand bindingskinetiek, de redox activiteit etc. te bepalen (Hoofdstuk4). Om dit te bereiken hebben we het globine domein recombinant tot expressie gebracht in *Escherichia coli*. We merkten echter dat het niet mogelijk was om het globine domein in een oplosbare vorm tot expressie te brengen in *E.coli*. Uiteindelijk hebben we het eiwit gezuiverd in zijn gedenatureerde vorm en hebben we het *in vitro* heropgevouwen. Zo konden we een absorptie spectrum nemen, waaruit bleek dat de heem groep van Adgb hexa-gecoördineerd is. In het algemeen worden penta-gecoördineerde globines geassocieerd met functies in zuurstof transport of opslag en worden hexa-gecoördineerde globines eerder geassocieerd met functies zoals zuurstof sensing, redox signalisatie, uitschakelen van schadelijke reactieve zuurstof of stikstof partikels etc. Een uitgebreidere *in vitro* karakterisering van het globine domein was niet mogelijk vanwege de instabiliteit van het eiwit. Om de oplosbaarheid en stabiliteit te verbeteren hebben we twee andere expressie systemen uitgeprobeerd om het globine recombinant tot expressie te brengen; *Pichia pastoris* en baculovirus-geïnfecteerde insect cellen. Dit zijn twee eukaryote expressie systemen, welke gekend zijn voor hun mogelijkheid tot complexere post-translationele modificaties en hun betere mechanismen om eiwitten correct te vouwen. Desalniettemin was het niet mogelijk om het globine domein in een stabiele vorm te zuiveren en moesten we concluderen dat het de context van het volledige eiwit nodig heeft. Het volledige Adgb-eiwit werd met succes recombinant tot expressie gebracht in baculovirus-geïnfecteerde insect cellen. We zagen echter dat na expressie het eiwit afgebroken werd, mogelijks door intracellulaire insect proteases of door het calpaïne-achtige protease domein van Adgb zelf. Het feit dat Adgb werd afgebroken na expressie verhinderde een verdere biochemische karaterisering van het volledige eiwit. In de toekomst zal deze afbraak vermeden proberen te worden door middel van site-directed mutagenese van de mogelijke katalytische residuen van het calpaïne-achtige domein.

Als derde hebben we de mogelijke eiwit interactiepartners van Adgb bepaald met een pull-down co-immunoprecipitatie (co-IP) experiment in muis testis weefsel, gevolgd door

massa spectrometer analyse (Hoofdstuk 5). Drie van deze mogelijke interacties (Tdrd6, Piwil1 en Spata20) werden vervolgens gevalideerd met reciproke co-IP en fluorescentie resonantie energie transfer (FRET) analyse. Tdrd6 en Piwil1 zijn beiden hoofdcomponenten van het chromatoid body, een gespecialiseerd organel van de mannelijke kiemcel. Men vermoedt dat het chromatoid body functioneert als een centrum waar mRNA wordt opgeslagen en verwerkt gedurende de late, transcriptie-loze fasen van de spermatogenese. We weten momenteel nog niet welke functie Adgb juist uitvoert in dit gespecialiseerde organel en verdere karakterisering van de mogelijke eiwit-eiwit interacties die hier werden geïdentificeerd is hiervoor vereist.

Ten laatste hebben we de rol van ADGB onderzocht in mannelijke infertiliteit en de mogelijkheid onderzocht of ADGB een potentiële kandidaat is als biomerker voor mannelijke infertiliteit (Hoofdstuk6). Dit deden we door een expressie analyse uit te voeren in sperma stalen en testis biopten van zowel vruchtbare als onvruchtbare mannen. De gen expressie van ADGB werd onderzocht met RT-qPCR in testis biopten van mannen met zowel een normaal functionerende als een abnormale spermatogenese. Hieruit konden we afleiden dat ADGB expressie vermindert of zelf afwezig is in testis biopten met een afwijkende of afwezige spermatogenese. Om verdere besluiten te trekken over de biologische functie van ADGB is een uitgebreidere studie nodig in meerdere biopten met diverse afwijkingen van de spermatogenese. Hoewel we gebruikmaakten van verschillende technieken (Western blot en MRM-analyse), konden we ADGB op eiwit-niveau niet detecteren in mature sperma cellen of in testis weefsel. Dit duidt mogelijks op een zeer tijdelijke rol voor ADGB tijdens de spermatogenese en/of het feit dat ADGB mogelijks niet meer aanwezig is in de mature spermacel. Gezien deze lage cellulaire concentraties van ADGB in spermacellen en testisweefsel moeten we concluderen dat ADGB geen geschikte kandidaat is als mogelijke diagnostische eiwit-biomerker voor mannelijke infertiliteit.

Tot slot kunnen we stellen dat de functie van Adgb geleidelijk aan opgehelderd wordt en dat de kennis, verworven in deze doctoraatsthesis, een goed startpunt vormt voor het verder ontrafelen van de moleculaire functie van Adgb. Uiteindelijk zal dit alles leiden tot een beter begrijpen van de moleculaire processen van de spermatogenese, wat belangrijk is gezien de stijgende frequentie van mannelijke onvruchtbaarheid de laatste decennia.

I. General Introduction

1 Globins and their structure

1.1 Globin fold

Globins are present in the three big domains of life; Bacteria, Archaea and Eukarya and thus can be called 'ancient' proteins[1]. However, until the beginning of the 21^{the} century there were only two types of globins known in vertebrates: hemoglobin (Hb) and myoglobin (Mb). Since then, comparative genomic studies have led to the identification of novel globin genes. The ability of high-throughput sequencing and the progress in bioinformatics stimulated the growth of the "globin superfamily". In vertebrates it led to the discovery of neuroglobin (Ngb)[2], cytoglobin (Cygb)[3–5], globine E (GbE)[6], globine X (GbX)[7], globine Y (GbY)[8] and androglobin (Adgb)[9]. Despite the great variety in function and amino acid sequence among the globin superfamily, they all share a similar 3D structure. The classic globin exists of approximately 150 amino acids and is characterized by a typical 3-over-3 α-helical Mb-fold, which was described for the first time, fifty years ago, by John Kendrew[10] (Fig1A). The Mb-fold is characterized by a heme group surrounded by eight α-helices, named from A till H, with a pattern of predominantly hydrophobic residues at 37 conserved, solvent-inaccessible positions. Besides the classical 3-over-3 α-helical folded globins, another group can be defined in the globin superfamily characterized by a shortened 2-over-2 α-helical folding (also called 'truncated Hbs'). These 2-over-2 globins can be found in unicellular eukaryotes, cyanobacteria, green algae, plants and in many bacteria, but not in vertebrates[11]. Another classification in the globin superfamily can be made based on their domain structure; globins can be divided in 'single domain globins' and 'chimeric globins'[12]. The latter possesses, besides the globin domain, also other functional domains. Examples of chimeric globins are the flavohemoglobins (FHBs) in *Escherichia coli* and yeast, consisting of an N-terminal Hb coupled to a ferredoxin reductase-like domain, and the globin coupled sensors (GCS), reported in Bacteria and Archaea, comprising an N-terminal Hb linked to variable gene regulatory domains. FHBs and GCS do not occur in vertebrates.

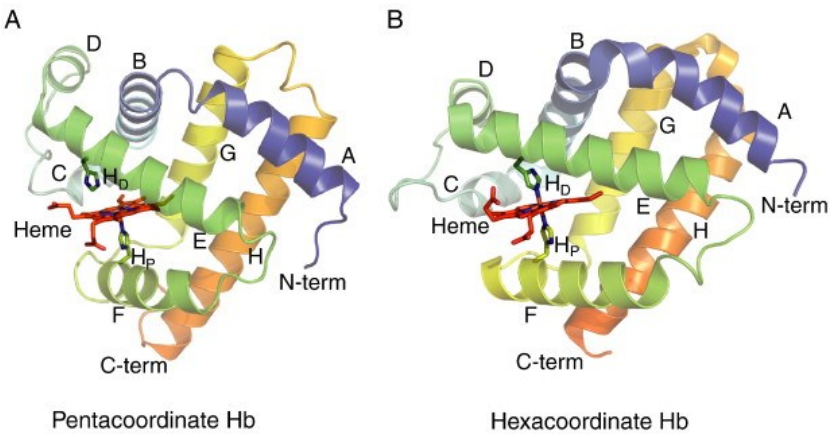


Figure 1: Penta-coordinated and hexa-coordinated hemoglobins. A) The structure of sperm whale myoglobin (2MBW.pbd) displays a penta-coordinated heme with an open distal binding site B) The structure of neuroglobin (1QIF.pdb) demonstrates hexa-coordination with the binding site occupied by the side chain of the distal histidine. In each structure, eight alpha-helices are labeled (A through H) along with showing the N-terminus (N-term), the C-terminus (C-term) and the distal (H_D) and proximal (H_P) histidines. Figure from[13].

1.2 Heme group

The ability of globins to bind oxygen (or other gaseous ligands) depends on the presence of a bound prosthetic group called 'heme' (Fig2). The heme group consists of an organic component and a central iron atom. The organic component, called protoporphyrin, is made up of four pyrrole rings linked by methine bridges to form a tetrapyrrole ring. The iron atom lies in the center of the protoporphyrin, coordinated to the four pyrrole nitrogen atoms. The iron atom can form two additional bonds, one on each site of the heme plane. These binding sites are called the fifth and sixth coordination sites. In globins like Hb and Mb, which are called 'penta-coordinated', the fifth coordination site is occupied by the imidazole ring of a histidine residue from the protein. This histidine is referred to as the proximal histidine. The sixth coordination site remains unoccupied in the deoxy form, available for binding with oxygen, or other gaseous ligands like nitrogen oxide or carbon oxide (FigIA). In globins like Ngb and Cygb, that are called 'hexa-coordinated globins', both the fifth and sixth coordination place of the Fe atom are occupied by the N-atom of the imidazole ring of a histidine of the protein, referred to as the proximal and distal histidine, respectively (FigIB). The proximal histidine is always found on position F8 (8^{the} amino acid of helix F) and is conserved through all globins.

The distal histidine is mostly found on position E7, in some globins the distal histidine is replaced by a glutamine residue. In hexa-coordinated globins, the binding with the external ligands (like oxygen, nitric oxide or carbon oxide) is in competition with the distal histidine (or glutamine).

The iron atom of the heme group can occur in three different oxidation states: Fe^{2+} (ferro), Fe^{3+} (ferri), or Fe^{4+} (ferryl). Only when the oxidation state is Fe^{2+}, a reversible binding with oxygen is possible. In practice, recombinant globins in the laboratory will adopt the Fe^{2+} and Fe^{3+} oxidation states, the Fe^{4+} state can only be obtained by exposure to hydrogen peroxide.

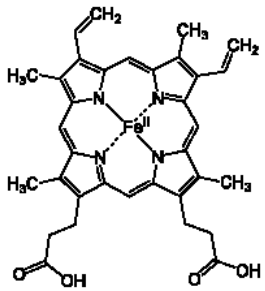


Figure 2: Structure of heme group in the Fe^{2+} (ferrous) oxidation state

2 Globins and their function

It is assumed that the early evolution of a globin ancestor that was able to react with gaseous ligands was accompanied by the acquisition of several functions including the detection, sequestration and detoxification of oxygen and oxygen-derived species (e.g nitric oxide and carbon oxide). Globins, early in the evolutionary tree, are mostly enzymes or less frequently, sensors. Later in evolution, when larger animals emerged, oxygen supply by simple diffusion became inefficient and in addition to a circulatory system and specialized respiratory organs, the former enzyme/sensor-globins evolved in proteins with a function in oxygen transport and/or storage[14].

2.1 Oxygen transport and storage

Most known globins fulfil respiratory functions, supplying the cell with adequate amounts of oxygen for aerobic energy production via the respiratory chain. The red color of blood results from the high amount of Hb in the erythrocytes. The hetero-tetrameric Hb transports oxygen from the respiratory surfaces (usually lungs, gills or skin) to the inner organs via the circulatory system. In the venous blood, deoxygenated Hb also carries carbon dioxide, which is bound to the N-terminal amino groups of the protein chains, for release at the respiratory surfaces. Mb on the other hand is a monomeric globin that is predominantly expressed in cardiac and striated muscle. The affinity of Mb for oxygen is higher than that of Hb, allowing the extraction of oxygen from the blood. Within the muscle cells, Mb enhances oxygen supply by facilitating the diffusion to the mitochondria or by storing oxygen for short-term or long-term hypoxic periods.

2.2 Role in nitric oxide metabolism

The reactions with nitric oxide (NO) are common to all hemeproteins and likely represent one of the earliest and most important globin functions[14]. (Fig3) For example, under temporary hypoxic conditions, deoxy-Hb mediates NO release which triggers blood vessel dilatation. Also deoxy-Mb may act as a nitrite reductase producing NO from NO_2^- in response to cellular hypoxia. On the other hand, oxygenated Mb is also instrumental in the decomposition of toxic NO, present in cells of high metabolic activity, by its dioxygenase enzyme activity. In pathogenic microorganisms, FHb provides protection from human macrophage NO-mediated killing and promotes the virulence of bacteria, e.g. *Salmonella*, and of the yeasts *Candida* and *Cryptococcus neoformans*, a worldwide pathogen causing pulmonary infection in animals and humans. The NO dioxygenase activity of FHbs is more than 20-fold higher than that of either Hb or Mb.[15]

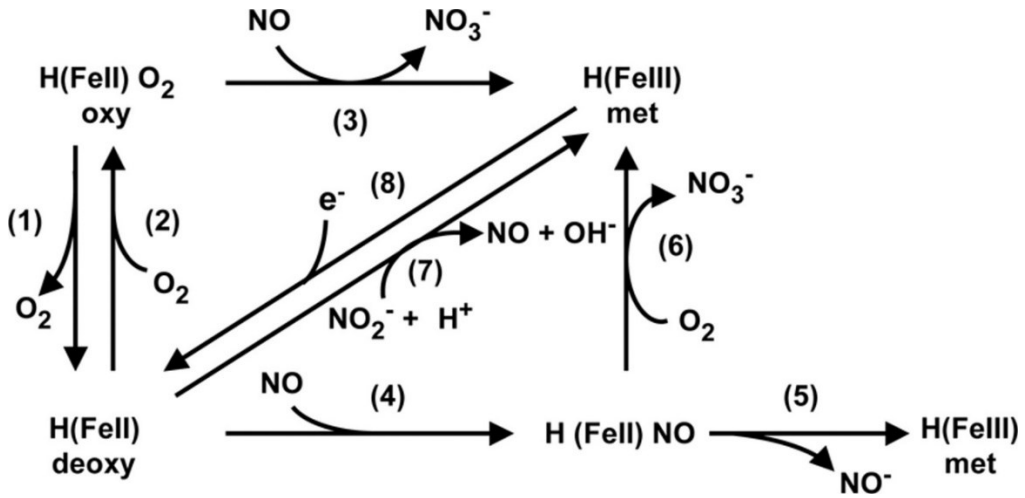


Figure 3: Reactions at heme group (H). (1) deoxygenation (2) oxygenation (3) NO dioxygenation (4) nitrosylation (5) NO reduction (6) O_2 nitrosylation (heme denitrosylation) (7) nitrite reduction (8) MetHb reduction. Figure from[14]

2.3 Oxygen sensing

Members of the GCS family are widespread among Bacteria and Archaea. GCS sensors are often chimeric multi-domain proteins responsible for the cellular adaptive responses to environmental changes. The signal transduction is mediated by the sensing capability of the globin domain, which transmits a usable signal to the cognate transmitter domain, responsible for providing the adequate answer. The GCSs can be divided into two subclasses: I) the aerotactic sensors; globins with a C-terminal domain playing a role in the aerotaxis of the bacterium (movement toward or away from oxygen) and II) the gene regulators; globins having C-terminal domains, some of which may regulate second messengers and others that have unknown functions.[16]

2.4 Peroxidase activity and reaction with free radicals

Hb and Mb can, under certain conditions, exhibit a peroxidase-like enzymatic activity and can catalyze the oxidation of biological molecules using hydrogen peroxide [14]. The first step is the formation of the ferric form either by autoxidation or by reaction with hydrogen peroxide. Next, the ferric form reacts with peroxides to produce a ferryl (Fe^{4+}) radical. This ferryl radical subsequently may oxidize appropriate substrates, e.g. membrane lipids. The oxidation of lipids produces active radicals, e.g. peroxyl and

hydroperoxides, which can react with the ferri (Fe^{3+}) Hb or Mb to form additional ferryl (Fe^{4+}) species. Hence, the Hb and Mb peroxidase activities initiate a cascade of lipid oxidations, particularly with polyunsaturated fatty acids, generating prostaglandin-like molecules with potent vasoconstrictor activity.

Some globins are suggested to play a role in the scavenging of reactive oxygen species (ROS), e.g. Ngb[14,17]. It has been suggested that Ngb plays a neuroprotective role during hypoxic stress. Its up-regulation in hypoxia, followed by a decrease upon re-oxygenation, implies scavenging of ROS.[18,19] It has been shown *in vitro* that Ngb has direct and distinct antioxidant capacity and can efficiently scavenge a variety of free radicals including superoxide anion, hydrogen peroxide and hydroxyl radical[20].

2.5 Role of hexa-coordinated globins

The function of hexa-coordinated globins, like Ngb and Cygb, remains in many cases unknown. They most probably exhibit another role than binding and releasing ligands (like most penta-coordinated globins). Hexa-coordinated globins facilitate electron transfer and are present in very low concentrations. Ligand binding to hexa-coordinated globins could trigger conformational and redox changes that regulate interaction with other signaling molecules [13]. One example of a hexa-coordinated globin with a well characterized function is Globin 12 (GLB 12) of *Caenorhabditis elegans*. It has been shown that GLB 12 participates in electron transfer and interacts with oxygen to generate superoxide. The produced superoxide serves as a biological messenger in signaling pathways involved in reproduction in *C.elegans*.[21]

3 The discovery of androglobin

In 2012 extensive *in silico* searches in deuterostome sequence data led to the discovery of a novel family of large, chimeric proteins containing putative calpain-like and globin-like domains[9]. These chimeric sequences were documented in a phylogenetically diverse array of metazoan taxa (including humans) and in choanoflagellates. This newly discovered chimeric globin family was named androglobin (Adgb), because of its predominant expression in testis tissue of mice and human. The query that was used in the initial *in silico* database searches was the sequence of human NGB and the polymeric

globin from the sea urchin (*Strongylocentrotus*). The Adgb gene was first found in the genome of the sea urchin and the lancelet. Using the sequences of these initially discovered Adgb genes as query for further BLAST searches, homologues of Adgb could be identified in the genomes of more than 30 Metazoan taxa, including 23 vertebrates, uro-, cephalo-, and hemi-chordates, lophotrochozoa, ecdysozoa, coelenterates, placozoa, and the choanoflagellate *Monosiga brevicollis*, which may represent the closest relative to metazoans (Fig4). Alongside the well-known Hb and Mb, and the more recently discovered Ngb and Cygb, Adgb is now already the fifth characterized human globin.

3.1 Domain structure of androglobin

The Adgb family is defined by its typical modular domain structure. The chimeric proteins characteristically comprise four domains, an N-terminal ~350 residues calpain-like cysteine protease domain, a region of ~300 residues without known motifs/domains, followed by an ~150 residue circularly permutated globin domain, and a second, uncharacterized ~750 residues C-terminal domain. (Fig5A)

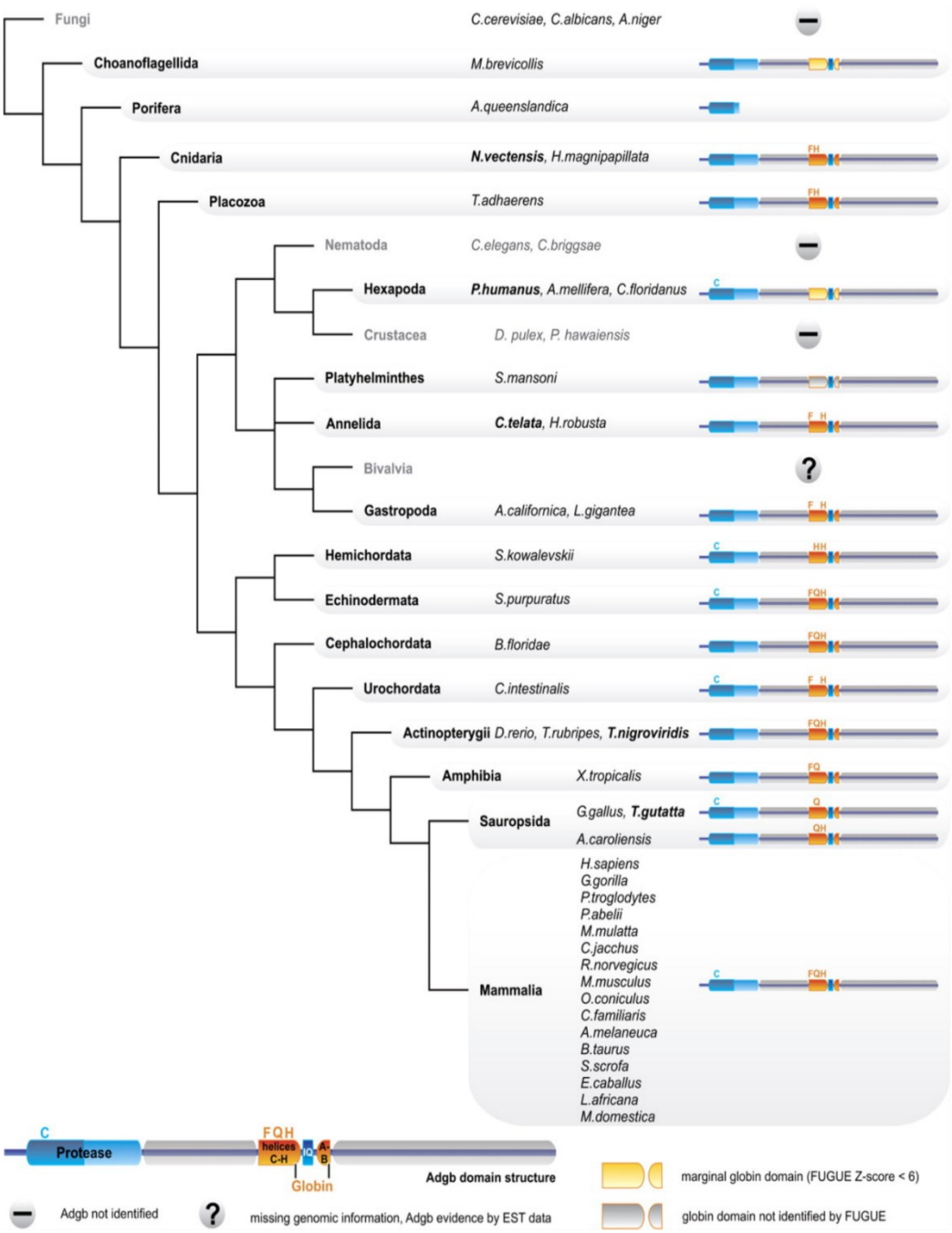


Figure 4: Diagrammatic representation of the phylogenetic distribution of Adgb orthologues. The Adgb domain structure, the conserved Cys (C) residues of the calpain-like domains and essential residues of the globin domain (CD1, E7, F8) are indicated by the amino acid one-letter code. If a taxon contains different Adgb variants, the Adgb structure is shown for the species displayed in bold letters. Figure from[9]

3.1.1 Globin domain

The globin domain of Adgb was initially recognized by "NCBI Conserved Domain" searches, subsequently the globin domain was manually verified for preservation of the Mb-fold, the pattern of predominantly hydrophobic residues at 37 conserved positions and the invariant His at F8. Highly conserved, functionally important residues such as the Phe in the inter-helical region CD and the proximal His F8, which coordinates the heme iron, are present in most Adgbs (Fig4). Adgb shares, with several other known globins, the substitution of the ligand-binding distal His E7 for a Gln (Fig5B). Interestingly, the globin domain, which normally consists of eight consecutive α-helices (named A-H), is circularly permutated and split into two parts within Adgb. The part containing helices A and B has been shifted in the C-terminal direction and is separated from the main globin sequence (helices C–H) by a calmodulin-binding IQ motif (Fig5A). Alignments of the split Adgb globin domain with mammalian Mb, Ngb, and Cygb sequences revealed that Adgb—despite its rearrangement—conforms to the criteria of the "globin fold" tertiary structure (Fig5B). This was confirmed by molecular modelling of the human ADGB globin domain 3D structure, showing that the helix C–H segment alone is able to produce a bona fide globin fold (Fig6).

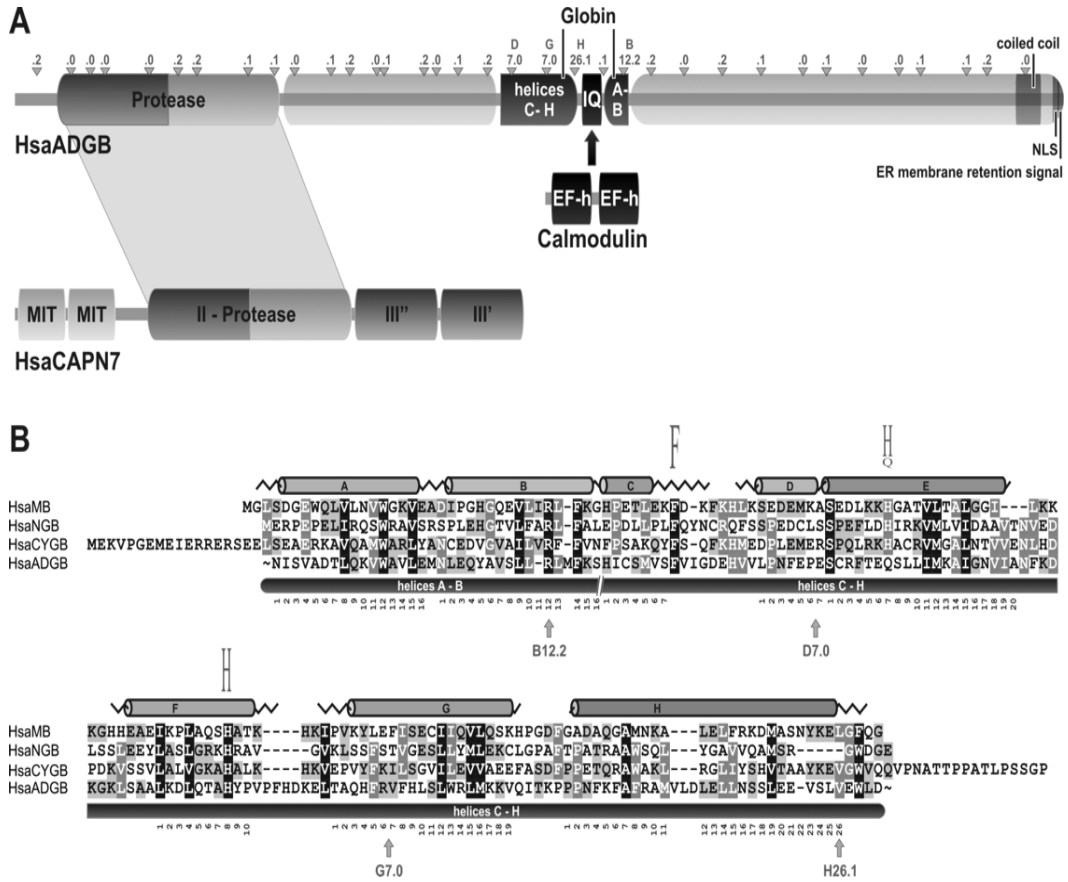


Figure 5: (A) The chimeric domain structure of human ADGB (HsaADGB). The calpain-like protease domain, the rearranged globin domain, the IQ motif, a C-terminal coiled-coil region, and candidate nuclear localization (NLS) and ER membrane retention signals are indicated. The ADGB IQ motif may mediate binding to calmodulin, which reacts to Ca^{2+} levels via its EF hand (EFh) motifs. Small triangles indicate the position of introns in the human ADGB gene and intron phases are given (e.g., '.1' indicates insertion of an intron in phase 1 between codon positions 1 and 2). For comparison, the structure of human calpain-7 (HsaCAPN7) is shown below (MIT, microtubule-interacting and trafficking domain; III'' and III', domains moderately similar to the Ca^{2+} binding domain of other calpains). (B) Amino acid sequence alignment of the concatenated human ADGB globin domain (helices A/B plus C-H) with human NGB, CYGB and Mb. The globin α-helical structure is drawn on top of the alignment. The functionally conserved Phe (F) in the C-D region and the proximal and distal His (H) residues at positions F8 and E7 are indicated. Note that ADGB contains a Gln (Q) instead of the distal His. Gray-scale shading indicates conservation of iso-functional amino acid residues. Intron positions within the globin domain of the ADGB gene are shown by arrows below the alignment. Figure from[9]

3.1.2 Molecular modeling of the human androglobin domain

The 3D structure of the full circular permutated globin domain as well as the shortened form of the globin domain (C till H helix) were predicted using computer modeling. The modeling of the C-H domain of Adgb resulted in a shortened globin folding of 6 α-helices, with a three-over-three alpha helical sandwich structure. This modeling was aligned with the 3D structure of human CYGB (Fig 6A) and it was found that helix A and D were missing in the C-H Adgb globin domain. Furthermore, the modeling was aligned with the structure of the mini-hemoglobin (miniHb) of *Cerebratulus lacteus* (Fig6B), which also lacks the A and D helix. The backbones of these two structures could not be aligned exactly but the general globin folding was similar to each other (the Adgb C-H globin domain had a longer F and H helix compared to the miniHb). These results indicate the presence of a stable, but short globin domain in ADGB. The 3D modeling of the full circular permutated globin domain showed a globin folding, lacking helix A and D, followed by 5 extra α-helices (Fig6C). The first three α-helices are part of the IQ-motif and the last two helices represent helix A and B of the original globin fold. The missing D helix can be explained by the rearrangement in the CD region, caused by the shift of the first two helices (A and B) towards the C-terminus. This rearrangement causes a shift in the structure, whereby helix D disappears and a new short helix B is created, which ensures the stability of the globin fold. This is not the first time that a missing D helix is observed in globins [22].

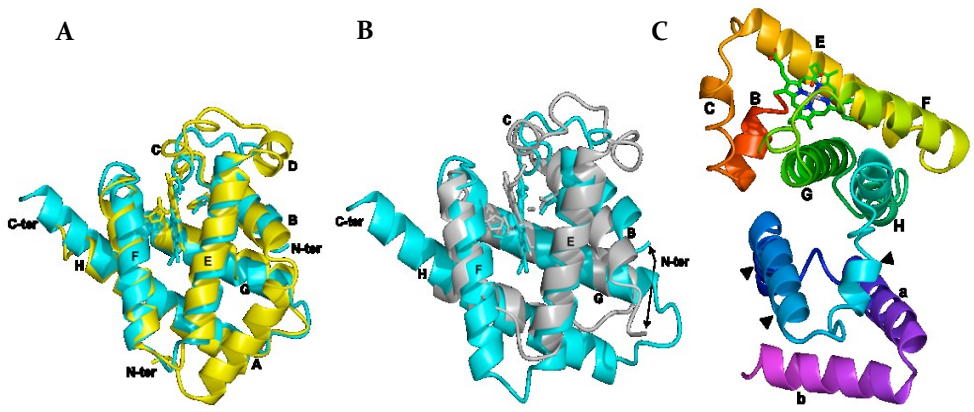


Figure 6: Computational three-dimensional modeling of the ADGB globin domain A) Overlay of the ADGB C-H predicted model (cyan) on the CYGB crystal structure (PDB: 1UT0) (yellow). The ADGB C-H model displays an overall conserved 3-over-3 α-helical globin fold, where the lack of helices A and D can be detected. B) The overlay of the ADGB C-H predicted model (cyan) on CerHb (PDB: 1KR7) (grey) displays a globin fold reminiscent of the so-called "mini-hemoglobins". Indeed, both globin types lack their helices A and D. Helices F and H are both substantially longer in the ADGB C-H model. C) Prediction model of the complete circular permutated ADGB globin sequence. Capital letters indicate the helices of the globin domain; small letters indicate the circular permutated helices A and B; arrows indicate the predicted IQ motif. Figure from[9]

3.2 Calpain domain

The calpain-like region of Adgb is homologues to the catalytic domain of the great subunit of the human calpain 7, a probably ancient member of the calpain family. Calpains are calcium-regulated cytoplasmic cysteine proteases, involved in the intracellular processing of proteins. They are ubiquitously spread in eukaryotes, inclusively in unicellular eukaryotes, fungi, plants and bacteria, but not in archaea. They function as modulators in several cellular processes and are involved in pathological conditions, such as molecular dystrophy, diabetes, multiple sclerosis and neuronal ischemia. The catalytic domain of calpain comprises three important active site residues (Cys, His and Asn) on conserved positions.[23] Alignments with human calpain proteins reveal that only the Cys-active site residue is observed at a strictly matching position in the calpain-like domain of Adgb[9]. However, the calpain domain of Adgb contains several other His and Asn residues at non-standard positions, suggesting that Adgbs may have a cysteine protease activity.

4 Evolutionary tree of androglobin

Phylogenetic reconstructions and the widespread taxonomic distribution clearly indicate that Adgbs are an ancestral protein lineage. It probably traces back to choanoflagellates, the closest unicellular relatives of metazoans, which contain a marginally recognizable globin domain of unknown functionality in Adgb (Fig4). Phylogenetic analyses of the Adgb globin domain and different globin lineages demonstrates that Adgb forms a well-supported monophyletic group, which shows affinity to the Ngb lineage[9]. Within animals, phylogenetic analyses agree that Ngb, GbX and Adgb represent major clades that are distinct from the other globin types, suggesting their early evolutionary origin. In fact, Ngb, GbX and Adgb genes have been identified in both vertebrates and invertebrates, showing that they emerged before the separation of Protostomia an Deuterostomia. The predecessors of Ngb, GbX and Adgb thus formed the minimal globin repertoire of the first animals, which can be taken as an indicator that they had distinct, non-overlapping function. Notably, all three globins are hexa-coordinated, which suggests that hexa-coordination was probably the original coordination state of the Fe atom in the heme group of deoxy-globins. Penta-coordination, which is typical for most globins with respiratory roles like, e.g. Hb and Mb, is therefore the derived state that evolved later.

5 Function of androglobin

Using RT-qPCR, Adgb expression was quantified in eight different mice tissues (heart, brain, testis, liver, lung, spleen, kidney and tongue)[9]. Expression is most abundant in testes, with a tenfold lower expression in lung, and an ~100-fold lower expression in the remaining tissues (Fig7A). This result was strongly supported by an *in silico* expression analysis of different human tissues using the NCBI UniGene[24], ONCOMINE[25], and R2 database[26].

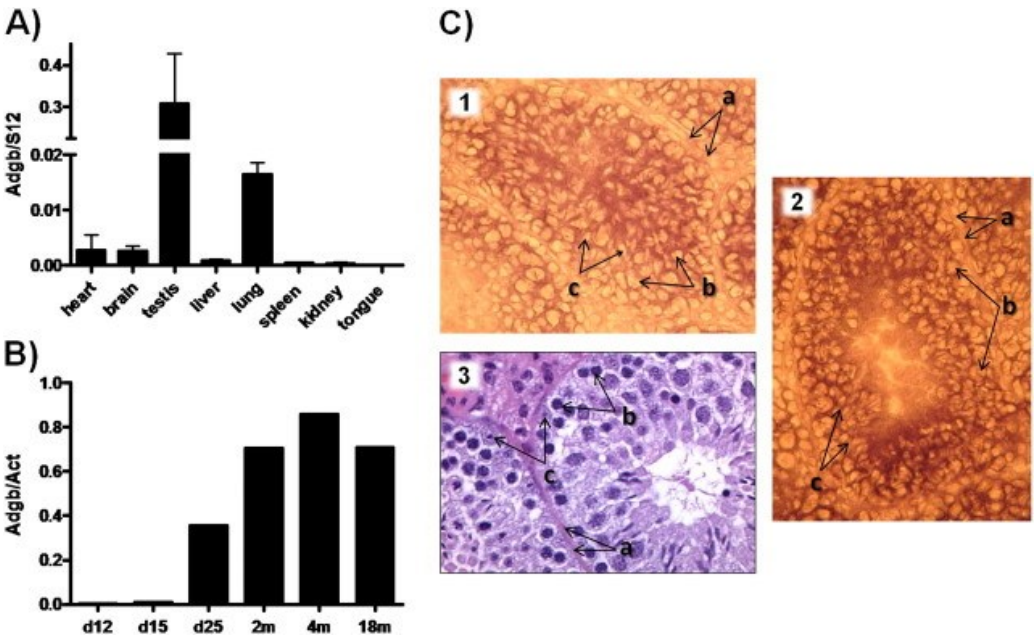


Figure 7: Adgb mRNA expression analysis. (A) Quantitative RT-PCR showing Adgb expression in normal mouse tissues. Bars indicate normalized mRNA levels with standard errors of three different mice. (B) Quantitative RT-PCR of Adgb in testes of younger and older mice (d = days, m = months after birth). (C) In situ hybridization of Adgb antisense RNA to mouse testis cryosections (1,2). Dark signal is concentrated towards the lumen of the seminiferous tubules. A hematoxylin/eosin-stained section (3) is shown for comparison. Smooth muscle cells (a), spermatogonia (b), and Sertoli cells (c) are indicated (magnification 40-fold). Figure from[9]

Adgb expression was analyzed during mouse testis development and an increase of Adgb expression was observed at postnatal day 25 when post-meiotic spermatids are abundant and this persisted into adulthood (Fig7B). Very low or almost no expression was detected in Leydig cells, Sertoli cells, spermatogonia and spermatocytes, indicating again that Adgb plays an important role in the late phases of spermatogenesis. mRNA in situ hybridization of an Adgb antisense RNA probe to mouse testis cryosections showed pronounced signals toward the lumen of the seminiferous tubes, confirming the cellular specificity and expression preference in late spermatogenesis (Fig7C). Furthermore, R2 database analysis suggested 4-fold higher expression levels in fertile vs. infertile males[27], indicating that ADGB plays an important role in spermatogenesis.

5.1 Knockout mouse model for androglobin

A knockout mouse model was created by the research group of David Hoogewijs at the university of Zürich (not yet published). Male Adgb-deficient mice are infertile. Analysis of these mice demonstrates absence of mature spermatozoa and developing elongating spermatids in the lumen of seminiferous tubules (Fig 8), indicating an Adgb-dependent arrest of spermatogenesis prior to spermatid differentiation at the round haploid spermatid stage.

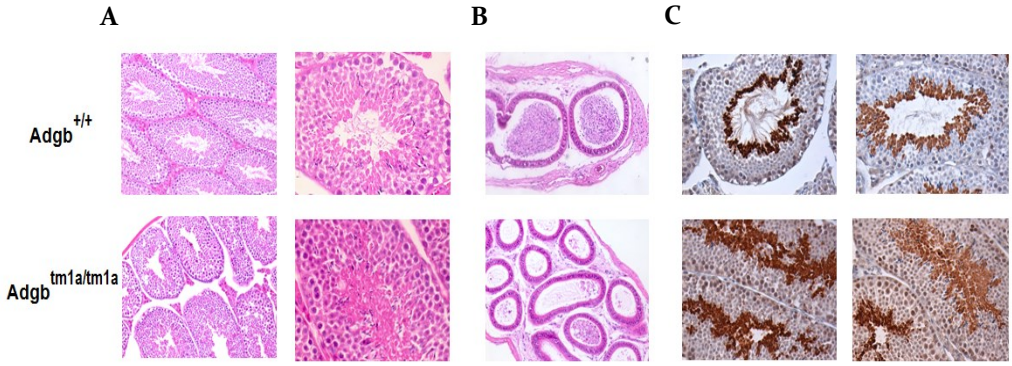


Figure 8: Histological analysis of testis (A) and epididymis (B) sections from $Adgb^{+/+}$ and $Adgb^{tm1a/tm1a}$ mice. (C) AQP-1l IHC specifically staining spermatids and residual bodies.

6 Spermatogenesis

Spermatogenesis is the process by which the male gametes, called spermatozoa, are created. This process occurs in the seminiferous tubules of the testes. Mature spermatozoa are derived from germ cells through a series of complex transformations. When seminiferous tubules are viewed in cross section, the least mature cells are located adjacent to the basement membrane, whereas the most differentiated germ cells are located nearest the lumen (Fig9). The primordial germ cells migrate into the gonad during embryogenesis,: these cells become immature germ cells or spermatogonia. Beginning at puberty and continuing thereafter throughout life, these spermatogonia divide mitotically.

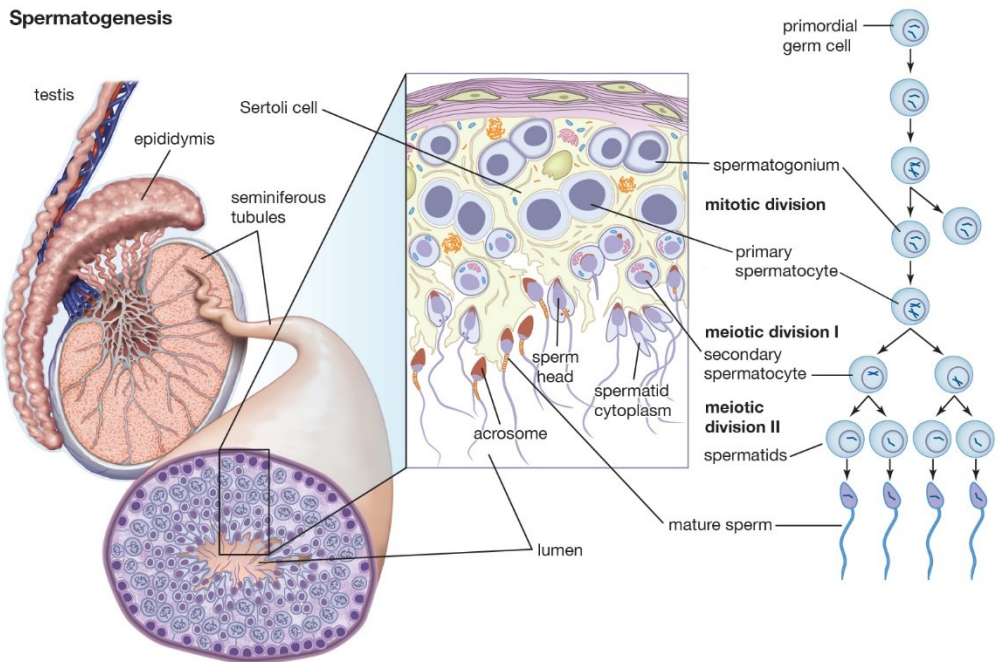


Figure 9: Schematic representation of the process of spermatogenesis in human testis. Figure was taken from Encyclopedia Britannica Inc. 2010 [28].

The spermatogonia have the normal diploid complement of 46 chromosomes (2N): 22 pairs of autosomal chromosomes plus one X and one Y chromosome. Some of the spermatogonia enter into their first meiotic division and become primary spermatocytes. At the prophase of this first meiotic division the chromosomes undergo crossing over. Paired homologous chromosomes are brought into closer juxtaposition by a mechanisms that depends on the programmed double-strand DNA breaks that occur in sister chromatids. A synaptonemal complex is formed in which homologous chromosomes become tightly linked to each other. The morphological changes that occur during the pairing of meiotic chromosomes are the basis for dividing prophase I into five sequential stages – leptotene, zygotene, pachytene, diplotene and diakinesis. As shown in Fig 10 prophase I starts with leptotene when homologs condense and pair, and genetic recombination begins. At zygotene the synaptonemal complex begins to assemble in local regions along the homologs; assembly initiates at sites where the homologs are closely associated and recombination events are occurring. At pachytene, the assembly process is complete, and the homologs are synapsed along their entire lengths. The pachytene stage can persist for days or longer, until desynapsis begins at diplotene with the

disassembly of the synaptonemal complexes and the concomitant condensation and shortening of the chromosomes. The homologs are now ready to begin the process of segregation. Prophase I ends with diakinesis, the stage of transition to metaphase I. (Alberts et al. [29])

At prophase I stage (=primary spermatocytes), each cell has a duplicated set of 46 chromosomes (4N): 22 pairs of duplicated autosomal chromosomes, a duplicated X chromosome and a duplicated Y chromosome.

