Charles University

Faculty of Science

Study programme: Physical Chemistry

Hagen Sülzen

Structural insights into innate immune evasion mechanisms of African trypanosomes and type C adenoviruses

Doctoral thesis

Supervisor: Dr. Sebastian Zoll

Prague, 2024

Charles University

Abstract

Structural insights into innate immune evasion mechanisms of African trypanosomes and type C adenoviruses

by Hagen Sülzen

The human body has evolved a plethora of intricate defence mechanisms to fend off any potentially pathogenic microorganisms, in turn prompting the latter to develop countermeasures. Obtaining a detailed understanding of the structural foundation underlying the immune evasion mechanisms employed by human-infective pathogens is crucial for development of effective drugs and therapeutics. In this work, a protocol for improvement of recombinant expression of proteins with complex fold and generation of a minimal construct amendable to structural characterisation is presented by example of invariant surface glycoprotein 65 (ISG65) from the blood-stream form of the human infective parasite *Trypanosoma brucei gambiense (Tbg)*. In addition, the findings shown here reveal the biological function of ISG65 as a receptor for human complement C3b, describe its role as a selective inhibitor of the alternative pathway of the complement system and present the first structures of *Tbg*ISG65 in complex with human C3 and C3b. To elucidate how other invariant surface receptors of the parasite may exert their function within the protective shield formed by the surrounding variant surface glycoproteins (VSGs), the conformational flexibility of the parasites blood-stream surface coat was assessed. While infection with African trypanosomes has severe consequences for patient health, infections with the widespread human adenovirus C5 on the contrary are usually mild and self-limiting. Despite its high prevalence around the world and great potential in gene therapy and vaccination applications, the molecular mechanisms governing the infection are still poorly understood. Results presented here provide structural insights into how species C adenoviruses may exploit human lactoferrin, an antimicrobial component of the human innate immune defence, to facilitate infection of the respiratory epithelium. This discovery holds promise for the development of novel antivirals or adjuvants for improved transduction efficacy in gene therapy applications.

Declaration

I solemnly declare that this dissertation, submitted for attainment of the degree of Doctor of Philosophy, has been written independently. All sources and literature utilised have been duly acknowledged through their citation. I affirm that neither this work nor any substantial part of it has been previously submitted to obtain any other academic degree.

In Prague, Date:

....
Hagen Sülzen

Acknowledgements

I would like to extend my deepest gratitude to everyone who supported me throughout my PhD journey.

I owe special thanks to my supervisor, Sebastian Zoll, not only for providing me with this incredible opportunity but also for your unparalleled guidance and continuous support. Your mentorship, unwavering trust in my abilities and the responsibilities you entrusted to me have been pivotal in my growth. In addition, your candid sharing of the complexities and inner workings of our profession has been a privilege and provided me with a clear-eyed view of the associated challenges, instrumentally shaping my future endeavours.

A heartfelt thank you goes to my early co-workers Jakub Began and Jitka Votrubová, whose foundational efforts have set the stage for my successes. Your unmatched dedication and collective support in the early days of the group have continuously inspired me.

I am also thankful to Alžběta Kadlecová for voluntarily administering my grant, providing invaluable assistance that allowed me to focus on my research.

To my Mum and Katka, your encouragement and infinite support have been my backbone during this PhD. You have accompanied me through every high and every low, offering support during challenging times and celebrating with me in moments of success. This journey would have been impossible without you, and for that, I could not be more grateful.

Contents

Abbreviations

AA	Ammonium acetate
AAT	Animal African trypanosomiasis
AdV	Adenovirus
AM	Alveolar macrophages
AP	Alternative pathway *[of the complement system]*
AVP	Adenovirus protease
Bb	Factor B fragment b
BBB	Blood-brain barrier
CAR	Coxsackie and adenovirus receptor
CNS	Central nervous system
CP	Classical pathway *[of the complement system]*
Cryo-EM	Cryo-electron microscopy
CSF	Cerebral spinal fluid
CTD	C-terminal domain
DAD	Disseminated adenoviral disease
DFMO	α-difluoromethylornithine
DMEM	Dulbecco's modified Eagle's medium
DSA	Di-(N-succinimidyl)adipate
E. coli	*Escherichia coli*
EGF	Epidermal growth factor
eIF-2α	Eukaryotic initiation factor 2α
ES	VSG expression site
ESAG	VSG expression site associated gene
fB, fD, fI, fP	Factor B, D, I, P (properdin)
FoB	Follicular B-cells
$FSC_{0.143}$	Fourier shell correlation *[at the 0.143 cut-off]*
FX	Coagulation factor X
gHAT	'gambiense' human African trypanosomiasis
Gla domain	γ-carboxyglutamic acid-rich domain
GONs	'Groups-of-nine'
GPI	Glycosylphosphatidylinositol
HAdV	Human adenovirus
HAT	Human African trypanosomiasis
Hb	Haemoglobin
HDL	High-density lipoprotein
HDP	Host defence protein
HDX-MS	Hydrogen-deuterium exchange mass spectrometry
hLF	Human lactoferrin
Hp	Haptoglobin
HSPGs	Heparan sulfate proteoglycans
HVR	Hypervariable region
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IEX	Ion-exchange chromatography
IL-8	Interleukin 8
IM	Intramuscular
ISG	Invariant surface glycoprotein

IV	Intravenous
JAM	Junctional adhesion molecule
kDNA	Kinetoplast DNA
Lfcin	Lactoferricin
LP	Lectin pathway *[of the complement system]*
MAC	Membrane attack complex
MARCO	Macrophage receptor with collagenous structure
MASP	MBL-associated serine protease
MBL	Mannan-binding lectin
MHC-1	Major histocompatibility complex-1
MPI cell	Max Planck Institute cell
MST	Microscale thermophoresis
mVSG	Metacyclic VSG
MZ	Marginal zone
NECT	Nifurtimox–eflornithine combinatorial therapy
NHS	Normal human serum
NK cells	Natural killer cells
NTD	N-terminal domain
PATTEC	Pan African Tsetse and Trypansomiasis Eradication Campaign
pIX	Protein IX
PKR	Protein kinase R
PRR	Pattern recognition receptor
PTRE	Post-treatment reactive encephalopathy
pVIII	Polypeptide VIII
Rab5	Ras analogue in brain protein 5
rHAT	'rhodesiense' human African trypanosomiasis
RU	Response unit
SAXS	Small-angle x-ray scattering
SEC	Size exclusion chromatography
SIF	Stumpy induction factor
SR	Scavenger receptor
SRA	Serum resistance antigen
TAC	Tripartite attachment complex
Tbb	*T. b. brucei*
Tbg	*T. b. gambiense*
TbHpHbR	*Trypanosoma brucei* haptoglobin-haemoglobin receptor
TgsGP	*T. b. gambiense*-specific glycoprotein
TLF	Trypanolytic factor
TLR	Toll-like receptor
TMD	Transmembrane domain
TP	Terminal pathway *[of the complement system]*
TPCC	Terminal pathway complete complex
VCAM-1	Vascular cell adhesion molecule-1
VSG	Variant surface glycoprotein
WHO	World Health Organization