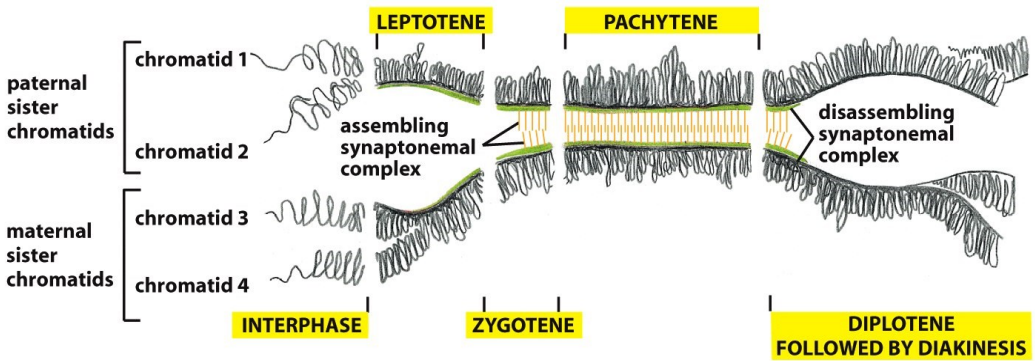


Figure 10: Homolog synapsis and desynapsis during the different stages of prophase I. At leptotene, the two sister chromatids coalesce, and their chromatid loops extend out together from a common axial core. The synaptonemal complex begins to assemble focally in early zygotene. Assembly continues through zygotene and is complete in pachytene. The complex disassembles in diplotene. Figure from [29]

After completing this first meiotic division, the daughter cells become secondary spermatocytes, which have a haploid number of duplicated chromosomes (2N): 22 duplicated autosomal chromosomes and either a duplicated X or a duplicated Y chromosome. These secondary spermatocytes enter their second meiotic division almost immediately. This division results in smaller cells called spermatids, which have a haploid number of unduplicated chromosomes (1N). Spermatids form the inner layer of the epithelium and are found in rather discrete aggregates inasmuch as the cells derived from a single spermatogonium tend to remain together and differentiate synchronously. Spermatids transform into spermatozoa in a process called spermiogenesis, which involves cytoplasmic reduction and differentiation of the tail pieces. Thus, as maturation progresses, developing male gametes decrease in volume. Conversely, maturation leads

to an increase in cell number, with each primary spermatocyte producing four spermatozoa, two with an X chromosome and two with a Y. (Boron and Boulpaep [30])

Ejaculated mammalian sperm are initially not competent to fertilize an oocyte. They must first be modified by conditions in the female reproductive tract. Because it is required for sperm to acquire the capacity to fertilize an oocyte, the process is called capacitation. Capacitation takes about 5–6 hours in humans and is completed only when the sperm arrive in the oviduct. The sperm undergo extensive biochemical and functional changes, including changes in glycoproteins, lipids, and ion channels in the sperm plasma membrane and a large change in the resting potential of this membrane (the membrane potential moves to a more negative value so that the membrane becomes hyperpolarized). Capacitation is also associated with an increase in cytosolic pH, tyrosine phosphorylation of various sperm proteins, and the unmasking of cell-surface receptors that help bind the sperm to the zona pellucida. Capacitation alters two crucial aspects of sperm behavior: it greatly increases the motility of the flagellum, and it makes the sperm capable of undergoing the acrosome reaction. (Alberts et al. [29])

During this doctoral thesis the molecular function of Adgb will be investigated in the context of these complex processes of spermatogenesis and capacitation.

7 References

1. Vinogradov, S. N. *et al.* A phylogenomic profile of globins. *BMC Evol. Biol.* **6,** 31 (2006).

2. Burmester, T., Weich, B., Reinhardt, S. & Hankeln, T. A vertebrate globin expressed in the brain. *Nature* **407,** 520–523 (2000).

3. Burmester, T., Ebner, B., Weich, B. & Hankeln, T. Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. *Mol. Biol. Evol.* **19,** 416–421 (2002).

4. Kawada, N. *et al.* Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells. *J. Biol. Chem.* **276,** 25318–25323 (2001).

5. Trent, J. T. 3rd & Hargrove, M. S. A ubiquitously expressed human hexacoordinate hemoglobin. *J. Biol. Chem.* **277,** 19538–19545 (2002).

6. Kugelstadt, D., Haberkamp, M., Hankeln, T. & Burmester, T. Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken. *Biochem. Biophys. Res. Commun.* **325,** 719–25 (2004).

7. Fuchs, C., Burmester, T. & Hankeln, T. The amphibian globin gene repertoire as revealed

by the Xenopus genome. *Cytogenet. Genome Res.* **112,** 296–306 (2006).

8. Roesner, A., Fuchs, C., Hankeln, T. & Burmester, T. A Globin Gene from Lower Vertebrates of Ancient Evolutionary Origin?: Evidence for Two Distinct Globin Families in Animals.

9. Hoogewijs, D. *et al.* Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes. *Mol. Biol. Evol.* **29,** 1105–14 (2012).

10. KENDREW, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181,** 662–666 (1958).

11. Wittenberg, J. B., Bolognesi, M., Wittenberg, B. A. & Guertin, M. Truncated hemoglobins: A new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. *J. Biol. Chem.* **277,** 871–874 (2002).

12. Vinogradov, S. N. *et al.* Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 11385–9 (2005).

13. Kakar, S., Hoffman, F. G., Storz, J. F., Fabian, M. & Hargrove, M. S. Structure and reactivity of hexacoordinate hemoglobins. *Biophys. Chem.* **152,** 1–14 (2010).

14. Vinogradov, S. N. & Moens, L. Diversity of globin function: enzymatic, transport, storage, and sensing. *J. Biol. Chem.* **283,** 8773–7 (2008).

15. Bonamore, A. & Boffi, A. Flavohemoglobin: Structure and reactivity. *IUBMB Life* **60,** 19–28 (2008).

16. Germani, F., Moens, L. & Dewilde, S. in 1–47 (2013). doi:10.1016/B978-0-12-407693-8.00001-7

17. Hankeln, T. *et al.* Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *J. Inorg. Biochem.* **99,** 110–9 (2005).

18. Herold, S., Fago, A., Weber, R. E., Dewilde, S. & Moens, L. Reactivity Studies of the Fe (III) and Fe (II) NO Forms of Human Neuroglobin Reveal a Potential Role against Oxidative Stress *. **279,** 22841–22847 (2004).

19. Fordel, E., Thijs, L., Moens, L. & Dewilde, S. Neuroglobin and cytoglobin expression in mice Evidence for a correlation with reactive oxygen species scavenging. **274,** 1312–1317 (2007).

20. Li, W. *et al.* scavenger. 115–125 (2010). doi:10.1002/prot.22863

21. De Henau, S. *et al.* A redox signalling globin is essential for reproduction in Caenorhabditis elegans. *Nat. Commun.* **6,** 8782 (2015).

22. Pesce, A. *et al.* The 109 residue nerve tissue minihemoglobin from Cerebratulus lacteus highlights striking structural plasticity of the alpha-helical globin fold. *Structure* **10,** 725–35 (2002).

23. Goll, D. E., Thompson, V. F., Li, H., Wei, W. E. I. & Cong, J. The Calpain System. **1990,** 731–801 (2003).

24. UniGene. Available at: https://www.ncbi.nlm.nih.gov/unigene. (Accessed: 29th October 2017)

25. Rhodes, D. R. *et al.* ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia* **6,** 1–6 (2004).

26. R2. Available at: https://hgserver1.amc.nl/cgi-bin/r2/main.cgi. (Accessed: 29th October 2017)

27. Platts, A. E. *et al.* Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Hum. Mol. Genet.* **16,** 763–73 (2007).

28. Encyclopedia Britannica. (2010). Available at: https://www.britannica.com/science/spermatogenesis. (Accessed: 26th December 2017)

29. Alberts, B. *et al. Molecular Biology of The Cell*. (2008).

30. Boron W. F. and Boulpaep E.L. *Medical Physiology updated edition*. (2005).

II. Aim of the thesis and study outline

Androglobin (Adgb) is now already the fifth mammalian globin type. This newly discovered chimeric globin is specifically expressed in testis tissue and it plays a crucial role in spermatogenesis, as Adgb-deficient male mice are infertile and display a maturation arrest in the late phases of spermatogenesis. Although we know that Adgb plays an essential role in male fertility, nothing is known about its molecular function. The general aim of this thesis is to unravel this molecular function of Adgb. This will be done through different strategies: the function of Adgb will be investigated on a biochemical, cellular and biological level. In obtaining this general aim we have four objectives:

Objective 1: Obtaining a clear view of the up-to-now characterized molecular defects underlying male infertility.

This will form an instrument to investigate the molecular function of Adgb in the context of spermatogenesis and male fertility.

Objective 2: *In vitro* biochemical characterization of the globin domain of human androglobin (ADGB).

The *in vitro* biochemical characterization of the ADGB globin domain will reveal information about its heme environment, ligand binding kinetics, redox activity, three-dimensional structure, etc. The obtained structural and functional information can be compared with other members of the globin family, of which the function is already known and this will eventually lead to a better understanding of the molecular (*in vitro*) function of Adgb. In order to perform this *in vitro* characterization, a recombinant form of the globin domain will be expressed in an overexpressing system, as was done for many other globins before.

Objective 3: Determining the cellular protein-interaction partners of Adgb in testis tissue.

The characterization of the protein-interaction partners of Adgb in testis tissue will reveal in which cellular pathways Adgb is involved in. These protein-protein interactions can subsequently further be investigated *in vitro*, in order to unravel the molecular function of Adgb in spermatogenesis.

Objective 4: Investigating the role of ADGB in human male infertility and exploring the clinical applications for ADGB.

Finally, we want to gain more insight in the role of ADGB in human male infertility and we will investigate the putative clinical applicability of ADGB as a diagnostic biomarker. This will be done by examining the expression pattern of ADGB in human sperm samples and testis tissue of fertile and infertile men.

III. A search for molecular mechanisms underlying male idiopathic infertility

An Bracke[1], Kris Peeters[2], Usha Punjabi[2], David Hoogewijs[3], Sylvia Dewilde[1]

[1]Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, Antwerp 2610, Belgium
[2]Center for Reproductive Medicine, University Hospital Antwerp, Wilrijkstraat 10, Edegem 2650, Belgium
[3]Department of Medicine/Physiology, University of Fribourg, Chemin du Musée 5, CH 1700 Fribourg, Switzerland

Published in Reproductive BioMedicine Online (https://doi.org/10.1016/j.rbmo.2017.12.005)

1 Introduction

According to the World Health Organization (WHO), the definition of infertility is "the inability of a sexually active, non-contracepting couple to achieve spontaneous pregnancy in one year" [1]. About 15% of all couples are infertile and seek medical treatment for fertility and in 50% a male-infertility-associated factor is found together with abnormal semen parameters[2]. Known causes of male infertility are: congenital or acquired urogenital abnormalities, malignancies, urogenital tract infections, increased scrotal temperatures, endocrine disturbances, immunological factors and some characterized genetic abnormalities. However, in 30 to 40% of the cases, the etiology of male infertility remains unknown and it is called idiopathic male infertility[2]. Despite the fact that these men have no history of diseases affecting fertility and show normal findings of physical examinations and endocrine, genetic and biochemical laboratory testing, their semen analysis often reveals abnormal parameters. The most common abnormalities in a routine semen analysis are: absence of spermatozoa (azoospermia); very low sperm count in the ejaculate (oligozoospermia); abnormal sperm morphology (teratozoospermia) and/or abnormal sperm motility (asthenozoospermia). These idiopathic sperm abnormalities are assumed to be caused by several factors, including reactive oxygen species (ROS), unknown genetic and epigenetic abnormalities and endocrine disruption by environmental pollution[3]. According to the human protein atlas, 74% of all human proteins (n=19692) are expressed in the testis and 1980 of these genes display an elevated expression in testis compared to other tissue types. An analysis of the genes with elevated expression in the testis shows that most of the corresponding proteins are involved in spermatogenesis[4–7]. Spermatogenesis covers a complex network of processes that occur in the seminiferous tubules of the testis. The subsequent processes of spermatogenesis can be described as: 1) proliferation of spermatogonia; 2) spermatogonial differentiation into spermatocytes; 3) meiotic division of round spermatids; and 4) the release of highly specialized mature spermatozoa into the testicular tubule lumen[8]. Mutations in any of these genes may directly or indirectly lead to abnormalities in one of the spermatogenic processes and may subsequently result in male infertility. Mutations that have a direct causal-effect on fertility are not transmitted, which renders appropriate characterization

of the genetic etiology of male infertility difficult. Notwithstanding these difficulties, many studies were undertaken to unravel the genetic and molecular background underlying the different male infertility phenotypes (azoospermia, oligozoospermia, teratozoospermia and asthenozoospermia).

The condition with an absence of spermatozoa in the ejaculate is referred to as **azoospermia** and is identified in 11% of infertile men. In 40% of the cases azoospermia is caused by a physical blockage to the male excurrent ductal system (= obstructive azoospermia, OA) and here the etiology is mostly known. The remaining cases are called non-obstructive azoospermia (NOA) and are frequently associated with testicular failure, characterized by a smaller testis volume and elevated luteinizing hormone and follicle-stimulating hormone levels.[9] However, often the cause of NOA remains unknown and might be due to genetic abnormalities. These genetic defects may result in a total absence of spermatogenesis (also named "Sertoli Cell Only (SCO) syndrome") or in a maturation arrest of spermatogenesis.

Ejaculate sperm count of less than 15 million/ml is termed **oligozoospermia**[1]. A recently published study on 1737 patients with reduced total sperm counts revealed that in 60% of the cases the primary causal factor could not be assigned and that almost 75% of all oligozoospermic cases remained idiopathic[10].

Teratozoospermia is a condition where more than 96% of the spermatozoa in the sperm sample have an abnormal morphology[1]. Two severe and rare phenotypes of teratozoospermia are recognized: macrozoospermia and globozoospermia. Macrozoospermia is characterized by the presence of a very high percentage of spermatozoa with enlarged and irregular shaped heads and multiple flagella. These spermatozoa display a high rate of polyploidy, many of them are tetraploid. Globozoospermia is characterized by the presence of a large majority of round spermatozoa lacking the acrosome in the ejaculate. These sperm cells are unable to adhere and penetrate the zona pellucida. These phenotypes are monomorphic since all the spermatozoa display the same abnormality. A third, heterogeneous phenotype of teratozoospermia has been described by Ben Khelifa et al.[11] and is called "multiple morphological anomalies of the flagella (MMAF). This phenotype includes spermatozoa with absent, short, bent, and coiled flagella and flagella of irregular width. These

morphological abnormalities of the sperm flagella are also leading to impaired sperm motility (asthenozoospermia).

WHO criteria delineate the normal percentage of motile sperm to be 40%. A lower percentage of motile sperm is referred to as **asthenozoospermia**[1]. Asthenozoospermia represents a common cause of male infertility and is observed in up to 81% of all abnormal semen analyses, frequently combined with oligo- and/or teratozoospermia. Only 18% of the cases are isolated asthenozoospermic (immotile sperm with a normal morphology and sperm count)[12]. Possible causes of decreased sperm motility include prolonged duration of sexual abstinence, unhealthy lifestyle, varicocele and/or infection. However, mostly the etiology of asthenozoospermia remains unknown and may be caused by genetic factors.

In some male infertility cases no abnormalities are detected after conventional semen analysis (sperm count, motility and morphology) and the etiology of infertility remains unclear. These cases are referred to as **normozoospermic infertility**. Mature mammalian spermatozoa require capacitation in the female reproductive tract before binding to and crossing the zona pellucida and finally fusing with the oocyte plasma membrane. Defects in these processes are not detectable during sperm analysis and may represent a possible cause of idiopathic normozoospermic male infertility. At the cell biology level, capacitation induces changes in the sperm motility pattern known as hyperactivated movement and prepares the sperm to undergo an exocytotic process known as acrosome reaction. At the molecular level, capacitation is associated with cholesterol loss from the sperm plasma membrane, increased membrane fluidity, changes in intracellular ion concentrations, hyperpolarization of the sperm plasma membrane, increased activity of the protein kinase A (PKA), and protein tyrosine phosphorylation[13]. Defects in any of these molecular mechanisms may result in male infertility.

In this review we investigate the possible causes of male idiopathic infertility by extensive literature searches of I) causal gene mutations linked to male infertility, II) proteome studies of sperm from idiopathic infertile men, III) the role of epigenetics, IV) post-translational modifications and V) sperm DNA fragmentation in infertile men. In this way we created an extensive overview of the knowledge obtained during the last decades about idiopathic infertility and its underlying molecular mechanisms. We believe that

this information is important to properly understand the etiology of male infertility and to develop adjusted genetic screenings and biomarkers towards a better diagnosis and more personalized treatment of male infertility.

2 Genetic causes of male infertility

Idiopathic infertility may often have a genetic basis. However, there is still little known about the mutations and the genes involved in male infertility. Initially, searches for genetic risk factors were done through a candidate gene approach; single nucleotide polymorphisms (SNPs) of certain, spermatogenic related, genes were studied in azoospermic and oligozoospermic men. Since 2009, novel high-throughput approaches such as genome wide association studies (GWAS), comparative genomic hybridization-arrays (array-CGH) and next generation sequencing (NGS) were carried out. The results of all these genetic studies are reviewed by Tüttelmann et al. and Krausz et al.[14,15]. Despite that these association studies provided interesting data, the Krausz group concluded that the 10-year effort research was largely unsuccessful in identifying recurrent genetic factors with potential clinical application. Even the high-throughput approaches were not able to identify overlapping SNPs in the different studies, confirming again that the underlying genetic defects of male infertility exists of numerous rare genetic events. However, the last ten years a handful of such rare genetic defects linked to male infertility were identified and they were recently published in two separate review papers. One deals with the single gene defects leading to azoospermia and oligozoospermia[16] and the second is focusing on genetic abnormalities leading to teratozoospermia and asthenozoospermia[17].

The genetic causes of male infertility remain still largely unknown at this moment. Recent technological advances like whole genome sequencing (WGS) predicts golden ages for the genetics of infertility, because WGS makes it possible to perform studies of large cohorts of patients that will enable the rapid identification of dozens or perhaps hundreds of new genetic causes of male infertility. It will also be possible to detect all copy number variations and to directly characterize their extent and localization[16].

Below we provide a concise overview of the genetic defects that are proven to have a major impact on male fertility and we will discuss their function related to spermatogenesis and sperm function.

2.1 Genes involved in non-obstructive azoospermia

We will focus on the genes involved in non-syndromic NOA, the etiology of OA is mostly known and will not be discussed here. Also the well-established recurrent genetic causes of NOA and SO, like Y chromosome microdeletions at the azoospermic factor (AZF) locus and other chromosomal abnormalities such as Klinefelter (47, XXY), are beyond the scope of this review.

Mitchell et al. reviewed the few characterized genetic causes of NOA[16]. One of the rare successes in identifying causal mutations has been the X-chromosome meiotic gene *TEX11*. *TEX11* is the first and thus far only gene in which causal mutations of NOA have been validated by the discovery of distinct loss of function mutations in unrelated men[18,19]. Additionally, five strong candidates for genetic causes of autosomal recessive NOA are identified through studies of consanguineous families with multiple azoospermic sons: *ZMYND15, TAF4B* (implicated in transcriptional regulation during spermatogenesis) and *SYCE1, MCM8* and *TEX15* (implicated in the synapsis of homologous chromosomes and meiotic homologous recombination) (Fig1) [20–23]. It has to be mentioned that these mutations are extremely rare and that for strict validation of these variants as monogenic causes of NOA, second independent cases are required.

SYCE1 is a component of the central element of the synaptonemal complex. The synaptonemal complex is a proteinaceous structure which physically links the pairs of sister chromatids and is visualized by electron microscopy as a zipper like structure with two lateral elements and the central element in between. When *Syce1* is disrupted in mice it results in incomplete DNA repair and in the absence of the synaptonemal complex and crossovers. *Syce1* null mice are azoospermic, displaying a maturation arrest at zygotene stage.[24]

The disruption of ***Tex11*** in mice causes chromosomal asynapsis and reduced crossover formation. A subset of chromosomes is completely asynapsed in Tex11 deficient

spermatocytes, and other chromosomes are synapsed. Spermatocytes with asynapsed autosomes undergo apoptosis at the pachytene stage, while those with only asynapsed sex chromosomes progress. The cells that survive the pachytene stage display chromosome nondisjunction at the first meiotic division, resulting in cell death and male infertility. In addition, it has been shown that Tex11 interacts with Sycp2, an integral component of the synaptonemal complex lateral element. Summarized, it can be suggested that Tex11 promotes initiation and/or maintenance of synapsis and formation of crossovers, and may provide a physical link between these two meiotic processes.[25]

Loss of ***Tex15*** function in mice causes early meiotic arrest in males, with complete lack of pachytene spermatocytes. Specifically, Tex15-deficient spermatocytes exhibit a failure in chromosomal synapsis. In mutant spermatocytes, DNA double-strand breaks are formed, but localization of the recombination proteins Rad51 and Dmc1 to meiotic chromosomes is severely impaired. Based on these data, it has been proposed that Tex15 regulates the loading of DNA repair proteins (like Rad51 and Dmc1) onto sites of DNA double-strand breaks and, thus, its absence causes a failure in meiotic recombination.[26]

Mcm8 forms a complex with Mcm9. The deficiency of these two genes impairs homologous recombination mediated DNA repair during gametogenesis and somatic cells cycles. *Mcm8* null mice are sterile because spermatocytes are blocked in meiotic prophase I at the zygotene stage[27]. It has been shown that the Mcm8-Mcm9 complex promotes Rad51 recruitment at DNA damage sites to facilitate homologous recombination[28]. The function of MCM8 seems to be similar to the function of TEX15.

ZMYND15 is exclusively expressed in spermatids, and acts as a transcription repressor through the recruitment of histone deacetylase enzymes. Inactivation of Zmynd15 in mice results in the disruption of haploid gene expression and consequently in late spermatid depletion, causing azoospermia and complete male infertility.[29]

TAF4B is a TATA-binding protein associated factor, a component of the TFIID complex of RNA polymerase II basal transcription machinery. Taf4b null mice exhibit a unique testicular phenotype that includes normal fertility at early ages followed by a complete loss of fertility by 12 weeks of age, characterized by spermiogenesis defects, loss of germ cells and testicular degeneration. Taf4b is required for mitosis specifically in gonocytes

and spermatogonia. Taf4b appears to be essential for germ cell stem cell differentiation and proliferation, which are required for the maintenance of fertility in adult males.[30]

2.2 Genes involved in teratozoospermia

The monomorphic phenotypes macrozoospermia and globozoospermia are for approximately 80% of the patients caused by a mutation in one single gene, *AURKC* for macrozoospermia and *DPY19L2* for globozoospermia[17]. In addition, study of globozoospermia in a consanguineous family led to the identification of a homozygous mutation in the *SPATA16* gene[31]. Mutations in *SPATA16* are more rare compared to mutations in *DPY19L2*, as they were only identified in two families[31,32]. Macrozoospermia and globozoospermia are however, extremely rare and are only seen very occasionally in *in vitro* fertilization (IVF) centers. The third teratozoospermic phenotype MMAF cannot be explained by one causal gene, the etiology is more heterogeneous. However, the analysis of non-syndromic MMAF patients allowed the identification of mutations in *DNAH1* in approximately 30% of the tested patients[11,33,34].

AURKC is a member of the Aurora subfamily of serine/threonine protein kinases and is a component of the chromosomal passenger complex (Fig2A). The chromosomal passenger complex is a pivotal regulator of mitotic events, including chromosome segregation and cytokinesis. It has essential functions at the centromere in ensuring correct chromosome alignment and segregation and is required for chromatin-induced microtubule stabilization and spindle assembly. Defects of the AURKC protein lead to perturbation in meiosis and to the production of chromosomally unbalanced gametes[35] resulting in a high rate of polyploidy, which is a typical characteristic of macrozoospermia. (Fig2A)

DPY19L2 is a transmembrane protein of the inner nuclear membrane and is necessary to anchor the acrosome to the nucleus. (Fig2B) A knock out mouse model for *Dpy19l2* showed that in the absence of Dpy19L2 the forming acrosome slowly separates from the nucleus before being totally removed from the sperm together with the cytoplasm[36]. (Fig2B)

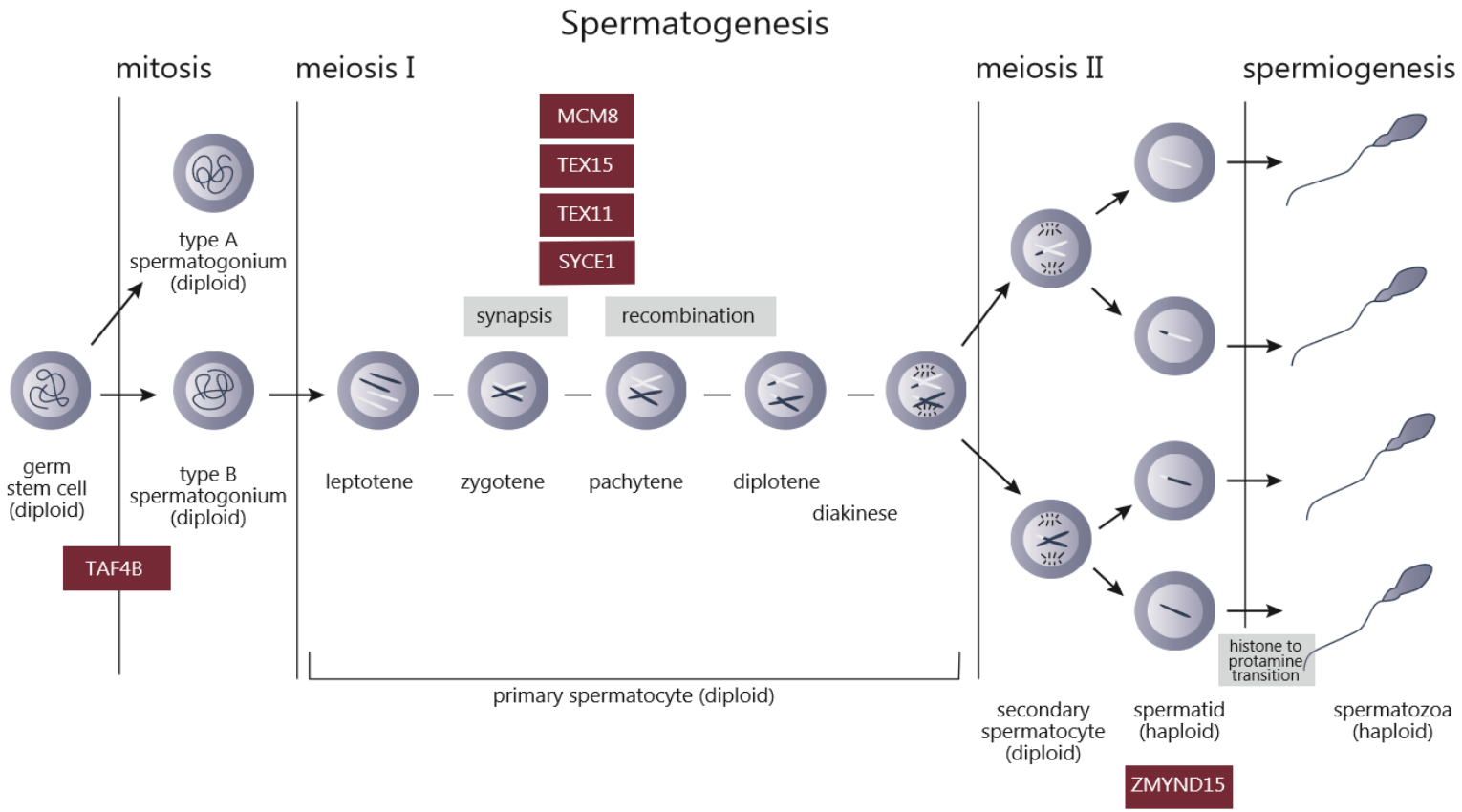


Figure 1: Schematic representation of human spermatogenesis. Six causal gene mutations have been described for azoospermia/oligozoospermia: *TEX11*, *TEX15*, *MCM8*, *SYCE1*, *ZMYND15* and *TAF4B*. These genes are placed in the scheme where they function in the spermatogenic process.

SPATA16 is localized at the Golgi apparatus and is involved in the transport of pro-acrosomal granules to the acrosome in round and elongated spermatids and thereby plays an essential role in the formation of the acrosome during spermatogenesis [37,38]. (Fig2B)

DNAH1 encodes an axonemal inner dynein heavy chain. Electron microscopy examination of spermatozoa of patients with mutations in the *DNAH1* gene show general axonemal disorganization including mislocalization of the microtubule doublets, absence of the central pair in about half of the analyzed cross sections, loss of the inner arm dyneins as well as severe disorganization of the fibrous sheath, the outer dense fibers and the mitochondrial sheath[11]. (Fig2C Red color)

2.3 Genes involved in astheno(terato)zoospermia

Only four genes have been linked to asthenozoospermia through genetic analyses: *CATSPER1*, *CATSPER2*, *SEPTI2* and *SLC26A8*. It has to be mentioned that the identified mutations are rare and not wide-spread.

Three heterozygous missense mutations have been identified in ***SLC26A8*** among three unrelated asthenozoospermic men[39]. *SLC26A8* encodes the testis anion transporter 1, a sperm-specific anion transporter localized to the annulus, a ring-shaped structure located beneath the plasma membrane that connects the midpiece and the principal piece of mammalian sperm flagellum (Fig2C). Slc26a8 null mice are sterile due to complete lack of sperm motility and reduced sperm fertilization potential[40]. SLC26A8 associates with the cystic fibrosis transmembrane conductance regulator (CFTR) and stimulates its activity. CFTR and SLC26A8 cooperate to regulate the anion fluxes required for correct sperm motility and capacitation.[41] Furthermore, this transmembrane protein is thought to play a structural role, anchoring the cytoplasmic components of the annulus to the plasma membrane.[40] (Fig2C Blue color)

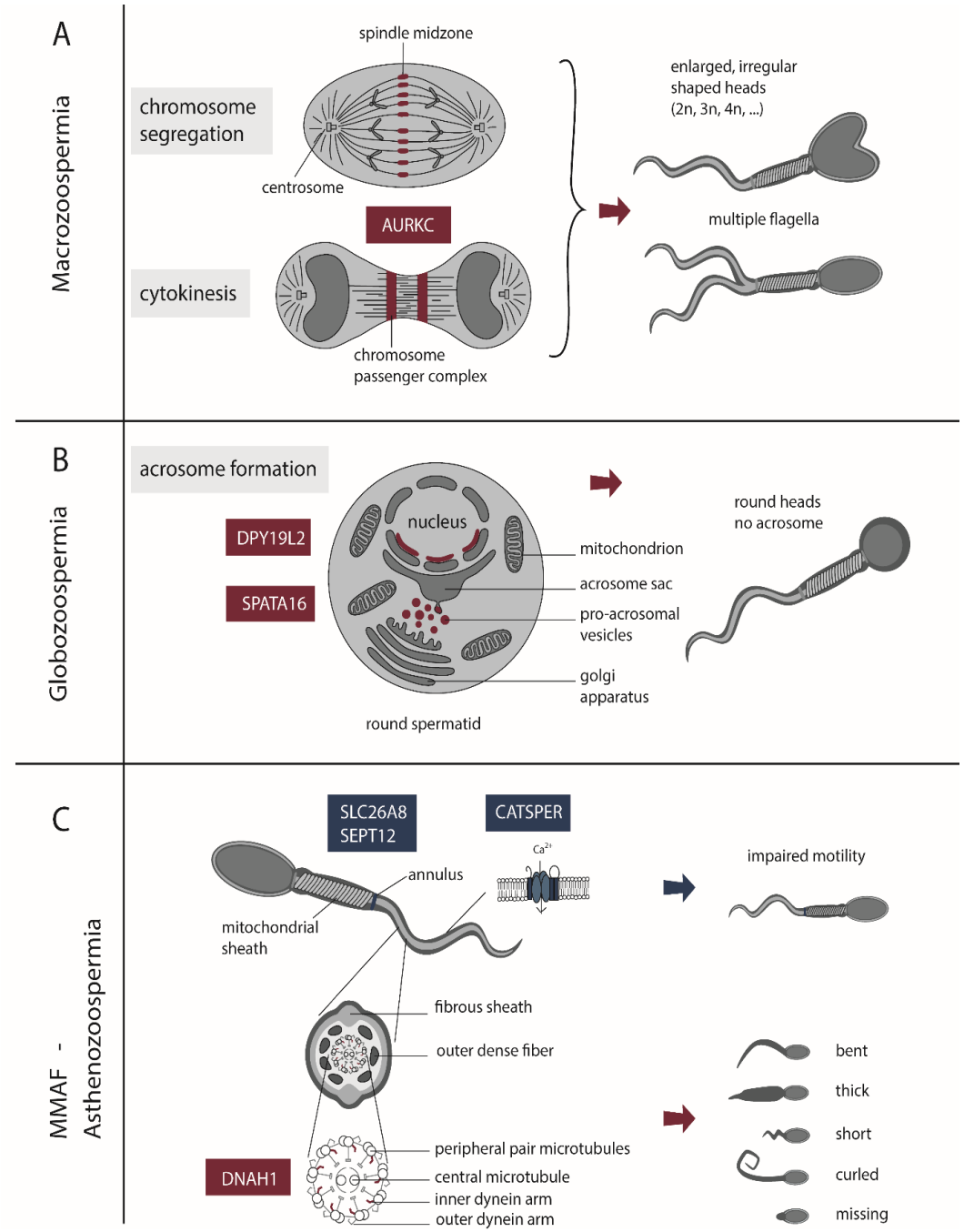


Figure 2: Representation of the genes linked to the three different teratozoospermic (RED) phenotypes and asthenozoospermia (BLUE). A) Macrozoospermia: sperm cells with enlarged and irregular shaped heads and multiple flagella. B) Globozoospermia: round sperm cells lacking the acrosome. C) MMAF: Multiple morphological anomalies of the flagella: spermatozoa with absent, short, bent, and coiled flagella and flagella of irregular width (RED). Asthenozoospermia: sperm cells with impaired motility (BLUE).

Two heterozygous mutations were identified in the ***SEPTI2*** gene in two unrelated patients displaying asthenoteratozoospermia[42]. *SEPTI2* encodes for SEPTIN 12, a GTP-binding protein and is a part of the SEPT ring structure at the sperm annulus[43]. Mice carrying a mutation in the *Septl2* gene that disrupts the Septl2 filament formation display a disorganized sperm annulus, bent tail, reduced motility and loss of the Sept ring structure[44]. *In vitro* studies in the two unrelated asthenozoospermic men, indicated that both mutated proteins impaired filament formation of the wild-type SEPTI2 in a dose-dependent manner; thus acting as dominant negative proteins. Altogether these findings indicate that mutations in SEPTI2 impair sperm structural integrity by disrupting Septin filaments formation at the annulus. (Fig2C Blue color)

CATSPER2 mutations or deletions have been identified in four unrelated families with asthenozoospermic members[45,46]. ***CATSPER1*** mutations have been identified in only one consanguineous family[47]. Catsper stands for 'cation channel sperm associated' and is a channel protein exclusively located to the plasma membrane of the principal piece in the sperm flagellum, where it is required for sperm hyperactivation. It is composed of four separate pore-forming alpha subunits, which are encoded by the *CATSERPI-4* genes, and three additional auxiliary subunits encoded by *CATSPER beta*, *CATSPER gamma* and *CATSPER delta*. The CATSPER channel mediations Ca^{2+} influx, necessary for sperm motility and hyperactivation, in a cAMP-and pH-dependent manner[47]. (Fig2C Blue color)

2.4 Genes involved in normozoospermic infertility

Up-to-now only a single gene has been identified to be mutated in normozoospermic infertile men, *PLCZ1*. Two mutations in the *PLCZ1* gene, on distinct paternal chromosomes, were identified in an infertile patient showing defects in oocyte activation[48]. Moreover, a recent whole exome sequencing was performed in two infertile brothers exhibiting normal sperm morphology with complete fertilization failure after intracytoplasmic sperm injection (ICSI). This led to the discovery of a missense homozygous mutation in *PLCZ1* resulting in the absence of the protein in sperm[49].

PLCZ1 (phospholipase C Zetal) is a member of a large family of enzymes able to bind lipids in membranes where they hydrolyze specific phospholipids, mostly phosphatidylinositol 4,5-bisphosphate (PIP_2). PLCZ1 is a sperm-specific protein and has

been characterized as the mediator of oocyte activation. At fertilization, oocyte activation is triggered by a specific calcium signaling, made of Ca^{2+} oscillations, necessary for embryo development. The cleavage of PIP_2 by PLCZ1 results into inositol 1,4,5-triposphate (IP_3) and diacylglycerol (DAG). IP_3 acts as a secondary messenger that is capable of traveling through the cytoplasm to the endoplasmic reticulum, where it stimulates the release of calcium into the cytoplasm.[50]

3 Differential proteome analyses

Numerous proteome studies have been performed on the human sperm cell. To understand the molecular mechanisms behind male infertility and to find putative biomarkers, many differential proteome studies focused on patients with altered semen analysis (altered sperm count, morphology and/or motility). The most commonly analyzed proteome is that of asthenozoospermic sperm and sperm from normozoospermic infertile men, because it is relatively straightforward to obtain sufficient protein from these phenotypes. Only one comparative proteome analysis between globozoospermic and normal spermatozoa has been reported[51] and resulted in the characterization of 19 differentially expressed proteins (Supplementary Table1). However, it is not possible to draw robust conclusions on targeted pathways and molecular processes in this specific phenotype based on a single study. Hence, we will only discuss the proteome studies performed linked to asthenozoospermia and normozoospermic infertility.

3.1 Proteome of asthenozoospermic sperm

Six differential proteome analyses were performed in asthenozoospermic patients[52–57] and the results of these proteome analyses were reviewed by several authors[58–60]. Most of these proteome studies are based on two-dimensional gel electrophoresis, followed by mass spectrometry and lead to the identification of several altered proteins in asthenozoospermic patients compared to fertile donors. In 2014 Amaral et al. performed a higher-throughput proteome analysis using TMT isobaric peptide labeling followed by LC-MS/MS protein identification and quantification, leading to the identification of more than hundred altered proteins in asthenozoospermic patients[56]. In total, all these proteome analyses in asthenozoospermic patients resulted in a list of 128 unique altered

proteins, listed in supplementary table 1. Three functional classes were mainly represented among these proteins: cytoskeleton, energy metabolism and protein folding and degradation (Fig3). The localization of these functional classes in the sperm cell are indicated in Fig4.

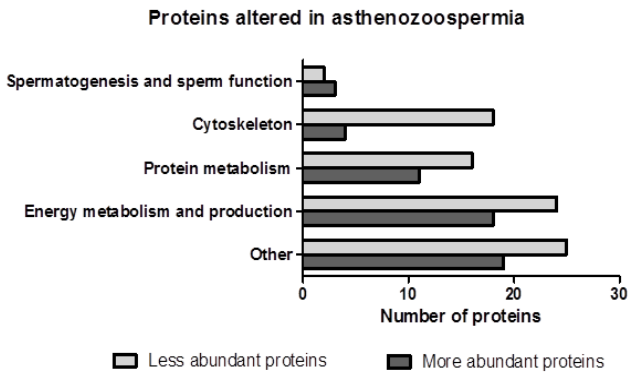


Figure 3: Comparative proteome analyses in asthenozoospermic men resulted in the identification of 128 altered proteins. The protein cellular functions were determined according to the information available at the UniProt website. Proteins were grouped as those that were present at higher abundance and those that were present at lower abundance in sperm from asthenozoospermic men compared to fertile men.

3.1.1 Cytoskeleton

Normal flagellar morphology is critical for sperm cell motility. The flagellum can be divided in four pieces: the connecting piece, the mid-piece, the principal piece and the end-piece. A central axoneme extends throughout the length of all four subdivisions of the flagellum and starts at the remnant of the centriole lying in the connecting piece just behind the nucleus. The axoneme consists of two central singlet microtubules surrounded by a ring of nine microtubule doublets. The active bending of the flagellum is caused by the sliding of adjacent microtubule doublets past on another, driven by dynein motor proteins, which use the energy of ATP hydrolysis. The mid-piece of the flagellum is defined by the presence of nine outer dense fibers (ODFs) surrounding each of the nine outer axonemal microtubule doublets and by a sheath of mitochondria that encloses the ODFs and the axoneme. The mitochondrial sheath is exclusive to the mid-piece. The annulus marks the termination of the mid-piece and the start of the principal piece. In the principal piece, two of the ODFs are replaced by two longitudinal columns

of fibrous sheath (FS), thus reducing the number of ODFs in the principal piece from nine to seven. Finally, the end-piece contains only the axoneme surrounded by the plasma membrane.[61] (Fig4) Several proteome analyses revealed a decreased expression of different components of the flagellum.

ODF2 was decreased in three out of the six proteome analyses[52,55,57]. A knock out mouse model for Odf2 displayed spermatozoa with tail defects, including the absence of one or more axonemal microtubule doublets and bent tails with abnormal motility.[62]

Several tektins (TEKT1, TEKT4 and TEKT5) were also decreased in the proteome of asthenozoospermic men[54–56]. Tektins are filament-forming proteins localized in microtubules in cilia, flagella, basal bodies, and centrioles. In humans five tektins (TEKT1,2,3,4 and 5) have been identified in testis and spermatozoa. They are localized in the flagellum of spermatozoa and show almost all predominant localization at the peri-axoneme structure of the flagella, i.e. the mitochondrial sheath, the outer dense fibers and the fibrous sheath[63]. TEKT3 and TEKT1 were also localized near the acrosome region of the spermatozoa[64,65].

TUBB2B, a major component of the microtubule, was found to be decreased in asthenozoospermic men in the proteome analyses performed by Siva et al.[54] and Hashemitabar et al.[57].

Finally, β-actin (ACTB) has found to be decreased in spermatozoa of asthenozoospermic patients[53,55]. ACTB is one of the major cytoskeletal proteins and participates in many crucial cellular processes like cell motility, cytokinesis, vesicle and organelle movement, cell signaling as well as the establishment and maintenance of cell junctions and cell shape[66].

3.1.2 Energy metabolism

Spermatozoa are highly specialized cells which require a large amount of energy. The beat frequency of the flagellum is directly related to the production rate of energy from ATP[67]. It is generally accepted that ATP is produced in spermatozoa via two main metabolic pathways: glycolysis in the head and principal piece and oxidative phosphorylation in the mitochondria (the mid piece)[68] (Fig 4). The proteome analyses in asthenozoospermic

men reveal that several ATP-producing cellular pathways are required for the flagellar movement of the sperm cell. Enzymes part of glycolysis, tricarboxylic acid cycle, pyruvate metabolism, ketone metabolism, oxidative phosphorylation and beta-oxidation of fatty acids were found to be downregulated in asthenozoospermic men. These findings indicate that several metabolic pathways may contribute to sperm motility regulation.