Chapter 1

Introduction

Part I – African trypanosomes

I.1   Sleeping sickness

I.1.1 Cause

African trypanosomiasis is a neglected tropical disease found exclusively in sub-Saharan Africa. When left untreated, the disease, infectious to both humans and animals, is fatal [1]. African trypanosomiasis is caused by uniflagellar, protozoan parasites of the species *Trypanosoma brucei* which is divided into three subspecies: *T. b. brucei*, *T. b. gambiense* and *T. b. rhodesiense* [2]. While both, *T. b. gambiense* and *T. b. rhodesiense* are causative of human African trypanosomiasis (HAT) [3], commonly referred to as 'sleeping sickness', *T. b. brucei* is not infective to humans but causes animal trypanosomiasis (AAT) instead, a disease also known as nagana (from Zulu 'N'gana', powerless/useless) [4]. *Trypanosoma brucei*, along with the animal infective species *T. vivax* and *T. congolense*, belong to the group of salivarian trypanosomes, a name referring to the mode of transmission [5]. All three species are predominantly transmitted via the saliva of the biological disease vector, the tsetse fly (*Glossina* spp.) (*see* I.2.2).

I.1.2 Disease burden and distribution

The geographical distribution of the *Trypanosoma brucei* subspecies is closely tied to the ecological requirements of the corresponding vector. Although all subspecies within the salivaria group are transmitted by members of the *Glossina* genus, *T. b. gambiense* is primarily transmitted by *Glossina palpalis, fuscipes* and *tachinoide*s, all three of which are found in West and Central Africa [6]. The term *'gambiense'* is derived from the Gambia river in the Republic of The Gambia, where the parasite was first identified [7]. On the contrary, *T. b. rhodesiense*, named after the region of Rhodesia (modern-day Zimbabwe) [8], is primarily transmitted by *Glossina*

species *moristans, pallidipes* and *swynnertoni,* all found in Eastern and Southern Africa [6]. In 2022, 837 total cases of HAT have been reported, 799 of which were caused by *T. b. gambiense* (gHAT) and 38 were caused by *T. b. rhodesiense* (rHAT) (Figure 1) [9]. Although continuous relief and prevention efforts by the World Health Organization (WHO) have led to a stark decrease in the number of annually reported infections (88.68% and 65.45% over the past 10 years for *T. b. gambiense* and *T. b. rhodesiense*, respectively) [9], millions of people are still at risk [10].
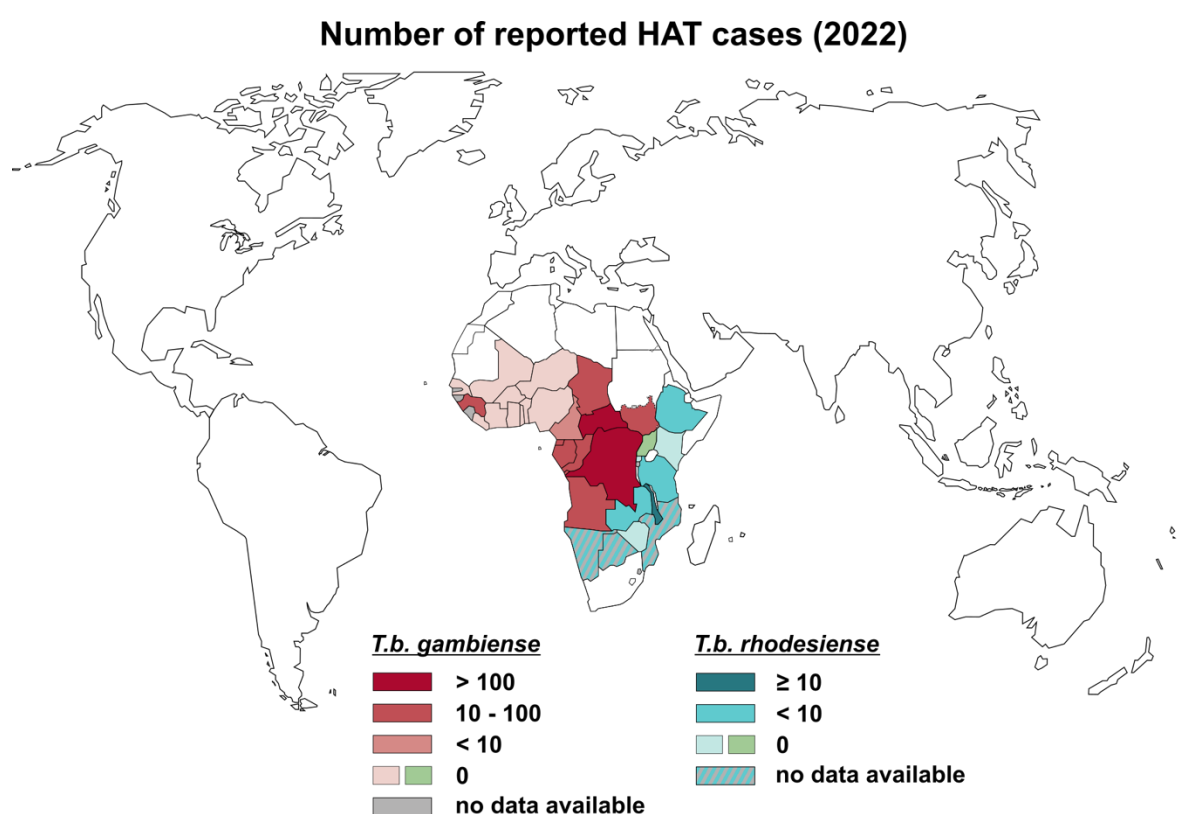
Figure 1 - Reported Human African Trypanosomiasis cases in the year 2022.
World map illustrating the number of reported HAT cases for individual countries in the year 2022. *T. b. gambiense* infections are displayed with a red colour gradient, *T. b. rhodesiense* cases are coloured with a blue gradient. Uganda, the only country endemic to both species, is highlighted in light green. It reported no new cases for either species in the year 2022. The figure was adapted from the Global Health Observatory data repository (WHO) [9].