3.1.3 Protein folding and degradation

Defect of sperm formation during spermatogenesis may also result in impaired motility. Altered expression of proteins involving protein folding and degradation were observed in several proteome analyses in asthenozoospermic men. Most of these proteins were heat shock proteins (HSP) which are molecular chaperones mediating protein folding and preventing protein aggregation. The exact relationship between altered expression of HSPs and impaired motility is not fully understood, but many studies could link HSPs to male infertility[69–71]. HSPA2 was altered in three different proteome analyses, but the results of these studies were not consistent: whereas Amaral et al. and Martínez-Heredia et al. reported increased expression of HSAP2, Siva et al. observed that HSPA2 expression was decreased[53,54,56]. HSPA2 is expressed predominantly in the testis and targeted disruption in mice results in germ cell apoptosis and failed meiosis[69]. In addition to HSPA2, several other HSPs (HSPD1, HSPA9, HSP70 and GRP78) displayed an altered expression pattern in asthenozoospermic men, including both down-regulation and up-regulation. Considering the function of HSPs in preventing protein misfolding during stress conditions (like elevated temperatures), a reduced expression associated with impaired motility seems more likely. Another group of proteins linked to asthenozoospermia are proteasomal proteins. [53,54,56]. The proteasome is a proteinase complex with ATP dependent proteolytic activity that mediates protein turnover of ubiquitinylated proteins. During spermatogenesis many proteins and organelles are degraded, and the ubiquitin-proteasome pathway plays a key role in formation of condensed sperm. Different subunits of the proteasome showed altered expression in spermatozoa with impaired motility (PSMB4, PSMB5, PSMB6 and PSMA3). It has to be mentioned that some of these proteasomal proteins are decreased whereas others are increased in sperm cells with impaired motility. In conclusion, their expression is altered and this seems to be associated with deregulated sperm motility.

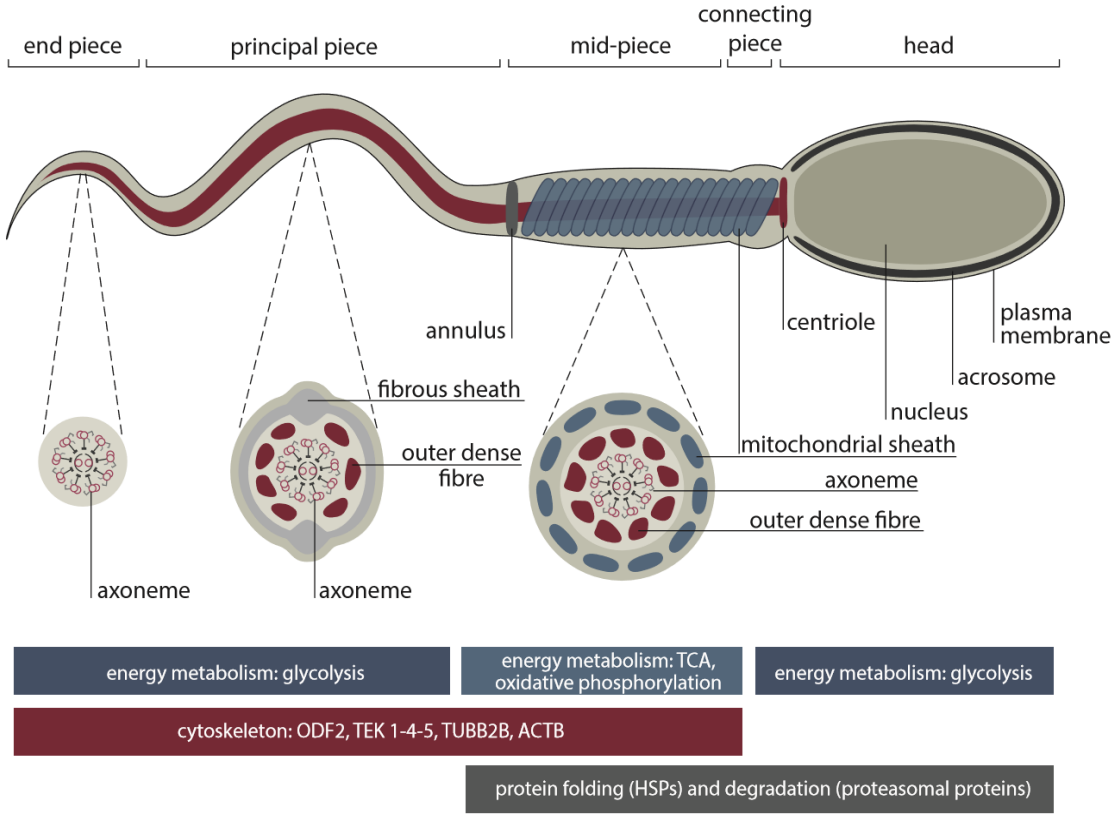


Figure 4: Schematic representation of a mature human sperm cell. Proteins that are altered in expression in asthenozoospermic men can be classified in three main functional classes: energy metabolism (glycolysis in head and principal piece and TCA and oxidative phosphorylation in mitochondria); cytoskeleton; and protein folding/degradation. Horizontal bars indicate the region of the sperm cell were these proteins are functioning.

3.2 Proteome of normozoospermic infertile sperm

During the last decade four differential proteome analyses were performed between infertile normozoospermic patients who experienced failure in IVF and fertile donors[72–75]. In total 123 proteins were identified showing an altered expression in sperm from normozoospermic infertile patients. Results of these studies were not always consistent and it is hard to draw relevant conclusions. However, gene ontology analyses of these differentially expressed proteins identified common functional clustering groups: sexual reproduction, metabolic process, cell growth and/or maintenance, protein metabolism and protein transport (Fig5). Interestingly, some proteins that are directly correlated with sperm capacitation and fertilization were down-regulated in normozoospermic infertile sperm cells: ZPBP (a zona pellucida binding protein), ELSPBP1 (plays a role in sperm capacitation), VCP, VPS13A, RAB5C (proteins involved in vesicular traffic and formation of the acrosome) and ACR (major protease of the acrosome). Additionally, both the Légaré et al. and the Azpiazu et al. study revealed that several histones were upregulated in the IVF-failure group (HIST1H2BA, HIST2H3A, HIST1H4A, HIST1H2BC, HIST1H4A, HIST2H2AC, HIST1H2AC, HIST1H2AA), which may indicate a deficient histone to protamine transition during spermiogenesis. These findings confirm the underlying importance of correct sperm epigenetic signatures for a successful fertilization[72,73].

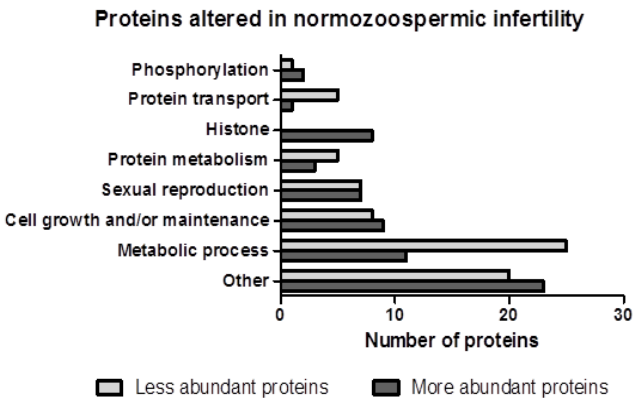


Figure 5: Comparative proteome analyses in normozoospermic infertile men resulted in the identification of 123 altered proteins. The protein cellular functions were determined according to the information available at the UniProt website. Proteins were grouped as those that were present at higher abundance and those that were present at lower abundance in sperm from normozoospermic infertile men compared to fertile men.

4 Epigenetics and Post-translational Modifications

In addition to genetic screenings and proteome analyses, the study of epigenetics and post-translational modifications (PTMs) is an up-coming and promising research field for studying idiopathic infertility. These alterations in epigenetics and PTMs are caused by genetic or environmental factors and can be used as biomarkers for male infertility.

4.1 Epigenetics

Epigenetics has been defined as "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence"[76]. The most important epigenetic events in sperm cells are: 1) the replacement of histones with protamines, 2) the retention of certain histones at key regulatory regions of the genome and 3) the hypomethylation of specific DNA regions. Additionally, non-coding RNAs can also be seen as essential epigenetic modifiers. It has been suggested that epigenetic dysregulation can be involved in the etiology of idiopathic male infertility[77]. However, the causative links between various epigenetic profiles and infertility remains unclear. The use of epigenetics as an indication, or cause, of male infertility is only beginning to be explored thoroughly and is currently not well understood. Jenkins et al. reviewed the potential clinical applications of sperm epigenetics in the study of male infertility and they concluded that there is emerging data suggesting that the predictive power of especially DNA methylation and RNA signatures in sperm likely exceeds the traditional assessment techniques of male infertility[78]. DNA methylation signatures remain remarkably stable throughout the process of spermatogenesis, what makes it an attractive technique for the assessment of health of spermatogenesis via observation of the mature sperm only (sperms cells are easily collected compared to a testis biopsy). However, a great deal of further work on this topic is required.

4.2 Post-translational modifications

In spermatozoa the protein synthesis machinery is silenced and the dynamic changes in the sperm proteome depend either on the acquisition of new proteins via vesicular transport or to several PTMs of existing proteins. PTMs have proved to play an important

role in regulating sperm function, including sperm maturation and acquisition of fertilizing potential[79]. One of the most studied PTMs in sperm is phosphorylation, which is required for epididymal maturation, motility, capacitation and acrosomal reaction of sperm cells. Two phosphoprotein proteome analyses have been performed in asthenozoospermic men and identified together 45 proteins displaying a differential phosphorylation pattern compared to normozoospermic men[80,81]. These altered proteins were mostly composed of HSPs, cytoskeletal proteins, proteins associated with the fibrous sheath, and those associated with energy metabolism. These functional classes are comparable with the ones characterized by the proteome analyses in asthenozoospermic men (section 3) and emphasize again the importance of these functional classes in regulating the motility of the sperm cell. Furthermore, a deficiency in tyrosine phosphorylation of tail proteins, especially those related to hyperactivated motility, was also reported to be associated to asthenozoospermia[82]. For more details and other PTMs linked to male infertility we refer to the review of Samanta et al.[79]. Comparable to the epigenetics of sperm DNA, PTMs form putative biomarkers in male infertility, but require additional research.

5 Oxidative stress and DNA fragmentation

Oxidative stress occurs when reactive oxygen species (ROS) are greatly elevated and overwhelm the body's natural anti-oxidant defense. Studies have shown that elevated ROS levels can be found in 40-80% of infertile men and in up to 11-78.5% of normozoospermic infertile patients.[83] ROS are unstable oxygen derived molecules that are formed as byproducts of oxidative metabolism, like e.g. hydroxyl ions, superoxide (free radicals) and hydrogen peroxide and lipid peroxide (non-free radicals). Seminal ROS can be produced by leukocytes as result of inflammation of the genital tract or by immature spermatozoa in the case of varicocele. Additionally, age and some life style factors like smoking, alcohol abuse and exposure to radiation and toxic chemicals can also increase seminal ROS[83]. In the sperm cell itself, ROS is generated through the NADPH oxidase system at the level of the sperm plasma membrane and through the NADH-dependent oxido-reductase system at the mitochondrial level. These physiological levels of ROS are essential in various aspects of the capacitation process, including the suppression of tyrosine phosphatase activity and the stimulation of cAMP

generation[84]. However, when ROS levels are elevated, resulting in oxidative stress, ROS cause damage to the sperm cell. Many studies have confirmed that oxidative stress affects the motility of sperm cells, their DNA integrity and their competence for sperm-oocyte fusion. Sperm cells are highly vulnerable to oxidative stress due to several reasons, they have limited volume and restricted distribution of cytoplasmic space where they can store ROS-metabolizing enzymes, such as catalase and glutathione peroxidase, resulting in a significant lack in antioxidant protection. Moreover, sperm cells contain exceptionally high levels of polyunsaturated fatty acids in their plasma membrane, which are very susceptible to oxidative stress. When polyunsaturated fatty acids are attacked by free radicals they generate lipid peroxides and aldehydes, which have a direct inhibitory action on sperm movement. The damage to the plasma membrane caused by oxidative stress will lead to failure to undergo sperm-oocyte fusion. Oxidative stress will eventually also lead to the generation of oxidized DNA base adducts. Throughout the cell cycle damage DNA is repaired by the cellular mechanism called 'base excision repair'. Spermatozoa only possess the first enzyme in the base excision repair pathway, 8-oxoguanine DNA glycosylase, leading to the generation of abasic sites at locations that have been affected by oxidative stress. Such abasic sites lead to destabilization of the ribose-phosphate backbone and eventually to DNA strand breaks[85]. The integrity of sperm DNA is considered to be vital for normal fertilization, embryo development, and for successful implantation and pregnancies in both natural and assisted reproduction. Recently, a systematic review and meta-analysis of 41 independent studies confirmed that sperm DNA damage affects clinical pregnancy following IVF and/or ICSI treatment[86]. Furthermore, Oleszczuk et al. showed that sperm DNA impairment can, at least partly, explain as many as 25% of previously unexplained cases[87]. Sperm DNA integrity tests are already frequently used to supplement the diagnostic evaluation, however, clinical practice guidelines do not recommend the routine use of these tests in the evaluation and treatment of the infertile couple.

For further reading on oxidative stress causing male infertility we refer to the very complete reviews on this topic:[83,85,88].

6 Discussion/Conclusion

During the last three decades the incidence of male infertility has increased steadily and semen quality has declined systematically[89,90]. In contrast to female infertility treatments, where hormonal manipulations to stimulate or enhance oocyte production are possible, spermatogenesis and sperm quality abnormalities are much more difficult to affect positively. In some cases, a healthy lifestyle can improve the fertility of men. However, if this is unsuccessful, assisted reproductive technologies (ART), such as IVF with ICSI, can be used to obtain pregnancy[91,92]. When these techniques are used, an adequate (epi)genetic diagnosis is of major importance to evaluate if a genetic abnormality will be transmitted to the offspring. The common routine to assess male fertility is a physical examination followed by two separate semen analyses, where standard parameters like cell count, morphology and motility are measured. However, this "golden-standard" semen analysis provides limited information and cannot discriminate fertile form infertile men on individual basis. Consequently, there is need for adjusted genetic screenings and improved seminal biomarkers to assess the likelihood that an individual will need to go through the IVF process or if less invasive therapeutic interventions may be effective. Overall our knowledge on the causes underlying idiopathic male infertility is limited. Only a few and rare genetic mutations have been linked to male infertility, mainly through genetic studies in consanguineous families. We believe that WGS will lead to a rapid increase in the identification of causal gene mutations for male infertility and will enhance the diagnosis of genetic factors underlying idiopathic infertility. Apart from genetic factors we also reviewed proteome differences between infertile and fertile sperm. These differential proteome analyses are mostly difficult to interpret, and are not leading to well defined causes of male infertility as is the case for genetic screenings. They rather reflect the physiological and molecular processes that are affected in infertile men. For asthenozoospermia, wherein most proteome analysis were performed, many cytoskeletal components and proteins involved in energy metabolism are downregulated and protein folding and degradation is altered. In normozoospermic infertility, it was more difficult to conclude which mechanisms are targeted. Two studies revealed that several histones were upregulated in normozoospermic infertile men, underlying the important value of correct sperm epigenetic signatures for successful fertilization. Considerable heterogeneity among these proteome analyses precludes the development of relevant

protein-based assays for male infertility. Epigenetic and post-translational modifications represent promising tools for the development of biomarkers in male infertility. However, these topics are still in their infancy and additional research is required in order to develop relevant diagnostic assays based on e.g. methylation patterns, RNA profiles or phosphorylation profiles. In contrast, the DNA integrity of sperm is at present a very attractive biomarker for unexplained male infertility as several assays exist to assess chromatin structure and DNA integrity of sperm. In conclusion we can say that during the coming years it is expected that many men who are currently diagnosed as having unexplained male infertility will be assigned genetic, epigenetic, or environmental causes[92].

7 References

1. WHO. Examination and processing of human semen. *Cambridge Univ. Press* **Edition, V,** 286 (2010).

2. Nieschlag, E., Behre, H. M. & Nieschlag, S. *Male Reproductive Health and Dysfunction*. (2010). doi:10.1007/978-3-540-78355-8

3. European Association of Urology. Guidelines on Male Infertility. (2015). doi:10.1007/978-1-60761-193-6

4. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).

5. Djureinovic, D. *et al.* The human testis-specific proteome defined by transcriptomics and antibody-based profiling. *Mol. Hum. Reprod.* **20,** 476–488 (2014).

6. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13,** 397–406 (2014).

7. Yu, N. Y.-L. *et al.* Complementing tissue characterization by integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic Acids Res.* **43,** 6787–6798 (2015).

8. Neto, F. T. L., Bach, P. V., Najari, B. B., Li, P. S. & Goldstein, M. Spermatogenesis in Humans and its affecting factors. *Semin. Cell Dev. Biol.* (2016). doi:10.1016/j.semcdb.2016.04.009

9. Wosnitzer, M., Goldstein, M. & Hardy, M. P. Review of Azoospermia. *Spermatogenesis* **4,** e28218 (2014).

10. Punab, M. *et al.* Causes of male infertility: a 9-year prospective monocentre study on 1737 patients with reduced total sperm counts. *Hum. Reprod.* **32,** 1–14 (2016).

11. Ben Khelifa, M. *et al.* Mutations in DNAH1, which encodes an inner arm heavy chain

dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *Am. J. Hum. Genet.* **94,** 95–104 (2014).

12. Curi, S. M. *et al.* Asthenozoospermia: analysis of a large population. *Arch. Androl.* **49,** 343–349 (2003).

13. Stival, C. *et al.* Sperm Capacitation and Acrosome Reaction in Mammalian Sperm. *Adv. Anat. Embryol. Cell Biol.* **220,** 93–106 (2016).

14. Tüttelmann, F., Meyts, E. R.-D., Nieschlag, E. & Simoni, M. Gene polymorphisms and male infertility – a meta-analysis and literature review. *Reprod. Biomed. Online* **15,** 643–658 (2007).

15. Krausz, C., Escamilla, A. R. & Chianese, C. Genetics of male infertility: From research to clinic. *Reproduction* **150,** R159–R174 (2015).

16. Mitchell, M. J. *et al.* Single gene defects leading to sperm quantitative anomalies. *Clin. Genet.* 1–9 (2016). doi:10.1111/cge.12900

17. Ray, P. F., Toure, A. & Mitchell, M. J. Genetic abnormalities leading to qualitative defects of sperm morphology or function. 217–232 (2017). doi:10.1111/cge.12905

18. Yatsenko, A. N. *et al.* X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *N. Engl. J. Med.* **372,** 2097–2107 (2015).

19. Yang, F. *et al.* TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. *EMBO Mol. Med.* **7,** 1198–210 (2015).

20. Ayhan, Ö. *et al.* Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J. Med. Genet.* **51,** 239–44 (2014).

21. Maor-Sagie, E. *et al.* Deleterious mutation in SYCE1 is associated with non-obstructive azoospermia. *J. Assist. Reprod. Genet.* **32,** 887–891 (2015).

22. Tenenbaum-Rakover, Y. *et al.* Minichromosome maintenance complex component 8 (MCM8) gene mutations result in primary gonadal failure. *J. Med. Genet.* **52,** 391–9 (2015).

23. Okutman, O. *et al.* Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. *Hum. Mol. Genet.* **24,** 5581–5588 (2015).

24. Bolcun-Filas, E. *et al.* Mutation of the mouse Syce1 gene disrupts synapsis and suggests a link between synaptonemal complex structural components and DNA repair. *PLoS Genet.* **5,** (2009).

25. Yang, F. *et al.* Meiotic failure in male mice lacking an X-linked factor Meiotic failure in male mice lacking an X-linked factor. 682–691 (2008). doi:10.1101/gad.1613608

26. Yang, F., Eckardt, S., Leu, N. A., McLaughlin, K. J. & Wang, P. J. Mouse TEX15 is essential for DNA double-strand break repair and chromosomal synapsis during male meiosis. *J. Cell Biol.* **180,** 673–679 (2008).

27. Lutzmann, M. *et al.* MCM8- and MCM9-Deficient Mice Reveal Gametogenesis Defects and Genome Instability Due to Impaired Homologous Recombination. *Mol. Cell* **47,** 523–534 (2012).

28. Park, J. *et al.* The MCM8-MCM9 Complex Promotes RAD51 Recruitment at DNA Damage Sites To Facilitate Homologous Recombination. *Mol. Cell. Biol.* **33,** 1632–1644 (2013).

29. Yan, W. *et al.* Zmynd15 encodes a histone deacetylase-dependent transcriptional repressor essential for spermiogenesis and male fertility. *J. Biol. Chem.* **285,** 31418–31426 (2010).

30. Falender, A. E. *et al.* Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. *Genes Dev.* **19,** 794–803 (2005).

31. Dam, A. H. D. M. *et al.* Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *Am. J. Hum. Genet.* **81,** 813–20 (2007).

32. ElInati, E. *et al.* A new mutation identified in SPATA16 in two globozoospermic patients. *J. Assist. Reprod. Genet.* 1–6 (2016). doi:10.1007/s10815-016-0715-3

33. Wang, X. *et al.* Homozygous DNAH1 frameshift mutation causes multiple morphological anomalies of the sperm flagella in Chinese. *Clin. Genet.* **91,** 313–321 (2017).

34. Amiri-Yekta, A. *et al.* Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new DNAH1 mutations. *Hum. Reprod.* **31,** 2872–2880 (2016).

35. Coutton, C., Escoffier, J., Martinez, G., Arnoult, C. & Ray, P. F. Teratozoospermia: Spotlight on the main genetic actors in the human. *Hum. Reprod. Update* **21,** 455–485 (2015).

36. Pierre, V. *et al.* Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. *Development* **139,** 2955–2965 (2012).

37. Xu, M. *et al.* Identification and characterization of a novel human testis-specific Golgi protein, NYD-SP12. *Mol. Hum. Reprod.* **9,** 9–17 (2003).

38. Lu, L., Lin, M., Xu, M., Zhou, Z.-M. & Sha, J.-H. Gene functional research using polyethylenimine-mediated in vivo gene transfection into mouse spermatogenic cells. *Asian J. Androl.* **8,** 53–59 (2006).

39. Dirami, T. *et al.* Missense mutations in SLC26A8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. *Am. J. Hum. Genet.* **92,** 760–766 (2013).

40. Touré, A. *et al.* The Testis Anion Transporter 1 (Slc26a8) is required for sperm terminal differentiation and male fertility in the mouse. *Hum. Mol. Genet.* **16,** 1783–1793 (2007).

41. Rode, B. *et al.* The testis anion transporter TAT1 (SLC26A8) physically and functionally interacts with the cystic fibrosis transmembrane conductance regulator channel: A potential role during sperm capacitation. *Hum. Mol. Genet.* **21,** 1287–1298 (2012).

42. Kuo, Y.-C. *et al.* SEPT12 mutations cause male infertility with defective sperm annulus. *Hum. Mutat.* **33,** 710–719 (2012).

43. Yeh, C.-H. *et al.* SEPT12/SPAG4/LAMINB1 complexes are required for maintaining the integrity of the nuclear envelope in postmeiotic male germ cells. *PLoS One* **10,** e0120722 (2015).

44. Lin, Y.-H. *et al.* SEPT12 deficiency causes sperm nucleus damage and developmental arrest of preimplantation embryos. *Fertil. Steril.* **95,** 363–365 (2011).

45. Avidan, N. *et al.* CATSPER2, a human autosomal nonsyndromic male infertility gene. *Eur. J. Hum. Genet.* **11,** 497–502 (2003).

46. Zhang, Y. *et al.* Sensorineural deafness and male infertility: a contiguous gene deletion syndrome. *BMJ Case Rep.* **2009,** (2009).

47. Avenarius, M. R. *et al.* Human male infertility caused by mutations in the CATSPER1 channel protein. *Am. J. Hum. Genet.* **84,** 505–510 (2009).

48. Kashir, J. *et al.* A maternally inherited autosomal point mutation in human phospholipase C zeta (PLCζ) leads to male infertility. *Hum. Reprod.* **27,** 222–231 (2012).

49. Escoffier, J. *et al.* Homozygous mutation of PLCZ1 leads to defective human oocyte activation and infertility that is not rescued by the WW-binding protein PAWP. *Hum. Mol. Genet.* **25,** 878–891 (2016).

50. Clapham, D. E. Calcium Signaling. *Cell* **131,** 1047–1058 (2007).

51. Liao, T.-T., Xiang, Z., Zhu, W.-B. & Fan, L.-Q. Proteome analysis of round-headed and normal spermatozoa by 2-D fluorescence difference gel electrophoresis and mass spectrometry. *Asian J. Androl.* **11,** 683–693 (2009).

52. Zhao, C. *et al.* Identification of several proteins involved in regulation of sperm motility by proteomic analysis. *Fertil. Steril.* **87,** 436–8 (2007).

53. Martínez-Heredia, J., de Mateo, S., Vidal-Taboada, J. M., Ballescà, J. L. & Oliva, R. Identification of proteomic differences in asthenozoospermic sperm samples. *Hum. Reprod.* **23,** 783–791 (2008).

54. Siva, A. B. *et al.* Proteomics-based study on asthenozoospermia: Differential expression of proteasome alpha complex. *Mol. Hum. Reprod.* **16,** 452–462 (2010).

55. Shen, S., Wang, J., Liang, J. & He, D. Comparative proteomic study between human normal motility sperm and idiopathic asthenozoospermia. *World J. Urol.* **31,** 1395–1401 (2013).

56. Amaral, A. *et al.* Identification of proteins involved in human sperm motility using high-throughput differential proteomics. *J. Proteome Res.* **13,** 5670–5684 (2014).

57. Hashemitabar, M., Sabbagh, S., Orazizadeh, M., Ghadiri, A. & Bahmanzadeh, M. A proteomic analysis on human sperm tail: Comparison between normozoospermia and asthenozoospermia. *J. Assist. Reprod. Genet.* **32,** 853–863 (2015).

58. Amaral, A., Castillo, J., Ramalho-Santos, J. & Oliva, R. The combined human sperm proteome: Cellular pathways and implications for basic and clinical science. *Hum. Reprod. Update* **20,** 40–62 (2014).

59. Carrell, D. T., Aston, K. I., Oliva, R., Emery, B. R. & De Jonge, C. J. The "omics" of human male infertility: integrating big data in a systems biology approach. *Cell Tissue Res.* **363,** 295–312 (2016).

60. Jodar, M., Soler-Ventura, A. & Oliva, R. Semen proteomics and male infertility. *J.*

Proteomics (2016). doi:10.1016/j.jprot.2016.08.018

61. Turner, R. M. Moving to the beat: A review of mammalian sperm motility regulation. *Reprod. Fertil. Dev.* **18,** 25–38 (2006).

62. Tarnasky, H. *et al.* Gene trap mutation of murine outer dense fiber protein-2 gene can result in sperm tail abnormalities in mice with high percentage chimaerism. *BMC Dev. Biol.* **10,** 67 (2010).

63. Amos, L. A. The tektin family of microtubule-stabilizing proteins. *Genome Biol.* **9,** 229 (2008).

64. Oiki, S., Hiyama, E., Gotoh, T. & Iida, H. Localization of Tektin 1 at both acrosome and flagella of mouse and bull spermatozoa. *Zoolog. Sci.* **31,** 101–7 (2014).

65. Takiguchi, H. *et al.* Characterization and subcellular localization of Tektin 3 in rat spermatozoa. *Mol. Reprod. Dev.* **78,** 611–620 (2011).

66. Machesky, L. M. & Insall, R. H. Signaling to actin dynamics. *J. Cell Biol.* **146,** 267–272 (1999).

67. Cardullo, R. A. & Baltz, J. M. Metabolic regulation in mammalian sperm: mitochondrial volume determines sperm length and flagellar beat frequency. *Cell Motil. Cytoskeleton* **19,** 180–188 (1991).

68. du Plessis, S. S., Agarwal, A., Mohanty, G. & van der Linde, M. Oxidative phosphorylation versus glycolysis: what fuel do spermatozoa use? *Asian J. Androl.* **17,** 230–5 (2015).

69. Dix, D. J. *et al.* Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 3264–3268 (1996).

70. Dix, D. J. *et al.* HSP70-2 is required for desynapsis of synaptonemal complexes during meiotic prophase in juvenile and adult mouse spermatocytes. *Development* **124,** 4595–4603 (1997).

71. Nixon, B. *et al.* The role of the molecular chaperone heat shock protein A2 (HSPA2) in regulating human sperm-egg recognition. *Asian J. Androl.* **17,** 568 (2015).

72. Azpiazu, R. *et al.* High-throughput sperm differential proteomics suggests that epigenetic alterations contribute to failed assisted reproduction. *Hum. Reprod.* **29,** 1225–1237 (2014).

73. Légaré, C. *et al.* Investigation of male infertility using quantitative comparative proteomics. *J. Proteome Res.* **13,** 5403–5414 (2014).

74. Xu, W. *et al.* Proteomic characteristics of spermatozoa in normozoospermic patients with infertility. *J. Proteomics* **75,** 5426–5436 (2012).

75. Pixton, K. L. *et al.* Sperm proteome mapping of a patient who experiencd failed fertilization at IVF reveals altered expression of at least 20 proteins compared with fertile donors: Case report. *Hum. Reprod.* **19,** 1438–1447 (2004).

76. Dupont, C., Armant, D. R. & Brenner, C. A. Epigenetics: definition, mechanisms and clinical perspective. *Semin. Reprod. Med.* **27,** 351–357 (2009).

77. Carrell, D. T. Epigenetics of the male gamete. *Fertil. Steril.* **97,** 267–274 (2012).

78. Jenkins, T. G., Aston, K. I., James, E. R. & Carrell, D. T. Sperm epigenetics in the study of male fertility, offspring health, and potential clinical applications. *Syst. Biol. Reprod. Med.* **63,** 69–76 (2017).

79. Samanta, L., Swain, N., Ayaz, A., Venugopal, V. & Agarwal, A. Post-Translational Modifications in sperm Proteome: The Chemistry of Proteome diversifications in the Pathophysiology of male factor infertility. *Biochim. Biophys. Acta - Gen. Subj.* **1860,** 1450–1465 (2016).

80. Chan, C. C. *et al.* Motility and protein phosphorylation in healthy and asthenozoospermic sperm. *J. Proteome Res.* **8,** 5382–5386 (2009).

81. Parte, P. P. *et al.* Sperm phosphoproteome profiling by ultra performance liquid chromatography followed by data independent analysis (LC-MSE) reveals altered proteomic signatures in asthenozoospermia. *J. Proteomics* **75,** 5861–5871 (2012).

82. Yunes, R., Doncel, G. F. & Acosta, A. A. Incidence of sperm-tail tyrosine phosphorylation and hyperactivated motility in normozoospermic and asthenozoospermic human sperm samples. *Biocell* **27,** 29–36 (2003).

83. Hamada, A., Esteves, S. C. & Agarwal, A. Unexplained male infertility : potential causes and management. *Hum. Androlog* **38,** 2–16 (2011).

84. Aitken, R. J., Baker, M. A. & Nixon, B. Are sperm capacitation and apoptosis the opposite ends of a continuum driven by oxidative stress? *Asian J. Androl.* **17,** 633–9 (2015).

85. Aitken, R. J., Gibb, Z., Baker, M. A., Drevet, J. & Gharagozloo, P. Causes and consequences of oxidative stress in spermatozoa. *Reprod. Fertil. Dev.* **28,** 1 (2016).

86. Simon, L., Zini, A., Dyachenko, A., Ciampi, A. & Carrell, D. T. A systematic review and meta-analysis to determine the effect of sperm DNA damage on in vitro fertilization and intracytoplasmic sperm injection outcome. *Asian J. Androl.* 1–11 (2016). doi:10.4103/1008-682X.182822

87. Oleszczuk, K., Augustinsson, L., Bayat, N., Giwercman, A. & Bungum, M. Prevalence of high DNA fragmentation index in male partners of unexplained infertile couples. *Andrology* **1,** 357–360 (2013).

88. Sanocka, D., Miesel, R., Jedrzejczak, P. & Kurpisz, M. K. Oxidative stress and male infertility. *J. Androl.* **17,** 449–454 (1996).

89. Andersson, A.-M. *et al.* Adverse trends in male reproductive health: we may have reached a crucial "tipping point". *Int. J. Androl.* **31,** 74–80 (2008).

90. Swan, S. H., Elkin, E. P. & Fenster, L. The question of declining sperm density revisited: an analysis of 101 studies published 1934-1996. *Environ. Health Perspect.* **108,** 961–966 (2000).

91. Tournaye, H., Krausz, C. & Oates, R. D. Concepts in diagnosis and therapy for male reproductive impairment. *lancet. Diabetes Endocrinol.* **8587,** 1–11 (2016).

92. Tournaye, H., Krausz, C. & Oates, R. D. Novel concepts in the aetiology of male reproductive impairment. *lancet. Diabetes Endocrinol.* **8587,** 1–10 (2016).

Supplementary table 1: List of proteins that display altered expression in asthenozoospermic, teratozoospermic and normozoospermic infertile men

UniProt accession number	Gene symbol	Full name of protein	Regulation	Reference	Cellular function
Asthenozoospermic sperm versus normozoospermic sperm					
P52565	ARHGDIA	Rho GDP dissociation inhibitor	Down	Zhao et al.2007	other
Q5BJF6	ODF	Outer dense fiber protein	Up	Zhao et al.2007	cytoskeleton
O75874	IDH1	Isocitrate dehyddrogenase subunit α	Down	Zhao et al.2007	energy production and metabolism
P15259	PGAM2	Phosphoglycerate mutase 2	Up	Zhao et al.2007	energy production and metabolism
P60174	TPI1	Triosephosphate isomerase	Up	Zhao et al.2007	energy production and metabolism
P17174	GOT1	Glutamate oxaloacetate transaminase-1	Up	Zhao et al.2007	energy production and metabolism
P00918	CA2	Carbonic anhydrase II	Up	Zhao et al.2007	energy production and metabolism
P04279	SEMG1	Semenogelin-1 precursor	Up	Zhao et al.2007	spermatogenesis and sperm function
P15104	GLUL	Glutamine synthetase	Up	Zhao et al.2007	energy production and metabolism
P62195	PSMC5	26S protease regulatory subunit	Down	Zhao et al.2007	protein metabolism
P60709	ACTB	Cytoskeletal actin	Down	Martínez-Heredia et al.2008	cytoskeleton
P08758	ANXA5	Annexin A5	Down	Martínez-Heredia et al.2008	other
Q7L1R4	COX6B	Cytochrome c oxidase subunit 6B	Down	Martínez-Heredia et al.2008	energy production and metabolism
Q93077	H2A	Histone H2A	Down	Martínez-Heredia et al.2008	dna binding
P12273	PIP	Prolactin induced protein	Down	Martínez-Heredia et al.2008	other
P12273	PIPpre	Prolactin-inducible protein precursor	Down	Martínez-Heredia et al.2008	other

P06702	S100A9	S100 calcium binding protein A9	Down	Martínez-Heredia et al.2008	other
P10909	CLUpre	Clusterin precursor	Up	Martínez-Heredia et al.2008	protein metabolism
P09622	DLDpre	Dihydrolipoamide dehydrogenase (DLD) precursor	Up	Martínez-Heredia et al.2008	spermatogenesis and sperm function
P07954	FHpre	Fumarate hydratase precursor	Up	Martínez-Heredia et al.2008	energy production and metabolism
P54652	HSPA2	Heat shock-related 70kDa protein 2	Up	Martínez-Heredia et al.2008	protein metabolism
P29218	IMPA1	Inositol-1(or-4)-monophosphatase	Up	Martínez-Heredia et al.2008	energy production and metabolism
P25325/Q13011	MPST/ECH1 pre	3-mercapto-pyruvate sulfurtransferase/Delta 3,5-delta 2,4-dienoyl-CoA isomerase precursor	Up	Martínez-Heredia et al.2008	other
P49720	PSMB3	Proteosome beta 3 subunit human	Up	Martínez-Heredia et al.2008	protein metabolism
P04279	SEMG1 pre	Semengolin I protein precursor	Up	Martínez-Heredia et al.2008	spermatogenesis and sperm function
Q9BXU0	TEX12	Testis-expressed sequence 12 protein	Up	Martínez-Heredia et al.2008	other
P11021	GRP78	78kDa glucose-regulated protein	Down	Shen et al.2013	protein metabolism
P34931		Heat shock protein 70 testis variant	Down	Shen et al.2013	protein metabolism
Q8WW24	TEKT4	Tektin4	Down	Shen et al.2013	cytoskeleton
Q5EK51	Lacto	Lactoferrin	Down	Shen et al.2013	other
Q5BJF6	ODF2	Outer dense fiber of sperm tails 2	Down	Shen et al.2013	cytoskeleton
Q9NS25	SPANXB	Sperm protein associated with the nucleus on the X chromosome B/F	Down	Shen et al.2013	spermatogenesis and sperm function
P07205	PGK2	Phosphoglycerate kinase 2	Down	Shen et al.2013	energy production and metabolism
Q2E173		Flagellin	Down	Shen et al.2013	cytoskeleton
Q53FA3	Hsp70	Heat shock 70kDa protein 1-like	Down	Shen et al.2013	protein metabolism
Q99497	DJ-1	Chain A, Human Dj-1 with sulfinic acid	Down	Shen et al.2013	other
P15927	RPA2	XPA binding protein2, isoform CRA_b	Down	Shen et al.2013	other
P36969	GPX4	Phospholipid hydroperoxide glutathione peroxidase	Up	Shen et al.2013	other
P84336	ACTB	Cytoskeletal actin	Down	Shen et al.2013	cytoskeleton

O14556	GAPDHS	Glyceraldehyde-3-phosphate-dehydrogenase, testis-specific	Up	Shen et al.2013	energy production and metabolism
P12273	PIP	Prolactin-inducible protein precursor	Up	Shen et al.2013	ohter
P25788	PSMA3	Proteasome subunit alpha type3	Down	Siva et al.2010	protein metabolism
P54652	HSPA2	Heat shock 70kDa related protein	Down	Siva et al.2010	protein metabolism
P68371	TUBB2C	Tubulin beta-2C chain	Down	Siva et al.2010	cytoskeleton
Q969V4	TEKT1	Tektin 1	Down	Siva et al.2010	cytoskeleton
P60174	TPI1	Triose phosphate isomerase	Up	Siva et al.2010	energy production and metabolism
Q14410	GKP2	Glycerol kinase testis specific 2	Up	Siva et al.2010	energy production and metabolism
P55809	OXCT1	Succinyl-CoA:3-Ketoacid co-enzyme A transferase I	Up	Siva et al.2010	energy production and metabolism
Q9P2T0	THEG	Testicular haploid expressed gene protein	Down	Amaral et al.2014	spermatogenesis and sperm function
Q13642	FHL1	Four and a half LIM domains protein1	Down	Amaral et al.2014	other
O95831	AIFM1	Apoptosis-inducing factor1, mitochondrial	Down	Amaral et al.2014	other
Q6UWQ5	LYZL1	Lysozyme-like enzyme protein1	Down	Amaral et al.2014	energy production and metabolism
P55060	CSE1L	Exportin-2	Down	Amaral et al.2014	other
Q96M32	AK7	Adenylate kinase 7	Down	Amaral et al.2014	energy production and metabolism
P10253	GAA	Lysosomal alpha-glucosidase	Down	Amaral et al.2014	energy production and metabolism
Q92526	CCT6B	T-complex protein 1 subunit zeta2	Down	Amaral et al.2014	protein metabolism
P48047	ATP5O	ATP synthase subunit O, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
Q3ZCQ8	TIMM50	Mitochondrial import inner membrane translocase subunit TIM50	Down	Amaral et al.2014	other
Q96KX0	LYZL4	Lysozyme-like protein 4	Down	Amaral et al.2014	other
Q99798	ACO2	Aconitate hydratase, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
P55786	NPEPPS	Puromycin-sensitive aminopeptidase	Down	Amaral et al.2014	protein metabolism
Q9H4K1	RIBC2	RIB43A-like with coiled-coils protein 2	Down	Amaral et al.2014	other
Q15172	PPP2R5A	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit alpha isoform PPP2R5A Diablo homolog, mitochondrial	Down	Amaral et al.2014	energy production and metabolism

Q9NR28	DIABLO	Diablo homolog, mitochondrial	Down	Amaral et al.2014	other
P13667	PDIA4	Protein disulfide-isomerase A4	Down	Amaral et al.2014	protein metabolism
Q8N7U6	EFHB	EF-hand domain-containing family member B	Down	Amaral et al.2014	other
P36776	LONP1	Lon protease homolog, mitochondrial	Down	Amaral et al.2014	protein metabolism
P06744	GPI	Glucose-6-phosphate isomerase	Down	Amaral et al.2014	energy production and metabolism
P06576	ATP5B	ATP synthase subunit beta, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
Q96M91	CCDC11	Coiled-coil domain-containing protein 11	Down	Amaral et al.2014	cytoskeleton
Q96MR6	WDR65	WD repeat-containing protein 65	Down	Amaral et al.2014	cytoskeleton
P10515	DLAT	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
Q5T655	CCDC147	Coiled-coil domain-containing protein 147	Down	Amaral et al.2014	cytoskeleton
P36957	DLST	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutaratedehydrogenase complex, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
O75190	DNAJB6	DnaJ homolog subfamily B member 6	Down	Amaral et al.2014	protein metabolism
Q5JVL4	EFHC1	EF-hand domain-containing protein 1	Down	Amaral et al.2014	other
P49841	GSK3B	Glycogen synthase kinase-3 beta	Down	Amaral et al.2014	energy production and metabolism
Q86VP6	CAND1	Cullin-associated NEDD8-dissociated protein 1	Down	Amaral et al.2014	protein metabolism
Q4G0X9	CCDC40	Coiled-coil domain-containing protein 40	Down	Amaral et al.2014	cytoskeleton
Q9BRQ6	CHCHD6	Coiled-coil−helix-coiled-coil−helix domain-containing protein 6, mitochondrial	Down	Amaral et al.2014	other
Q86Y39	NDUFA11	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11	Down	Amaral et al.2014	energy production and metabolism
P40939	HADHA	Trifunctional enzyme subunit alpha, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
P41250	GARS	Glycine-tRNA ligase	Down	Amaral et al.2014	protein metabolism
P14314	PRKCSH	Glucosidase 2 subunit beta	Down	Amaral et al.2014	energy production and metabolism
Q969V4	TEKT1	Tektin-1	Down	Amaral et al.2014	cytoskeleton
Q14697	GANAB	Neutral alpha-glucosidase AB	Down	Amaral et al.2014	energy production and metabolism
Q6BCY4	CYB5R2	NADH-cytochrome b5 reductase 2	Down	Amaral et al.2014	energy production and metabolism