For *T. b. gambiense*, the most effective containment strategy has proven to be 'case finding and treatment' in active and passive screening campaigns, thereby minimising the human reservoir and subsequently decreasing transmission rates [11]. Control of *T. b. rhodesiense* infections poses a far greater challenge as both domestic cattle and wild animals are thought to be epidemiologically relevant animal reservoirs [12]. Their

close proximity to humans makes blanket treatment and insecticide application on these animals imperative for containment of trypanosome infections [3]. Additionally, albeit to a lesser degree, vector control measures such as bush clearing, release of sterile, male tsetse flies and aerial or ground spaying of insecticides have been employed to limit the transmission rates [3]. Total eradication of *T. b. gambiense* is challenged primarily by sustenance of the commitment to disease control measures by population and national authorities despite receding infection numbers [3]. Historically, countries endemic to the parasite have been politically unstable; war and civil unrest may disrupt ongoing prevention measures and cause resurgence of infection rates. Successful suppression of *T. b. rhodesiense* infections is largely limited by the difficulties associated with the control of wild animal reservoirs.

Similarly to HAT, animal African trypanosomiasis (AAT) is caused by multiple *Trypanosoma* species (*T. b. brucei, T. congolense, T. vivax*), all of which are transmitted by the bite of the tsetse fly [13]. Consequently, the disease occurs across the entire habitat of *Glossina* spp., an estimated area of 10 million sq. kilometres. Although infectious to a range of wild and domestic animals, the diseases impact is most severe for cattle. Annually, an estimated 3 million livestock perish from nagana, causing enormous economic losses and food insecurities in the affected areas [14]. While intensive screening and monitoring campaigns have resulted in high morbidity but low mortality of AAT, emerging drug resistance of the parasites and a lack of technological advances in treatment of the disease raises concerns about the sustainability of the current approach [14]. While vector control and eradication of the tsetse fly, as outlined by the Pan African Tsetse and Trypanosomiasis Eradication Campaign (PATTEC), would circumvent these problems, the practicability of this method has been questioned [15]. Alternatively, the use of farm animals with increased immunity towards the parasite could alleviate the disease burden. While a range of 'trypanotolerant' livestock breeds are known, the genetic foundation that confers resistance to the parasite is poorly understood, thereby jeopardizing efforts to selectively breed these species for high productivity [16].

I.1.3 Clinical manifestation

Human African trypanosomiasis progresses through two stages, each with a commonly associated subset of clinical manifestations. A first sign of infection, more commonly observed in *T. b. rhodesiense* infection, is the development of a painful skin chancre at the site of the tsetse fly bite as a result of a local immune response, 5-15 days post-infection [17]. From the site of infection, the parasites migrate into the blood and lymphatic system, marking the beginning of the first (or early) clinical stage, also referred to as haemolymphatic stage [18]. Early-stage symptoms are unspecific and variable, but often include intermittent fever with rigors, weight loss, fatigue, malaise, arthralgia, haemolytic anaemia, hepatomegaly and abnormal liver function, splenomegaly and generalised or regional lymphadenopathy [18]. A typical clinical manifestation of early *T. b. gambiense* infection is posterior triangle cervical lymphadenopathy, also referred to as 'Winterbottom's sign' [18].

Migration of the parasite across the blood-cerebral spinal fluid (CSF) and blood-brain barrier (BBB) into the central nervous system (CNS) initiates the second stage of the disease, also referred to as late or 'encephalitic' stage [18]. The majority of patients develop a characteristic sleeping disorder characterised by daytime somnolence and nocturnal insomnia [18], coining the term 'sleeping sickness' [3]. Other psychiatric manifestations may include emotional instability or apathy. The late-stage disease is also accompanied by a wide range of motor symptoms which may include, but are not limited to, akinesia and involuntary movements including tremors, fasciculation, cerebellar ataxia and muscular hyper- or hypotonia [3]. Other symptoms that may occur encompass headaches, myelopathy, and pathological changes in cranial nerve pairs leading to, e.g. ocular alterations such as optic neuritis and diplopia [18]. Clinical manifestations of both stages of the disease have been extensively summarised and reviewed elsewhere [18, 19].

The pathology of sleeping sickness varies with the trypanosome species causing the disease [3]. Based on the distinct geographical localisation of the two human-infective subspecies, HAT is differentiated into East and West African Trypanosomiasis, caused by *T. b. rhodesiense* and *gambiense,* respectively [6]. Despite constituting only a small fraction of reported trypanosome infections (3-5%), the manifestation of the Eastern form of the disease is acute and progression to the second stage occurs within weeks, usually accompanied by severe clinical manifestations [18, 20]. The predominant

Western form of the disease on the contrary is chronic and progresses more slowly, averaging an estimated mean duration of 3 years [21]. Although progression of the disease can be clearly distinguished between the Eastern and Western form, the specific phenotype of sleeping sickness may vary even among infections with the same clinical variant in different regions of the same country [22]. Moreover, past studies have suggested that neurological symptoms, typically associated with late stage sleeping sickness, can also occur during early-stage HAT, challenging the traditional disease staging, diagnosis, and subsequent treatment [23].

I.1.4 Treatment

Although *T. brucei* has been identified as the causative agent of the African sleeping sickness already at the turn of the 20th century, only a handful of efficacious drugs have been approved for treatment of the disease since. As a result of a severe lack in development of novel treatments for HAT, anti-trypanosomal drugs administered to this day are accompanied by a wide range of disadvantages, including toxicity of the drugs, lengthy and complicated administration plans, inefficient drug delivery, poor bioavailability, and emerging drug resistances [24]. Due to the lack of an 'umbrella drug', capable of effectively treating HAT regardless of disease stage or clinical variant, identification of the correct treatment necessitates an accurate diagnosis [25]. As current challenges in HAT treatment and recent advances have been extensively reviewed in recent literature [24, 26, 27], the following chapter aims to provide only a brief overview.