P46459	NSF	Vesicle-fusing ATPase	Down	Amaral et al.2014	other
P11177	PDHB	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
Q14203	DCTN1	Dynactin subunit 1	Down	Amaral et al.2014	cytoskeleton
Q93009	USP7	Ubiquitin carboxyl-terminal hydrolase 7	Down	Amaral et al.2014	protein metabolism
Q9NSE4	IARS2	Isoleucine-tRNA ligase, mitochondrial	Down	Amaral et al.2014	protein metabolism
Q9Y371	SH3GLB1	Endophilin-B1	Down	Amaral et al.2014	other
O00330	PDHX	Pyruvate dehydrogenase protein X component, mitochondrial	Down	Amaral et al.2014	energy production and metabolism
Q8N1V2	WDR16	WD repeat-containing protein 16	Down	Amaral et al.2014	cytoskeleton
P22234	PAICS	Multifunctional protein ADE2	Down	Amaral et al.2014	energy production and metabolism
Q15631	TSN	Translin	Down	Amaral et al.2014	other
Q96M29	TEK5	Tektin-5	Down	Amaral et al.2014	cytoskeleton
P29218	IMPA1	Inositol monophosphatase 1	Up	Amaral et al.2014	energy production and metabolism
Q9GZT6	CCDC90B	Coiled-coil domain-containing protein 90B, mitochondrial	Up	Amaral et al.2014	other
P67870	CSNK2B	Casein kinase II subunit beta	Up	Amaral et al.2014	oterh
P10606	COX5B	Cytochrome c oxidase subunit 5B, mitochondrial	Up	Amaral et al.2014	energy production and metabolism
Q96DE0	NUDT16	U8 snoRNA-decapping enzyme	Up	Amaral et al.2014	other
P54652	HSPA2	Heat shock-related 70 kDa protein 2	Up	Amaral et al.2014	protein metabolism
P80303	NUCB2	Nucleobindin-2	Up	Amaral et al.2014	other
P20933	AGA	N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase	Up	Amaral et al.2014	protein metabolism
P57105	SYNJ2BP	Synaptojanin-2-binding protein	Up	Amaral et al.2014	other
P30041	PRDX6	Peroxiredoxin-6	Up	Amaral et al.2014	energy production and metabolism
P60174	TPI1	Triosephosphate isomerase	Up	Amaral et al.2014	energy production and metabolism
Q8N0U8	VKORC1L1	Vitamin K epoxide reductase complex subunit 1-like protein 1	Up	Amaral et al.2014	energy production and metabolism
Q8NA82	MARCH10	Probable E3 ubiquitin-protein ligase MARCH10	Up	Amaral et al.2014	protein metabolism
Q8NBU5	ATAD1	ATPase family AAA domain-containing protein 1	Up	Amaral et al.2014	other
P28070	PSMB4	Proteasome subunit beta type-4	Up	Amaral et al.2014	protein metabolism

Q96AJ9	VTI1A	Vesicle transport through interaction with t-SNAREs homolog 1A	Up	Amaral et al.2014	other
Q9H4Y5	GSTO2	Glutathione S-transferase omega-2	Up	Amaral et al.2014	energy production and metabolism
Q8IZS6	TCTE3	Tctex1 domain-containing protein 3	Up	Amaral et al.2014	cytoskeleton
Q6UW68	TMEM205	Transmembrane protein 205 OS=Homo sapiens	Up	Amaral et al.2014	other
P01034	CST3	Cystatin-C	Up	Amaral et al.2014	protein metabolism
P60981	DSTN	Destrin	Up	Amaral et al.2014	cytoskeleton
P24666	ACP1	Low molecular weight phosphotyrosine protein phosphatase	Up	Amaral et al.2014	ohter
Q9P0L0	VAPA	Vesicle-associated membrane protein-associated protein A	Up	Amaral et al.2014	other
O43674	NDUFB5	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 5, mitochondrial NDUFB5	Up	Amaral et al.2014	energy production and metabolism
P26885	FKBP2	Peptidyl-prolyl cis−trans isomerase FKBP2	Up	Amaral et al.2014	protein metabolism
Q8WY22	BRI3BP	BRI3-binding protein	Up	Amaral et al.2014	other
P32119	PRDX2	Peroxiredoxin-2	Up	Amaral et al.2014	other
P02649	APOE	Apolipoprotein E	Up	Amaral et al.2014	other
P61088	UBE2N	Ubiquitin-conjugating enzyme E2 N	Up	Amaral et al.2014	protein metabolism
P15309	ACPP	Prostatic acid phosphatase	Up	Amaral et al.2014	other
Q9BVA1	TUBB2B	Tubulin beta-2B	Down	Hashemitabar er al.2015	cytoskeleton
P45880	VDAC2	Voltage dependent anion selective channel protein 2	Down	Hashemitabar er al.2015	other
P14854	MT-CO1	Cytochrome c oxidase subunit	Down	Hashemitabar er al.2015	energy production and metabolism
P36969	GPX4	Phospholipid hydroperoxidase (glutathione peroxidase), mitochondrial	Down	Hashemitabar er al.2015	other
Q7L266	ASRGL1	Isoaspartyl peptidase/L-asparaginase	Up	Hashemitabar er al.2015	protein metabolism
Q9BVA1	ODF2	Outer dense fiber protein 2	Down	Hashemitabar er al.2015	cytoskeleton
P45880	AKAP4	A-kinase anchor protein 4	Down	Hashemitabar er al.2015	cytoskeleton
P36969	GAPDHS	Glyceraldehyde-3-phosphate dehydrogenase, testis specific	Down	Hashemitabar er al.2015	energy production and metabolism
Q7L266	CLU	Clusterin	Down	Hashemitabar er al.2015	cytoskeleton

P21266	GSTM3	Glutathione S-transferase Mu3	Up	Hashemitabar er al.2015	other
P54652	HSPA2	Heat shock-related 70kDa protein 2	Down	Hashemitabar er al.2015	protein metabolism
P38646	HSPA9	Stress-70 protein, mitochondrial	Down	Hashemitabar er al.2015	protein metabolism
Q8TAD1	SPANXE	Sperm protein associated with the nucleus on the X chromosome E	Down	Hashemitabar er al.2015	other
P04264	KRT	Keratin, type II cytoskeletal	Up	Hashemitabar er al.2015	cytoskeleton
		Teratozoospermic sperm versus normozoospermic sperm			
Q5JQC9	AKAP4	A-kinase (PRKA) anchor protein 4	Up	Liao et al.2009	
P10909	CLU pre	Clusterin precursor	Up	Liao et al.2009	
P02768	ALB	Serum albumin	Up	Liao et al.2009	
Q8WYR4	RSPH1	Testis-specific gene A2 protein	Up	Liao et al.2009	
		Sperm protein	Up	Liao et al.2009	
P21266	GSTM3	Glutathione S-transferase Mu 3	Down	Liao et al.2009	
	RcTPI1	RcTPI1 (fragment)	Down	Liao et al.2009	
P60709	ACTB	Actin, cytoplasmic	Up	Liao et al.2009	
Q8TDY3	ACTRT2	Actin-related protein T2	Down	Liao et al.2009	
Q5BJF6	ODF2	Outer dense fiber protein 2	Down	Liao et al.2009	
P12273	PIP pre	Prolactin-inducible protein precursor	Up	Liao et al.2009	
P60174	TPI	Triosephosphate isomerase	Dwon	Liao et al.2009	
Q5JQC9	AKAP4	A-kinase anchor protein 4 precursor	Down	Liao et al.2009	
Q7L266	ASRGL1	Asparaginase-like 1 protein	Down	Liao et al.2009	
Q13748	TUBA2	Tubulin a-2 chain	Down	Liao et al.2009	
Q9NY87	SPANXC	Sperm protein associated with the nucleus on the X chromosome C	Down	Liao et al.2009	
Q9NS26	SPANXA	Sperm protein associated with the nucleus on the X chromosome A	Down	Liao et al.2009	
Q9NS25	SPANXB	Sperm protein associated with the nucleus on the X chromosome B	Down	Liao et al.2009	
Q9HBV2	SPACA1 pre	Sperm acrosome membrane-associated protein 1 precursor	Down	Liao et al.2009	
		Normozoospermic infertile versus normozoospermic fertile sperm			
P04279	SEMG1	Semengolin 1	Up	Xu et al. 2012	sexual reproduction
P04279	SEMG1pre	Semenogelin1 preprotein	Up	Xu et al. 2012	sexual reproduction

Q02383	SEMG2	Semenogelin 2	Up	Xu et al. 2012	sexual reproduction
Q02383	SEMG2pre	Semenogelin 2 precursor	Down	Xu et al. 2012	sexual reproduction
P25786	PSMA1	Proteasome (prosome macropain) subunit, alpha type 1	Down	Xu et al. 2012	metabolic process
Q9NS25	SPANX	Sperm protein associated with the nucleus on the X chromosome B/F	Down	Xu et al. 2012	sexual reproduction
P15309	ACPP	Acid phosphatase, prostate	Up	Xu et al. 2012	other
Q61CB4	FAM109B	Family with sequence similarity 109, member B	Up	Xu et al. 2012	protein transport
Q13748	TUBA3D	Tubulin alpha-3D chain	Down	Xu et al. 2012	other
P09466	PAEP	Progestagen-associated endometrial protein	Up	Xu et al. 2012	cell growth and/or maintenance
P04075	ALDOA	Fructose-bisphosphate aldolase A	Up	Xu et al. 2012	metabolic process
P12272	PIPpre	Prolactin-inducible protein precursor	Down	Xu et al. 2012	other
P10909	CLU	Clusterin	Up	Xu et al. 2012	cell growth and/or maintenance
P10909	CLUpre	Apolipoprotein J precursor	Up	Xu et al. 2012	cell growth and/or maintenance
P07288	PSA	Prostate-specific antigen	Up	Xu et al. 2012	other
P07288	PSApre	Prostate-specific antigen precursor	Up	Xu et al. 2012	other
P36969	GPX4pre	Phospholipid hydropreoxide glutathione peroxidase, mitochondrial isoform A precursor	Down	Xu et al. 2012	cell growth and/or maintenance
P30044	PRDX5	Chain A, human peroxiredoxin 5	Up	Xu et al. 2012	other
Q5JQC9	AKAP4	A kinase (PRKA) anchor protein 4	Down	Xu et al. 2012	other
Q5JQC9	AKAP4pre	A kinase (PRKA) anchor protein 4 precursor	Down	Xu et al. 2012	other
P10606	COX5B	Cytochrome C oxidase subunit 5B	Up	Xu et al. 2012	metabolic process
P08758	ANXA5	Annexin A5	Up	Xu et al. 2012	other
Q14990	ODFP	Outer dense fiber protein	Down	Xu et al. 2012	other
	MSF-FN70	Migration stimulation factor	Down	Xu et al. 2012	other
P04279	SEMG1	Semenogelin 1	Up	Légaré et al.2014	sexual reproduction
P15309	ACPP	Prostatic acid phosphatase	Up	Légaré et al.2014	other
A6NM11	LRRC37A2	Leucine-rich repeat-containing protein 37A2	Up	Légaré et al.2014	other
P49913	CAMP	Cathelicidin antimicrobial peptide	Up	Légaré et al.2014	other
P15144	ANPEP	Aminopeptidase N	Up	Légaré et al.2014	cell growth and/or maintenance
P13796	LCP1	Plastin-2		Légaré et al.2014	other
P12272	PIP	Prolactin-inducible protein	Up	Légaré et al.2014	other

O14556	GAPDHS	Glyceraldehyde-3-phosphate dehydrogenase, testis-specific	Down	Légaré et al.2014	metabolic process
Q96BH3	ELSPBP1	Epididymal sperm-binding protein1	Down	Légaré et al.2014	sexual reproduction
P21266	GSTM3	Glutathione S-transferase Mu3	Down	Légaré et al.2014	other
Q9BS86	ZPBP	Zona pellucida-binding protein 1	Down	Légaré et al.2014	zona pellucida binding protein
P07205	PGK2	Phosphoglycerate kinase 2	Down	Légaré et al.2014	metabolic process
P06744	GPI	Glucose-6-phosphate isomerase	Down	Légaré et al.2014	metabolic process
P60174	TPI1	Triphosphate isomerase	Down	Légaré et al.2014	metabolic process
P00338	LDH	L-Lactate dehydrogenase	Down	Légaré et al.2014	metabolic process
P00918	CA2	Carbonic anhydrase 2	Down	Légaré et al.2014	metabolic process
P14618	PKM	Pyruvate kinase isozymes M1/M2	Down	Légaré et al.2014	metabolic process
P10323	ACR	Acrosin	Down	Légaré et al.2014	protein metabolism
P49221	TGM4	Protein-glutamine gamma-glutamyltransferase 4	Up	Légaré et al.2014	sexual reproduction
P14618	PKM	Pyruvate kinase isozymes M1/M2	Up	Légaré et al.2014	metabolic process
P35579	MYH9	Myosin-9	Up	Légaré et al.2014	cell growth and/or maintenance
P30101	PDIA3	Protein disulfide-isomerase A3	Up	Légaré et al.2014	protein metabolism
P06744	GPI	Glucose-6-phosphate isomerase	Up	Légaré et al.2014	metabolic process
O14556	GAPDHS	Glyceraldehyde-3-phosphate dehydrogenase, testis-specific	Up	Légaré et al.2014	metabolic process
P62805	HIST1H4A	Histone H4	Up	Légaré et al.2014	Histone
P10909	CLU	Clusterin	Up	Légaré et al.2014	cell growth and/or maintenance
Q16531	DDB1	DNA damage-binding protein 1	Up	Légaré et al.2014	other
Q71DI3	HIST2H3A	Histone H3.2	Up	Légaré et al.2014	Histone
P12272	PIP	Prolactin-inducible protein	Up	Légaré et al.2014	other
Q96A08	HIST1H2BA	Histone H2B type 1-A	Up	Légaré et al.2014	Histone
P05109	S100A8	Protein S100-A8	Up	Légaré et al.2014	cell growth and/or maintenance
P02751	FN1	Fibronectin isoform 1 preproprotein	Up	Légaré et al.2014	cell growth and/or maintenance
P04279	SEMG1	Semenogelin-1	Up	Légaré et al.2014	sexual reproduction
Q9H0B3	KIAA1683	Uncharacterized protein KIAA1683	Down	Légaré et al.2014	other
Q00059	TFAM	Transcription factor A, mitochondrial	Down	Légaré et al.2014	other
P15259	PGAM2	Phosphoglycerate mutase 2	Down	Légaré et al.2014	metabolic process

P28074	PSMB5	Proteasome subunit beta type5	Down	Légaré et al.2014	metabolic process
Q13561	DCTN2	Dynactin subunit 2	Down	Légaré et al.2014	cell growth and/or maintenance
P06733	ENO1	Alpha-enolase	Down	Légaré et al.2014	metabolic process
P09496	CLTA	Clatherin light chain A	Down	Légaré et al.2014	other
P28838	LAP3	Cytosol aminopeptidase	Down	Légaré et al.2014	metabolic process
P14314	PRKCSH	Glucosidase 2 subunit beta	Down	Légaré et al.2014	metabolic process
Q9Y265	RUVBL1	RuvB-like 1	Down	Légaré et al.2014	cell growth and/or maintenance
P21912	SDHB	Succinate dehydrogenase [ubuiquinone] iron-sulfur subunit, mitochondrial	Down	Légaré et al.2014	metabolic process
P04406	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	Down	Légaré et al.2014	metabolic process
P55072	VCP	Transitional endoplasmic reticulum ATPase	Down	Légaré et al.2014	protein transport
P25705	ATP5A1	ATP synthase subunit alpha, mitochondrial	Down	Légaré et al.2014	metabolic process
P48643	CCT5	T-complex protein 1 subunit epsilon	Down	Légaré et al.2014	protein metabolism
P06576	ATP5B	ATP synthase subunit beta, mitochondrial	Down	Légaré et al.2014	metabolic process
P50502	ST13	Hsc 70-interacting protein	Down	Légaré et al.2014	protein metabolism
P07205	PGK2	Phosphoglycerate kinase 2	Down	Légaré et al.2014	metabolic process
		Secretory actin-binding protein	Up	Pixton et al.2004	other
Q8IY3Y	DDX37	DEAD/DEAH box helicase	Missing	Pixton et al.2004	other
		Bromodomain PHD finger transcription factor	Missing	Pixton et al.2004	other
Q5BJF6	ODF2	Outer dense fiber protein 2	Up	Pixton et al.2004	other
B0QY90	EIF3L	Eukaryotic translation initiation factor 3 subunitI	Down	Azpiazu et al.2014	protein metabolism
P31947	SFN isoform2	Isoform 2 of 14-3-3 protein sigma	Down	Azpiazu et al.2014	cell growth and/or maintenance
Q02383	SEMG2	Semenogelin-2	Down	Azpiazu et al.2014	sexual reproduction
B4D170		cDNA FLJ5309, highly similar to Galectin-3-binding protein	Down	Azpiazu et al.2014	other
Q6XZB0	LIPI	Lipase member I	Down	Azpiazu et al.2014	metabolic process
P04279-2	SEMG1 isoform2	Isoform 2 of Semenogelin-1	Down	Azpiazu et al.2014	sexual reproduction
P02768	ALB	Serum albumin	Down	Azpiazu et al.2014	other
B3KRV7	LPL	Lipoprotein lipase	Down	Azpiazu et al.2014	metabolic process
P54886-2	ALDH18A1 isoform	Isoform short of Delta-I-pyrroline-5-carboxylate synthase	Down	Azpiazu et al.2014	metabolic process

Q99988	GDF15	Growth/differentiation factor 15	Down	Azpiazu et al.2014	cell growth and/or maintenance
Q6BCY4-2	CYB5R2 isoform2	Isoform 2 of NADH-cytochrome b5 reductase 2	Down	Azpiazu et al.2014	metabolic process
O43592	XPOT	Exportin-T	Down	Azpiazu et al.2014	protein transport
O75610	LEFTY1	Left-right determination factor 1	Down	Azpiazu et al.2014	cell growth and/or maintenance
P0C7M6	IQCF3	IQ domain-containing protein F3	Down	Azpiazu et al.2014	other
Q8NB37	PDDC1	Parkinson disease 7 domain-containing protein 1	Down	Azpiazu et al.2014	other
P051109	S100A8	Protein S100-A8	Down	Azpiazu et al.2014	cell growth and/or maintenance
Q8N427	NME8	Thioredoxin domain-containing protein 3	Down	Azpiazu et al.2014	sexual reproduction
Q96RL7-4	VPS13A isoform 4	Isoform 4 of vacuolar protein sorting-associated protein 13A	Down	Azpiazu et al.2014	protein transport
Q5H943	CT83	Kita-kyushu lung cancer antigen 1	Down	Azpiazu et al.2014	other
P02649	APOE	Apolipoprotein E	Down	Azpiazu et al.2014	other
P35237	SERPINB6	Serpin B6	Down	Azpiazu et al.2014	other
Q9UJS0	SLC25A13	Calcium-binding mitochondrial carrier protein Aralar2	Down	Azpiazu et al.2014	metabolic process
Q92599-3	SETP8 isoform3	Isoform 3 of Septin-8	Down	Azpiazu et al.2014	other
P0CG04	IGLC1	Ig lambda-I chain C regions	Down	Azpiazu et al.2014	other
O95336	PGLS	6-phosphogluconolactonase	Down	Azpiazu et al.2014	metabolic process
O14980	XPO1	Exportin-1	Down	Azpiazu et al.2014	protein transport
B4DY72	HSPH1	Heat shock protein 105kDa	Down	Azpiazu et al.2014	protein metabolism
P51148	RAB5C	Ras-related protein Rab-5C	Down	Azpiazu et al.2014	protein transport
O15084	ANKRD28	Serine/threonine-protein phosphatase 6 regulatory Ankyrin repeat subunit A	Down	Azpiazu et al.2014	phosphorylation
Q96KX2	CAPZA3	F-actin-capping protein subunit alpha-3	Down	Azpiazu et al.2014	cell growth and/or maintenance
B4DPM9	SHMT	Serine hydroxymethyltransferase	Down	Azpiazu et al.2014	metabolic process
Q53HC9	TSSC1	Protein TSSC1	Up	Azpiazu et al.2014	other
F8VV32	LYZ	Lysozyme C	Up	Azpiazu et al.2014	other
Q9H095-2	IQCG isoform2	Isoform 2 of IQ domain-containing protein G	Up	Azpiazu et al.2014	sexual reproduction
Q0P670		Uncharacterized protein C17orf74	Up	Azpiazu et al.2014	other

Q8N5K1	CISD2	CDGSH iron-sulphur domain-containing protein 2	Up	Azpiazu et al.2014	cell growth and/or maintenance
P05164-2	isoform H14 MPO	Isoform H14 of myeloperoxidase	Up	Azpiazu et al.2014	other
P04083	ANXA1	Annexin A1	Up	Azpiazu et al.2014	other
O14521	SDHD	Succinate dehydrogenase[ubiquinone] cytochrome b small subunit, mitochondrial	Up	Azpiazu et al.2014	metabolic process
P24752		Uncharacterized protein C7orf61	Up	Azpiazu et al.2014	other
Q81Z16		Uncharacterized protein C6orf10	Up	Azpiazu et al.2014	other
Q5SRN2	HIST1H2BC	Histone H2B type 1-A	Up	Azpiazu et al.2014	Histone
Q96A08	HIST1H4A	Histone H4	Up	Azpiazu et al.2014	Histone
P62805	IDH1 isoform A	Isoform A of isocitrate dehydrogenase [NAD] subunit beta, mitochondrial	Up	Azpiazu et al.2014	metabolic process
Q5TEC6		Histone cluster 2, H3, pseudogene 2	Up	Azpiazu et al.2014	Histone
B7Z112	GOT	Aspartate aminotransferase	Up	Azpiazu et al.2014	metabolic process
P10644	PRKAR1A	cAMP-dependent protein kinase type 1-alpha regulatory subunit	Up	Azpiazu et al.2014	phosphorylation
P16152	CBR1	Carbonyl reductase [NADPH] 1	Up	Azpiazu et al.2014	other
Q96SB4	SRPK1	SRSF protein kinase 1	Up	Azpiazu et al.2014	phosphorylation
E7EBP3		60S ribosomal protein kinase 1	Up	Azpiazu et al.2014	phosphorylation
Q5TGZ0	MINOS1	Mitochondrial inner membrane organizing system protein 1	Up	Azpiazu et al.2014	other
P13646-3	KRT13 isoform3	Isoform 3 of Keratin, type 1 cytoskeletal 13	Up	Azpiazu et al.2014	other
E5RJX2	RPS20	40S ribosomal protein S20	Up	Azpiazu et al.2014	other
Q961X5		Up-regulated during skeletal muscle growth protein 5	Up	Azpiazu et al.2014	other
Q16777	HIST2H2AC	Histone H2A type 2-C	Up	Azpiazu et al.2014	Histone
O43464-3	HTRA2 isoform3	Isoform 3 of serine protease HTRA2, mitochondrial	Up	Azpiazu et al.2014	protein metabolism
P49721	PSMB2	Proteasome subunit beta type-2	Up	Azpiazu et al.2014	protein metabolism
P14174	MIF	Macrophage migration inhibitory factor	Up	Azpiazu et al.2014	other
Q93077	HIST1H2AC	Histone H2A type 1-C	Up	Azpiazu et al.2014	Histone
O14841	OPLAH	5-oxoprolinase	Up	Azpiazu et al.2014	metabolic process
Q96QV6	HIST1H2AA	Histone H2A type 1-A	Up	Azpiazu et al.2014	Histone
Q5T5M1	AQP7	Aquaporin 7	Up	Azpiazu et al.2014	other
O14949	UQCRQ	Cytochrome b-c1 complex subunit 8	Up	Azpiazu et al.2014	metabolic process

F8WAS3	NDUFA5	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	Up	Azpiazu et al.2014	metabolic process
O76024	WFS1	Wolframin	Up	Azpiazu et al.2014	other

IV. Exploring different expression systems for recombinant expression of globins

Application to human neuroglobin and androglobin

An Bracke[1], David Hoogewijs[2], Sylvia Dewilde[1]

[1]Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, Antwerp 2610, Belgium
[2]Department of Medicine/Physiology, University of Fribourg, Chemin du Musée 5, CH 1700 Fribourg, Switzerland

Published in Analytical Biochemistry: Methods in the Biological Sciences (https://doi.org/10.1016/j.ab.2017.11.027)

1 Introduction

Globins can be defined as small heme-containing proteins which can mostly reversibly bind oxygen. They occur in the three kingdoms of life: Archaea, Prokaryotes and Eukaryotes and exist as single domain or chimeric proteins[1,2]. The best known single domain globins are the vertebrate hemoglobin (Hb) and myoglobin (Mb), playing respectively a role in oxygen transport and storage. Examples of chimeric globins are the flavohemoglobins (FHbs) in *Escherichia coli* and yeast, consisting of an N-terminal globin domain coupled to a ferredoxin reductase-like domain [3], and the globin coupled sensors (GCS), reported in Bacteria and Archaea, comprising an N-terminal globin domain linked to variable gene regulatory domains [4]. FHbs and GCS do not occur in vertebrates. The recent and rapid accumulation of genomic information has resulted in a substantial increase in newly recognized globins. Bioinformatics genome surveys revealed that more than 50% of the bacterial, approximately 20% of the archaeal and almost 90% of the eukaryote genomes contain genes encoding globins[1,5–7] Despite their great variety in function and amino acid sequence, most of them share a similar three-dimensional structure, referred to as the globin fold, characterized by a three-over-three alpha helical folding. Few globins, however, display a two-over-two alpha helical folding and are called 'truncated globins'.[8]

Globins are among the most investigated proteins in biological and medical sciences and represent a major tool for the study of gene evolution and structure-function relationship of proteins. Sperm whale Mb was the first protein of which the three-dimensional structure was revealed by X-ray crystallography [9]. The structure of human Hb followed five years later [10]. In the early times of protein crystallization and structure determination, the investigated proteins were purified from their natural context, e.g. Mb out of sperm whale muscle tissue and Hb out of human erythrocytes. However, under normal physiological conditions, most proteins are present in low cellular concentrations and cannot be purified out of the tissue itself. Since the early eighties it is has been possible to express proteins recombinantly in heterologous host cells like the bacterium *Escherichia coli (E.coli)* [11,12] or the yeast *Saccharomyces cerevisiae (S.cerevisiae)* [13]. These host cells can produce a large amount of recombinant protein, that enables the characterization of proteins with a low cellular abundance. In addition, this technique

facilitates the study of protein-function relationships by site-directed mutagenesis. This revolution in protein biochemistry caused an exponential increase in knowledge on the structure and function of proteins in general.

Neuroglobin (Ngb) was first described in 2000 as a vertebrate globin, preferentially expressed in brain and other nerve tissue[14]. The estimated amount of Ngb in the vertebrate brain under non-pathological conditions is only in the micromolar range. It was possible to perform a biochemical and structural characterization of NGB subsequent to the recombinant expression of the protein in *E.coli* [15].

Androglobin (Adgb) was described in 2012 as the fifth mammalian globin type (in addition to Hb, Mb, Ngb and cytoglobin (Cygb)) and it displays predominant expression in testis tissue in mice and human [16]. Adgb characteristically comprises four domains, an N-terminal ~350 residues calpain-like cysteine protease domain, a region of ~300 residues without known motifs/domains, followed by an ~150 residue circularly permutated globin domain, and a second, uncharacterized ~750 residues C-terminal domain. Interestingly, the globin domain, which normally consists of eight consecutive α-helices (named A-H), is circularly permutated and split into two parts within Adgb. The part containing helices A and B has been shifted in the C-terminal direction and is separated from the main globin sequence (helices C–H) by a calmodulin-binding IQ motif. Alignments of the split Adgb globin domain with mammalian Mb, Ngb, and Cygb sequences revealed that Adgb—despite its rearrangement—conforms to the criteria of the "globin fold" tertiary structure [16]. This was confirmed by molecular modelling of the human Adgb (ADGB) globin domain 3D structure, showing that the helix C–H segment alone is able to produce a *bona fide* globin fold [16]. *In vitro* structural and functional analysis of ADGB is necessary to better understand its molecular function during spermatogenesis, for this a recombinant form of ADGB is required.

In this paper, we explore three different expression systems for the expression of globins: a bacterial (*E.coli*), a yeast (*Pichia pastoris (P.pastoris)*) and an insect cell expression system (baculovirus infected *Sf9*). These three expression systems are widely used and in previous studies they have been compared for their recombinant expression of e.g. G-protein-coupled receptors[17], endostatin[18] and rabbit liver carboxylesterases[19]. We investigate their advantages and disadvantages for the expression of globins in general.

Furthermore, we used these three different expression systems to express the two different human globin types that were described above: I) the single domain globin neuroglobin (NGB), which has already been biochemically and structurally characterized and II) the uncharacterized chimeric globin ADGB.

2 Material and methods

2.1 Construction of the recombinant expression vectors

The cDNA sequence of the globin domain of human ADGB (referred to as CH-IQ-AB) and human NGB were cloned into the recombinant expression vectors pET23a (*E.coli*), pPicZα (*P.pastoris*) and pDEST10 (*Spodoptera frugiperda 9 (Sf9)*). The cDNA of full length ADGB was only cloned into the pDEST10 vector (Table1). Full length ADGB was RT-PCR amplified, with NcoI and XbaI overhangs (Table2) from human testis cDNA and cloned into the pENTR4 vector. The CH-IQ-AB domain was PCR-amplified for further cloning from the pENTR4 ADGB vector. The NGB domain was PCR-amplified from pET3a NGB construct, which was already available in our laboratory [15]. For the cloning of CH-IQ-AB in the pPicZα vector, the cDNA of the fragment was synthesized by Invitrogen GeneArt, with optimized codons for expression in *P.pastoris*.

Table I: Overview of the different expression vectors used in this paper

Vector	Tag	Promotor	Expression host
pET23a	6xHis, C-terminal	T7	*Escherichia coli* BL21(DE3)pLysS
pPicZα	6xHis, C-terminal	AOX	*Pichia pastoris* X33
pDEST10	6xHis, N-terminal	Polyhedrin	*Spodoptera frugiperda* Sf9

The amplifications of the fragments coding for the CH-IQ-AB and NGB were performed by PCR. PCR amplification was performed as follows: initial denaturation at 94 °C for 5 min, followed by 35 cycles (94 °C, 1 min; 55 °C, 1 min; 72 °C, 1min) and a final elongation at 72 °C for 10 min. The forward and reverse primers' sequences, including restriction sites/attB recombination sites, were in house designed and are listed in Table2. Recombinant pET23a CH-IQ-AB , pET23a NGB, pPicZα CH-IQ-AB and pPicZα NGB were

created using standard restriction enzyme cloning. Two mutations were made in the pET23a CH-IQ-AB construct resulting in the replacement of two cysteines (on position C3 and E2) using the QuikChange™ site-directed mutagenesis method (Stratagene). The recombinant pDEST10 CH-IQ-AB and pDEST10 NGB vector were created using the Gateway cloning technique: first a pENTR CH-IQ-AB and pENTR NGB vector were created by a BP recombination reaction between attB CH-IQ-AB/NGB PCR product and pDONR 221 vector and second the pDEST10 NGB and pDEST10 CH-IQ-AB were created by a LR recombination reaction between pENTR CH-IQ-AB/pENTR NGB vector and the pDEST10 vector. The pDEST10 ADGB (full length) was created using the Gateway cloning technique: LR recombination reaction between pENTR4 ADGB and the pDEST10 vector. The constructed recombinant expression vectors were verified via direct sequencing.

Table 2: Overview of the in-house designed primers for the cloning of the different recombinant expression vectors

Primer	**Sequence**	**Restriction site / attB**
ADGB pENTR4F	CATG**CCATGG**GCTCAGAGCTCAGCCCTACA	**NcoI**
ADGB pENTR4R	TGC**TCTAG**AGCATGGACATCCCCTGGTTACTT	**XbaI**
NGB pET23a F	GGGAATTC**CATATG**GAGCGCCCGGAGCCCGAG	**NdeI**
NGB pET23a R	CCG**CTCGAG**CTCGCCATCCCAGCCTCG	**XhoI**
CH-IQ-AB pET23a F	GGGAATTC**CATATG**CACATCTGCAGC ATGGTG	**NdeI**
CH-IQ-AB pET23a R	CCG**CTCGAG**GCTTTTAAACATTAGTCT	**XhoI**
NGB pPicZα F	CCG**CTCGAG**AAAAGAGAGGCTGAAGCTATGGAGCGCCC GGAGCCCGAGCTG	**XhoI** KeX2 cleavage site
NGB pPicZα R	CTAG**TCTAGA**GGCTCGCCATCCCAGCCTCGAC	**XbaI**
CH-IQ-AB pPicZα R	CCG **CTCGAG**AAAAGAGAGGCTGAAGCTCACATCTGCAGC ATGGTG	**XhoI** KeX2 cleavage site
CH-IQ-AB pPicZα R	CTAG**TCTAGA**GCTTTTAAACATTAGTCT	**XbaI**
NGB pDONR F	**GGGGACAAGTTTGTACAAAAAAGCAGGCT**TCATGGAGC GCCCGGAGCCCGAGCTG	**attBI**

NGB pDONR R	**GGGGACCACTTTGTACAAGAAAGCTGGGT**TTACTCGCC ATCCCAGCCTCGACTCATG	**attB2**
CH-IQ-AB pDONR F	**GGGGACAAGTTTGTACAAAAAAGCAGGCT**TCCACATCT GCAGCATGGTGTCATTTG	**attB1**
CH-IQ-AB pDONR R	**GGGGACCACTTTGTACAAGAAAGCTGGGT**TTAGCTTTT AAACATTAGTCTTAAG	**attB2**
CH-IQ-AB Cys C3 F	CATATGCACATC**TCC**AGCATGGTGTCA	**Cys → Ser**
CH-IQ-AB Cys C3 R	TGACACCATGCT**GGA**GATGTGCATATG	**Cys → Ser**
CH-IQ-AB Cys E2 F	GAACCAGAGAGC**TCC**CGATTTACGGAA	**Cys → Ser**
CH-IQ-AB Cys E2 R	TTCCGTAAATCG**GGA**GCTCTCTGGTTC	**Cys → Ser**

2.2 Expression in *Escherichia coli*

pPET23a CH-IQ-AB and pPET23a NGB were transformed into competent *E.coli* BL21(DE3)pLysS cells. Subsequently, cells were grown in a shaking incubator (160rpm) at 25°C in TB medium (1.2% bactotryptone, 2.4% yeast extract, 0.4% glycerol, 72mM potassium phosphate buffer, pH7.5) containing 200µg/ml ampicillin, 30µg/ml chloramphenicol, and 1mM δ-amino-levulinic acid. The expression of the C-terminal His-tagged fusion proteins was induced at A_{550nm} = 0.8 by the addition of isopropyl-1-thio-D-galactopyranoside (IPTG) to a final concentration of 0.4mM, and expression was continued overnight. The cells were harvested and resuspended in ~1/20 of the initial volume (resuspension buffer = 50mM Tris-HCl pH8.0, 300mM NaCl). The cells were then exposed to three freeze-thaw steps and were sonicated until completely lysed. The extract was clarified by centrifugation (12000xg, 10min, 4°C) and both supernatant and pellet fraction were screened for expression of recombinant protein (CH-IQ-AB, NGB) using SDS-PAGE (section 2.7).

2.3 Expression in *Pichia pastoris*

pPicZα CH-IQ-AB and pPicZα NGB were linearized with restriction endonuclease *SacI* and subsequently 5µg plasmid was transformed into *P.pastoris* X33 electrocompetent cells by electroporation: cells were pulsed using the manufacturer's instructions for *Saccharomyces cerevisiae* (Bio-rad electroporation system). Transformed cells were

plated out on YPDS agar plates (1% yeast extract, 2% peptone, 2% dextrose, 1M sorbitol, 2% agar) containing 100µg/ml Zeocin and were incubated 4 days at 30°C. The Mut Phenotype was determined by streaking the colonies out on both MD (1,34% YNB, 4x10-5% biotin, 2% dextrose) and MM (1,34% YNB, 4x10-5% biotin, 0,5% methanol) agar plates containing 100µg/ml Zeocin. 8 Mut^+ colonies, of which the presence of the insert was confirmed by PCR using the recommended sequencing primers, were inoculated in 25ml BMGY medium (1% yeast extract, 2% peptone, 100mM potassium phosphate pH6.0, 1.34% Yeast Nitrogen Base, $4x10^{-5}$% biotin, 1% glycerol) and grown overnight at 30°C in a shaking incubator (200 rpm) until they reached an OD_{600} = 2-5. Subsequently, the cells were harvested by centrifugation (2000xg, 10min, RT) and the cell pellet was resuspended to an OD_{600} of 1 in BMMY medium (1% yeast extract, 2% peptone, 100mM potassium phosphate pH6.0, 1.34% Yeast Nitrogen Base, $4x10^{-5}$% biotin, 0.5% methanol) containing 1mM δ-amino-levulinic acid and further grown at 30°C in a shaking incubator (200rpm). Methanol will induce the expression of the C-terminal His-tagged proteins, so every 24h 100% methanol was added to final concentration of 0.5% to maintain induction. 1ml of expression culture was collected every 24h to analyze expression levels with Western blot (see 2.7). For this, the cells were centrifuged (4000xg, 20min, RT) ; the supernatant was screened for secreted expression of recombinant protein; the cell pellet was resuspended in SDS-loading buffer and screened for intracellular expression. After 96h all cells were harvested by centrifugation (4000xg, 20min, RT).

2.4 Expression in *Spodoptera frugiperda 9*

pDEST10 CH-IQ-AB, pDEST10 NGB and pDEST10 ADGB were transformed into MAX Efficiency® DH10Bac™ competent *E.coli* (ThermoScientific™) to generate recombinant baculovirus DNA. The presence of the insert in the baculovirus DNA was verified by PCR using the recommended sequencing primers. 1µg of pure recombinant baculovirus DNA was subsequently transfected into *Sf9* insect cells (ThermoScientifi^{cTM}) using Cellfectin® II reagent (ThermoScientifi^{cTM}) following the manufacturers guidelines. *Sf9* cells were maintained and passaged adherently in cell culture flasks at 28°C with loose caps to allow gas exchange, the cells were cultured in Grace's Insect Cell Culture Medium (ThermoScientifi^{cTM}), supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin/Streptomycin. After transfection, cells were incubated at 28°C for 72 hours

until signs of viral infections could be seen. Recombinant virus particles were harvested by collecting the medium. The recombinant baculoviral stock was amplified three times (P3). Finally, *Sf9* cell cultures (Grace's Insect Cell medium supplemented with 10% FBS, 1% Penicillin/Streptomycin, 1mM δ-amino-levulinic acid, 14µM E64, 0.15µM Pepstatin A, 4µM CalpainII Inhibitor) were infected with P3 recombinant virus particles and were incubated for 72h at 28°C. Proteins were extracted by resuspending cells in native lysis buffer (25mM sodium phosphate (pH 7.0), 150mM NaCl, 1% NP40), followed by centrifugation (12000g, 20min, 4°C). Expression was screened using Western blot (see 2.7).

2.5 Purification

2.5.1 Purification NGB

20mM imidazole was added to the clarified cell lysates of *E.coli* or to the cell medium of *P.pastoris,* prior to loading onto a Ni Sepharose™ High Performance column (GE Healthcare), pre-equilibrated with equilibration buffer (50mM Tris-HCl pH7.5, 300mM NaCl, 20mM imidazole). Bound proteins were eluted with elution buffer (50mM Tris-HCl pH7.5, 300mM NaCl, 250mM imidazole). The fractions containing the pure recombinant protein were pooled, dialyzed against the proper buffer (50mM Tris-HCl pH8.5, 150mM NaCl, 0.5mM EDTA) and concentrated.

2.5.2 Purification CH-IQ-AB

CH-IQ-AB was purified under denaturing conditions. 8M urea was added to the cell lysate of *E.coli* and the recombinant protein was purified under denaturing conditions (8M urea was added to equilibration and elution buffer). Subsequently, the immobilized nickel affinity chromatography was carried out as described in 2.5.1. Afterwards, the pure unfolded recombinant CH-IQ-AB was refolded *in vitro* by dialysis towards a buffer without urea (50mM Tris-HCl pH8.5, 150mM NaCl, 0.5mM EDTA) and by adding an excess of free heme before starting dialysis (end concentration of 1M hemin). Subsequently, non-incorporated heme was removed by gel filtration (Sephaycryl S200 column, running buffer = 50mM Tris-HCl pH8.5, 150mM NaCl, 0.5mM EDTA).

2.6 Optical absorption spectra

To verify a correct heme-incorporation of the purified recombinant globin, UV/visible spectra were taken (200nm – 700nm). Spectral measurements were made with a spectrophotometer (Shimadzu UV-2101PC). The ferrous deoxy sample was obtained by adding an excess of sodium dithionite.

2.7 SDS-Page and Western Blot analysis

Samples were first denatured and reduced by the addition of an equal amount of loading buffer (4% SDS, 10% v/v β-mercapto-ethanol, 125mM Tris-HCl pH6.8, 20% glycerol) and by incubating them at 95°C for 5min. Subsequently, equal mass samples were separated by SDS-PAGE (15% acrylamide gel for NGB and CH-IQ-AB and 7,5% acrylamide for ADGB). For SDS-PAGE analysis, proteins were stained with Coomassie Brilliant Blue G-250. For Western Blot analysis, proteins were transferred to polyvinylidene difluoride (PVDF) membranes according to the manufacturer's instructions (Immobilon®-P PVDF membrane, pore size 0.45µm, Millipore). The membranes were probed with mouse-derived anti-His Tag antibody (6x-His Tag Monoclonal Antibody (4E3D10H2/E3) Invitrogen) or anti-ADGB antibody (polyclonal anti-ADGB antibody HPA036340 SIGMA) (1:5000 in 5% Milk TBS buffer); overnight at 4°C, followed by an incubation at room temperature for 1h with an HRP-conjugated antibody (1:5000 in TBS) (polyclonal goat anti-mouse/rabbit Immunoglobulins – HRP, DAKO). Subsequently, Luminata Forte Western HRP substrate (Millipore) was added to the membrane and the targeted proteins were visualized using a SynGene imager.

3 Results and discussion

3.1 Recombinant expression in *Escherichia coli*

The by far most popular host for recombinant protein expression is *E.coli*: 86% of the protein structures entered in the protein databank (PDB)[20] were expressed recombinantly in *E.coli*. This can be explained by the fact that, for a large scale protein

production, the *E.coli* system is the easiest, quickest and cheapest method. There are many commercial and non-commercial expression vectors available with different N- and C-terminal tags and many different *E.coli* strains, which are optimized for special applications. The first globins recombinantly expressed in *E.coli* were human Mb [21] and Hb [12], followed rapidly by many other globin types of many different organisms. In most cases, globins are small globular proteins, with no specific requirements in folding or post-translational processing and thus *E.coli* is often the perfect host for their large scale recombinant expression.

However, the high degree of accumulation of soluble protein during recombinant expression is not always accepted by the metabolic system of the host. Sometimes, the recombinantly expressed protein is toxic to the host cell and causes a cellular stress response, which results in poor and irregular protein expression. An example of such a situation are the expression of microbial FHbs. These chimeric globins possess the ability to generate superoxide and peroxide in the presence of reductant and NAD(P)H and the accumulation of these reactive oxygen species are toxic to the host cell. These problems can be overcome by expressing the FHbs under the control of an inducible promoter. Lewis et al. used the pBAD expression system, which allows tightly controlled, titratable expression of a protein through the regulation of arabinose concentration[22].

Another often encountered problem is the accumulation of target proteins into insoluble aggregates known as inclusion bodies. These proteins are misfolded and biologically inactive. Aggregation prone proteins require the existence of a number of molecular chaperones that interact reversibly with nascent polypeptide chains to prevent aggregation during the folding process. The formation of inclusion bodies could therefore result either from accumulation of high concentrations of folding intermediates or from inefficient processing by molecular chaperones[23]. Refolding from inclusion bodies is in many cases considered undesirable as it results most of the time in a poor recovery of bio-active protein[24]. Most used strategies for shifting the recombinant expression from inclusion bodies to soluble protein are: changing the expression temperature, media and transcription rates, trying various *E.coli* strains, including molecular chaperones, including tRNA complementation plasmids, expressing fragments, using fusion

technology, stabilizing mRNA or screening for soluble variants. For further reading on these strategies we refer to the review paper of Sorensen et al. [23]

The formation of inclusion bodies during recombinant globin expression is a frequently encountered problem. The first recombinantly expressed globins (Mb and Hb) were expressed as inclusion bodies and refolded *in vitro* into their functional form[12,21]. Nevertheless, by changing expression conditions most globins can be expressed and purified as soluble proteins[25].

Expression of NGB and CH-IQ-AB in E.coli

For the recombinant expression of human NGB and the globin domain of human ADGB (referred to as CH-IQ-AB) in *E.coli* we cloned them into the pET23a expression vector, which uses the strong inducible T7 promotor. The recombinant pET23a constructs were subsequently transformed in BL21(DE3)pLysS *E.coli*, where the recombinant expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). pET23a expresses the proteins with a C-terminal His-tag, which binds to several types of immobilized metal ions, like nickel. His-tagged proteins can thus easily be purified using nickel affinity chromatography. Both the soluble and insoluble fractions of the cell lysate were screened using SDS-PAGE for the recombinant over-expression of NGB and CH-IQ-AB (Fig1A).

NGB was expressed in a soluble form and was purified with nickel affinity chromatography. After purification of NGB, optical absorption spectra were recorded, which were characteristic for NGB[15]. NGB is a hexacoordinated globin which means that under deoxy conditions the protein forms two coordinative bonds with the iron atom of the heme group. The optical absorption spectrum of a hexacoordinated heme group has large amplitudes for the alpha band (562nm) and the Soret band (428nm). This observation indicates that the heme group was correctly incorporated in the polypeptide chain of NGB (Fig1B), consistent with a previous report[15].

CH-IQ-AB was expressed in *E.coli* as insoluble aggregates, also known as inclusion bodies and it was not possible to purify CH-IQ-AB as a soluble protein, even after altering the conditions (Fig1A). Different strategies were used to shift the expression from inclusion bodies to stable soluble protein. By changing time of induction and lowering the

temperature during protein expression we were able to avoid inclusion body formation, however CH-IQ-AB remained unstable and precipitated over time. We tried to increase the solubility of the globin domain by the addition of reducing agents or detergents and by changing the pH or salt concentrations of the resuspension buffer. Because CH-IQ-AB possesses two cysteine residues, which can possibly cause intermolecular disulfide bridge formation and hence precipitation, we mutated the cysteines into serines. Despite these efforts, CH-IQ-AB remained unstable. As such, CH-IQ-AB was expressed as inclusion bodies, purified under denaturing conditions (8M urea) and the recombinant protein was refolded *in vitro* by dialysis and adding free heme, as previously done for human CYGB[26]. In contrast to the predictions of the computer three-dimensional modeling[16] the refolded globin domain was not stable and precipitated over time. Despite the precipitation, we were able to measure an UV-visible spectrum of the refolded globin domain. The optical absorption spectra of the refolded ferrous deoxy CH-IQ-AB domain displayed the typical spectra of a hexacoordinated globin (high α band (557nm) and Soret band (421nm), low beta band (527nm)), like NGB (Fig1C). However, the purification of the refolded CH-IQ-AB was not efficient and only low yields could be obtained. The refolded globin domain remained unstable, precluding biophysical characterizations, including ligand binding kinetic determination and crystallization.

Folding studies of apomyoglobin (apoMb) [27] and plant hemoglobins [28] showed that the folding of a globin happens based on the contact order of not only local-residue contacts (as is seen with very small proteins), but also on long-range amino acid interactions. A stepwise-folding process of apoMb has been suggested in the following order; helices G, A, H, B, E, C and D. During this folding pathway, there are two folding intermediates, if one of these intermediates is misfolded, aggregation occurs. We can imagine that the folding of the circular permutated globin of ADGB, where helices A and B are placed after helices C-H, is more complicated than the folding of a typical globin with a normal consecution of the helices, like NGB.

Previously, the effect of circular permutation in Mb was investigated by Ribeiro et al.[29]. A circular permutated Mb was engineered, where helices A and B were shifted towards the C-terminus (similar to the globin domain of ADGB). Subsequently, this protein was expressed in *E.coli* as inclusion bodies and it was afterwards refolded *in vitro*. The refolded

circular permutated myoglobin displayed absorption spectra similar to the WT Mb. However, the circular permutation caused a significant decrease in the stability of the protein and a large fraction formed stable aggregates, similar to our experience with the circular permutated globin domain of ADGB.

Most plausible, the globin domain of ADGB needs long-range interactions from the amino-acid sequence of the full-length ADGB to form a stable globin fold. Another possibility is that the human globin domain is simply not able to fold correctly in the prokaryotic *E.coli* system. Therefore, we further investigated if the globin domain can be natively folded in an eukaryotic yeast expression system.

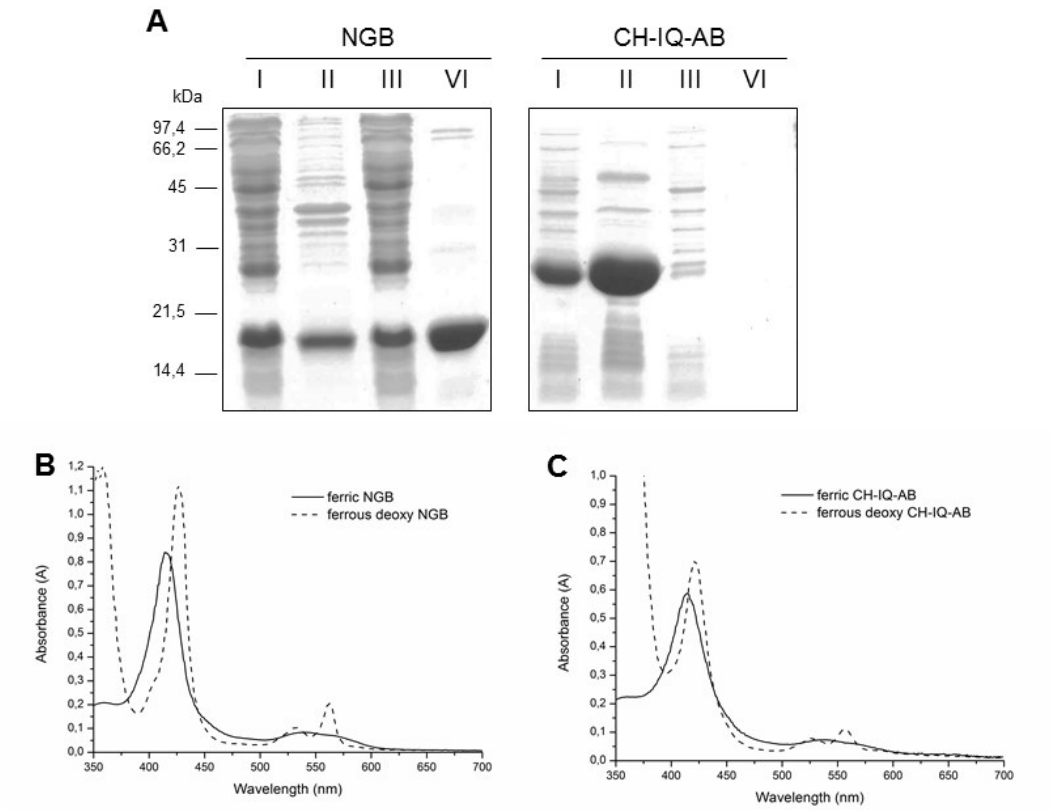


Figure1: A) SDS-PAGE (15%): Recombinant expression in *E.coli* of CH-IQ-AB (~25kDa) and NGB (~18kDa). I) *E.coli* total cell lysate II) Insoluble fraction of cell lysate III) Soluble fraction of cell lysate IV) Fraction after native purification with Nickel column. B) Optical absorption spectra of NGB expressed in *E.coli* (pET23a) and C) optical absorption spectra of refolded CH-IQ-AB, expressed in *E.coli* (pET23a).

3.2 Recombinant expression in *Pichia pastoris*

E.coli as expression system might be the obvious first choice, but it is not always the best choice, as we have seen for CH-IQ-AB. *E.coli* is a prokaryotic system and does not contain the subcellular machinery of eukaryotes needed for post-translational modifications, including proteolytic processing, folding, disulfide bond formation, glycosylation and phosphorylation. Many eukaryotic proteins need to be modified following translation in order to become active and/or adapt the proper structure. Yeast is an eukaryotic organism and, similar to *E.coli*, can be grown to very high densities, which makes them especially useful for large scale protein production. The two most common yeast strains used for recombinant protein expression are *S.cerevisiae* and the methylotrophic yeast *P.pastoris*.

Only few cases are reported where globins are recombinant expressed in a yeast system [30–32]. An example of a globin that has been successfully expressed in *P.pastoris* is the *Crocodylus siamensis* hemoglobin. This crocodile Hb displays potent antibacterial and anti-inflammation activity, as well as strong antioxidant properties[33–35]. Furthermore, it comprises six cysteine residues in the α-globin chain and three cysteine residues in the β-globin chain[35]. Each of these cysteine residues can form disulfide bonds both of the intermolecular and intramolecular formation. The expression in *E.coli* resulted in low expression yields and precipitation after purification[32]. To overcome these problems the crocodile Hb was successfully expressed in *P.pastoris*.

As the CH-IQ-AB globin domain of ADGB displays similar problems of precipitation after purification and comprises also three cysteine residues (two in the α-helices and one in the IQ motif). We recombinantly expressed CH-IQ-AB in the *P.pastoris* system.

Expression of NGB and CH-IQ-AB in P.pastoris

For the expression of NGB and CH-IQ-AB we chose for methylotrophic yeast *P.pastoris* because it is widely used, cost-effective and suitable for large-scale protein production. cDNA of NGB and CH-IQ-AB were cloned into the pPicZα vector, which places a native *S.cerevisiae* α-secretion signal before the recombinant protein. The advantage of a secreted expression is that *P.pastoris* secretes very low levels of native proteins, facilitating further purification steps. The pPicZα vector uses the AOX promoter for a

high-level, methanol-induced expression. After induction of recombinant expression, every 24h a sample was taken for analysis of expression of CH-IQ-AB and NGB using Western blot. NGB is expressed and secreted by *P.pastoris* (Fig2A). 96h post-induction, NGB was purified out of the expression medium and optical absorption spectra were taken (Fig2B). These spectra evidence proper folding and incorporation of heme in NGB in line with previous descriptions [31]. Despite our efforts to optimize the cDNA sequence of CH-IQ-AB to the codon usage of *P.pastoris*, we could not detect recombinant expression of CH-IQ-AB in *P.pastoris* (Fig2A), neither in the secreted fraction nor in the intracellular fraction.

This may be explained by the possibility that CH-IQ-AB cannot be expressed as a stable soluble form because it requires the context of the full length ADGB protein for proper folding as was already suggested in section 3.1. In *E.coli*, misfolded proteins are stored as inclusion bodies. In eukaryotic systems, protein folding occurs in a different way; proteins are folded and processed in the endoplasmic reticulum (ER) and/or the Golgi apparatus and sometimes secreted into the extracellular environment through vesicular transport (as is in the case of recombinant expression of CH-IQ-AB in *P.pastoris*). In eukaryotes aberrant folding properties of the target protein and/or high level production can lead to the accumulation of unfolded or even aggregated protein in the ER, which can initiate the unfolded protein response (UPR) and ER-associated degradation [36–40]. This can thus possibly explain the fact that we cannot detect expression of CH-IQ-AB in *P.pastoris*.

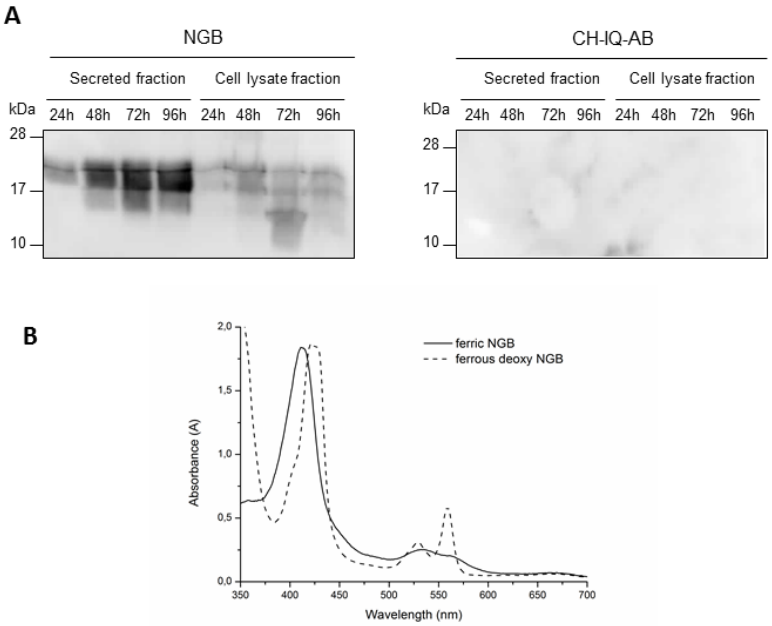


Figure2: A) Western Blot analysis (anti His-tag) (15% polyacrylamide) of recombinant expression of NGB (~19kDa) and CH-IQ-AB (~26kDa) in X33 *P.pastoris* (pPicZα expression vector). Expression was analyzed every 24h post induction. Both secreted and intracellular fractions were analyzed. B) Optical absorption spectra of NGB, expressed in *P.pastoris* (pPicZα).

3.3 Recombinant expression in *Sf9* insect cells

Finally, we opted for a third expression system, one that is much closer to mammalian cells; the baculovirus-insect cell system. This expression system is a binary system, consisting of a recombinant baculovirus as the vector and lepidopteran insect cells as the host[41]. Insect cells represent a higher eukaryotic system than yeast; they are able to carry out more complex post-translational modifications and have a better folding machinery for mammalian proteins. Furthermore, baculoviral vectors can accommodate large DNA inserts, which allows for the production of multi-subunit protein complexes or even high molecular weight virus-like particles [42–44]. A potential disadvantage of the baculovirus-insect system is that it is a high-cost and time-consuming method of recombinant protein expression.

So far, there are no reports of globins recombinantly expressed in insect cells. Since the baculovirus-insect cell expression system is widely used for the recombinant expression

of high molecular weight proteins, we used it to express the 190kDa full length ADGB protein.

Expression of NGB, CH-IQ-AB and full length ADGB in Sf9 cells

NGB, CH-IQ-AB and full length ADGB cDNA were cloned into the pDEST10 vector, which possesses the strong transcriptional promoter derived from the very late baculovirus gene *polyhedrin*, able to produce milligrams of proteins. *Sf9* insect cells were infected with recombinant NGB, CH-IQ-AB and ADGB baculovirus particles. 72h after infection, the expression of recombinant protein was analyzed with Western blot (Fig3). Expression was detected for NGB and ADGB. The expression level of CH-IQ-AB was however very low, again possibly reflecting that the globin domain would require the context of the full length ADGB for proper folding.

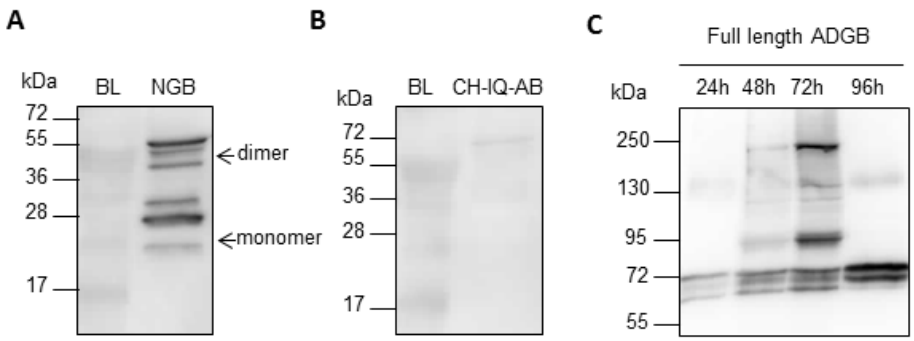


Figure3: A) Western blot analysis (anti His tag) (15% polyacrylamide) of *Sf9* cell lysate (soluble fraction) infected with recombinant NGB baculovirus particles. Expected molecular weight of NGB is ~21kDa. Multiple bands are detected, most probably caused by dimerization and formation of different stable protein conformations or by post-translational modifications. B) Western blot analysis (anti His tag) (15% polyacrylamide) of *Sf9* cell lysate (soluble fraction) infected with recombinant CH-IQ-AB baculovirus particles. Expected molecular weight of CH-IQ-AB is ~28kDa. C) Western blot analysis (anti ADGB) (7,5% polyacrylamide) of *Sf9* cell lysate (soluble fraction) infected with recombinant ADGB baculovirus particles. Expected molecular weight of ADGB is ~194kDa. ADGB might be (auto)cleaved by internal proteases or by the N-terminal calpain domain of ADGB. BL = Blanc, uninfected *Sf9* cell lysate.

Surprisingly, for NGB multiple bands were detected with Western blot analysis (Fig3A). It seems that NGB a) can be detected as dimer and b) appears in three apparent molecular mass forms, in *Sf9* insect cells. The dimerization could not be prevented by heat denaturation or addition of reducing agents, like β-mercaptoethanol or DTT. Dimerization was also seen for NGB when it was expressed in *E.coli*, but not in *P.pastoris* (Fig4). In *P.pastoris* NGB is secreted into the medium, and is thus present in lower concentrations than when it is intracellularly expressed in *E.coli* or *Sf9* , dimer formation may in this way be concentration dependent. This phenomenon of dimerization, resistant to heat denaturation and reduction, has been observed previously for Ngb in whole protein extracts from mouse brain and retinas[45–48]. Furthermore, it has been shown that three forms of apparent molecular masses (~17, ~19 and ~21 kDa), specifically recognized by anti-Ngb antibodies, were identified in mitochondrial enriched fractions of mouse retina[49]. This pattern of three different molecular mass forms is similar to the one that is seen after expression of NGB in *Sf9* insect cells (Fig3A). These forms might represent atypical migrations of stable protein conformations and it has been hypothesized that rupture/formation of the intramolecular (CD7-B5) disulfide bond of NGB[50] can give rise to different apparent molecular mass forms [49]. Post-translational modifications could also occur, such as covalent links of cofactors or phosphorylation. Indeed, it has previously been displayed that Ngb is phosphorylated during hypoxia[51]. As insect cells are known for their capabilities of protein N-glycosylation, we analyzed the possibility that Ngb is N-glycosylated using the NetNGlyc1.0 prediction server[52] and by the addition of PNGaseF, which enzymatically cleaves N-glycans, to the cell lysate. No N-glycosylation sites were predicted in NGB and PNGaseF did not cause a shift in molecular weight on Western blot, ruling out the possibility that NGB is N-glycosylated by the insect cell system. It seems that NGB only forms these different molecular mass forms after expression in the higher eukaryotic baculoviral insect cell system and not in *E.coli* or *P.pastoris*, assuming that the latter do not provide the cellular environment for formation of these different conformations or post-translational modifications.

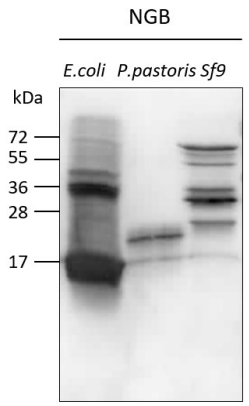


Figure 4: Western blot analysis (anti-His tag) (15% polyacrylamide) of *E.coli, P.pastoris* and *Sf9* cell lysates were NGB was recombinantly in expressed. Molecular weight of NGB in *E.coli* should be ~18kDa, in *P.pastoris* ~19kDa and in *Sf9* ~21kDa, due to different C-terminal or N-terminal His-tags. Dimerization of NGB resistant to heat denaturation and reducing agents is seen in both *E.coli* and *Sf9*, not in *P.pastoris*. In *Sf9* NGB appears in three apparent molecular weight forms, possibly due to the formation of different stable protein conformations or by post-translational processing. This is not seen in *E.coli* or *P.pastoris*.

In contrast to the CH-IQ-AB domain, the full length ADGB was significantly expressed in *Sf9* cells, confirming the hypothesis that the full length protein is required for stable expression. However, we observed a time-dependent truncation of the full length protein (Fig3C) and even after scaling-up the expression, the amount of full length ADGB was insufficient for further biophysical characterization. This truncation is most probably a result of proteolysis by intracellular proteases or by auto-cleavage by the N-terminal calpain-like domain, a phenomenon commonly observed among calpains[53–56]. To further evaluate if ADGB autolytic activity would be the source of this potential auto-cleavage we added three different protease inhibitors to the growth medium; the cysteine protease inhibitor E64, the acid protease inhibitor Pepstatin A and the calpain specific Calpain II inhibitor. However, administration of these inhibitors did not prevent the truncation of ADGB. The ADGB calpain-like domain is homologous to calpain 7, for which no specific protease inhibitors exists. Interestingly, substantial autolytic activity has also been demonstrated for calpain-7[57]. Further research is necessary to understand this putative auto-cleavage of ADGB.

4 Conclusion

In this paper we explored three different expression systems for the recombinant expression of globins; the prokaryote expression system *E.coli* and the two eukaryote expression systems *P.pastoris* and baculovirus infected *Sf9*. In general, globins are small globular proteins, which do not need post-translation modification for a proper folding and in this regard *E.coli* is the obvious first choice for recombinant expression. However, the major disadvantage of recombinant expression in *E.coli* is the formation of insoluble aggregates, inclusion bodies.

We attempted to express the circular permutated globin domain of ADGB in *E.coli*, *P.pastoris* and *Sf9* cells, however, a soluble expression was not possible. This suggests that the globin domain of ADGB needs the full context of the full length protein in order to form a stable globin fold. Furthermore, we displayed that ADGB was truncated, possibly due to auto-cleavage by the internal calpain-like domain. Our study demonstrates that it is not always possible to express globins in *E.coli* and that switching to a (higher) eukaryotic expression system might represent a valid alternative for more complex, multi-domain globins, like the chimeric globin ADGB.

When NGB was recombinantly expressed in insect cells, three apparent molecular mass forms and dimerization were observed, this was not observed in *E.coli* or *P.pastoris*. Interestingly, the same multiple band pattern on Western blot has been observed in mitochondrial enriched fractions of mouse retina and in brain tissue. It seems that the baculoviral insect cell expression system provides a cellular environment for folding and processing of proteins that is more similar to the natural context of Ngb (brain and retina) compared to the cellular environments of *E.coli* or *P.pastoris*. In this regard, it might be worthwhile to switch to a (higher) eukaryotic expression systems for eukaryotic proteins in general, even if they are easily expressed in a bacterial or yeast expression system.

5 References

1. Vinogradov, S. N. *et al.* A phylogenomic profile of globins. *BMC Evol. Biol.* **6,** 31 (2006).

2. Vinogradov, S. N. *et al.* Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 11385–9 (2005).

3. Bonamore, A. & Boffi, A. Flavohemoglobin: Structure and reactivity. *IUBMB Life* **60,** 19–28 (2008).

4. Germani, F., Moens, L. & Dewilde, S. in 1–47 (2013). doi:10.1016/B978-0-12-407693-8.00001-7

5. Vinogradov, S. N. *et al.* A model of globin evolution. *Gene* **398,** 132–42 (2007).

6. Vinogradov, S. N. *et al.* Microbial eukaryote globins. *Adv. Microb. Physiol.* **63,** 391–446 (2013).

7. Vinogradov, S. N., Tinajero-Trejo, M., Poole, R. K. & Hoogewijs, D. Bacterial and archaeal globins - A revised perspective. *Biochim. Biophys. Acta - Proteins Proteomics* **1834,** 1789–1800 (2013).

8. Wittenberg, J. B., Bolognesi, M., Wittenberg, B. A. & Guertin, M. Truncated hemoglobins: A new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. *J. Biol. Chem.* **277,** 871–874 (2002).

9. KENDREW, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181,** 662–666 (1958).

10. Muirhead, h. & perutz, m. F. Structure of haemoglobin. A three-dimensional fourier synthesis of reduced human haemoglobin at 5-5 a resolution. *Nature* **199,** 633–638 (1963).

11. Shine, J., Fettes, I., Lan, N. C., Roberts, J. L. & Baxter, J. D. Expression of cloned beta-endorphin gene sequences by Escherichia coli. *Nature* **285,** 456–463 (1980).

12. Nagai, K. & Thogersen, H. C. Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in Escherichia coli. *Nature* **309,** 810–812 (1984).

13. Valenzuela, P., Medina, A., Rutter, W. J., Ammerer, G. & Hall, B. D. Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298,** 347–350 (1982).

14. Shi, X. & Jarvis, D. L. Protein N-glycosylation in the baculovirus-insect cell system. *Curr. Drug Targets* **8,** 1116–25 (2007).

15. Dewilde, S. *et al.* Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J. Biol. Chem.* **276,** 38949–55 (2001).

16. Hoogewijs, D. *et al.* Androglobin: a chimeric globin in metazoans that is preferentially

expressed in Mammalian testes. *Mol. Biol. Evol.* **29,** 1105–14 (2012).

17. Sarramegna, V., Talmont, F., Demange, P. & Milon, A. Heterologous expression of G-protein-coupled receptors: Comparison of expression systems from the standpoint of large-scale production and purification. *Cell. Mol. Life Sci.* **60,** 1529–1546 (2003).

18. Mohajeri, A. *et al.* The Challenges of Recombinant Endostatin in Clinical Application: Focus on the Different Expression Systems and Molecular Bioengineering. *Adv. Pharm. Bull.* **7,** 21–34 (2017).

19. Morton, C. L. & Potter, P. M. Comparison of Escherichia coli, Saccharomyces cerevisiae, Pichia pastoris, Spodoptera frugiperda, and COS7 cells for recombinant gene expression. Application to a rabbit liver carboxylesterase. *Mol. Biotechnol.* **16,** 193–202 (2000).

20. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P. E. B. The Protein Data Bank. *Nucleic acids research (28)* 235–242 (2000). Available at: http://www.rcsb.org/pdb/home/home.do. (Accessed: 23rd October 2017)

21. Varadarajan, R., Szabo, A. & Boxer, S. G. Cloning, expression in Escherichia coli, and reconstitution of human myoglobin (gene expression/heme protein). *Biochemistry* **82,** 5681–5684 (1985).

22. Lewis, M. E. S., Corker, H. A., Gollan, B. & Poole, R. K. *A Survey of Methods for the Purification of Microbial Flavohemoglobins*. *Methods in Enzymology* **436,** (Elsevier Masson SAS, 2008).

23. Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb. Cell Fact.* **4,** 1 (2005).

24. Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **99,** 303–310 (2005).

25. Natarajan, C. *et al.* Expression and purification of recombinant hemoglobin in escherichia coli. *PLoS One* **6,** 1–7 (2011).

26. de Sanctis, D. *et al.* Crystal structure of cytoglobin: the fourth globin type discovered in man displays heme hexa-coordination. *J. Mol. Biol.* **336,** 917–927 (2004).

27. Ishimura, B. C. N. Review Folding of apomyoglobin : Analysis of transient intermediate structure during refolding using quick hydrogen deuterium exchange and NMR. **93,** 10–27 (2017).

28. Nakajima, S., Álvarez-Salgado, E., Kikuchi, T. & Arredondo-Peter, R. Prediction of folding pathway and kinetics among plant hemoglobins using an average distance map method. *Proteins Struct. Funct. Genet.* **61,** 500–506 (2005).

29. Ribeiro, E. A. & Ramos, C. H. I. Circular permutation and deletion studies of myoglobin indicate that the correct position of its N-terminus is required for native stability and solubility but not for native-like heme binding and folding. *Biochemistry* **44,** 4699–4709 (2005).

30. Mould, R., Hofmann, O. & Brittain, T. Production of human embryonic haemoglobin

(Gower II) in a yeast expression system. *Biochem. J* **298 Pt 3,** 619–22 (1994).

31. Ye, Q. *et al.* Pichia pastoris Production of Tat-NGB and Its Neuroprotection on Rat Pheochromocytoma Cells. *Mol. Biotechnol.* **58,** 22–29 (2016).

32. Anwised, P. *et al.* Cloning, Expression, and Characterization of Siamese Crocodile (Crocodylus siamensis) Hemoglobin from Escherichia coli and Pichia pastoris. *Protein J.* **35,** 256–268 (2016).

33. Jandaruang, J. *et al.* The effects of temperature and pH on secondary structure and antioxidant activity of Crocodylus siamensis hemoglobin. *Protein J.* **31,** 43–50 (2012).

34. Phosri, S. *et al.* An investigation of antioxidant and anti-inflammatory activities from blood components of Crocodile (Crocodylus siamensis). *Protein J.* **33,** 484–492 (2014).

35. Srihongthong, S. *et al.* Complete amino acid sequence of globin chains and biological activity of fragmented crocodile hemoglobin (Crocodylus siamensis). *Protein J.* **31,** 466–476 (2012).

36. Vanz, A. L., Nimtz, M. & Rinas, U. Decrease of UPR- and ERAD-related proteins in Pichia pastoris during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. *Microb. Cell Fact.* **13,** 23 (2014).

37. Whyteside, G. *et al.* Native-state stability determines the extent of degradation relative to secretion of protein variants from Pichia pastoris. *PLoS One* **6,** e22692 (2011).

38. Hesketh, A. R., Castrillo, J. I., Sawyer, T., Archer, D. B. & Oliver, S. G. Investigating the physiological response of Pichia (Komagataella) pastoris GS115 to the heterologous expression of misfolded proteins using chemostat cultures. *Appl. Microbiol. Biotechnol.* **97,** 9747–9762 (2013).

39. Vanz, A. L. *et al.* Physiological response of Pichia pastoris GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. *Microb. Cell Fact.* **11,** 103 (2012).

40. Welihinda, A. A., Tirasophon, W. & Kaufman, R. J. The cellular response to protein misfolding in the endoplasmic reticulum. *Gene Expr.* **7,** 293–300 (1999).

41. Van Oers, M. M., Pijlman, G. P. & Vlak, J. M. Thirty years of baculovirus-insect cell protein expression: From dark horse to mainstream technology. *J. Gen. Virol.* **96,** 6–23 (2015).

42. Kost, T. A., Condreay, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23,** 567–575 (2005).

43. Hefferon, K. L. Baculovirus late expression factors. *J. Mol. Microbiol. Biotechnol.* **7,** 89–101 (2004).

44. Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nature biotechnology* **22,** 1583–1587 (2004).

45. Fabrizius, A. *et al.* Critical re-evaluation of neuroglobin expression reveals conserved

patterns among mammals. *Neuroscience* **337,** 339–354 (2016).

46. Schmidt, M. *et al.* How does the eye breathe? Evidence for neuroglobin-mediated oxygen supply in the mammalian retina. *J. Biol. Chem.* **278,** 1932–1935 (2003).

47. Jin, K., Mao, Y., Mao, X., Xie, L. & Greenberg, D. A. Neuroglobin expression in ischemic stroke. *Stroke* **41,** 557–559 (2010).

48. Last visited 04/08/2017. *Last visited 04/08/2017* Available at: https://www.mybiosource.com/prods/Antibody/Polyclonal/NGB/datasheet.php?products_id=129771#QLREF.

49. Lechauve, C. *et al.* Biochimica et Biophysica Acta Neuroglobin involvement in respiratory chain function and retinal ganglion cell integrity. **1823,** 2261–2273 (2012).

50. Hamdane, D. *et al.* The Redox State of the Cell Regulates the Ligand Binding Affinity of Human Neuroglobin and Cytoglobin. *J. Biol. Chem.* **278,** 51713–51721 (2003).

51. Jayaraman, T. *et al.* 14-3-3 Binding and phosphorylation of neuroglobin during hypoxia modulate six-to-five heme pocket coordination and rate of nitrite reduction to nitric oxide. *J. Biol. Chem.* **286,** 42679–42689 (2011).

52. R. Gupta, E. J. and S. B. Prediction of N-glycosylation sites in human proteins. (2004). Available at: http://www.cbs.dtu.dk/services/NetNGlyc/. (Accessed: 25th October 2017)

53. Ono, Y. *et al.* PLEIAD/SIMC1/C5orf25, a novel autolysis regulator for a skeletal-muscle-specific calpain, CAPN3, scaffolds a CAPN3 substrate, CTBP1. *J. Mol. Biol.* **425,** 2955–2972 (2013).

54. Ono, Y. *et al.* The N- and C-terminal autolytic fragments of CAPN3/p94/calpain-3 restore proteolytic activity by intermolecular complementation. *Proc. Natl. Acad. Sci.* **111,** E5527–E5536 (2014).

55. Murphy, R. M. & Lamb, G. D. Endogenous calpain-3 activation is primarily governed by small increases in resting cytoplasmic [Ca2+] and is not dependent on stretch. *J. Biol. Chem.* **284,** 7811–7819 (2009).

56. Ermolova, N., Kramerova, I. & Spencer, M. J. Autolytic activation of calpain 3 proteinase is facilitated by calmodulin protein. *J. Biol. Chem.* **290,** 996–1004 (2015).

57. Osako, Y. *et al.* Autolytic activity of human calpain 7 is enhanced by ESCRT-III-related protein IST1 through MIT-MIM interaction. *FEBS J.* **277,** 4412–26 (2010).

V. The protein interaction partners of androglobin suggest a function in the chromatoid body of the male germ cell

An Bracke[1], María Suarez Alonso[3], Sara Santambrogio[2], Sylvia Dewilde[1], David Hoogewijs[3]

[1]Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, Antwerp 2610, Belgium
[2]Institute of Physiology and Zürich Center for Integrative Human Physiology, University of Zürich, Zürich, Switzerland
[3]Department of Medicine/Physiology, University of Fribourg, Chemin du Musée 5, CH 1700 Fribourg, Switzerland

This manuscript is in preparation

1 Introduction

Globins are small oxygen binding proteins present in the three domains of life: Archaea, Bacteria and Eukaryotes. In mammalians, five different globin types are characterized: hemoglobin, myoglobin, neuroglobin, cytoglobin and the recently discovered androglobin.[1] The latter was discovered by comparative genomic studies in deuterostome sequence data, searching for globin-like genes[2]. Androglobin (Adgb) is a large protein with a modular domain structure; it comprises four domains, I) an N-terminal ~350 residues calpain-like cysteine protease domain, II) a region of ~300 residues without known motifs/domains, III) followed by an ~150 residue circularly permutated globin domain, and IV) a second, uncharacterized ~750 residues C-terminal domain.

Initial expression analysis of Adgb in mice tissue displayed a predominant expression in testis tissue, with a pronounced expression pattern in the lumen of the seminiferous tubules, where mature spermatozoa are formed. This process is called spermatogenesis and it covers a complex network of processes. During embryogenesis, the primordial germ cells migrate into the gonad and become immature germ cells, called spermatogonia. Beginning at puberty, these spermatogonia divide mitotically, and some of them enter into their first meiotic division and become primary spermatocytes. After completing this first meiotic division, the daughter cells become secondary spermatocytes, which have a haploid number of duplicated chromosomes. The second meiotic division of the secondary spermatocytes results in smaller cells called spermatids, which have a haploid number of unduplicated chromosomes. Spermatids transform into spermatozoa in a process called spermiogenesis, which involves cytoplasmic reduction and differentiations of the tail pieces.[3]

A more detailed expression analysis in mice testis tissue displayed an increase of Adgb expression during post-meiotic phases of spermatogenesis, when sperm cells are in the spermatid stage[2]. Additionally, RT-qPCR analysis in

different mouse testis-derived cell lines corresponding to Leydig cells, Sertoli cells, spermatogonia, and spermatocytes, displayed very low expression levels for Adgb, which is consistent with the hypothesis that Adgb functions in the late phases of spermatogenesis. *In silico* expression analysis of different human tissues also strongly supports a testis-specific expression pattern for ADGB and more interestingly, an R2 database-based analysis suggested 4-fold higher expression levels in fertile vs. infertile males[4]. Collectively, this data support the hypothesis that Adgb plays a crucial role in the late phases of spermatogenesis, however, the molecular function of Adgb remains currently unknown.

As proteins act rarely alone, a critical step towards revealing the biological function of a protein is to determine its molecular interaction partners. Thus, in order to understand the physiological role of Adgb in spermatogenesis and define the potential pathways Adgb is involved in, we explored the Adgb-dependent interactome by performing a pull down co-immunoprecipitation (co-IP) in mouse testis tissue, followed by mass spectrometry analysis. Subsequently, some of these detected proteins (based on their abundancy and function) were validated for their interaction with Adgb using reciprocal co-IP experiments and Fluorescence Resonance Energy Transfer (FRET) analysis.

2 Material and Methods

2.1 Tissue immunoprecipitation followed by mass spectrometry analysis

Img adult mouse testis/brain tissue was homogenized in a non-denaturing lysis buffer (50mM Tris-HCl pH 7.5, 300mM NaCl, 5mM EDTA, 1% NP-40, protease inhibitors), followed by a pre-clearing step where the tissue lysate was incubated with 30µL of Protein G sepharose beads for 2 hours at 4°C. Tissue lysate and beats were afterwards centrifuged for 5 min at low speed. Immunoprecipitation was performed in the supernatant with a custom-made

specific monoclonal Adgb antibody (Abmart) or an unspecific IgG antibody as negative control. Subsequently, the Adgb-complex was recovered by binding on 30µL Protein G sepharose beads (overnight, 4°C). After extensive washing with PBS, the Adgb-complexes were eluted with SDS loading buffer and loaded on an 8% SDS-PAGE. Gels were stained with Coomassie Brilliant blue staining and gel lanes were carefully cut into 9 small pieces. These gel pieces were again cut in small pieces and washed 2x with 100µL 100µM NH_4HCO_3/50% acetonitrile and 1x with 50µL acetonitrile. All three supernatants were discarded. Subsequently, 20µL trypsin (5ng/µL in 10mM Tris-HCl pH8.2, 2mM $CaCl_2$) and 50µL buffer (10mM Tris-HCl pH8.2, 2mM $CaCl_2$) was added and heated for 30min at 60°C. Finally, supernatant was removed and digested peptides were extracted with 150µL 0.1% TFA/50% acetonitrile. Samples were dried and dissolved in 20µL 0.1% formic acid. 3µL was injected for LC-ESI/MS/MS and analysis was performed by the protein analysis group of the Functional Genomics Center of Zurich. Database searches were performed by using the Mascot (SwissProt, all species) search program.

2.2 (Reciprocal) co-Immunoprecipitation followed by immunoblotting

Co-immunoprecipitation (co-IP) in mouse testis tissue was performed as described above (2.1) with a custom made specific monoclonal Adgb antibody (Abmart). Reciprocal co-IP was performed with specific antibodies against Tdrd6, Spata20 and Piwil1 and independent Adgb antibody (Sigma) as positive control (see 2.3 for information about antibodies). The eluate after co-IP was loaded on 8% SDS-PAGE gels and electroblotted onto nitrocellulose membranes. The nitrocellulose membranes were blocked for 1 h in 5% nonfat milk in TBS (10mM Tris-HCl, pH 7.5, 200mM NaCl) followed by overnight incubation with primary antibodies at 4°C (1:1000 dilution). After 1h incubation with secondary antibody (anti-rabbit IgG HRP conjugated, 1:10000 dilution in TBS), the bound antibodies were visualized by Lumi-Phos WB Chemiluminescent Substrate (ThermoScientificTM). Reciprocal co-IP was

carried out with specific antibodies for TDRD6, PIWIL1 and SPATA20 (2.3) and afterwards blotted with an independent custom-made antibody against Adgb.

2.3 Antibodies

Androglobin: I) custom made specific Adgb mouse monoclonal antibody (Abmart) II) anti-ADGB rabbit polyclonal HPA036340 SIGMA. Tdrd6: anti-TDRD6 rabbit polyclonal antibody, ABE231 Merck Millipore. Piwil1: anti-PIWL1 rabbit polyclonal antibody, Abcam ab12337. Spata20: anti-SPATA20 rabbit polyclonal antibody, OriGene TA32424.

2.4 Gateway Entry Constructs

Gateway entry vectors were cloned or purchased for human ADGB, SPATA20, PIWIL1 and TDRD6. ADGB was RT-PCR amplified, with NcoI and XbaI overhangs (Primer F: 5'- GCTCA GAGCTCAGCCCTACA-3', Primer R: 5'-TGCTCTAGAGCATGGACATCCCC TGGTTACTT-3'), from human testis cDNA and cloned into the pENTR4 vector. pENTR233 SPATA20 and pENTR233 PIWIL1 were purchased with DNASU. pENTR4 TDRD6 was purchased with Genscript.

2.5 Fluorescence Resonance Energy Transfer (FRET) analysis

Gateway Entry vectors for ADGB, SPATA20, PIWIL1 and TDRD6 were recombined with pECFP-C1-DEST or pEYFP-C1-DEST using LR recombinase (Invitrogen) to obtain the expression vector for a cyan or yellow fluorescent fusion protein, respectively. HeLa cells were transiently transfected with the pECFP-ADGB and pEYFP-PIWIL1, pEYFP-SPATA20 or pEYFP-TDRD6 constructs, using Rotifect (Roth). Fluorescence Resonance Energy Transfer (FRET) analysis was performed using a standard inverted confocal microscope (Nikon GmbH, Düsseldorf Germany) as was described before[5,6]. FRET was monitored 24h post-transfection. FRET efficiencies for CFP-ADGB and YFP-

PIWIL1, YFP-SPATA20 or YFP-TDRD6 fusion proteins pairs were calculated from 30 to 40 randomly selected cells.

3 Results

3.1 Tissue immunoprecipitation followed by mass spectrometry analysis

The Adgb interactome was determined by performing a pull down co-immunoprecipitation in adult mouse testis tissue with a custom made Adgb-specific antibody, followed by mass spectrometry. As negative control, the same pull down immunoprecipitation was performed in mouse brain tissue, which does not express Adgb[2]. Peptides were identified using the Mascot search engine. A complete list of proteins, which were detected exclusively in the testis immunoprecipitation, can be found in the supplementary table 1 and a Genome Ontology (GO)-analysis was performed on this complete list of proteins (Fig1). The three main biological processes that were found in the GO analysis are I) metabolic processes II) cellular processes and III) localization.

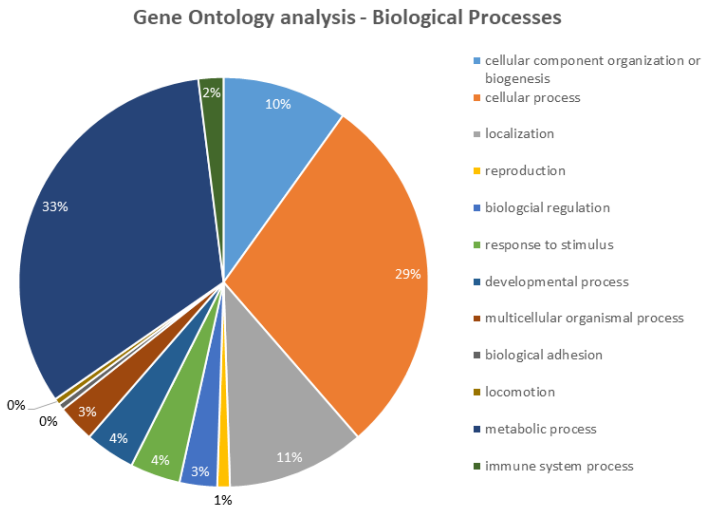


Figure 1: Genome Ontology (GO) analysis of proteins found in mass spectrometer analysis after a Adgb pull down co-immunoprecipitation in mice testis tissue. GO analysis was performed using PANTHER 12.0[7]

This GO-analysis does not provide conclusive results, because the biological processes that are enriched are very general and are not pointing consistently to a single cellular pathway or biological process. Hence, we narrowed the list of interaction partners down to only the proteins with ≥ 10 peptide hits in mass spectrometer analysis (Table1) and analyzed the functions of these proteins using the UniProt database. Interestingly, among the candidate cellular interaction partners of Adgb many RNA binding proteins are present, functioning in RNA processing, RNA splicing and post-transcriptional regulation. A few proteins can directly be linked to spermatogenesis and sperm function: Tdrd6, Piwil1 and Spata20. These three proteins are predominantly expressed in testis tissue, like Adgb. Therefore, we validated the interaction between Adgb and these three proteins. Moreover, Tdrd6 and Spata20 were found to interact both with Piwil1[8].