The haemolymphatic stage of sleeping sickness is treated with suramin and pentamidine for infection with *T. b. rhodesiense* and *T. b. gambiense*, respectively. While Suramin effectively treats the early stage of the disease caused by both clinical variants, it is used only for treatment of rHAT due to its high toxicity and adverse side-effects including renal dysfunction, peripheral neuropathy, anaphylactic reactions, and bone marrow toxicity [24]. Additionally, suramin has to be administered intravenously (IV) in a lengthy and complex treatment schedule [28]. Infection with *T. b. gambiense* is treated by intramuscular injection (IM) of pentamidine instead [24]. Although accompanied with a range of reported side effects such as hypotension, abnormalities

of glucose metabolism, renal dysfunction, and gastro-intestinal symptoms, it is generally well tolerated and therefore the preferred treatment for Western HAT [3].

Melarsorpol, an organic arsenical, is used in treatment of the encephalitic stage of both, Eastern- and Western HAT [24]. Intravenous administration of the drug is painful and associated with extreme toxicity, resulting in a 5-9% mortality rate caused by post-treatment reactive encephalopathy (PTRE) [24]. Other side effects include agranulocytosis, skin rashes, peripheral neuropathy, and cardiotoxicity [24].

Due to its significantly reduced toxicity in comparison to melarsorpol, eflornithine, or α-difluoromethylornithine (DFMO), has become the predominant treatment for late-stage *T. b. gambiense* infection [24]. While DFMO monotherapy is effective, by 2010 it had been largely replaced by nifurtimox–eflornithine combinatorial therapy (NECT) [26]. Co-administration of nifurtimox, an independent drug used for the treatment of American trypanosomiasis (Chagas disease) [26], decreases overall toxicity and reduces the number of required infusions from 56 to 14 while retaining equal or improved efficacy over DMFO monotherapy [28].

A century after the discovery of the first anti-trypanosomal drugs, the first orally administered drug for treatment of *T. b. gambiense* infection was developed [24]. Since 2019, fexinidazole is the WHO's recommended first choice treatment for early and non-severe late-stage infections, eliminating the need for systematic lumbar punctures for disease staging and injection-based treatments for a subset of patients [29]. As an oral treatment, administration of fexinidazole is significantly less labour intensive than IV or IM infusions, does not require systematic hospitalisation and reduces costs and logistical challenges for the health care system.

Although the introduction of fexinidazole is likely to aid in achieving the WHO's goal of eliminating sleeping sickness by the year 2030, some critical challenges remain. While several emerging drug resistances have been identified in trypanosomes [30], their incidence in endemic areas and consequently the risk posed by drug resistant strains has not been elucidated yet. Regardless, reports of fexinidazole and nifurtimox cross-resistance are cause for concern as the efficacy of NECT in fexinidazole resistant strains may be diminished [31]. A lack of both active and passive screening campaigns in remote areas of endemic countries as well as a growing body of evidence for the existence of asymptomatic carriers [18] suggest that wide-spread prevention of infection rather than treatment of HAT may be key to successful eradication of the sleeping sickness. Despite significant efforts, no vaccine against HAT is available to

this date [32, 33]. Elaborate immune evasion mechanisms employed by the parasite (*see* I.3.3) have stalled vaccine development, posing an obstacle that has yet to be overcome.

I.2 The African trypanosome

I.2.1 Structure and morphology of *Trypanosoma brucei*

Throughout their lifecycle, *Trypanosoma brucei* spp. pass through a wide range of different environments across both mammalian and insect hosts, each posing a new challenge to parasite survival. To overcome these environmental changes, different developmental stages are accompanied by distinct morphological adaptations (*see* I.2.2). Certain key features, however, are conserved throughout the entire lifecycle (Figure 2).
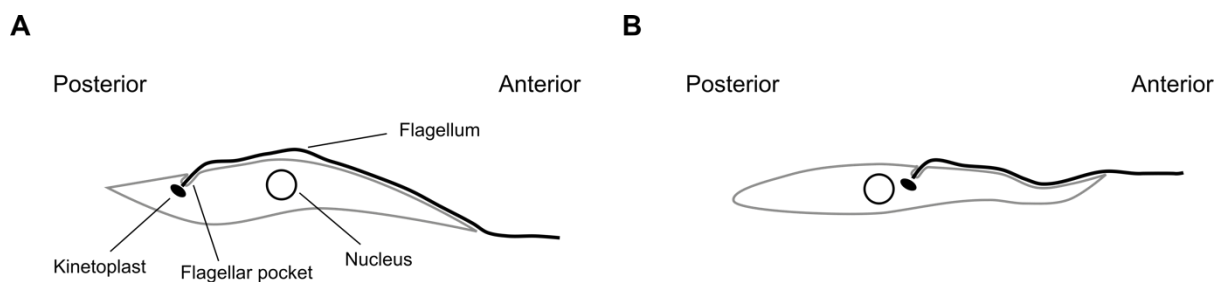
Figure 2 – Schematic illustration of a trypanosome cell.
A Schematic drawing of a bloodstream (slender) trypomastigote. Key organelles conserved throughout the lifecycle, namely flagellum, kinetoplast, flagellar pocket and the nucleus, are illustrated and labelled accordingly. **B** Schematic drawing of an epimastigote stage trypanosome as found in the salivary gland of the insect vector. The same organelles as in panel A are illustrated. Contrary to the trypomastigote stage, the kinetoplast is located anterior to the nucleus in the epimastigote stage. This figure was adapted from [34].

The arguably most prominent feature of the trypanosome cell is its singular flagellum, granting the parasite motility [35]. The flagellum emerges from the flagellar pocket, an invagination of the parasite's membrane, and is attached along the entire length of the cell [35]. The flagellar pocket is the only site for endo- and exocytosis on the parasite membrane, playing a crucial role in acquisition of nutrients and immune evasion mechanisms [36] (*see* I.3.3 & I.4). Localised adjacent to the flagellar pocket and connected to the basal body of the flagellum via the tripartite attachment complex (TAC) is the diagnostic feature of the kinetoplastida order, the kinetoplast [37]. The

kinetoplast is a specialised region of the mitochondrion, harbouring the kinetoplast DNA (kDNA), a complex, disk-shaped mass of circularly arranged mitochondrial DNA, catenated into a single network [38]. The relative position of the kinetoplast to the parasites nucleus varies throughout the parasite's life cycle, imposing distinct morphologies (Figure 2) [34]. In the bloodstream stage, the mitochondrion adapts a smooth, tubular configuration devoid of cristae, indicating the lack of mitochondrial respiration [38]. Indeed, in the glucose-rich environment of the hosts bloodstream, the parasites depend entirely on glycolysis as their sole source of energy [39]. By confining glycolysis into specialised, peroxisome-related organelles called glycosomes [39], *T. brucei* achieve glycolysis rates 50-fold higher than mammalian cells [40].