Table I: Potential protein-interaction partners of Adgb, determined by an Adgb pull-down co-immunoprecipitation experiment in mouse testis tissue, followed by mass spectrometry analysis. Proteins exclusively detected in the testis tissue precipitation, with ≥ 10 hits in mass spectrometry analysis are listed.

UniProt ID	Full protein name	Function	Hits
EIF3A_MOUSE	Eukaryotic translation initiation factor 3 subunit A	RNA binding component of eukaryotic translation initiation factor 3 complex; initiation of protein synthesis	59
HNRPM_MOUSE	Heterogeneous nuclear ribonucleoprotein M	Pre-mRNA binding protein; Involved in splicing	47
ADGB_MOUSE	Androglobin	Unknown	44
GRP78_MOUSE	78 kDa glucose-related protein	Protein folding and degradation of misfolded proteins	33
TDRD6_MOUSE	Tudor domain-containing protein 6	Involved in spermatogenesis and chromatoid body formation	26
TERA_MOUSE	Transitional endoplasmic reticulum ATPase	Role in fragmentation and reassembly of Golgi stacks	25
HSP72_MOUSE	Heat shock-related 70 kDa protein 2	Chaperone function	23
ALDH2_MOUSE	Aldehyde dehydrogenase, mitochondrial	Alcohol metabolism	22
LMNB1_MOUSE	Lamin-B1	Component of the nuclear lamina	21
GRP75_MOUSE	Stress-70 protein, mitochondrial	Chaperone protein	20
HXK1_MOUSE	Hexokinase-1	Energy metabolism	16
CRBG3_MOUSE	Beta/gamma crystalline domain-containing protein 3	Unknown	16
VINC_MOUSE	Vinculin	Involved in cell-matrix adhesion and cell-cell adhesion	16

UniProt ID	Full protein name	Function	Hits
AKA12_MOUSE	A-kinase anchor protein 12	Mediates the subcellular compartmentation of protein kinase A and protein kinase C	16
RBM14_MOUSE	RNA-binding protein 14	Transcription regulation	16
RENT1_MOUSE	Regulator of nonsense transcripts 1	RNA-dependent helicase; involved in nonsense-mediated mRNA decay	15
ALDOA_MOUSE	Fructose-biphosphate aldolase A	Energy metabolism (glycolysis)	15
EF1D_MOUSE	Elongation factor 1-delta	Regulation of transcription	15
PIWIL1_MOUSE	Piwi-like protein 1	Central role in spermatogenesis. Role in piRNA metabolism	15
PSMD2_MOUSE	26S proteasome non-ATPase regulatory subunit 2	Component of the ubiquitin-proteasome pathway	15
U520_MOUSE	U5 small nuclear ribonucleoprotein 200kDa helicase	Role in mRNA splicing	14
KINH_MOUSE	Kinesin-1 heavy chain	Microtubule-dependent motor protein	14
DDX5_MOUSE	Probable ATP-dependent RNA helicase DDX5	Involved in the alternative regulation of pre-mRNA splicing	14
ODP2_MOUSE	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	Energy metabolism	14
ATPA_MOUSE	ATP synthase subunit alpha, mitochondrial	Energy metabolism	13
ACLY_MOUSE	ATP-citrate synthase	Energy metabolism	13
DHX8_MOUSE	ATP-dependent RNA helicase DHX8	Facilitates nuclear export of spliced mRNA	13
CAND1_MOUSE	Cullin-associated NEDD8-dissociated protein 1	Role in protein ubiquitination	13

UniProt ID	Full protein name	Function	Hits
QCR2_MOUSE	Cytochrome b-c1 complex subunit 2, mitochondrial	Component of mitochondrial respiratory chain	13
FXR1_MOUSE	Fragile X mental retardation syndrome-related protein 1	RNA-binding protein; may regulate intracellular transport and local translation of certain mRNAs	13
TPP2_MOUSE	Tripeptidyl-peptidase 2	Component of the ubiquitin-proteasome pathway	12
HYOU1_MOUSE	Hypoxia up-regulated protein 1	Role in cytoprotective cellular mechanisms triggered by oxygen deprivation	12
VP13A_MOUSE	Vacuolar protein sorting-associated protein 13A	May play a role in the control of protein cycling through the trans-Golgi network, lysosomes and plasma membranes	12
WDR62_MOUSE	WD repeat-containing protein 62	Role in centriole replication and mitotic spindle organization	12
DHX30_MOUSE	Putative ATP-dependent RNA helicase DHX30	Role in assembly of the mitochondrial large ribosomal subunit	12
ECHA_MOUSE	Trifunctional enzyme subunit alpha, mitochondrial	Lipid metabolism	12
DHX9_MOUSE	ATP-dependent RNA helicase A	RNA-binding; promotes mRNA stability	11
SUCB1_MOUSE	Succinate—CoA ligase [ADP-forming] subunit beta, mitochondrial	Energy metabolism; TCA cycle	11
NASP_MOUSE	Nuclear autoantigenic sperm protein	Required for DNA replication, normal cell cycle progression and cell proliferation	11
HMCS2_MOUSE	Hydroxymethylglutaryl-CoA synthase, mitochondrial	Cholesterol metabolism	11
THIM_MOUSE	3-ketoacyl-CoA thiolase, mitochondrial	Fatty acid metabolism	11

UniProt ID	Full protein name	Function	Hits
STK31_MOUSE	Serine/threonine-protein kinase 31	Protein phosphorylation	11
IF4G1_MOUSE	Eukaryotic translation initiation factor 4 gamma 1	Role in initiation of translation	10
EFIG_MOUSE	Elongation factor 1-gamma	Role in protein biosynthesis	10
PGK2_MOUSE	Phosphoglycerate kinase 2	Essential for sperm motility and male fertility; Role in pyruvate metabolism	10
ENOA_MOUSE	Alpha-enolase	Energy metabolism	10
CCAR2_MOUSE	Cell cycle and apoptosis regulator protein 2	Role in mRNA processing and mRNA splicing	10
HSP7C _MOUSE	Heat shock cognate 71 kDa protein	Chaperone. Acts as a repressor of transcriptional activation	10
TIF1B_MOUSE	Transcription intermediary factor 1-beta	Role in transcription regulation	10
SPT20_MOUSE	Spermatogenesis-associated protein 20	May play a role in fertility regulation	10
ACON_MOUSE	Aconitate hydratase, mitochondrial	Energy metabolism	10

3.2 Validation of the interaction of Adgb with Piwil1, Tdrd6 and Spata20

The interaction between Adgb and Piwil1, Tdrd6 and Spata20 was validated through co-immunoprecipitation (co-IP) and reciprocal co-IP, followed by Western blot and fluorescence resonance energy transfer (FRET) analysis. The co-IP experiments were performed in mouse testis tissue and showed that mouse Adgb interacts with Piwil1, Tdrd6 and Spata20 (Fig2).

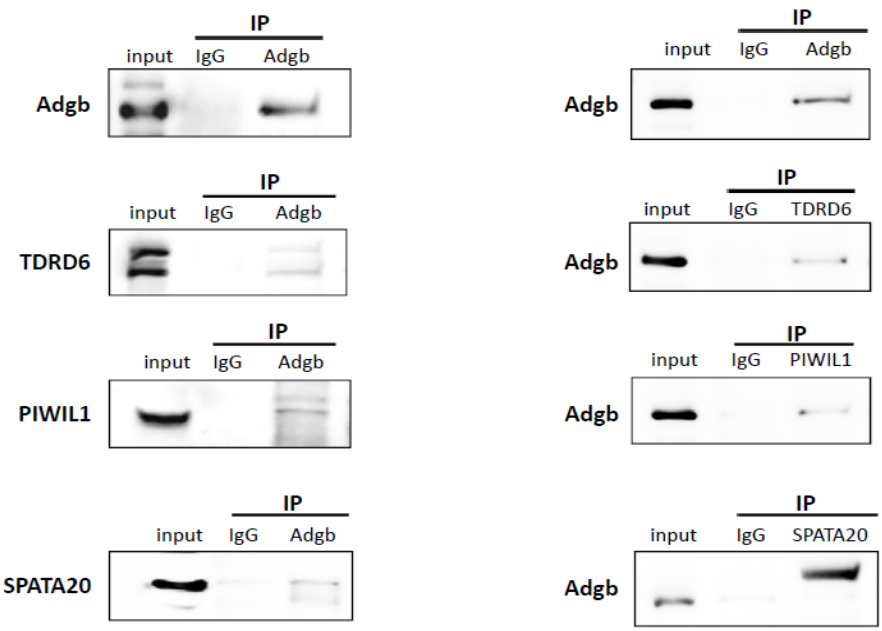


Figure 2: Identification of Adgb-interacting proteins by co-IP (left) and reciprocal co-IP (right) followed by western blot. Left panel: Immunoprecipitation was performed with a specific Adgb antibody and an unspecific IgG as negative control. Subsequent to complex binding on Protein G sepharose beads, elution and sample loading on 8% SDS-PAGE, immunoblotting was performed with specific antibodies against Tdrd6, PiwiL1 and Spata20 and an independent Adgb antibody. Right panel: Immunoprecipitation was performed with specific antibodies against Tdrd6, PiwiL1 and Spata20 and Adgb and an unspecific IgG as negative control. Subsequent to complex binding on Protein A sepharose beads, elution and sample loading on 8% SDS-PAGE, immunoblotting was performed with a specific antibody against Adgb.

FRET analysis was performed by co-expression of human ADGB coupled to enhanced Cyan Fluorescent Protein (eCFP) and PIWIL1 or SPATA20 coupled to enhanced Yellow Fluorescent Protein (eYFP) in HeLa cells. FRET analysis robustly showed that human ADGB associates with both PIWIL1, and SPATA20 (Fig3). At the time of submission of this thesis optimization of the FRET analysis with ADGB and TDRD6 fusion proteins was still ongoing. This double validation, *in vivo* as well as *in vitro*, indicates consistent interaction between Adgb and Piwil1, Tdrd6 and Spata20 in both mice and human.

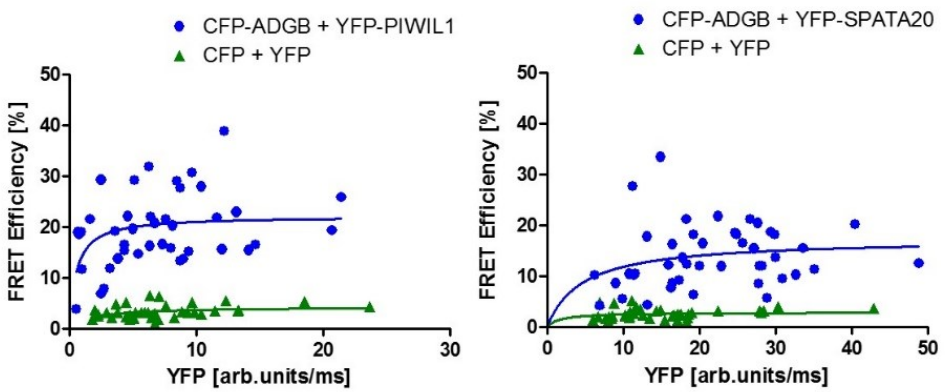


Figure 3: HeLa cells were transiently transfected with the pECFP-ADGB and pEYFP-PIWIL1 or pEYFP-SPATA20 constructs and fluorescence resonance energy transfer (FRET) was monitored 24h post-transfection. FRET efficiencies for CFP-ADGB and YFP-PIWIL1 (left panel) or YFP-SPATA20 (right panel) fusion protein pairs were calculated from 30 to 40 randomly selected cells. Blue filled circles represent FRET efficiency values from individual cells which result in robust mean FRET efficiency as indicated by the blue curve for ADGB-PIWIL1 (left panel) and ADGB-SPATA20 (right panel). The green curves represent the random FRET measured with empty vectors for ECFP and EYFP where unspecific FRET signal (or false-positive FRET signal) results from random collision of the dyes.

3.3 Chromatoid body

Both Tdrd6 and Piwill play a central role in spermatogenesis and are main components of the chromatoid body (CB)[8,9], a specialized germ cell organelle. Remarkably, 27 out of the 88 established CB components[10] were detected in the pull down Adgb immunoprecipitation, this is 30% of the chromatoid body content (Fig4A). In addition, six out of the nine main components of the CB were found in the Adgb pull down experiment (Fig4B). This supports the hypothesis that Adgb might play a crucial role in the CB of the male germ cell, with Tdrd6 and Piwill as important interaction partners.

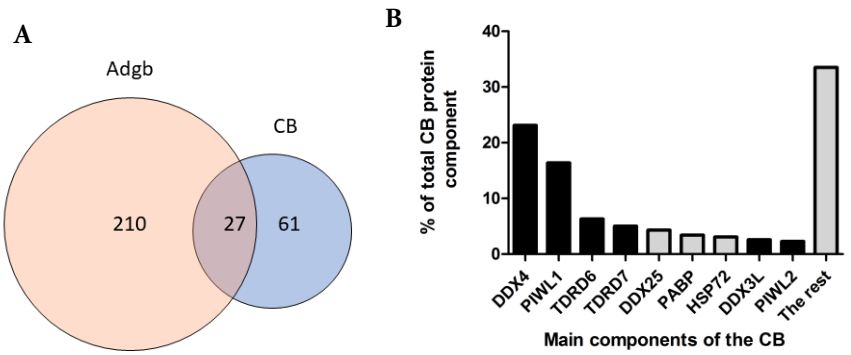


Figure 4: A) Overlap of CB protein components (blue)[10] and putative Adgb interaction partners (orange), which were detected after Adgb-immunoprecipitation in mouse testis tissue with pull down co-IP in mouse testis tissue. B) Main components of the chromatoid body[10]. Black bars represent proteins also detected after Adgb-immunoprecipitation in mouse testis tissue.

4 Discussion

Interestingly, 30% of the CB protein content was detected among the putative Adgb interacting partners. Piwil1 and Tdrd6 are both main components of the CB. We can thus assume that Adgb also might play a crucial role in the CB. The CB is an irregularly-shaped cytoplasmic structure in the male germ cell, which starts to appear in late pachytene spermatocytes and becomes prominent in round spermatids during spermatogenesis (Fig5). In round spermatids, the CB is a dynamic granule that moves actively within the vicinity of the nuclear pores. During spermiogenesis, the CB moves caudally to the base of the flagellum at the opposite site of the acrosome region, where it disperses or it is degraded[11]. Many RNA binding proteins and RNA strands are found in the CB[10]. Although the function of the CB remains elusive, many hypothetical roles have been proposed based on the components of this specialized organelle. In late phases of spermatogenesis, transcription is silenced by the compaction of the genome through histone-to-protamine transition[12]. During these steps, however, translation of several proteins necessary during spermiogenesis, is still ongoing. Based on its structural features and composition it can be assumed that the CB serves as a center for mRNA storage and processing in these late, transcriptional-silent phases of spermatogenesis. CBs move in a microtubule-dependent fashion and associate with nuclear pores, organelles and RNPs, suggesting that CBs also play a role in transport[13].

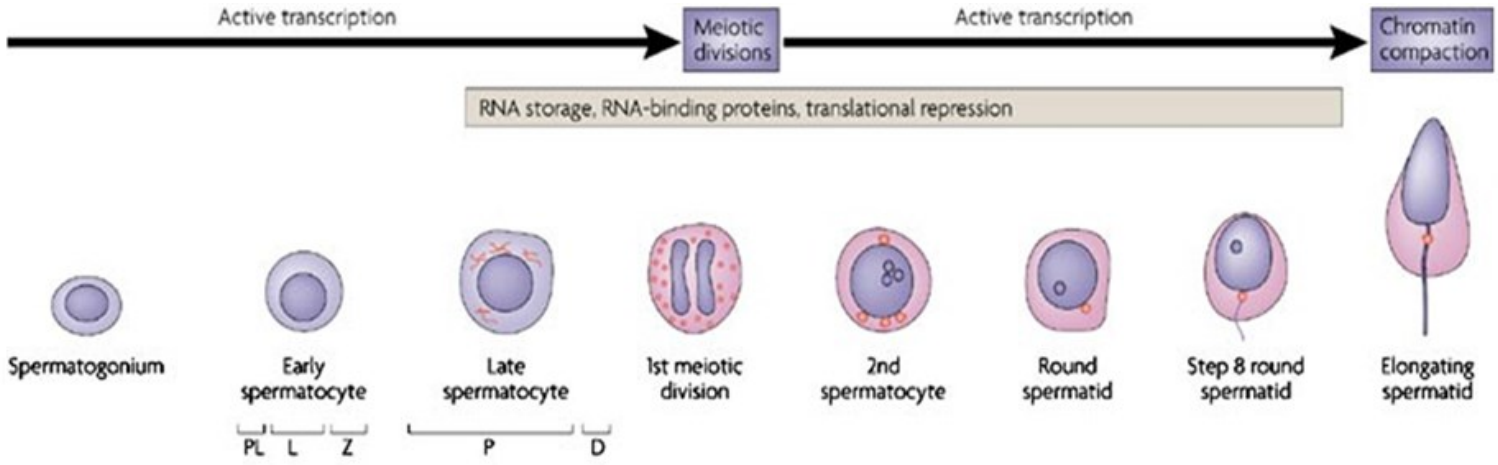


Figure 5: Waves of transcriptional upregulation during spermatogenesis are associated with mitotic growth of spermatogonia, the onset of meiosis and entry into post-meiotic cell maturation. After meiotic divisions in post-meiotic haploid male germ cells, transcription is active until the silencing of the haploid genome owing to compaction of the genome by the replacement of histones with protamines. Due to a drastic inhibition of transcriptional activity at late stages of germ-cell maturation, mRNA storage and translational regulation by RNA-binding proteins have an important role in the control of the synthesis of many spermatid and spermatozoa proteins. Chromatoid body (CB) material is first apparent in late pachytene spermatocytes prior to the first meiotic division as intermitochondrial fibrous structures (in red) that disperse during meiotic division. The CB is condensed to its final shape immediately after meiosis, and remains a distinctive feature in the cytoplasm of post-meiotic round spermatids. It diminishes in size and then disappears in elongating spermatids. D, diplotene; L, leptotene; P, pachytene; PL, preleptotene; Z, zygotene. Figure was taken from[14].

Piwil1 is a member of the PIWI proteins, a class of proteins specifically expressed in animal germ cells. In mice, PIWI proteins consist of three members: Piwil1 (Miwi), Piwil2 (Mili) and Miwi2. The expression of these genes during male germ cell differentiation is temporally and spatially regulated and is essential for spermatogenesis. PIWI proteins bind piRNAs, a class of 24-31-nt small noncoding RNAs. Recently, it has been shown that Piwil1 exerts a piRNA-guided mRNA degradation in late pachytene spermatocytes and post-meiotic spermatids[15]. Immunoprecipitation of endogenous Piwil1 complexes revealed that Tudor-domain family proteins (Tdrd1, Tdrd6, Tdrd7, Tdrkh and Stk31) are physiological interaction partners of Piwil1[8]. Tudor domain containing proteins bind on the methylated arginine residues in the N-terminal RG repeats of Piwil1.[8] Piwil1 is a RNA-binding protein and many other putative Adgb interaction partners are RNA-binding proteins with hence a role in post-transcriptional and/or translational regulation. To investigate a putative RNA-binding role for Adgb we searched for RNA binding motifs in the sequence of Adgb: no classical RNA-binding domains, like RNA recognition motif (RRM), K homology (KH) domain, DEAD motif, double-stranded RNA-binding motif (DSRM) or zinc-finger were recognized in the primary sequence of Adgb. However, interactome studies identified >1300 experimentally confirmed human RNA-binding proteins (RBP), many of these newly identified RBPs do not harbor a classical RNA-binding domain as mentioned above, but rather contain disordered and low amino acid complexity sequences, such as the RG/RGG motif[16]. Adgb was not identified as a RBP in these experiments and no RGG/RG motifs are found in Adgb, suggesting that Adgb most probably does not possess an RNA-binding property. However, a cryptic RNA-binding domain (d2adca1) is predicted by the PHYRE2 Protein Fold Recognition Server[17], although with moderate probability, in the more C-terminal region of the protein. This finding remains to be validated.

Tdrd6 is a TUDOR domain containing protein and it is well-established that CBs are enriched in TUDOR domain containing proteins (Tdrd6, Tdrd7,

Tdrd5, Tdrd9, Tdrd11). It has been proposed that the TUDOR domains of these proteins provide interaction interfaces and create a scaffold to organize the CB structure[9]. It has been shown that Tdrd6 is essential for a correct CB architecture: the development from round into elongated spermatids is abrogated in Tdrd6 null mice and their round spermatids have "ghost" CBs, whose architecture is greatly disrupted[18]. Tdrd6 contains 6 TUDOR domains and may therefore serve as a scaffold for multipoint contacts with PIWI proteins (like Piwil1) and thereby coordinate the formation and operation of the CB. It has been shown that Tdrd6 binds to methylated RK motifs in Piwil1[8]. Adgb however does not possess RK motifs and the Tdrd6 binding occurs most probably in a different way. During the transition from meiosis I to meiosis II, Tdrd6 is C-terminally cleaved from a 250kDa form to a 230kDa form. It has been suggested that these two forms have distinct functions during spermatogenesis. The 250kDa Tdrd6 is active in prophase I spermatocytes and functions in spliceosome maturation and mRNA splicing[19]. The 230kDa Tdrd6 in contrast, seems to be essential in the extended 3'UTR-triggered nonsense-mediated mRNA decay (NMD) pathway during spermiogenesis. This function depends on Tdrd6-promoted assembly of mRNA and decay enzymes in CBs[20,18]. The cleavage occurs at a predicted Caspase I-cleavage site, by an unknown protease. Calpain and caspases are both cysteine proteases, which often target the same proteins[17]. It is therefore tempting to hypothesize that the calpain-like domain of Adgb might play a role in the C-terminal cleavage of Tdrd6. The fact that Adgb expression starts in the pachytene phase, when the cleavage occurs, further supports this hypothesis. Notably, RENT1 an established interactor of Tdrd6[19] is also present among the newly identified Adgb interactors (Table 1).

Lack of Piwil1 or Tdrd6, results in male infertility in mice[9,21]. Piwil1 null mice show a maturation arrest of the spermatogenesis at the round spermatid stage; no spermatid undergoes cellular elongation or nuclear condensation. The same infertility phenotype was characterized in Tdrd6 null mice. Spermiogenesis is the final stage of spermatogenesis, which constitutes the maturation of

spermatids into mature, motile spermatozoa. It involves modification of several organelles in addition to the formation of several structures, including the flagellum and the acrosome. While over 400 mutations have been reported to affect spermatogenesis in the mouse, relatively few result in complete arrest of development at the end of the round spermatid stage[22,23]. Most of these mutations are in genes encoding for proteins functioning in RNA processing and/or regulation (like *Tdrd6* and *Piwil1*). Four of these mutation are in genes encoding transcription factors (*Crem*, *Tbpl1*, *Taf7l*, *Rfx2*) essential for the initiation of spermiogenesis[23–26]. We can hypothesize that the CB plays a role in the storage and processing of the mRNA of these transcription factors, which are necessary for the initiation of spermiogenesis and thus that a non-functional CB results in a maturation arrest at the round spermatid stage of spermatogenesis (as observed in Piwil1 null and Tdrd6 null mice). Unexpectedly, Adgb was not detected in the interactomes of Piwil1[8] or Tdrd6[18], nor in the components of the CB[10]. The Adgb mRNA, on the other hand, was detected in the RNA content of the CB[10]. This may be explained by a transient and fast association between Adgb and the CB and/or by a low cellular abundancy of Adgb in the male germ cell.

The molecular function of Spata20 has not been characterized. Computer modeling showed that Spata20 possess a conserved thioredoxin-like domain[27]. Thioredoxins play multivalent cellular roles: they act as reductases in redox control, protect proteins from oxidative aggregation and inactivation, help the cells coping with various environmental stresses (reactive oxygen species (ROS), peroxynitrite, arsenate), promote protein folding and regulate programmed cell death via denitrosylation[28]. Spata20 is specifically expressed in testis tissue, the expression starts from the spermatid phase and proceeds through adulthood. Spata20 was identified as interaction partner of Piwil1 by co-immunoprecipitation[27]. Although the function of Spata20 is not known, the presence of the thioredoxin domain suggests a role in the disulfide reductase system, which can provide electrons to a large range of enzymes and is found to be crucial in defense against oxidative stress. It has been suggested that

globins with a fast autoxidation rate and strong hexa-coordination, may have a redox function[29]. Adgb contains an internal globin domain, which has been shown to be hexa-coordinated[2]. It is well conceivable that electron transfer occurs between the globin domain of Adgb and the thioredoxin domain of Spata20 and that they together play a role in the defense against oxidative stress during spermatogenesis.

5 Conclusion and Future Perspectives

We determined the cellular protein interaction partners of Adgb through a pull down co-IP followed by mass spectrometry. Through functional analysis of these putative interaction partners of Adgb we found that many of them were RNA binding proteins, linked to the regulation of transcription or translation. We validated the interaction between Adgb and Tdrd6, Piwil1 and Spata20 through reciprocal co-IP and FRET analysis (optimization of FRET analysis between Adgb and Tdrd6 is still ongoing at the moment of submission of this thesis). Both Tdrd6 and Piwil1 are major components of the CB, suggesting that Adgb plays a role in this male-germ cell specific cell organelle. However, some additional validations are necessary to confirm the role of Adgb in the CB. First of all, the localization of Adgb at the CB has to be confirmed, this can be done by isolation of CBs out of mouse testis tissue, as was previously be described[10], followed by Western blot against Adgb. Secondly, the role of the calpain-like protease domain of Adgb in the truncation of Tdrd6 has to be investigated. For this purpose site-directed mutagenesis of the catalytic sites in the calpain-like domain of Adgb could be performed, followed by co-overexpression of Adgb (WT versus mutated) and Tdrd6 in a mammalian cell line. Truncation of Tdrd6 can subsequently be screened with Western blot. These additional validations and experiments will lead, together with the findings described in this chapter, towards a better understanding of the molecular function of Adgb in spermatogenesis.

6 References

1. Burmester, T. & Hankeln, T. Function and evolution of vertebrate globins. *Acta Physiol.* **211,** 501–514 (2014).

2. Hoogewijs, D. *et al.* Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes. *Mol. Biol. Evol.* **29,** 1105–14 (2012).

3. Neto, F. T. L., Bach, P. V., Najari, B. B., Li, P. S. & Goldstein, M. Spermatogenesis in Humans and its affecting factors. *Semin. Cell Dev. Biol.* (2016). doi:10.1016/j.semcdb.2016.04.009

4. Platts, A. E. *et al.* Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Hum. Mol. Genet.* **16,** 763–73 (2007).

5. Konietzny, R. *et al.* Molecular imaging: Into in vivo interaction of HIF-1α and HIF-2α with ARNT. *Ann. N. Y. Acad. Sci.* **1177,** 74–81 (2009).

6. Prost-Fingerle, K., Hoffmann, M. D., Schützhold, V., Cantore, M. & Fandrey, J. Optical analysis of cellular oxygen sensing. *Exp. Cell Res.* (2017). doi:10.1016/j.yexcr.2017.03.009

7. Mi, H. *et al.* PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45,** D183–D189 (2017).

8. Chen, C. *et al.* Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 20336–41 (2009).

9. Hosokawa, M. *et al.* Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: Domain composition, intracellular localization, and function in male germ cells in mice. *Dev. Biol.* **301,** 38–52 (2007).

10. Meikar, O. *et al.* An atlas of chromatoid body components. *RNA* **20,** 483–95 (2014).

11. FAWCETT, D. O. N. W., EDDY, E. M. & PHILLIPS, D. M. Observations on the Fine Structure and Relationships of the Chromatoid Body in Mammalian Spermatogenesis. *Biol. Reprod.* **2,** 129–153 (1970).

12. Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296,** 2176–2178 (2002).

13. Parvinen, M. The chromatoid body in spermatogenesis. *Int. J. Androl.* **28,** 189–201 (2005).

14. Kotaja, N. & Sassone-Corsi, P. The chromatoid body: a germ-cell-specific RNA-processing centre. *Nat Rev Mol Cell Biol* **8,** 85–90 (2007).

15. Zhang, P. *et al.* MIWI and piRNA-mediated cleavage of messenger RNAs in

mouse testes. *Cell Res* **25,** 193–207 (2015).

16. Lambard, S. Expression of aromatase in human ejaculated spermatozoa: a putative marker of motility. *Mol. Hum. Reprod.* **9,** 117–124 (2003).

17. PHYRE2 Protein Fold Recognition Server. Available at: http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index. (Accessed: 11th September 2017)

18. Vasileva, A., Tiedau, D., Firooznia, A. & Müller-reichert, T. Expression. **19,** 630–639 (2010).

19. Akpınar, M. *et al.* TDRD6 mediates early steps of spliceosome maturation in primary spermatocytes. *PLOS Genet.* **13,** e1006660 (2017).

20. Fanourgakis, G., Lesche, M., Akpinar, M., Dahl, A. & Jessberger, R. Chromatoid Body Protein TDRD6 Supports Long 3' UTR Triggered Nonsense Mediated mRNA Decay. *PLoS Genet.* **12,** 1–27 (2016).

21. Deng, W. & Lin, H. Miwi, a Murine Homolog of Piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. *Dev. Cell* **2,** 819–830 (2002).

22. Matzuk, M. & Lamb, D. The biology of infertility: research advances and clinical challenges. *Nat. Med.* **14,** 1197–1213 (2008).

23. Kistler, W. S. *et al.* RFX2 Is a Major Transcriptional Regulator of Spermiogenesis. *PLoS Genet.* **11,** e1005368 (2015).

24. Martianov, I. *et al.* Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol. Cell* **7,** 509–515 (2001).

25. Zhou, H. *et al.* Taf7l cooperates with Trf2 to regulate spermiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 16886–16891 (2013).

26. Nantel, F. *et al.* Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. *Nature* **380,** 159–162 (1996).

27. Shi, H.-J. *et al.* Cloning and characterization of rat spermatid protein SSP411: a thioredoxin-like protein. *J. Androl.* **25,** 479–93 (2004).

28. Collet, J.-F. & Messens, J. Structure, function, and mechanism of thioredoxin proteins. *Antioxid. Redox Signal.* **13,** 1205–1216 (2010).

29. Vinogradov, S. N. & Moens, L. Diversity of globin function: enzymatic, transport, storage, and sensing. *J. Biol. Chem.* **283,** 8773–7 (2008).

Supplementary table 1: Proteins detected by mass spectrometer analysis after Adgb pull-down co-immunoprecipitation in mice testis tissue.

UniProt ID	Hits
EIF3A_MOUSE	59
HNRPM_MOUSE	47
ADGB_MOUSE	44
GRP78_MOUSE	33
TDRD6_MOUSE	26
TERA_MOUSE	25
HSP72_MOUSE	23
ALDH2_MOUSE	22
LMNB1_MOUSE	21
GRP75_MOUSE	20
HXK1_MOUSE	16
CRBG3_MOUSE	16
VINC_MOUSE	16
AKA12_MOUSE	16
RBM14_MOUSE	16
RENT1_MOUSE	15
ALDOA_MOUSE	15
EF1D_MOUSE	15
PIWL1_MOUSE	15

UniProt ID	Hits
PSMD2_MOUSE	15
U520_MOUSE	14
KINH_MOUSE	14
DDX5_MOUSE	14
ODP2_MOUSE	14
ATPA_MOUSE	13
ACLY_MOUSE	13
DHX8_MOUSE	13
CAND1_MOUSE	13
QCR2_MOUSE	13
FXR1_MOUSE	13
TPP2_MOUSE	12
HYOU1_MOUSE	12
VP13A_MOUSE	12
WDR62_MOUSE	12
DHX30_MOUSE	12
ECHA_MOUSE	12
DHX9_MOUSE	11
SUCB1_MOUSE	11

UniProt ID	Hits
NASP_MOUSE	11
HMCS2_MOUSE	11
THIM_MOUSE	11
STK31_MOUSE	11
IF4G1_MOUSE	10
EF1G_MOUSE	10
PGK2_MOUSE	10
ENOA_MOUSE	10
CCAR2_MOUSE	10
HSP7C_MOUSE	10
TIF1B_MOUSE	10
SPT20_MOUSE	10
ACON_MOUSE	10
PYC_MOUSE	9
ECHB_MOUSE	9
SYVC_MOUSE	9
TT21B_MOUSE	9
F10A1_MOUSE	9
MYO7A_MOUSE	9

UniProt ID	Hits
DDX3L_MOUSE	9
HS90A_MOUSE	9
EF1A1_MOUSE	8
DJB11_MOUSE	8
IPO5_MOUSE	8
PERQ2_MOUSE	8
DDX42_MOUSE	8
PRS6A_MOUSE	8
UD16_MOUSE	8
PRS6B_MOUSE	8
DNPEP_MOUSE	8
DYHC2_MOUSE	8
MATR3_MOUSE	8
PRS7_MOUSE	8
UBP7_MOUSE	8
DHAK_MOUSE	8
RPN1_MOUSE	8
ACTB_MOUSE	8
HNRPK_MOUSE	8

UniProt ID	Hits
ACBG1_MOUSE	8
H12_MOUSE	8
HNRPU_MOUSE	7
ADRO_MOUSE	7
WDR60_MOUSE	7
BHMT1_MOUSE	7
SCRIB_MOUSE	7
U5S1_MOUSE	7
IPO4_MOUSE	7
SYEP_MOUSE	7
DCTN1_MOUSE	7
SRRT_MOUSE	7
XPO2_MOUSE	7
IMB1_MOUSE	7
RL4_MOUSE	6
CCAR1_MOUSE	6
IF4A2_MOUSE	6
IF4G3_MOUSE	6
GDIB_MOUSE	6
HNRPF_MOUSE	6
KIF15_MOUSE	6

UniProt ID	Hits
QCR1_MOUSE	6
FLNB_MOUSE	6
ASSY_MOUSE	6
CPSF1_MOUSE	6
PDIA6_MOUSE	6
RAI14_MOUSE	6
CP11A_MOUSE	6
RUVB2_MOUSE	6
SEPT2_MOUSE	6
TDRD7_MOUSE	6
NDUS1_MOUSE	6
AP2B1_MOUSE	6
SYFB_MOUSE	6
CNOT1_MOUSE	5
PRC2A_MOUSE	5
ODPAT_MOUSE	5
HYEP_MOUSE	5
IF122_MOUSE	5
COPA_MOUSE	5
MAP7_MOUSE	5
PGK1_MOUSE	5

UniProt ID	Hits
RSSA_MOUSE	5
TADBP_MOUSE	5
CISY_MOUSE	5
IQGA1_MOUSE	5
LPPRC_MOUSE	5
THIL_MOUSE	5
YTDC2_MOUSE	5
ACO10_MOUSE	5
CF2I1_MOUSE	5
CYFP1_MOUSE	5
GAPD1_MOUSE	5
GOGA2_MOUSE	5
HNRH1_MOUSE	5
PUR8_MOUSE	5
SC31A_MOUSE	5
UBA6_MOUSE	5
SHP1L_MOUSE	5
FUBP2_MOUSE	5
HNRPL_MOUSE	5
KHDR1_MOUSE	5
MYEF2_MOUSE	5

UniProt ID	Hits
DDX17_MOUSE	5
COPB_MOUSE	5
DYN2_MOUSE	5
GIT1_MOUSE	5
TRAD1_MOUSE	5
CCD39_MOUSE	4
EWS_MOUSE	4
DDX4_MOUSE	4
GANAB_MOUSE	4
SDHA_MOUSE	4
TBB4A_MOUSE	4
GUAA_MOUSE	4
HSP74_MOUSE	4
TRXR3_MOUSE	4
VTNC_MOUSE	4
DPEP3_MOUSE	4
ASPC1_MOUSE	4
CABS1_MOUSE	4
CSDE1_MOUSE	4
FIP1_MOUSE	4
ILF3_MOUSE	4

UniProt ID	Hits
KIFA3_MOUSE	4
STIP1_MOUSE	4
SYAC_MOUSE	4
SYK_MOUSE	4
TCPQ_MOUSE	4
TKTL2_MOUSE	4
TRAP1_MOUSE	4
TSKS_MOUSE	4
VATA_MOUSE	4
VINEX_MOUSE	4
AL1L1_MOUSE	3
CPSF6_MOUSE	3
AKAP4_MOUSE	3
CLPX_MOUSE	3
DPYL3_MOUSE	3
EIF3C_MOUSE	3
HS71L_MOUSE	3
K6PP_MOUSE	3
NUP93_MOUSE	3
PSPC1_MOUSE	3
PYGB_MOUSE	3

UniProt ID	Hits
SEPT9_MOUSE	3
PIWL2_MOUSE	3
ABCF2_MOUSE	3
AKAP3_MOUSE	3
AT1A1_MOUSE	3
CEP72_MOUSE	3
FUBP1_MOUSE	3
KIF3A_MOUSE	3
MUM1_MOUSE	3
TGM2_MOUSE	3
VPS35_MOUSE	3
CUL5_MOUSE	3
TCPZ_MOUSE	3
KLH10_MOUSE	3
DDX1_MOUSE	2
PCCA_MOUSE	2
ERF3A_MOUSE	2
ABCE1_MOUSE	2
ACE_MOUSE	2
ADAM2_MOUSE	2
CLGN_MOUSE	2

UniProt ID	Hits
ACBG2_MOUSE	2
ACOX3_MOUSE	2
AIFM1_MOUSE	2
ANM5_MOUSE	2
ATAD3_MOUSE	2
CDC5L_MOUSE	2
CSTF3_MOUSE	2
DNLI3_MOUSE	2
DPYL5_MOUSE	2
GPDM_MOUSE	2
KLC2_MOUSE	2
LRC47_MOUSE	2
MCM3_MOUSE	2
NCOA5_MOUSE	2
NXF1_MOUSE	2
OXLA_MOUSE	2
PDIA4_MOUSE	2
PDXD1_MOUSE	2
PLPL9_MOUSE	2
RGPS2_MOUSE	2
SC23A_MOUSE	2

UniProt ID	Hits
SNW1_MOUSE	2
SNX27_MOUSE	2
SRC8_MOUSE	2
STAT4_MOUSE	2
SYRC_MOUSE	2
TDRD3_MOUSE	2
TRFE_MOUSE	2
XPO1_MOUSE	2
DPP3_MOUSE	2
AKAP8_MOUSE	2
PRP6_MOUSE	2
SAE2_MOUSE	2
K6PL_MOUSE	2
CPSF2_MOUSE	2

VI. Expression analysis of androglobin in human spermatozoa and testis tissue

An Bracke[1], Geert Van Raemdonck[2], Kris Peeters[3], Usha Punjabi[3], Sylvia Dewilde[1]

[1]Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, Antwerp 2610, Belgium
[2]Center for Proteomics, University of Antwperp, Groenenborgerlaan 171, Antwerp 2610, Belgium
[3]Center for Reproductive Medicine, University Hospital Antwerp, Wilrijkstraat 10, Edegem 2650, Belgium

1 Introduction

Androglobin (Adgb) is a recently discovered globin type that is predominantly expressed in testis tissue of vertebrates. Adgb is characterized by its typical modular domain structure. The chimeric proteins characteristically comprise four domains, an N-terminal ~350 residues calpain-like cysteine protease domain, a region of ~300 residues without known motifs/domains, followed by an ~150 residue circularly permutated globin domain, and a second, uncharacterized ~750 residues C-terminal domain.[1] RT-qPCR expression analysis of Adgb in mice displayed a predominant expression in testis tissue. A more detailed analysis in different stages of mouse testis development showed an increase of Adgb expression from postnatal day 25, when post-meiotic spermatids are abundant, until adulthood. Very low or almost no expression was detected in Leydig cells, Sertoli cells, spermatogonia and spermatocytes (Fig1B), indicating again that Adgb plays an important role in the late phases of spermatogenesis. This was confirmed by mRNA *in situ* hybridization of an Adgb antisense RNA probe to mouse testis cryosections showing pronounced signals toward the lumen of the seminiferous tubes.[1] In humans, an *in silico* expression analysis using the NCBI UniGene[2], ONCOMINE[3], and R2 database[4] also strongly supported the findings of a testis-specific expression for Adgb. (Fig2) Furthermore, R2 database analysis suggested 4-fold higher expression levels in fertile vs. infertile males, indicating that ADGB plays an important role in spermatogenesis (Fig1A).[4–6]

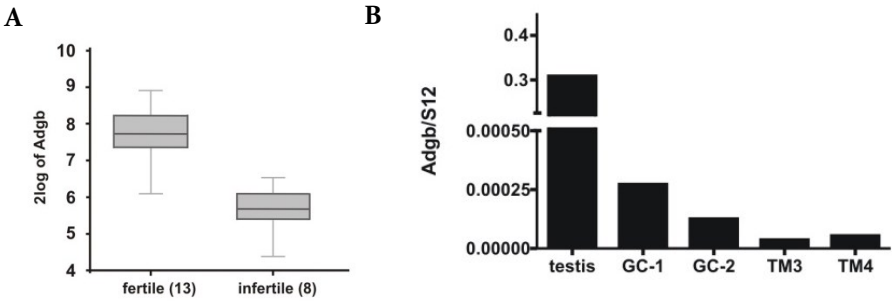


Figure 1: A) *In silico ADGB* expression analysis in spermatozoa from fertile vs. infertile human individuals[6] B) RT-qPCR expression analysis of *Adgb* mRNA in mouse cell lines derived from spermatogonia (GC-1), spermatocytes (GC-2), Leydig cells (TM3) and Sertoli cells (TM4), relative to mouse testis. Figure was taken from[1]

The last decades, infertility is an increasing problem and assisted reproductive techniques (ART), such as *in vitro* fertilization (IVF) with intracytoplasmic sperm injection (ICSI) are often used to obtain pregnancy. When these techniques are used, an adequate diagnosis is of major importance and good biomarkers are required. In the far future, ADGB might perhaps serve as a predictive biomarker for male infertility. In order to confirm the *in silico* expression analyses and to investigate the putative role as biomarker, we performed an expression analysis for ADGB in human sperm samples and human testis tissue. We analyzed the expression of ADGB in testis biopsies with different pathologies, e.g. maturation arrest in early meiotic stages, or late meiotic stages, etc. In this way, more insight will be gained about the stages of human spermatogenesis ADGB is involved in. In addition, we investigated the potential of ADGB to function as a seminal biomarker by performing a differential expression analysis for ADGB in semen samples, which are much easier and non-invasively obtained compared to testis biopsies.