I.2.2 Life cycle

Throughout their development, *Trypanosoma brucei* undergo various significant changes in their biology, including gene expression, metabolism, and cell morphology (Figure 3). At the peak of parasitaemia within the mammalian hosts, the typically slender bloodstream form of the parasite undergoes a transformation to the cell-cycle arrested, stumpy trypomastigote morphology, conducive to tsetse-fly infection (Figure 3). Upon ingestion by the insect vector during the bloodmeal on an infected host, the parasite undergoes differentiation in the midgut of the fly, giving rise to proliferating procyclic trypomastigotes (Figure 3) [41]. A subset of these procyclic cells adopts an elongated, non-proliferative morphology, denoted as mesocyclic trypomastigotes, initiating migration toward the anterior region of the midgut (Figure 3) [41]. Upon reaching the proventriculus (cardia), these cells become slimmer and transform into epimastigote morphology (Figure 3) [41], characterised by the relocation of the nucleus to the posterior side of the kinetoplast (Figure 2) [34]. From the cardia, the epimastigotes migrate to the salivary glands via the foregut of the fly, engaging in asymmetric division and generation of short epimastigotes within the salivary gland lumen (Figure 3) [41]. The short epimastigotes subsequently affix themselves to the salivary gland epithelium via the flagellum, where they proceed to proliferate [41]. Although two distinct modes of proliferation occur simultaneously, the prevailing mode is determined by the state of advancement of the infection [42]. During the early stage of infection, the epimastigotes undergo symmetric division, producing two equivalent

cells attached to the salivary gland epithelium, thereby precipitating a rapid amplification of the parasite population [42]. In later stages of infection asymmetric division ensues, producing a daughter cell that matures into the human infective metacyclic trypomastigote form which is subsequently released into the saliva [42]. When an infected fly is taking a bloodmeal, the metacyclic trypomastigotes from the saliva are deposited into the hosts dermis (Figure 3) [17]. The parasites travel to the afferent lymphatic vessels from where they are subsequently transported into the draining lymph nodes [17]. Once the parasite reaches the bloodstream, it differentiates into the long and slender bloodstream form and proliferates by binary fission [38]. As parasitaemia increases, a yet to be determined parasite-derived factor, denominated stumpy induction factor (SIF), accumulates [43]. In a form of quorum sensing[44], large amounts of SIF induce cell-cycle arrest and morphological transformation into the stumpy trypomastigote morphology at the peak of the parasitaemia wave, preparing the parasite for uptake by the insect vector and thus completing the lifecycle [44].
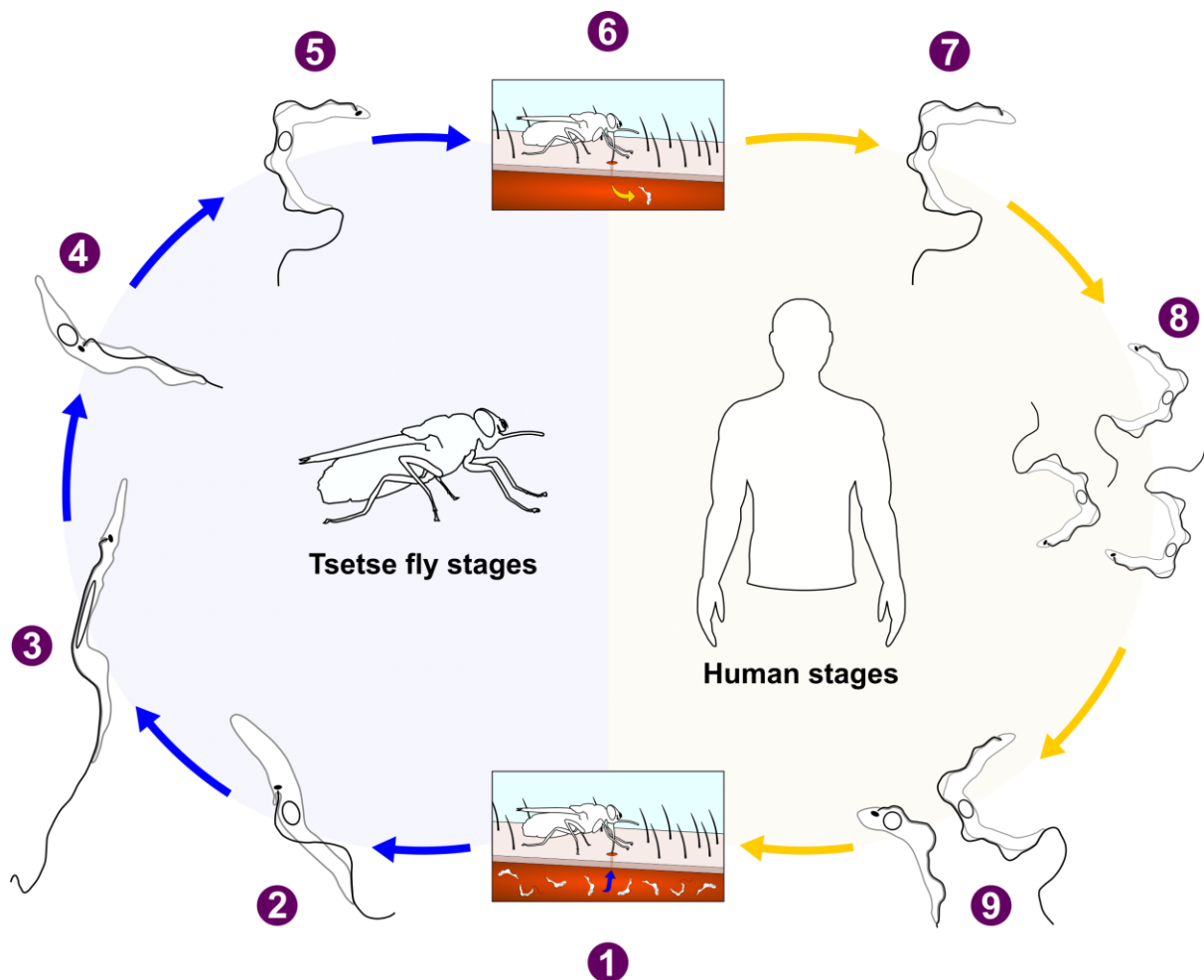
Figure 3 – Lifecycle of human infective African trypanosomes.