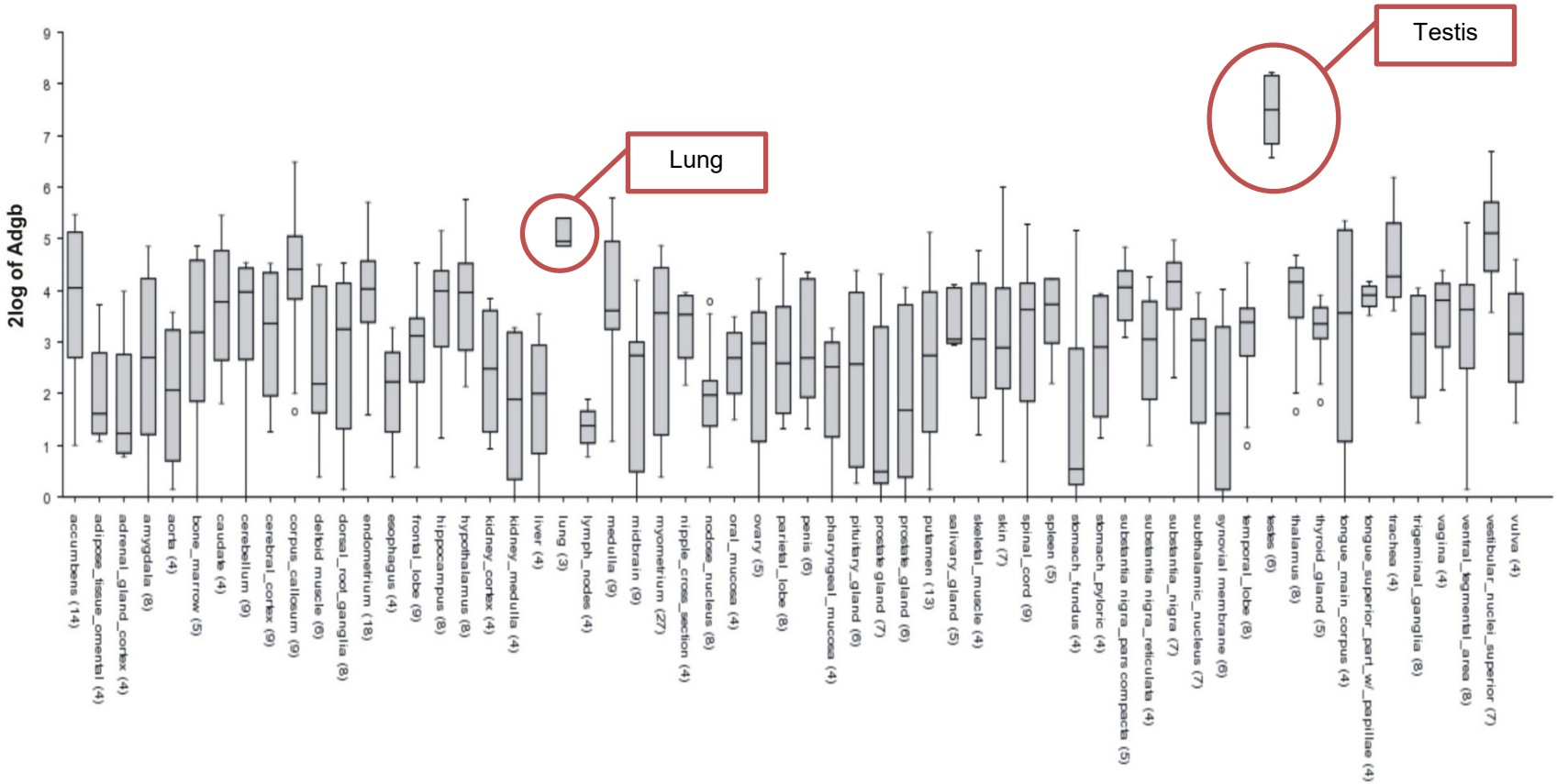


Figure 2: *In silico ADGB* expression analysis in normal human tissue. The 25th–75th percentiles of expression levels are represented by a closed box, with the median indicated by a line; different degrees of outliers are indicated by the whiskers and the points, as defined for standard 'box and whisker' plots. Only tissues represented by more than 3 measurements are displayed. As can be seen, ADGB expression is the highest in lung and testis tissue (encircled in red). Figure was adapted from[1]

2 Material and Methods

2.1 Sample and tissue collection

Sperm samples and testis biopsies were obtained from the centrum of reproductive medicines of the university hospital of Antwerp (UZA). Men participating in this study gave informed consent to use their semen or testis biopsy for research. Standard semen parameters (cell count, motility and morphology) were evaluated according to published recommendations (WHO 2010)[7]. Testis biopsies were screened for abnormalities in the spermatogenic process. These analyses were performed by the center of reproductive medicines of UZA.

2.2 RNA isolation

Sperm samples: To rule out the possibility of any somatic cell contamination, a 50% PureSperm100 (Nidacon) density gradient centrifugation was performed (centrum of reproductive medicines, UZA), followed by two phosphate-buffered saline (PBS) washes. After this initial purification, the samples were treated with somatic cell lysis buffer (0.1% SDS, 0.5% Triton X-100), as described by Goodrich et al.[8] Sperm cells were subsequently pelleted by centrifugation (220 g, 10min, 4°C) and resuspended in lysis buffer of the Qiagen RNeasy™ Mini kit. RNA isolation was performed as was described in manufacturer's recommendations. The purity and concentration of the RNA was checked spectrophotometrically at 260 and 280nm using a NanoDrop™ spectrophotometer.

Testis biopsies: Testis tissues were kept in RNA later buffer (Qiagen) at -80°C until they were processed for RNA extraction. Testis tissue was homogenized using the TissueRuptor (Qiagen). RNA extraction was performed using the RNA PureLink™ mini kit (ThermoScientificTM). The purity and concentration of the RNA was checked spectrophotometrically at 260 and 280nm using a NanoDrop™ spectrophotometer.

2.3 RT-qPCR

Using 500ng (sperm samples) or 1000 ng (testis tissue) RNA, cDNA was prepared using SuperScript™ II Reverse Transcriptase (ThermoScientificTM). Real-time RT-PCR was

performed using Power SYBR Green Master mix (ThermoScientificTM) in a 20µL total reaction volume, containing 10ng cDNA and 150nM of each sense and antisense primer (Table1). RT-qPCR was carried out in a StepOnePlus™ Real-Time PCR instrument (ThermoScientificTM). PCR was performed by initial denaturation at 95°C for 10min, followed by 40 cycles of 30 s at 95°C, 60 s at 60°C, and 60s at 72°C. Amplicon specificity was checked by melting curve analysis. The primers were in-house designed using NCBI/Primer-Blast, primers were chosen spanning on exon-exon junctions. The primer sequences and the expected lengths of the resulting PCR products are summarized in Table1. The GeNorm algorithm was used to analyze the stability of the selected reference genes, using qBasePlus (BioGazelle) software[9].

Table 1: List of genes that were testes in RT-qPCR analysis and their in-house designed primers with their corresponding PCR product length (base pairs).

Gene	Primer Forward	Primer Reverse	Expected length PCR product (bp)
ADGB	GCATTACCTTAGCGGGTTCA	ATTGCCCTTCTTCCACACC	200
CNOT8	GGTCCCGATCAGACCAAACA	CTGGCCCACACTTCACAGAT	132
DNAJB12	GAAGCAGCAGAAGGAAGGCT	CCCTGGCTCATACTTGGACC	176
TMEM92	TGTGGTCTCATCCTTGCCTG	GCCACAGATGCAAAAGACGG	141
NOTCH2	TACAGATGCGAGTGTGTCCCA	AATGCCCTGGATGGAAAATGG	184
CASP1	GCCTGTTCCTGTGATGTGGA	TTCACTTCCTGCCCACAGAC	175
PIGK	GGCCGCTAGTCATATCGAGG	GTACACACCAGAACAGCCCA	81
ERCC4	GGAGTTGAACACCTCCCTCG	ACACCAAGATGCCAGTAATTAAATC	163
STAT3	ACGACCTGCAGCAATACCAT	GGTGAGGGACTCAAACTGCC	123
DDX6	TTGGGAGCTTCGAGTCAACA	ACTCACGTCAATCCCAAGCC	103
PCDH7	CCTGACCTGGCAAGGCATTA	ACGAAATGGCTGTTTGCTGT	216
CSF3R	AATCATGGAGGAGGATGCCTT	TGCCCATAAAGGACCTGATCG	234
RNF7	GTCCAGGTGATGGTGGTCTG	GCCTTTGTAGGGCACTGGAT	206
ARID5B	ACTCGGTGCTAGTCACTGCT	AGATTTGGCGCCATTCCAGTC	227
ACTB	GTTGTCGACGACGAGCG	GCACAGAGCCTCGCCTT	93
GAPDH	AAGACCTTGGGCTGGGACTG	TGGCTCGGCTGGCGAC	168

2.4 Protein extraction

Sperm cells were first purified from somatic cells by a 50% PureSperm (Nidacon) gradient density centrifugation, followed by two PBS washes. Testis tissue was homogenized using the TissueRuptor (Qiagen). Proteins were extracted in 50mM ammonium bicarbonate

buffer containing 0.5% sodium deoxycholate, 50mM dithiothreitol (DTT), and protease inhibitor cocktail. Samples were sonicated (40Hz, 1min, puls1) then centrifuged at 10 000g for 15 min. The supernatant was precipitated overnight with 5 volumes of ice cold acetone followed by overnight incubation at -20°C. Samples were resuspended in 500µL PBS for determination of protein concentration using the Pierce™ BCA Protein Assay Kit. Subsequently, 100µg of protein was precipitated with 5 volumes of ice-cold acetone and stored at -20°C until further processing. In all cases, lo-bind Eppendorf tubes were used to ensure high recovery rates of protein and peptides.

2.5 Tryptic digestion

Protein samples were resuspended in 50mM Tris-HCl/6M urea/5mM DTT/10% beta-mercaptoethanol (25µL/100µg protein) at pH8.7. For the denaturation and reduction process all samples were incubated at 65°C during 1h. Subsequently, proteins were diluted in 50mM Tris-HCl/1mM $CaCl_2$ (75µL/100µg protein) and alkylated by adding 200mM iodoacetamide (10µL/100µg protein) during 1h at room temperature. Finally, proteomics-grade modified trypsin (Promega) was added at a 30:1 protein-to-enzyme ratio. After incubation at 37°C for 18h the digestion was stopped by freezing the fractions at -80°C. Protein digests were desalted by Solid Phase Extraction (SPE) GracePure SPE C18-Max (50mg) (W.R Grace&Co) RP cartridges and a vacuum manifold. SPE cartridges were conditioned with 100 % methanol and equilibrated with 100 % LC/MS grade H_2O and 0.1 % formic acid (FA). After loading the complete acidified (0.1 % FA) tryptic digest, peptides were washed with 10 % methanol and eluted with 40 % methanol/40 % acetonitrile (ACN) and 0.1 % FA. Eluted peptides were lyophilized and frozen at -20 °C until further analysis. Immediately before analysis, lyophilized digests were resuspended in 5 % ACN/0.1 % FA.

2.6 Selection of the proteotypic peptides

Proteotypic peptides (PTPs) were selected for ADGB, ACTB and TDRD6. First an *in silico* tryptic digest was performed using the Expasy 'peptide mass' tool. Second, these peptides were screened for usability as PTPs using ESP predictor software[10]. Finally, the selected peptides were screened for their uniqueness with NCBI protein blast (blastp). Synthetic

peptides were ordered at ThermoScientificTM Peptide Synthesis Services and reconstituted in 0.1% FA.

Table2. List of proteotypic peptides of ADGB, ACTB and TDRD6 and their molecular weight.

Proteotypic peptide	Sequence	Molecular weight (g/mol)
ADGB P1	ATSQGNTASQVILGK	1474.65
ADGB P2	RPQDILFSQTPVVVK	1727.05
ADGB P3	ETVITDEAQELIVK	1587.8
ACTB P1	EITALAPSTMK	1161.39
ACTB P2	VAPEEHPVLLTEAPLNPK	1954.27
TDRD6 P1	SALPYENIDSEIK	1478.62
TDRD6 P2	TNLVTQYQDSVGNK	1566.69

2.7 Multiple Reaction Monitoring (MRM) assay optimization and mass spectrometry analysis

Optimization of each PTP was performed on a triple quadruple mass spectrometer (Waters Xevo TQ) in order to obtain the most intense transitions. The capillary voltage was tuned to approximately 2kV with a source temperature of 150°C. Desolvation temperature was set at 400°C with a nitrogen gas flow of 800 L/h. Cone voltage, collision energy and dwell times were optimized for each of the PTPs. All PTPs were dissolved in mobile phase A (MP-A), containing 5% ACN (LC/MS grade) and 0.1% FA. For each of the peptides individually, the limit of detection was determined by performing a dilution series in MP-A. Based on these concentrations, a mixture of all PTPs was made. Chromatographic separation of the protein extracts from sperm samples or testis biopsy was performed on a RP-C18 UPLC column (Waters, CSH 150 x 2.1mm, 1.7µM at 35°C) connected to an Acquity UPLC system (Waters Corporation). In order to separate all peptides as best as possible, an optimized linear gradient of Mobile Phase B (MP-B) (0.1% FA in 100% ACN) was applied: 3% MP-B during 1 min and from 3 to 45% MP-B in 5min, followed by a steep increase to 100% MP-B in 2min, all at a flow rate of 300µL/min. At least three transitions (ion pairs) were selected for each peptide of interest. For each scheduled MRM analysis, 50µg of peptides (injection loop of 5µL) per sperm/testis sample were loaded onto the analytical column. Data acquisition was controlled by MassLynx version 4.1 (Waters Corporation). MRM analysis was performed in

collaboration with Dr. Geert Van Raemdonck at the Center for Proteomics of the university of Antwerp.

3 Results and discussion

3.1 RNA expression analysis of androglobin in human spermatozoa and testis tissue

3.1.1 RT-qPCR in spermatozoa

In late phases of spermatogenesis, transcription is silenced by the compaction of the genome through histone-to-protamine transition. Thus, only a small fraction of RNAs remains in the mature spermatozoa[11]. These RNAs have presumably an important function in spermiogenesis or in fertilization. There is a growing interest for these RNAs as they could be valuable diagnostic indicators of sperm survival, fertilization and early embryogenesis, and could serve as a predictor of the *in vitro* fertilization prognosis[12,13]. However, the isolation of the RNA content of human spermatozoa entails a unique set of challenges. First, the intrinsic heterogeneous population of cells present in the ejaculate necessitates the introduction of a purifying step in which only the spermatozoa are isolated. A second challenge is the small quantity of RNA present in spermatozoa (50 fg of RNA per cell)[14].

Here, we have investigated ADGB mRNA expression in human spermatozoa. We obtained 19 sperm samples from the center of reproductive medicines of the university hospital of Antwerp. Standard semen analysis was performed and only three samples showed abnormal parameters (marked in red in Table3), one sample displayed too low cell count (<15 million/ml = oligozoospermia) and impaired motility (<20% progressive motility = asthenozoospermia) and two samples displayed impaired motility (<20% progressive motility asthenozoospermia). All the sperm samples fulfilled WHO criteria regarding morphology (>96% of sperm cells showing abnormal morphology = teratozoospermia).

Before RNA isolation, spermatozoa were purified from somatic cells by a 50% density gradient centrifugation. Density gradient centrifugation separates sperm cells based on their density. A 50% density gradient will separate spermatozoa from leukocytes, epithelial cells, cell debris, bacteria and seminal fluid. To be sure that all somatic cells are removed from the sperm sample, the sample was afterwards treated with somatic cell lysis buffer as has been described by Goodrich et al.[8] RNA was extracted and the presence of ADGB mRNA in total sperm RNA could be confirmed by RT-PCR (Fig3A). We noticed that the yield of extracted RNA per sperm sample was variable and often very low (Table3) and only seven samples yielded enough RNA for further RT-qPCR analysis.

In order to perform a statisticallly relevant RT-qPCR analysis, reference genes have to be selected which are stably expressed in the different samples (fertile and infertile). For this, we selected a set of genes using the raw data of three independent differential micro-array transcriptome experiments of sperm from fertile versus infertile men. One study compared the transcriptome of sperm with normal morphology versus abnormal morphology[6], the two other studies compared the transcriptome of sperm with normal motility versus impaired motility[15,16]. We selected the 400 most stablly expressed transcripts form all three studies and made an overlap of these transcripts. This resulted in a set of 13 transcripts that displayed stable expression in sperm from both fertile and infertile men (taking into account both impaired sperm motility as morphology). The Gene Ontology terms were investigated for these 13 genes, demonstrating that they belonged to different functional classes, ruling out potential co-regulation of these genes. In addition to these 13 reference genes, selected from experimental micro-array data, we also included two standard reference genes: ACTB and GAPDH.

RT-qPCR was performed in 7 sperm samples (which yielded enough RNA), showing normal semen parameters, for ADGB and the 15 selected reference genes (Table1). The stability of the reference genes was analyzed using the GeNorm algorithm[9]. Despite all the efforts in selecting good candidate reference genes, none of these genes were stably expressed in our RT-qPCR analyses. Therefore, we have to conclude that it was not possible to perform a statistically relevant differential RT-qPCR expression analysis of ADGB in individual sperm samples.

Table 3: RNA isolation of 19 sperm samples. The purity and concentration of the isolated RNA was checked spectrophotometrically at 260 and 280nm using a NanoDrop instrument. Samples that yielded enough RNA for RT-qPCR analysis are marked in green. Standard semen parameters (cell count, motility grade and morphology) were evaluated according to published recommendations (WHO 2010)[7]. Motility was graded from a to d, according to the WHO manual criteria (a = fast progressive, b = slow progressive, c = non-progressive, d = immotile). Abnormal parameters are marked in red.

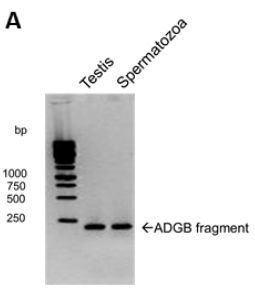
Sample	RNA Concentration (ng/µL)	260/280 ratio	Sperm parameters (native sample)		
			Cell count (Million/ml)	Motility grade (a/b/c/d)	% Abnormal morphology
1	4.7	1.76	38	54/10/12/24	6
2	37.0	1.85	246	50/10/10/30	14
3	1.5	1.97	40.2	33/10/12/44	9
4	14.6	1.98	57.4	44/14/11/32	4
5	24.9	1.89	36.6	42/11/146	?
6	26.4	1.82	53.3	56/6/4/34	14
7	3.8	/	107	48/6/10/36	10
8	4.1	1.99	124.5	15/7/4/72	4
9	9.7	/	120	48/10/9/33	?
10	6.6	/	88.5	54/8/4/33	3
11	59.2	1.88	105	57/3/7/33	11
12	100.2	1.89	188	51/0/0/49	9
13	12.53	1.48	42.6	58/2/0/40	1
14	9.0	1.51	7.4	16/25/15/44	5
15	4.7	1.31	98.7	41/19/11/28	11
16	181.9	2.08	346	52/8/7/34	11
17	95.4	2.06	197	48/4/3/44	6
18	94.4	2.00	213	60/7/7/26	11
19	44.6	1.49	21.6	41/8/10/41	10

3.1.2 RT-qPCR analysis in testis tissue

The problems we encountered in sperm cells, were not encountered in testis biopsies; RNA extraction out of testis tissue is more straightforward and resulted in higher yields of pure RNA. ADGB mRNA expression in testis tissue was confirmed using RT-PCR (Fig3A). Standard reference genes ACTB and GAPDH were stably expressed and could be used as endogenous control. Two preliminary RT-qPCR experiments were performed on respectively four and five testis biopsies (Fig 3C and 3D). Pathological information about these nine testis biopsies is described in Fig3B. These two preliminary experiments show,

independently, similar results: ADGB expression levels are high in testis tissue displaying normal spermatogenesis versus testis biopsies with an abnormal spermatogenesis and ADGB is completely absent in the biopsies with absence of spermatozoa. Test biopsy 5 is an exception, this biopsy has a normal spermatogenesis, but ADGB expression is lower compared to the other biopsies with normal spermatogenesis. This may be explained by inter-individual variation in ADGB expression. The RT-qPCR data suggest that ADGB expression is correlated with the amount of spermatozoa present in the testis biopsy. For further conclusions about the exact stages of spermatogenesis in which ADGB is functioning, more testis biopsies (with different types of pathologies) are needed to be included in the analysis. Due to limited availability of testis biopsies, this was not possible in the scope of this thesis.

A



B

Testis biopsy	Spermatogenesis	Spermatozoa in seminiferous tubules
1	Normal	Yes
2	Rare fully developed spermatozoa	Few
3	Lack of spermatogenesis	None
4	Maturation arrest in spermatogenesis	Few/None
5	Normal	Yes
6	Normal	Yes
7	Normal	Yes
8	Tumor – disturbed spermatogenesis	None
9	Tumor – disturbed spermatogenesis	Few

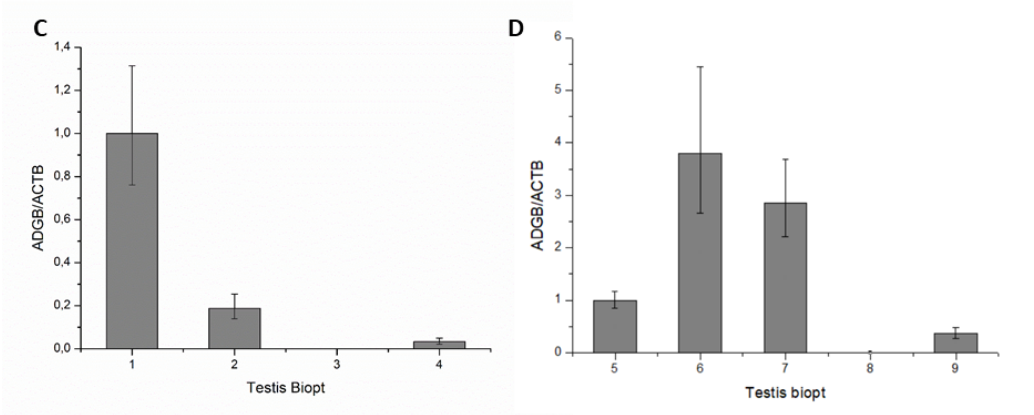


Figure 3: A) RT-PCR for ADGB in testis tissue with normal spermatogenesis and spermatozoa with normal motility and morphology parameters. ADGB PCR fragment is 200bp. B) Table with pathological information about 9 testis biopsies which were used in two independent RT-qPCR experiments (C) and (D). C) and D) show relative expression of ADGB in nine testis biopsies. ACTB was used as endogenous control and expression was normalized on testis biopsy 1 (C) and testis biopsy 5 (D).

3.2 Protein expression analysis of androglobin in human spermatozoa and testis tissue

As the preliminary RT-qPCR analyses in testis biopsies suggested a link between ADGB expression and male infertility, we want to confirm this hypothesis on protein level. However, Western blot analysis using the commercially available ADGB Prestige antibody from Sigma (epitope see Fig5), could not provide evidence for the presence of ADGB in spermatozoa, nor in testis tissue. To overcome the shortcomings of antibody-based techniques, an alternative validation method based on mass spectrometry was used, called multiple reaction monitoring (MRM) using a triple quadrupole mass spectrometer (Fig4)[17]. This technique is more sensitive compared to Western blot.

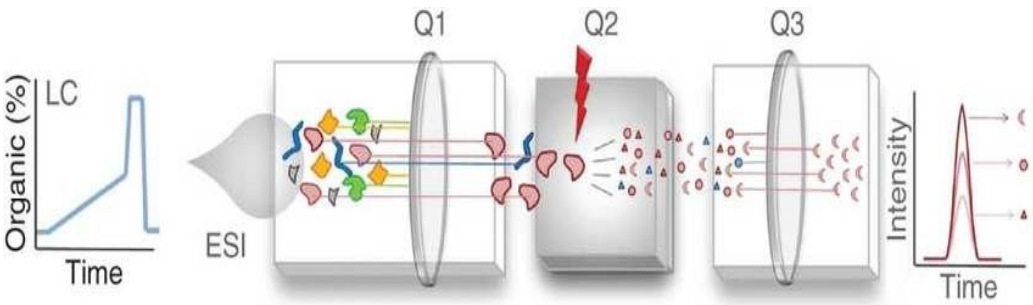


Figure 4: Schematic working principle of multiple reaction monitoring. Peptides, separated by liquid chromatography (LC), are ionized by electrospray ionization (ESI) and enter the triple quadrupole mass spectrometer. Several peptides (each represented in a specific color) will be filtered by the first quadrupole (Q1). Only specific peptides (red) are passed and will be fragmented I the collision cell (Q2). Mass filtering at a second level (Q3) will only pass specific fragments. Figure from[18]

MRM allows mass filtering at two levels, allowing the researchers to fine-tune an instrument to specifically look for peptides, or protein fragments, of interest. So, unlike traditional mass spectrometry, which attempts to detect all proteins in a biological sample in an unfocused fashion, MRM is highly selective. In a first step, a specific precursor will be selected at the first quadrupole (Q1). After fragmentation in a collision cell (Q2), only specific fragment ions will pass the third quadrupole (Q3) and reach the detector. Combinations of precursor and fragment ions are called transitions[18]. Before proteins can be analyzed using MRM, so called proteotypic peptides (PTPs) need to be

defined[19]. These are peptides that fulfil some essential criteria like being unique for a given protein and being able to ionize[10]. Analyzing peptides in MRM-mode has several advantages: 1) the fast and continuously monitoring of specific ions reduces the detection limit of peptides leading to a high sensitivity (up to low ng/ml, i.e attomolar ranges in plasma) 2) mass filtering at two levels results in an elevated specificity because background ions are filtered out and 3) the possibility to monitor several transitions during the same run[18]. The (relative) quantification of a target protein can be performed with MRM by integrating one or more of the peaks of the peptide fragment ions of the target protein[20]. In this way the differential expression of ADGB can be studied in sperm samples and testis tissue using MRM analysis and thus provides a valuable and even better alternative for RT-qPCR, as it measures the ADGB expression on protein level.

An MRM analysis was performed for ADGB and two control proteins ACTB and TDRD6. The highly abundant ACTB was included as positive control. As second control we included TDRD6, an interaction partner of ADGB. Three proteotypic peptides (PTPs) were selected for ADGB and two for ACTB and TDRD6 (Table 2 and fig 5). MRM transitions were optimized for each PTP (Supplementary data1). Two sperm samples displaying normal cell count, motility and morphology were lysed and proteins were extracted and digested with trypsin. MRM analysis of the proteome of these two sperm samples showed that ACTB was highly present, which ensures us of an efficient protein extraction and MRM analysis. For ADGB, only one of the three proteotypic peptides was detected in the sperm proteomes, though with low quantity. (MRM chromatograms are shown in section 6: Supplementary data) Here, the MRM analysis of only two sperm samples is described. However, more MRM analyses were performed of which the results are not shown. In most cases, none of the three PTPs of ADGB could be detected.

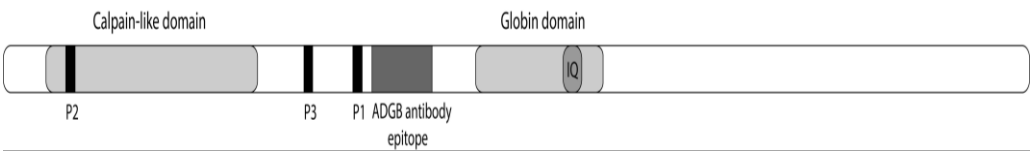


Figure 5: Schematic representation of the chimeric androglobin. Black rectangles represent the proteotypic peptides of ADGB that were used for MRM analysis. Dark grey rectangle represent the epitope of the ADGB Prestige antibody from Sigma.

TDRD6 was in a previous study localized at the chromatoid body of spermatids with immunohistochemistry (IHC)[21]. Interestingly, with MRM analysis TDRD6 was undetectable in the proteome of spermatozoa (similar to ADGB). A possible explanation can be found in the fact that the chromatoid body is an irregularly-shaped cytoplasmic structure in the male germ cell, which starts to appear in late pachytene spermatocytes and disappears during spermiogenesis[22]. As ADGB is an interaction partner of TDRD6, it is thus possible that both ADGB and TDRD6 are absent in mature spermatozoa and hence not or only partially detectable in MRM analysis. In mature spermatozoa transcription is silenced, but the mRNA of ADGB may perhaps still be present from expression during previous stages of spermatogenesis. This can explain the fact that ADGB expression was confirmed in mature spermatozoa with RT-PCR (section 3.1.1).

We also performed an MRM analysis on the proteome of a testis biopsy displaying normal spermatogenesis. The same result was obtained as in the sperm samples, ADGB and TDRD6 were not/partially detectable, ACTB on the other hand was clearly present in the testis proteome. This suggests that ADGB most likely plays a transient role in spermatogenesis and that the physiological concentrations of ADGB are very low. Another explanation could be that ADGB and TDRD6 are post-translationally modified, which causes a shift in molecular mass and are thus not detectable in MRM analysis.

It has to be mentioned that it cannot be excluded that ADGB is (auto)-cleaved during spermatogenesis and that the C-terminal part (including the globin domain) is still present in mature spermatozoa. Because the PTPs and the epitope of the ADGB antibody are all located in the N-terminal part of ADGB (Fig5), it can be that both Western blot and MRM could not detect the part of ADGB where the globin domain is located. This has to be examined further in future experiments.

4 Conclusion and future perspectives

We confirmed the presence of ADGB mRNA in human testis tissue and spermatozoa. Due to low presence of RNA in sperm samples and the lack of stable reference genes, it was not possible to perform a differential RT-qPCR analysis in sperm samples from fertile versus infertile men. Two preliminary RT-qPCR experiments were performed in respectively four and five testis biopsies with different infertility pathologies and we

concluded that ADGB expression is correlated with the presence of spermatozoa in the testis. An extended expression analysis in more testis biopsies, displaying different maturation arrests of spermatogenesis, is needed to gain more insight in which phases of spermatogenesis ADGB is functional. Furthermore, we performed an MRM analysis for ADGB in two sperm samples having normal parameters (cell count, motility and morphology) and in one testis biopsy displaying normal spermatogenesis. We could not detect ADGB with MRM analysis in spermatozoa nor in testis tissue, comparable to its interaction partner TDRD6, which was also not detectable using MRM. TDRD6 is a component of the chromatoid body, which is removed in mature spermatozoa during spermiogenesis. Our findings suggest thus a transient role in spermatogenesis and/or low abundant cellular concentrations for ADGB in sperm cells and testis tissue. From these results, it is clear that ADGB would not form a good biomarker for fertility.

5 References

1. Hoogewijs, D. *et al.* Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes. *Mol. Biol. Evol.* **29,** 1105–14 (2012).
2. UniGene. Available at: https://www.ncbi.nlm.nih.gov/unigene. (Accessed: 29th October 2017)
3. Rhodes, D. R. *et al.* ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia* **6,** 1–6 (2004).
4. R2. Available at: https://hgserver1.amc.nl/cgi-bin/r2/main.cgi. (Accessed: 29th October 2017)
5. Abu-Halima, M. *et al.* Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. *Fertil. Steril.* **99,** (2013).
6. Platts, A. E. *et al.* Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Hum. Mol. Genet.* **16,** 763–73 (2007).
7. WHO | WHO Manual for the Standardized Investigation, Diagnosis and Management of the Infertile Male.
8. Goodrich, R. J., Anton, E. & Krawetz, S. A. in *Spermatogenesis Methods and Protocols* 385–397 (2013).
9. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3,** RESEARCH0034 (2002).
10. Fusaro, V. A., Mani, D. R., Mesirov, J. P. & Carr, S. A. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat. Biotechnol.* **27,** 190–198 (2009).
11. Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296,** 2176–2178 (2002).

12. Sendler, E. *et al.* Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res.* **41,** 4104–4117 (2013).

13. Georgiadis, A. P. *et al.* Potential Application in Clinical Testing. **193,** 352–359 (2015).

14. Krawetz, S. A. Paternal contribution: new insights and future challenges. *Nat. Rev. Genet.* **6,** 633–642 (2005).

15. Jodar, M., Kalko, S., Castillo, J., Ballescà, J. L. & Oliva, R. Differential RNAs in the sperm cells of asthenozoospermic patients. *Hum. Reprod.* **27,** 1431–8 (2012).

16. Pacheco, S. E. *et al.* Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One* **6,** e20280 (2011).

17. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4,** 222 (2008).

18. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9,** 555–566 (2012).

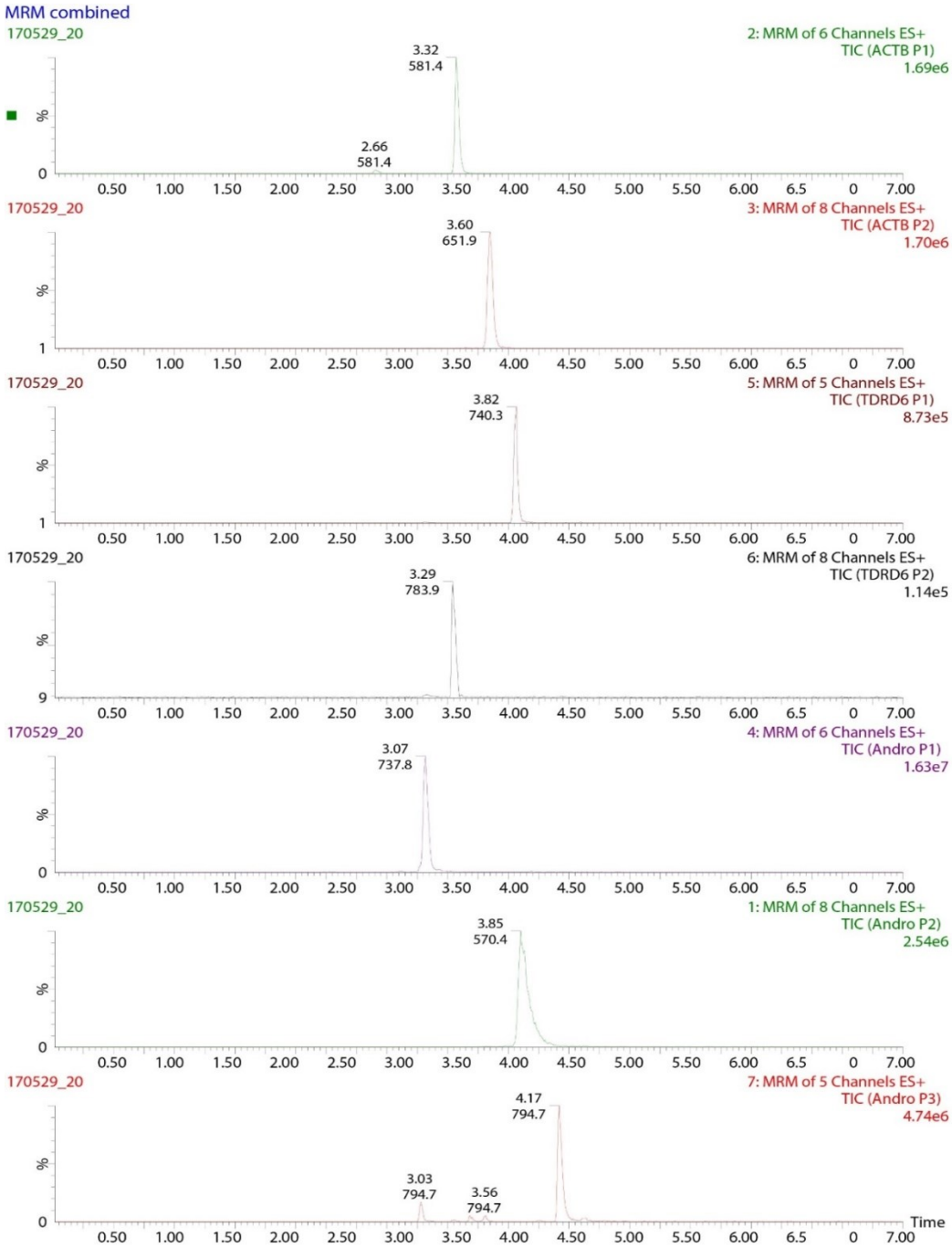
19. Gillette, M. A. & Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* **10,** 28–34 (2013).

20. Zhang, G. *et al.* Protein quantitation using mass spectrometry. *Methods Mol. Biol.* **673,** 211–222 (2010).

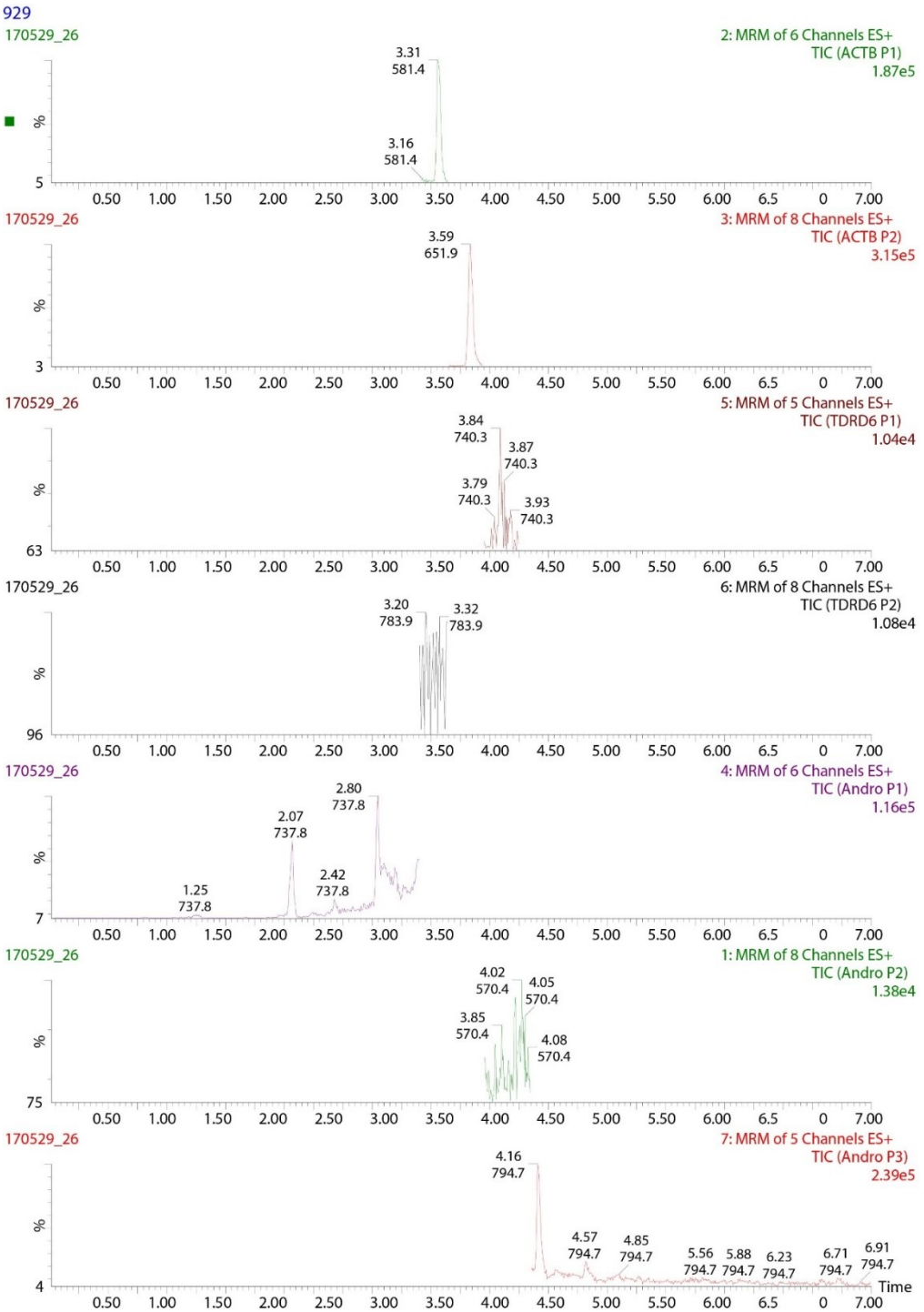
21. Hosokawa, M. *et al.* Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: Domain composition, intracellular localization, and function in male germ cells in mice. *Dev. Biol.* **301,** 38–52 (2007).

22. Parvinen, M. The chromatoid body in spermatogenesis. *Int. J. Androl.* **28,** 189–201 (2005).

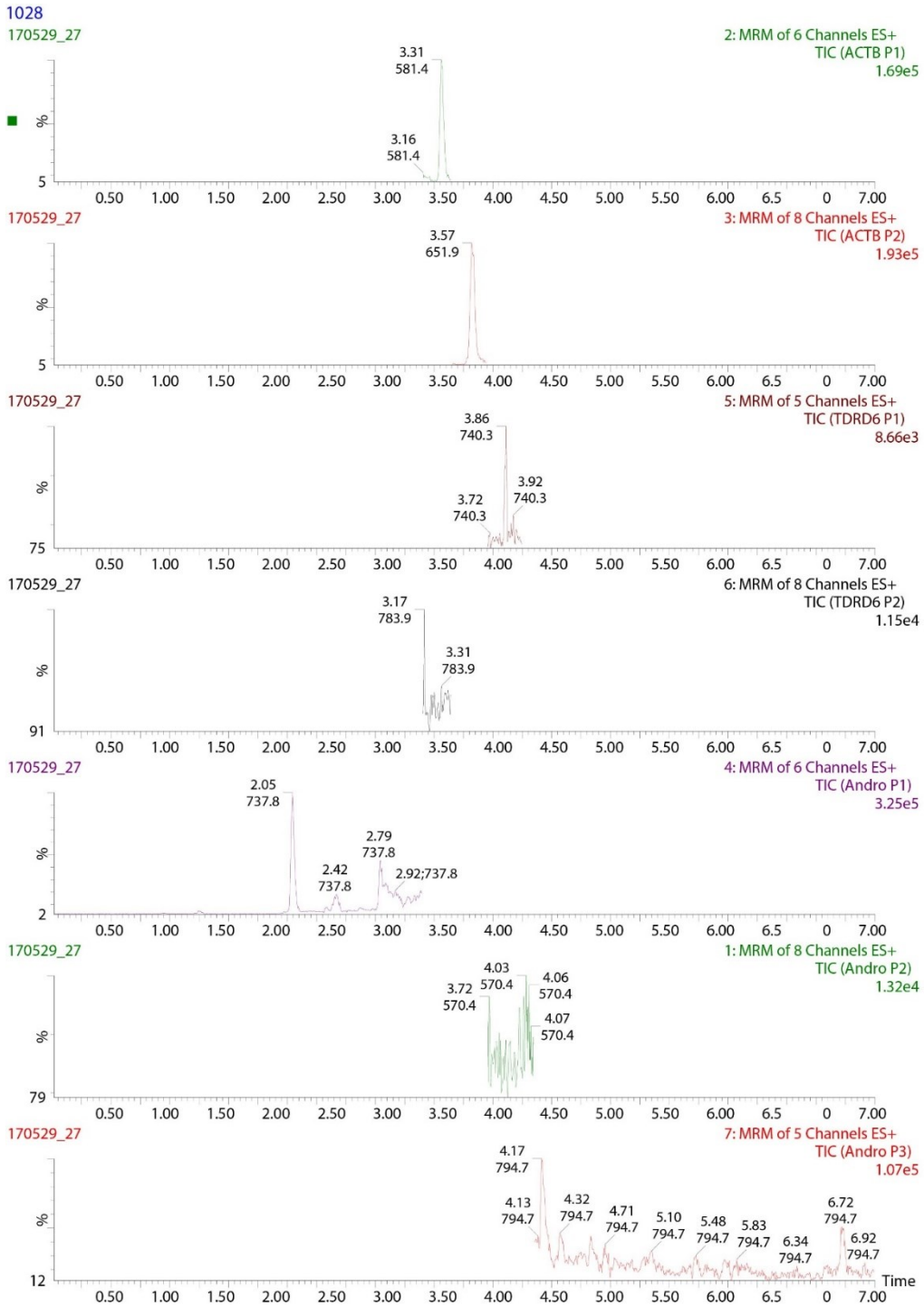
6 Supplementary data



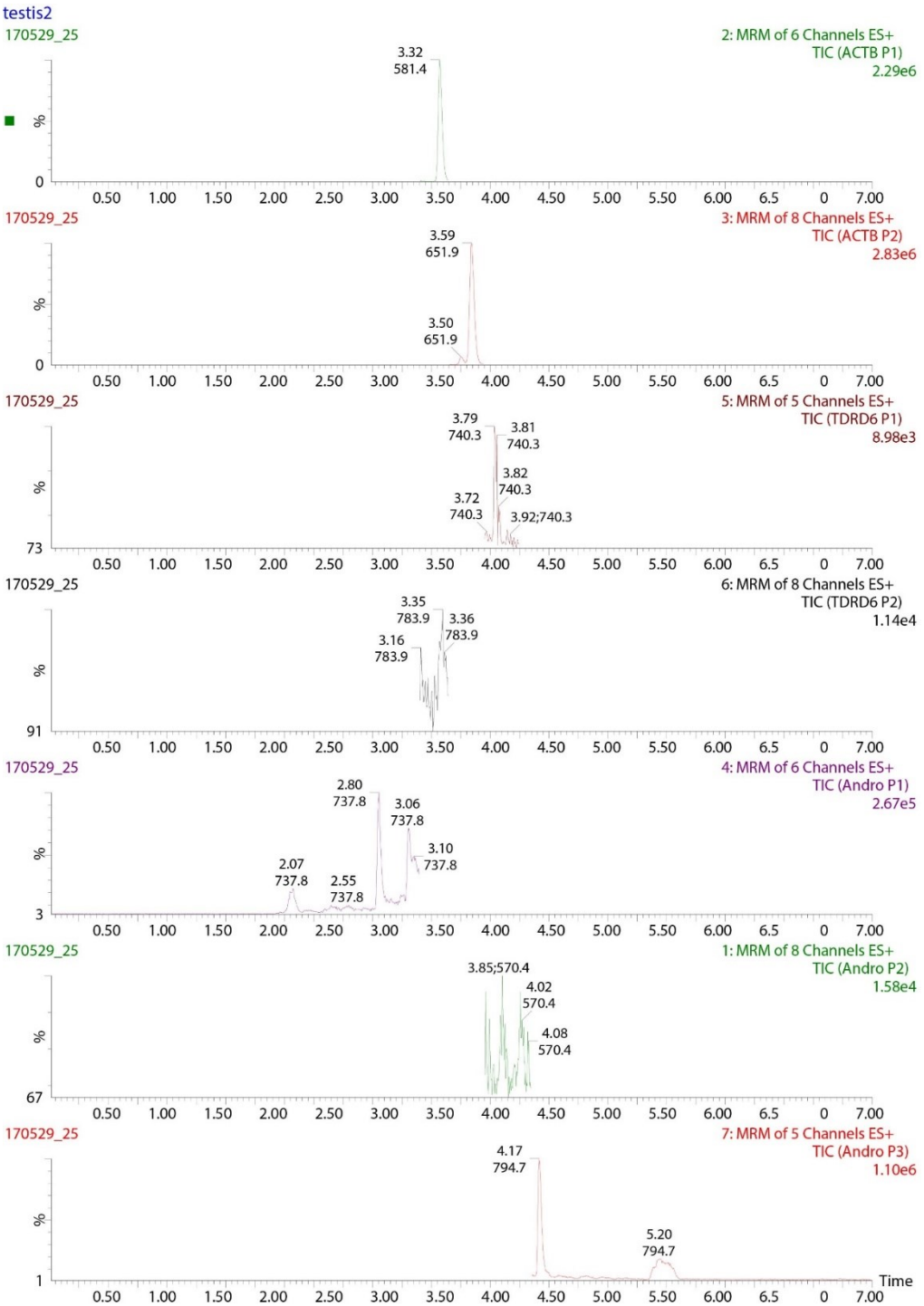
Supplementary data 1: MRM of synthetic PTPs of ACTB, TDRD6 and ADGB.