When a tsetse fly takes a bloodmeal on an infected mammalian host, it ingests stumpy trypomastigotes (1). In the midgut of the fly, the parasites differentiate into proliferating procyclic trypomastigotes (2). A subset of this cell population develops into mesocyclic trypomastigotes (3) and initiate migration to the anterior midgut. In the cardia, the trypomastigotes differentiate into the epimastigote morphology (4). The epimastigotes migrate to the salivary gland, where after successful establishment of infection, asymmetric division produces human infective metacyclic trypomastigotes (5), which are released into the saliva. When bitten by an infected fly, metacyclic trypomastigotes are injected into the mammalian host (6). Once the parasites reach the bloodstream, they differentiate into the long, slender BSF (7) and proliferate via binary fission (8). At the peak of parasitaemia, accumulation of parasite-derived SIF induces cell-cycle arrest and morphological transformation into stumpy trypomastigotes (9), completing the cell cyle. This diagram was adapted from [45]. Illustration of trypanosome morphology was adapted from [41].

I.3   Immune evasion mechanisms

As exclusively extracellular parasites [3], African trypanosomes continuously face the full pressure of the host's immune system throughout their entire lifecycle, in both their vertebrate and invertebrate hosts. Consequently, they had to evolve a wide range of sophisticated immune evasion mechanisms to ensure survival and propagation. As the research presented in this work has evolved primarily around the immune-evasion mechanisms of the bloodstream form of *T. b. gambiense*, the following chapter will focus only on the human immune system, it's response when challenged with *T. brucei* infection and the corresponding immune evasion strategies employed by the parasite. The interplay between the immune response across different hosts and other tissues as well as the parasites tissue specific survival strategies has been reviewed extensively elsewhere [17, 46-50].

I.3.1 Serum resistance

Humans are inherently resistant to infection by *Trypanosoma congolense*, *vivax*, *evansi* and the *Trypanosoma brucei brucei* subspecies. This intrinsic protection mechanism is facilitated by two serum complexes, termed trypanolytic factor (TLF) 1 and TLF2 [51]. TLF1 is a high-density lipoprotein (HDL) particle found in the HDL3 fraction [52], whereas TLF2 is an IgM-rich, lipid-poor (<2%) apolipoprotein A1 complex [53]. Both TLF1 and TLF2 owe their toxicity to the action of apolipoprotein L1 (ApoL1) [54]. Following internalisation, the protein complexes are shuffled to the parasite's lysosome for degradation, where ApoL1 inserts into the lysosomal membrane and forms a pore [55]. The resulting disruption in the ionic balance of the cell leads to lysis and subsequent cell death [55]. While this model has been widely accepted [56], alternative mechanisms based on ApoL1 pore formation at the plasma membrane have been suggested [57, 58].

TLF1 uptake into the parasite is mediated by the *Trypanosoma brucei* haptoglobin (Hp) - haemoglobin (Hb) receptor (TbHpHbR) [59]. The HDL particles contain a close haptoglobin homologue, the Hp-related protein (Hpr), which binds haemoglobin and forms the Hpr-Hb complex [60]. Binding of this complex to the TbHpHbR facilitates endocytosis, targeting the TLF1 particle to the lysosome where ApoL1 is released [60].

Although Hpr is also found in TLF2, studies have shown that uptake of TLF2 occurs independently of the TbHpHbR [61]. The molecular mechanism underlying TLF2 uptake has not been elucidated to this date [62]. This inherent resistance mechanism does not confer protection against *T. b. gambiense* and *T. b. rhodesiense.*

I.3.1.1 Serum resistance of *T. b. rhodesiense*

Resistance of *T. b. rhodesiense* to lysis by normal human serum (NHS) is linked to expression of the 'serum resistance associated' (SRA) gene [63]. Genes encoding the variant surface glycoproteins (VSGs), which constitute the vast majority of the protective protein surface coat in bloodstream form parasites (*see* I.3.3), are located in telomeric loci called VSG expression sites (ESs) [64]. The *T. brucei* spp. genomes contain 15 to 20 highly similar, although not identical, ESs [64, 65]. In addition to the genes encoding VSGs, 12 additional ES associated genes (ESAGs) are found in the VSG expression sites [66]. Not every ESAG is present across all ESs, the SRA gene, for example, is found exclusively in a specific ES known as the R-ES [67]. When metacyclic trypomastigotes differentiate into the long, slender BSF, different VSG ESs may be activated. Only cells transcribing the R-ES are resistant to lysis by human serum and thus will survive in the bloodstream [67]. In absence of this selection pressure, *T. b. rhodesiense* resistance to lysis by NHS is quickly lost in non-human hosts due to transcriptional switching between the R-ES and previously silenced ESs [67, 68]. Structurally, SRA resembles a truncated VSG, lacking the characteristic surface loops of the N-terminal head domain [69]. Contrary to its evolutionary ancestors, SRA is not targeted to the plasma membrane but is exclusively located in the endosome [69]. Inhibition of cell lysis was initially attributed to binding of ApoL1 to SRA via a coiled α-helix interaction, thought to preclude pore formation in the lysosome [54, 70]. Contrary to the previously proposed hydrophobic interaction, a structural study subsequently demonstrated that the interaction between the proteins is of electrostatic nature instead [69]. The discovery of hitherto unidentified protein contact sites indicated a more complex binding mechanism than previously anticipated, the molecular details of which remain to be elucidated [69].

I.3.1.2 Serum resistance of *T. b. gambiense*

Resistance to TLF mediated lysis of *T. b. gambiense* is conferred by expression of the *T. b. gambiense*-specific glycoprotein (TgsGP) [71]. TgsGP, although unrelated to SRA, is also a modified VSG found in the endocytic compartments [71]. Binding of TgsGP to ApoL1 via a hydrophobic beta-sheet inhibits TLF toxicity and was proposed to induce stiffening of membranes upon interaction with lipids [71]. It has been postulated that additional factors such as elevated cathepsin activity, due to a decrease in endosomal pH, and activity reducing mutations in the TbHpHbR could facilitate *T. b. gambiense* resistance to lysis by NHS further [71].

I.3.2 Complement

I.3.2.1 The complement system – an overview

The complement system is part of the innate immune system and serves as the first line of defence against invading pathogens in the human bloodstream [72]. It comprises a series of proteolytic events, resulting in multi-component enzyme complexes called convertases. The complement system can be divided into three 'early' pathways, the classical (CP), lectin (LP) and alternative pathway (AP), which all converge in the 'late' or terminal pathway (TP) (Figure 4) [72].

The classical pathway is initiated by binding of complement component C1, either to the pathogen surface directly or to surfaces opsonised by immunoglobulin (Ig) G or IgM [72, 73]. C1 is composed of a single molecule of serum protein C1q and two molecules each of C1r and C1s [73]. Binding of C1q to a pathogen surface or an antibody-antigen complex leads to proteolytic activation of C1s, generating an active serine protease [72, 73]. Activated C1s subsequently cleaves complement component C4 to C4b, thereby revealing an active thioester which allows covalent attachment to the pathogen surface [73]. Immobilised C4b then binds complement C2, rendering it susceptible to cleavage by C1s into C2a, an active serine protease in itself [73]. The complex of membrane bound C4b and activated C2a forms the C3 convertase of the classical (and lectin) pathway, C4bC2a (Figure 4) [72]. Although the lectin pathway, also referred to as the mannan-binding lectin (MBL) pathway, leads to the formation

of a C3 convertase identical to the classical pathway convertase, it is initiated by the binding of MBL to mannose-containing carbohydrates on pathogens [73]. Analogous to the C1 complex, MBL is associated with two proteases highly homologous to C1r and C1s, the mannan-binding lectin serin protease (MASP) 1 and MASP-2 [73]. Following the binding of MBL to the pathogen surface, the MASPs are activated to cleave C4 and C2, producing the same C3 convertase as formed in the classical pathway (Figure 4) [73].

The active C3 convertase C4bC2a proceeds to cleave its substrate, native complement factor C3, to C3b, the central hub and main effector of the complement system [74]. Cleavage of C3 to C3b exposes a reactive thioester, facilitating covalent attachment of C3b to the pathogen surface in a similar fashion to C4b [73]. The C3a fragment, proteolytically liberated during the activation of C3b, is a potent mediator of inflammation [75]. The newly generated C3b molecules can then associate with C4bC2a, creating the C5 convertase of the CP/LP, C4bC2aC3b (Figure 4) [73]. Cleavage of its substrate complement component C5 to C5b marks the beginning of the terminal pathway and the point of convergence of CP, LP and AP [72].

Unlike the CP and LP, the alternative pathway is initiated in absence of pathogen binding in a process termed 'tickover' [76, 77]. C3, at approximately 1.2 mg/mL the most abundant of the serum proteins [78], can undergo spontaneous hydrolysis to $C3(H_2O)$, inducing a conformational change believed to closely resemble the structural changes caused by proteolytic activation to C3b [79]. Binding of $C3(H_2O)$ by Factor B (fB) induces a conformational change in the latter, enabling proteolytic activation by the serine protease Factor D (fD) [72, 73]. The cleavage product of fB, an active serine protease denominated Bb, binds to $C3(H_2O)$, forming the so-called fluid phase C3 convertase, $C3(H_2O)Bb$ (Figure 4) [73]. Function and substrate of $C3(H_2O)Bb$ are identical to the CP/LP C3 convertase, converting large amounts of C3 to C3b. While most of the activated C3b will be hydrolysed and thus rendered inactive, a fraction will covalently attach to nearby cell surfaces of both, foreign and host cells (Figure 4) [72]. To prevent progression of the complement cascade and subsequent lysis of host cells, humans possess a range of soluble and cell-surface associated regulators of the complement. The membrane bound decay-accelerating factor (DAF) competitively inhibits fB binding to C3b and can also displace Bb from previously activated convertases [73]. The soluble plasma protein Factor H (fH) fulfils a regulatory role identical to DAF [73, 80]. Additionally, fH serves as a cofactor to the plasma protease

Factor I (fI) which prevents convertase formation by cleavage of membrane bound C3b to inactive iC3b [73, 81]. Factor H has a high affinity for sialic acid, thereby preferentially binding to and inhibiting convertase formation on vertebrate cells [73]. Other fI cofactors include the complement receptor 1, CD46 and the C4b-binding protein cells [73]. In absence of these negative regulators, C3(H_2O)Bb generated C3b will covalently attach to pathogen surfaces, where in a manner identical to fluid phase convertase formation, binding of fB will result in formation of the active AP C3 convertase, C3bBb (Figure 4) cells [73]. In some cases, covalent attachment of C3b to pathogen surfaces can be facilitated further by the soluble plasma protein properdin, also referred to as Factor P (fP), the only positive regulator of the complement (Figure 4) [82]. Properdin can act as an initiator, recruiting C3b to the cell surface, or exert a stabilising function on the AP C3 convertase, increasing the convertase half-life by 5 to 10-fold [83].