Supplementary data 2: MRM analysis of sperm sample 1, for ACTB, TDRD6 and ADGB.



Supplementary data 3: MRM analysis of sperm sample 2, for ACTB, TDRD6 and ADGB.



Supplementary data 4: MRM analysis of testis tissue, for ACTB, TDRD6 and ADGB.

VII. General discussion and future perspectives

1 Hypotheses about the function of androglobin

At this moment, it is still too early to draw conclusions about the molecular function of Adgb. However, based on the functional annotations of the different domains of Adgb (calpain-like protease domain, globin domain and two uncharacterized regions) and the findings described in the previous chapters, we will postulate some hypotheses for the molecular function of Adgb.

1.1 The globin domain of Androglobin: function in redox signaling?

Phylogenetic reconstructions and analysis of orthologs showed that Adgb is an evolutionary ancient globin gene. Adgb orthologs have been identified in the genomes of almost all metazoan lineages (definition of metazoan = 'all animals with differentiated tissues, including nerves and muscles). In most orthologous the circular permutated globin domain of Adgb is conserved, indicating that it has a preserved function during evolution.[1]

Alignments of the split Adgb globin domain with mammalian myoglobin (Mb), neuroglobin (Ngb) and cytoglobin (Cygb) sequences revealed that Adgb – despite its rearrangement – conforms to the criteria of the globin fold tertiary structure. This was confirmed by computer molecular modelling of the human ADGB globin domain 3D structure.[1] Due to the instability problems of the globin domain after recombinant expression, we were not able to perform X-ray crystallography to obtain the 3D structure of the globin domain. Though, we were able to record UV-visible spectra of the refolded globin domain of ADGB (Chapter 4). These spectra resembled the spectra of hexa-coordinated globins, such as Ngb and Cygb. Contrary to penta-coordinated globins, which mostly have respiratory roles like myoglobin and hemoglobin, hexa-coordinated globins are linked to other roles, for example: O_2 sensing, redox signaling, detoxification of reactive oxygen/nitrogen species (ROS/RNS) or O_2-dependent enzymatic functions.

Other hexa-coordinated globins known for their role in reproduction are globin 2 (Glob2) and globin 3 (Glob3) of *Drosophila melanogaster*[2] and globin 12 (GLB 12) of

Caenorhabditis elegans[3]. Interestingly, in both *D.melanogaster* and *C.elegans* the Adgb gene is missing. It is therefore tempting to say that Glob2/3, GLB 12 and Adgb might have similar functions related to reproduction. Glob2 and 3 are not yet biochemically characterized, but the function of GLB 12 on the other hand was fully described in 2015 by De Henau et al.[3] *C.elegans* has thirty-four globin-like proteins and an RNAi study for GLB 12 reported abnormal egg laying and embryonic lethality, indicating that GLB 12 plays an essential role in the reproduction of *C.elegans*[4]. Biochemical characterization showed that GLB 12 is a hexa-coordinated globin with redox properties (Fig1). GLB 12 participates in electron transfer and directly interacts with oxygen (O_2) to generate superoxide (0_2^{-}) . This superoxide signal is modulated by an intracellular and an extracellular superoxide dismutase (SOD), creating a transmembrane H_2O_2 gradient that acts as a redox signal. This signal then modulates reproduction, including p38/JNK MAPK-dependent germ cell apoptosis.[3]

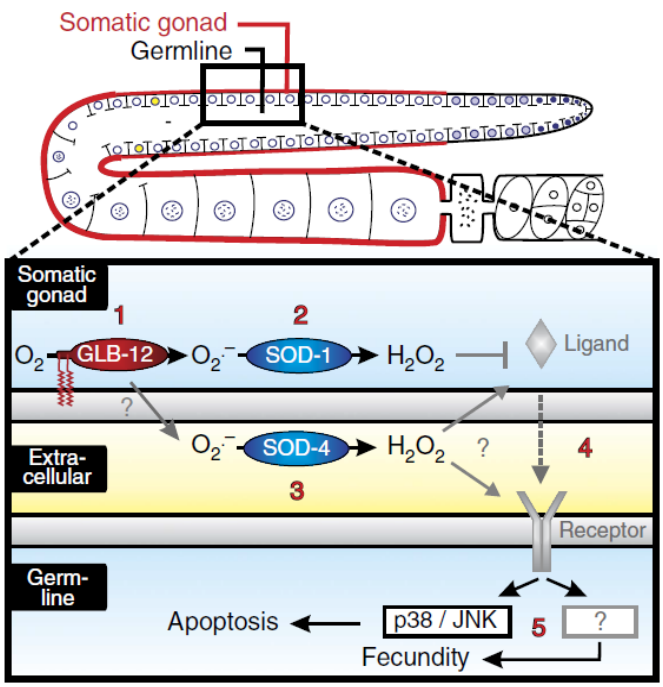


Figure 1: GLB 12 working model. GLB 12 is expressed in the somatic gonad and is capable of converting O2 to 0_2^{-} (1) after which this signal is modulated by an intracellular SOD (2) and influenced by an extracellular SOD (3). Combined GLB 12, SOD-1 and SOD-4 influence germline functioning, including p38 and JNK MAPK-mediated apoptosis (4-5). Figure from[3]

Since Adgb is involved in spermatogenesis and its globin domain is hexa-coordinated, we can hypothesize that the globin domain may have a similar function in redox signaling as GLB 12. Human spermatogenesis is of course entirely different from the reproduction in *C.elegans*. Nevertheless, it is known that ROS production and redox signaling are crucial for a correct functioning of mammalian spermatozoa[5]. ROS promotes the capacitation of spermatozoa, which is a series of morphological and metabolic changes necessary for the spermatozoon to achieve fertilizing ability[5–8] (Fig2). At the cell biology level, capacitation induces changes in the sperm motility pattern known as hyperactivated movement and prepares the sperm to undergo an exocytotic process known as acrosome reaction. At the molecular level, capacitation is associated with cholesterol loss from the sperm plasma membrane, increased membrane fluidity, changes in intracellular ion concentrations, hyperpolarization of the sperm plasma membrane, increased activity of the protein kinase A (PKA), and eventually protein tyrosine phosphorylation (Fig2)[9]. Increasing concentrations of superoxide anion ($O_2^{\cdot-}$), hydrogen peroxide (H_2O_2), nitric oxide ($NO^{\cdot}$) and peroxynitrite ($ONOO^-$) are produced over-time during capacitation. The roles of ROS during capacitation are diverse and complex, and involve the activation of several targets located on the plasma membrane and in other sperm compartments. As an early event, $O_2^{\cdot-}$ and $NO^{\cdot}$ activate adenylyl cyclase that produces cAMP during sperm capacitation. cAMP activates on his turn protein kinase A (PKA), which is essential to trigger the cascade of the late tyrosine phosphorylation associated with sperm capacitation. (Fig2) Interestingly, the identity of the sperm oxidase involved in the production of $O_2^{\cdot-}$ or $NO^{\cdot}$ during capacitation, remains elusive. The globin domain of Adgb could be a putative candidate for this role. The fact that Spata20 was validated as protein-interaction partner of Adgb can support this hypothesis, since Spata20 possess a thioredoxin-like domain. Thioredoxins are ubiquitous proteins involved in redox signaling through the reversible oxidation of two cysteine thiol groups to a disulfide, accompanied by the transfer of two electrons and two protons.[10]

Furthermore, the circular permutated globin domain of Adgb contains an internal IQ-calmodulin binding motif. Calmodulin is an intracellular messenger protein, which can be activated through the binding of Ca^{2+}. The globin domain may thus be regulated by Ca^{2+} signaling, which is of great importance in sperm function and fertilization: Ca^{2+} regulates the acrosome reaction, flagellar-beat mode, chemotaxis and capacitation.[11] The

redox signaling by the globin domain can thus be regulated through Ca^{2+} signaling as well.

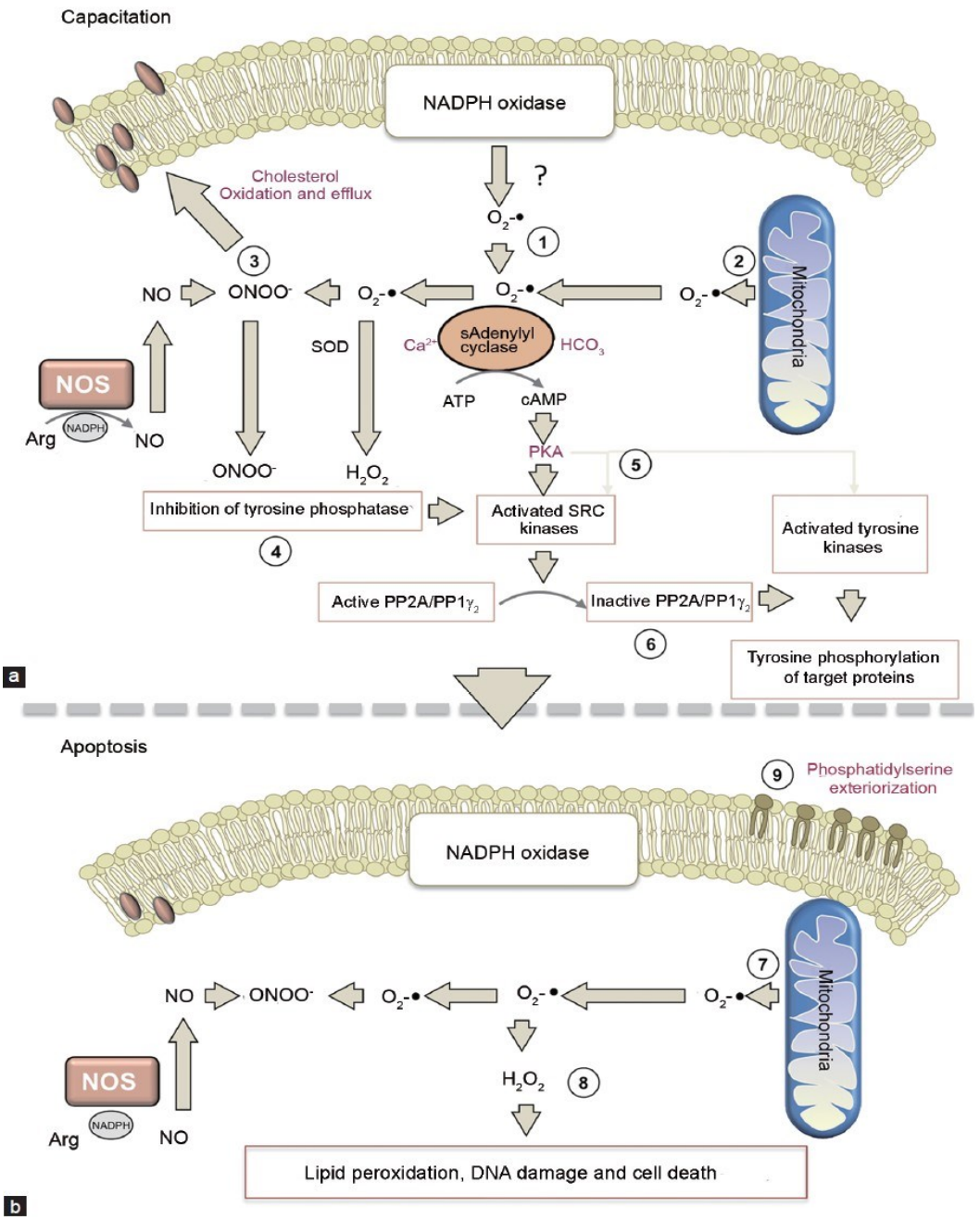


Figure 2: Proposed pathways mediating the redox regulation of (a) sperm capacitation and (b) the initiation of apoptosis. The globin domain might function in redox signaling by producing $O_2^{\cdot-}$ (indicated by the question mark in the proposed pathway). This figure was taken from[8]

However, there are some counter-arguments for a putative redox function of the Adgb globin domain during capacitation. First, the Adgb-deficient mice have a maturation arrest in spermatogenesis at the spermatid phase, long before capacitation takes place. Second, Adgb could not be detected in mature spermatozoa using MRM-analysis, which suggests that Adgb is functioning earlier in the spermatogenic process. And third, this hypothesis does not fit with the hypothesis that Adgb is involved in the CB, which disappears in mature spermatozoa. It is possible that Adgb plays a redox-function earlier in the spermatogenic process, or that Adgb is cleaved and that only the globin domain remains present in the mature spermatozoa (the proteotypic peptides selected for ADGB that were used in MRM-analysis, were not located in the globin domain). It has to be mentioned that other functions, such as oxygen sensing, ROS scavenging or oxygen-dependent enzyme activity, are also plausible for the globin domain of Adgb. Further research is required and an *in vitro* biochemical characterization of the globin domain is absolutely necessary to reveal the molecular function of Adgb.

1.2 The calpain-like domain of androglobin: Auto-cleavage? Truncation of Tdrd6?

Adgb possesses an N-terminal calpain-like protease domain. Calpains are a protein family of calcium-regulated cytoplasmic cysteine proteases involved in the intracellular processing of proteins. The two most-studied mammalian calpains (μ-calpain and m-calpain) are composed of a catalytic large subunit and a regulatory small subunit.[12] Both subunits have C-terminal penta-EF-hand domains, which contribute to the activation of the protease by Ca^{2+} binding. A phylogenetic analysis of the Adgb calpain-like domain demonstrates that it is most closely related to human calpain-7.[1] Calpain 7 is classified as an atypical calpain because it lacks penta-EF domains. Beside a C2-like domain and the catalytic domain, it also possesses a tandem repeat of microtubule-interacting and transport (MIT) domains at the N-terminus (Fig3). The MIT domains of calpain 7 interact with a subset of endosomal sorting complex required for transport (ESCRT)-III related proteins. One of them is the increased sodium tolerance-1 (ISF1) protein, which interacts with calpain 7 with its MIT-interacting motifs (MIMs). Calpain 7 displays autolytic activity enhanced by ISF1 through the MIM-MIT interaction.[13] The N-terminal calpain-like domain of Adgb is homologous to the catalytic domain of calpain-7, but it does not

possess C2-like domains or MIT domains[1] (Fig3). Similar to calpain 7, Adgb displayed truncation upon recombinant expression (chapter 4). However, we are not certain if this is due to autolysis by the calpain-like domain. The catalytic domain of calpain 7 comprises three active site residues: Cys 290, His 458 and Asn 478 (Fig3). Adgb has retained only the Cys-active site residue, but it contains several His and Asn residues at non-standard positions. Autolysis of calpain 7 could be prevented by site-directed mutagenesis of the catalytic Cys290 residue[13]. In future, the proteolytic activity of the Adgb calpain-like domain can be investigated in a similar way, by site-directed mutagenesis of the homologous Cys-residue. Adgb does not have a MIT domain, but similar to calpain 7 it is a multi-domain protein. It is thus plausible that the proteolytic activity of the Adgb calpain-like domain can be regulated through these domains, for instance through oxygen binding at the heme group of the globin domain or by calmodulin binding at the IQ motif of the globin domain.

We have determined Tdrd6 as interaction partner of Adgb and it is known that Tdrd6 is C-terminally cleaved during the transition from meiosis I to meiosis II. Considering that this is the stage of spermatogenesis where Adgb seems to be active, we can suggest that the calpain-like domain of Adgb might be involved in the cleavage of Tdrd6. The cleavage occurs at a predicted caspase I-cleavage site, by an unknown protease[14]. Calpain and caspases are both cysteine proteases with often the same targets. This hypothesis can be further investigated by the co-overexpression of Adgb and Tdrd6 in mammalian cells (e.g. HEK293 or HeLa) and site-directed mutagenesis of the active site residue of the calpain-like domain.

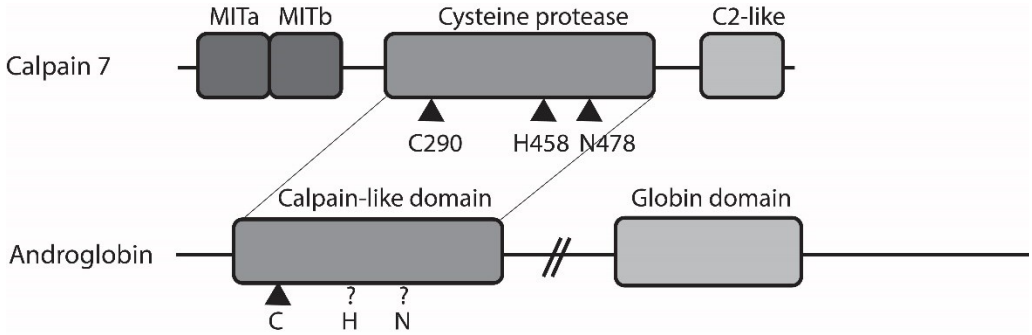


Figure 3: Schematic representation of calpain 7 and androglobin. The calpain-like domain of androglobin is homologous to the catalytic cysteine protease domain of calpain 7. Triangles are indicating the catalytic residues of the protease domain (Cys, His and Asn). Only the putative catalytic Cys residue is conserved in the calpain-like domain of androglobin.

1.3 The uncharacterized domains of androglobin: putative scaffolding function?

Androglobin is a large chimeric multi-domain protein and in addition to the globin domain and the calpain-like domain it also comprises two domains with unannotated functions. In the C-terminal uncharacterized domain lies a coiled-coil motif, possibly involved in protein dimerization, a putative nuclear localization signal, and an overlapping candidate ER membrane retention signal. The localization of Adgb in sperm cells is not yet investigated, because of the lack of a suitable antibody. Hence, we do not know if these nuclear localization signal or ER membrane retention signal are functional. The analysis of orthologous Adgb sequences from primates revealed higher conservation in the calpain and the globin domain than in these unannotated parts of the coding sequence.

A possible hypothesis for these uncharacterized domains is that they may function as scaffolding domains in the sperm cell (in the CB). A scaffold protein is a protein that simultaneously binds two or more proteins, and organizes binding partners into a functional unit to enhance signaling efficiency and fidelity. The seminiferous tubules are avascular and O_2 reaches the luminal region solely by diffusion, making this tissue hypoxic. In this environment globin proteins, which are able to sense O_2 would be advantageous. The binding with O_2 may cause conformational changes in the uncharacterized regions of Adgb and in this way transmit signals towards proteins that

can bind on these putative-scaffold domains. However, these are only speculations and it is too early to make hypotheses. Further research on the function of these unannotated domains is required.

2 Critical considerations and future perspectives

2.1 Choice of expression system: *Escherichia coli* or baculovirus-insect cells?

Currently, uncharacterized proteins are mostly discovered by comparative genomic database searches, implicating that their biochemical characterization starts from the information encoded in their DNA sequence. Computer modeling and alignments with known functional proteins can subsequently predict their potential structure and function. In this way Adgb was characterized as a multi-domain protein with a globin-like domain and a calpain-like domain. However, it must be stressed that this is only an *in silico* characterization, the effective functionality of Adgb and its domains has to be characterized *in vitro* and *in vivo*. Most of the recently discovered proteins, such as Adgb, are expressed in physiologically low concentrations and a pure form of the protein has to be obtained using a heterologous over-expression system. The problems we encountered with the recombinant expression of Adgb in *E.coli* and *P.pastoris* are frequently seen with more complex mammalian proteins. In such cases, I believe that many efforts can be saved by starting directly in a higher eukaryotic system with the appropriate cellular environment for a correct folding and processing of your protein of interest. Also for mammalian proteins that are easily expressed and purified in a bacterial expression system it might be interesting to switch to a higher eukaryotic system, as we have seen that Ngb displayed different characteristics on Western blot when it was expressed in insect cells compared to when it was expressed in *E.coli* or *P.pastoris*. A functional characterization of the baculoviral expressed Ngb might reveal some new insights in the molecular function of this protein.

2.2 The need for a suitable antibody against androglobin

The lack of a suitable antibody for Adgb complicates the *in vivo* characterization of Adgb. At this moment there are only antibodies available generated against recombinant peptides of Adgb. These antibodies do unfortunately not work in immunohistochemistry (IHC) experiments in testis tissue, complicating localization studies of Adgb. Immunostainings for Adgb in testis tissue would reveal crucial information about the biological function of Adgb. We would be able to define in which phases of spermatogenesis Adgb is involved and where Adgb is located in the sperm cell. The hypothesis that Adgb plays a role in the chromatoid body could be confirmed with a suitable antibody. Therefore, as future perspective, it is crucial to invest time and money in the development of a IHC-suitable antibody for Adgb, against different epitopes regarding the fact that Adgb seems to be (auto)-lytically processed. The recombinant expression of Adgb in the baculovirus insect cell system will be a fundamental factor in developing a valid antibody for Adgb.

2.3 Androglobin expression in lung tissue and fallopian tube

According to the human protein atlas, the anti-ADGB Antibody (HPA036340) produced by Atlas antibodies displayed poor results when used in IHC staining in testis tissue[15], but did however display strong positive staining for ADGB in glandular cells of the fallopian tube[16] and in respiratory epithelial cells of the bronchus[17] and nasopharynx[18]. Interestingly, both these cell types possess cilia structures, which are similar in structure and movement to the flagella of spermatozoa. This information strongly suggests that Adgb is involved in the formation of ciliated structures or in the regulation of their flagellar movement. The initial expression analysis in mice showed expression for Adgb in lung tissue, but did not reveal elevated expression in the fallopian tube. Expression of Adgb in these two cell types has to be investigated further. This can be done by screening for ADGB expression in cell lines derived from respiratory epithelial cells. This would form a major addition to the research of Adgb as there are no cell lines available for the haploid sperm cells. To initiate a new boost in unraveling the molecular and biological function of Adgb I believe it is necessary to research the function of Adgb in lung tissue (and fallopian tube), complementary to the characterization in sperm cells.

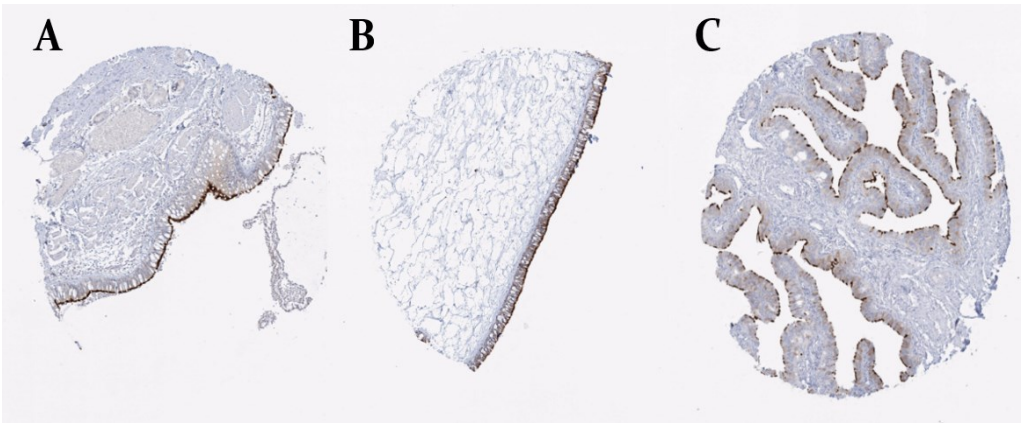


Figure 4: IHC staining with anti ADGB Antibody (HPA036340) produced by Atlas antibodies. Images are taken from v2017.proteinatlas.org. Cilia in respiratory cells of bronchus[17] (A) and nasopharynx[18] (B) and of glandular cells in fallopian tube[16] (C) are distinctly stained.

2.4 Animal model for the *in vivo* characterization of androglobin

In the future, it would be helpful to have a model were we can investigate easily the structure-function relationship of Adgb. The manipulation of Adgb in mice takes a lot of effort and is quite expensive. Therefore, it would be interesting to use a more simpler animal model, e.g. *Danio rerio*. *C.elegans* or *D.melanogaster* cannot be used as model, as they do not possess an Adgb orthologues. Different residues in the functional domains of Adgb can then be mutated and their effect on the fertility can be investigated. In addition, a GFP-coupled form of Adgb could be created, which will reveal crucial information about the localization and regulation of Adgb.

2.5 How is androglobin regulated?

To end this discussion I want to mention that it in order to fully understand the function of Adgb, it is important not only to understand the molecular function of Adgb, but also how Adgb is regulated. The regulation of Adgb can be investigated on transcriptional, translational and post-translational level. The transcriptional regulation of the Adgb gene can be investigated by studying which transcription factors bind on the Adgb promotor region and if they up or down regulate the Adgb gene expression. In addition, it has to be

examined which conditions cause an up or down regulation of Adgb gene expression, e.g hypoxia, oxidative stress, etc. Interestingly, a long intergenic non-coding RNA (linc RNA) has been identified overlapping the putative Adgb promotor region. This linc RNA may be involved in the regulation of Adgb expression, also this has to be investigated. On post-translational level, the Adgb protein might be regulated through binding with small molecules, such as Ca^{2+}, O_2, NO, H_2O_2, etc., by certain protein-protein interactions, or by post-translational modifications, such as phosphorylation, etc. These are all questions that have to be answered in future in order to fully understand how Adgb is functioning.

3 References

1. Hoogewijs, D. *et al.* Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes. *Mol. Biol. Evol.* **29,** 1105–14 (2012).

2. Gleixner, E., Herlyn, H., Zimmerling, S., Burmester, T. & Hankeln, T. Testes-specific hemoglobins in Drosophila evolved by a combination of sub- and neofunctionalization after gene duplication. *BMC Evol. Biol.* **12,** 34 (2012).

3. De Henau, S. *et al.* A redox signalling globin is essential for reproduction in Caenorhabditis elegans. *Nat. Commun.* **6,** 8782 (2015).

4. Kamath, R. S. *et al.* Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421,** 231–237 (2003).

5. Aitken, R. J., Jones, K. T. & Robertson, S. A. Reactive Oxygen Species and Sperm Function--In Sickness and In Health. *J. Androl.* **33,** 1096–1106 (2012).

6. O'Flaherty, C. Redox regulation of mammalian sperm capacitation. *Asian J. Androl.* **17,** 583–90 (2015).

7. Aitken, R. J., Gibb, Z., Baker, M. A., Drevet, J. & Gharagozloo, P. Causes and consequences of oxidative stress in spermatozoa. *Reprod. Fertil. Dev.* **28,** 1 (2016).

8. Aitken, R. J., Baker, M. A. & Nixon, B. Are sperm capacitation and apoptosis the opposite ends of a continuum driven by oxidative stress? *Asian J. Androl.* **17,** 633–9 (2015).

9. Stival, C. *et al.* Sperm Capacitation and Acrosome Reaction in Mammalian Sperm. *Adv. Anat. Embryol. Cell Biol.* **220,** 93–106 (2016).

10. Shi, H.-J. *et al.* Cloning and characterization of rat spermatid protein SSP411: a thioredoxin-like protein. *J. Androl.* **25,** 479–93 (2004).

11. Publicover, S., Harper, C. V & Barratt, C. [Ca2+]i signalling in sperm--making the most of what you've got. *Nat. Cell Biol.* **9,** 235–42 (2007).

12. Croall, D. E. & Ersfeld, K. The calpains: modular designs and functional diversity. *Genome Biol.* **8,** 218 (2007).

13. Osako, Y. *et al.* Autolytic activity of human calpain 7 is enhanced by ESCRT-III-related

protein IST1 through MIT-MIM interaction. *FEBS J.* **277,** 4412–26 (2010).

14. Vasileva, A., Tiedau, D., Firooznia, A., Müller-Reichert, T. & Jessberger, R. Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression. *Curr. Biol.* **19,** 630–9 (2009).

15. Protein Atlas - ADGB epidydimis. (2017). Available at: https://www.proteinatlas.org/ENSG00000118492-ADGB/tissue/epididymis. (Accessed: 29th October 2017)

16. Protein atlas - ADGB fallopian tube. (2017). Available at: https://www.proteinatlas.org/ENSG00000118492-ADGB/tissue/fallopian+tube. (Accessed: 29th October 2017)

17. Protein atlas - ADGB bronchus. (2017). Available at: https://www.proteinatlas.org/ENSG00000118492-ADGB/tissue/bronchus. (Accessed: 29th October 2017)

18. Protein atlas - ADGB nasopharynx. (2017). Available at: https://www.proteinatlas.org/ENSG00000118492-ADGB/tissue/nasopharynx. (Accessed: 29th October 2017)

Curriculum vitae

Personalia

Name	An Bracke
Adress	Ravelsbergstraat 70 bus 1, 2100 Deurne
Date of birth	24/01/1989
Nationality	Belgian
Phone	+32 487 292605
Email	an_bracke@hotmail.com

Relevant work experience

Oct 2017 – present

Research associate
University of Antwerp
Promotor: Prof. Dr. Anne-Marie Lambeir
Research group: Medicinal Biochemistry

Oct. 2012 – Sep 2017

PhD student Biochemistry and Biotechnology
University of Antwerp
Promotor: Prof. Dr. Sylvia Dewilde
Research group: Protein science, Proteome analysis and Epigenetic signaling
Doctoral thesis: Unraveling the molecular function of androglobin: a testis-specific globin

Financing:
Oct. 2013 – Present: FWO scholarship
Oct. 2012 – Sept 2013: UA BOF project

Aug 2015 – Dec 2015

Research stay
University of Duisburg-Essen
Institute for Physiology
Research group of Prof. David Hoogewijs

Financing:
FWO Travel grant long stay abroad

Education

Sept 2010 – Jun 2012

Master Biochemistry and Biotechnology: Molecular and Cellular Gene Biotechnology
University of Antwerp
Summa cum laude

Master thesis: Biochemical characterization of Androglobin: a recently discovered chimeric globin, preferentially expressed in testis.
Promotor: Prof. Dr. Sylvia Dewilde
Research group: : Protein science, Proteome analysis and Epigenetic signaling

Internship: Interaction between PREP and α-synuclein
Promotor: Prof. Dr. Anne-Marie Lambeir
Research group: Medical Biochemistry

Sept 2007 – Jun 2010

Bachelor Biochemistry and Biotechnology
University of Antwerp
Cum laude

Bachelor thesis: Expression and purification of myoglobinI and 2
Promotor: Prof. Dr. Sylvia Dewilde
Research group: Protein science, Proteome analysis and Epigenetic signaling

Scientific techniques

Protein expression and purification
Bacterial expression systems (*E.coli*), Yeast expression systems (*P.pastoris*), Insect cell expression systems (Sf9, baculovirus), Affinity/size exclusion/reverse phase chromatography with FPLC and HPLC

Protein Biochemistry
SDS-PAGE, Western Blot, UV/Vis, Flash Photolysis, Stopped Flow

Molecular Biology
Gateway cloning, Restriction Enzyme-cloning, PCR, RT-qPCR

Cell culture
Transient and stable transfection of cancer cell lines, Luciferase assays, Lentiviral infection, Mammalian-two-Hybrid

Computer skills
Word, Excel, Power point, CloneManager, qBasePlus, Origin, UCSC Genome Browser, UniProt, Expasy, Mendely

Additional courses

Courses Antwerp Doctoral school
Time management
Word: Long Documents
English: Academic writing
English for Communicating Effectively in an Academic Context
E-Sources

Awards

Poster Price
Awarded by the Faculty of 1000 XVIII International Conference of Oxygen Binding and Sensing Proteins
University of Sheffield, UK
July 2014

Best Master Student Biochemistry and Biotechnology
Awarded by Koninklijke Vlaamse Chemische Vereniging (KVCV)
June 2012

Best Bachelor Student Biochemistry and Biotechnology
Awarded by the faculty of Pharmaceutical, Biomedical and Veterinary Sciences
June 2010

Published Articles

Bracke A., Hoogewijs D., Dewilde S. (2017) "Exploring three different expression systems for recombinant expression of globins: Escherichia coli, Pichia pastoris and Spodoptera frugiperda" Analytical Biochemistry. (https://doi.org/10.1016/j.ab.2017.11.027)

Bracke A., Peeters K., Punjabi U., Hoogewijs D., Dewilde S. (2017) "A search for molecular mechanisms underlying male idiopathic infertility" Reproductive biomedicine online. (https://doi.org/10.1016/j.rbmo.2017.12.005)

Abstracts

Androglobin: a newly discovered globin preferentially expressed in testes
An Bracke, Kris Peeters, Usha Punjabi, Thomas Hankelen, David Hoogewijs, Sylvia Dewilde.

9the European Congress of Andrology
Rotterdam, The Netherlands
21-23/09/2016

19th International Conference on "Oxygen binding and sensing proteins"
Hamburg, Germany
11-14/09/2016

Unraveling the function of Androglobin through expression analysis of human sperm and testis and biochemical characterization of the Androglobin globin domain
An Bracke, Herald Berghmans, Thomas Hankeln, David Hoogewijs, Sylvia Dewilde.

18th International meeting on "Oxygen-binding and sensing proteins"
Sheffield, UK
29/08-01/09/2014

Dankwoord

Dankwoord

Tot slot had ik nog graag enkelen mensen willen bedanken, zonder hun hulp en steun had deze thesis niet kunnen zijn wat het nu is.

Eerst en vooral wil ik mijn promotor Prof. dr. Sylvia Dewilde bedanken. Sylvia, jij gaf me de kans om dit doctoraat uit te voeren. Mijn onderzoek naar androglobine liep niet van een leien dakje en het ging niet altijd zoals ik het me eerst had voorgesteld. Jij leerde me hoe ik toch een positieve draai kon geven aan al deze "negatieve resultaten en mislukkingen". Dit heeft uiteindelijk geleid tot een mooie paper waar we trots op kunnen zijn.

Prof. em. dr. Luc Moens, bedankt voor de interesse die u toonde in mijn onderzoek en de wijze raad die u me steeds spontaan meegaf. Uw uitspraak dat "onderzoek voor 80% bestaat uit proberen en mislukken en slechts 20% uit echt resultaten generen", zal ik nooit vergeten en heeft me vaak geholpen om de moed er in te houden.

Prof. dr. David Hoogewijs, bedankt om steeds zo enthousiast te blijven over androglobine. Jouw enthousiasme is zeer aanstekelijk. Ik heb genoten van het onderzoek dat ik bij jou in Essen mocht komen doen. Bedankt om steeds mijn geschreven teksten kritisch na te lezen en verbeteringen aan te brengen daar waar nodig.

dr. Kris Peeters en dr. Usha Punjabi, jammer dat het uiteindelijk zo moeilijk bleek te zijn om androglobine te detecteren in spermastalen. Toch heb ik veel geleerd uit de studies op menselijke spermastalen en testis biopten. Dit was mogelijk dankzij jullie tijd en moeite, bedankt hiervoor.

dr. Geert Van Raemdonck, bedankt om me zo goed te helpen met de MRM experimenten. Ik bewonder jouw passie en expertise die je hebt voor massa spectrometrie. Ik wens je nog veel succes met je verdere carrière.

Prof. dr. Anne-Marie Lambeir, Prof. dr. Dirk Snyders, Prof. dr. Bart Braeckman, Prof. dr. Jo Leroy en Prof. dr. Thomas Hankeln, bedankt om jullie taak als doctoraatsjury zo zorgvuldig uit te voeren. Jullie hebben mijn werk kritisch en tot in de details nagelezen, dit waardeer ik enorm.

Mijn collega's en ex-collega's van de PPES groep bedank ik graag voor de leuke werksfeer die ze creëerden gedurende de afgelopen vijf jaar, op T1 en daarna ook op S4.

Herald, bedankt om me zoveel bij te leren over eiwitexpressie en zuivering. Jij bent van onschatbare waarde voor het labo. Jij schoot me te hulp wanneer ik gisten in een bioreactor wilde groeien, wanneer er weer eens een pipet kapot was, wanneer ELN niet goed werkte, wanneer ik over de bindingskinetiek van hexa-gecoördineerde globines aan het nadenken was, en zo kan ik nog wel even doorgaan.

Marie-Louise, bedankt om ervoor te zorgen dat er orde en structuur was in het labo, dit maakte het werken steeds aangenaam. Het laatste jaar heb je me nog geholpen met verschillende kloneringen die ik plots nog moest doen, zonder jouw hulp hierin had ik dit niet op tijd kunnen afronden. Nu kan je van je welverdiende pensioen genieten! (ik ben er zeker van dat ze je al missen).

Stijn, we hebben gelijktijdig ons doctoraat afgewerkt, met een lach en een zwans konden we zo goed als alles relativeren en hebben we het tot een succesvol einde gebracht. Evi, jij kan iedereen aan het lachen doen brengen met je zotte verhalen of je imitatie-kunsten, dit zorgde voor een positieve sfeer in het labo. Zoë, jouw eerlijkheid en passie voor wetenschap stralen van jouw af, houden zo! Joey, jij blijft in elke omstandigheid jezelf, dat siert je. Het was fijn je als bureaugenoot te hebben. Dietmar, I appreciate your calm, it was pleasant to share a desk with you. Zainab, jij bent zo attent en betrokken met iedereen, mooi om te zien. Eva en Karen, bij jullie in het practicum kon ik steeds terecht voor een Western blot te detecteren of een qPCR uit te voeren, bedankt hiervoor.

Mops, Ruben en Martin, de mannen van de proteomics. Jullie stonden altijd klaar om me te helpen wanneer ik iets wilde meten met de massa spectrometer. Daarnaast kon ik steeds terecht bij jullie voor fijne babbels, grappige Youtube filmpjes of wijze spreekwoorden.

Jolien, Ken, Chandra en Farnaz, de mannen en vrouwen van de epigenetica, bedankt voor de fijne babbels onder de middag en de leuke werksfeer die jullie mee creëerden op T1 en S4.

Maarten en Marìa, my colleagues in Essen. Thanks for the warm welcome during my research stay in Germany. Maarten, het was heel fijn om samen met jouw elk weekend van Essen over en weer naar Antwerpen te rijden. Bedankt voor de fijne babbels tijdens deze autoritten, ik wens je nog heel veel succes met het beëindigen van je doctoraat.

Prof. dr. Anne-Marie Lambeir, bedankt voor de kans die je me gaf om te komen werken voor het labo medische biochemie en de kennis over enzym kinetiek die je met me wil delen. Mijn nieuwe collega's van de MedBio wil ik graag bedanken voor hun warm onthaal, jullie deden me vanaf dag 1 direct thuis voelen op S5.

Gwendolyn, partner-in-crime in de insect cellen, het was fijn om dit samen met jou te ontdekken en op punt te zetten. Ik wens je nog veel succes met je doctoraat.

Yannick en Martin, bedankt voor de fijne en verrijkende discussies over wetenschap, techniek, film, muziek en nog veel meer. We zien elkaar nog!

Roos, bij jou kon ik steeds terecht om iets te 'komen lenen', een vector, een enzym, een gist of buffer, … noem maar op. Daarnaast had je altijd nieuwe inzichten in de aanbieding of gewoonweg een paar aanmoedigende woorden, wanneer ik het (weer) even moeilijk had. Hierdoor vond ik steeds opnieuw de moed om er met volle zin terug in te vliegen en door te zetten.

Els, Leen en Lieve, jullie zijn stuk voor stuk top zussen. Els, bedankt om altijd klaar te staan met wijze raad als oudste zus. Jij gaf me het mooiste cadeau ooit: meter mogen zijn van jullie Karel, zijn schattigheid deed vaak alle doctoraatszorgen even vergeten. Leen, bedankt om me aan allerlei artikels te helpen waar de UA geen toegang tot had en de KUL wel. Als mede-biochemicus kon jij mijn frustraties in het labo altijd goed begrijpen en je stond steeds klaar met een luisterend oor. Lieve, zonder jou hadden de figuren in deze thesis er nooit zo professioneel en mooi uitgezien, bedankt hiervoor!

Mijn vrienden wil ik graag bedanken voor de fijne avonden in 'den Trol' met of zonder een spaghetti erbij, de leuke weekends in de Ardennen of gewoon gezellige avonden bij elkaar thuis. Bij jullie kan ik me helemaal ontspannen en mezelf zijn, dit is van onschatbare waarde!

Linde en Annelies, jullie weten wat het is om te doctoreren en welke stress dat er bij komt kijken wanneer je moet gaan verdedigen. Het was fijn dat ik bij jullie af en toe mijn frustraties en kopzorgen kon komen ventileren.

Roel, jij bent een vaste waarde in mijn leven, waar je ook ergens zit op de wereldbol, wij blijven steeds contact houden. Met jouw is het fijn te filosoferen over de kleine of grote zaken in het leven! Bedankt trouwens om me te helpen in de keuze van dit lettertype.

Vokke en Moeke, jullie hebben me alle kansen gegeven die ik me maar kon wensen. Jullie weten steeds perfect hoe ik me voel of wat ik denk, zonder jullie steun en liefde zou ik dit niet hebben kunnen verwezenlijken.

Ten slotte,

Koen, lief liefje. Jij staat altijd klaar voor mij, jij weet me altijd gerust te stellen, te troosten of te motiveren. Zo trots dat je bent op mij, ben ik trots op jou. Jij maakt me gelukkig en samen kunnen wij de wereld aan!

Bedankt allemaal

An Bracke
Antwerpen 2018