The activation of the complement system is greatly amplified through the AP, regardless of which pathway was initiated originally [73]. Both CP/LP and AP C3 convertases generate novel C3b, which, in turn, forms more AP C3 convertases. Newly generated C3b also binds to either of the two C3 convertases, thereby forming the C5 convertases C4bC2aC3b and C3bBbC3b, respectively. C5 convertases cleave C5 to C5b, exposing a binding site for complement C6. The liberated C5a fragment, like C3a, functions as an inflammatory regulator [75]. This step marks the beginning of the terminal pathway (Figure 4) [73]. C5bC6 reversibly binds to the pathogen surface, providing the base for the formation of the membrane attack complex (MAC). Association of C7 to the complex induces a conformational change, exposing a hydrophobic domain of C7 which then inserts into the pathogen's lipid bilayer [73]. C8 is recruited and inserted into the membrane in a similar fashion. Lastly, C8 induces polymerisation and membrane insertion of 10 to 16 copies of C9, completing formation of the MAC (also known as terminal pathway complete complex (TPCC)) [73]. The complex forms a large pore, up to 100 Å in diameter, thereby disrupting cellular homeostasis, ultimately resulting in cell lysis and death [84].
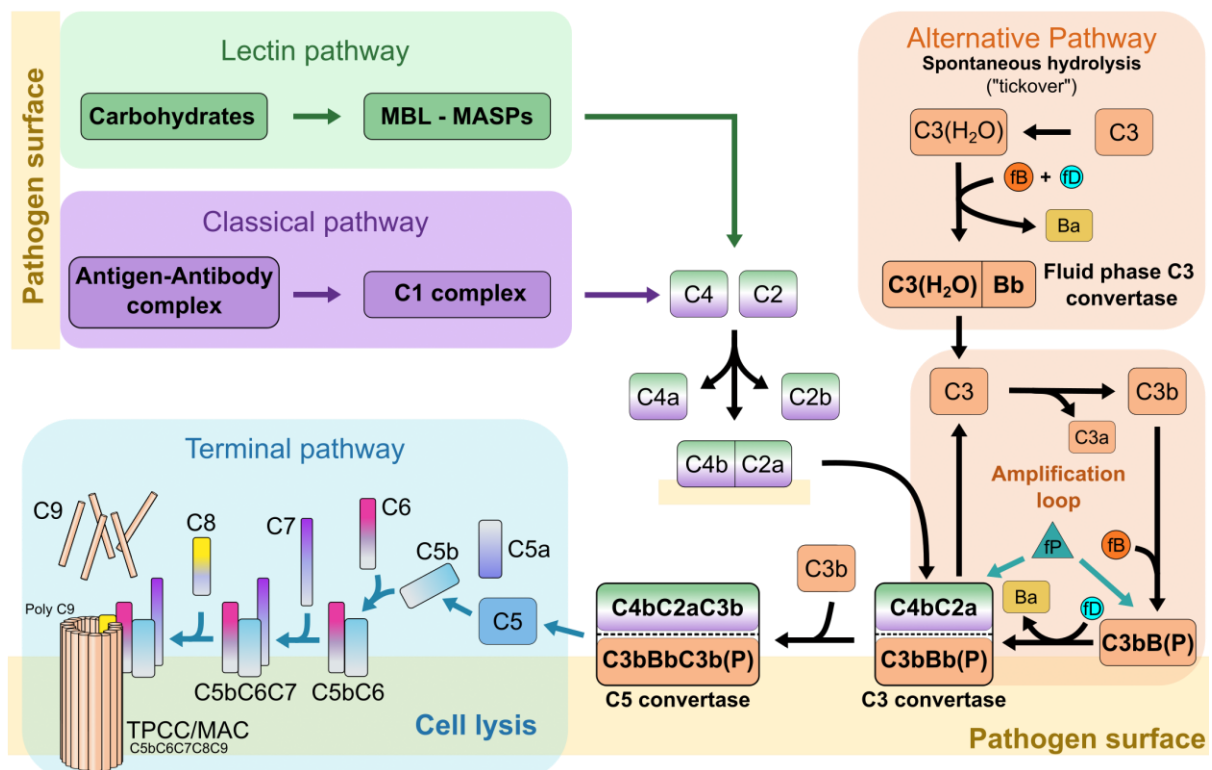
Figure 4 – Schematic overview of the three complement pathways and their components.
The LP and CP are initiated by recognition of specific pathogen surface antigens, subsequently resulting in proteolytic cleavage of C2 and C4, which form the CP/LP C3 convertase, C4bC2a. In addition to serving as an amplification loop of the CP/LP, the AP can be initiated by spontaneous hydrolysis ('tickover') of C3, resulting in formation of the fluid-phase C3 convertase. Analogously to CP/LP C3 convertase formation, covalent attachment of C3b to the pathogen surface forms the basis of the AP C3 convertase (C3bBb). Although structurally distinct, all C3 convertases cleave C3 to C3b, which in turn can bind to the convertases, forming the C5 convertases C4bC2aC3b and C3bBbC3b, respectively. Properdin may stabilise the AP C3 and C5 convertases. C5 cleavage by the C5 convertases initiates the terminal pathway, leading to sequential binding of C6, C7, C8 and poly-C9, triggering insertion into the pathogen membrane where it forms a lytic pore called the membrane attack complex. MBL *Mannose-binding lectin*, MASP *MBL-associated serine protease,* fB *factor B*, Ba *factor B fragment a*, Bb *factor B fragment b*, fD *factor D,* fP *factor P (Properdin),* TPCC *terminal pathway complete complex,* MAC *membrane attack complex*. This figure was reprinted from [85] with permission from the authors.

I.3.2.2 *T. brucei* evasion of complement mediated lysis

As blood-dwelling parasites, trypanosomes had to evolve a mechanism to either avoid activation of the complement system or prevent subsequent lysis. In 1983, A. Ferrante and A. C. Allison demonstrated that *T. congolense* TC35U, a strain devoid of the protein surface coat characteristic for BSF trypanosomes (*see* I.3.3), specifically activates the AP and triggers complement mediated cell lysis in presence of normal human serum [86]. Trypanolytic activity of the serum could not

be observed for the *T. congolense* parent strain which possesses a fully developed surface coat [86]. Similarly, *T. b. brucei* BSF parasites were protected from complement mediated lysis whereas procyclic trypanosomes were lysed as a result of AP activation [86]. These findings initially suggested that presence of the protective VSG coat limits serum accessibility to activators of the complement system on the parasites surface. By demonstrating deposition of complement component C3 on the surface of *T. b. gambiense* via radioimmunoassays in human serum with selectively inhibited complement pathways, Devine et al. confirmed specific activation of the AP [87]. This finding was further corroborated by the absence of any detectable IgG on the parasites surface, indicating that the CP is not initiated [87]. While the authors were able to detect factor B on the cell surface, they could not detect the presence of complement component C5 [87]. In synthesis, these findings indicated that rather than preventing the activation of the complement system altogether, the AP is indeed activated but does not progress beyond the formation of the AP C3 convertase [87]. Although this pointed to a dedicated mechanism of complement inactivation on the surface of trypanosomes, no receptor or molecular mechanism was identified at the time. Almost 40 years later, ISG65 was simultaneously identified as the trypanosome surface receptor for human complement C3 in *T.* brucei by multiple, independent research groups (*see* 4.2) [85, 88].

I.3.3 Variant surface glycoproteins

After transmission from the insect vector to the mammalian host and differentiation into the slender BSF forms, *T. brucei* spp. start to express a dense surface protein coat comprised of 5 to 10 million copies (~10% of the cell's total protein) of variant surface glycoproteins (VSGs), covering the entire cell surface [89]. The VSG coat plays a pivotal role in the immune evasion of the parasite by (i) creating a physical barrier against complement mediated attacks [86], (ii) mediating rapid antibody clearance from the parasite surface via VSG recycling [90], (iii) providing an immunomodulatory decoy on the surface [91] or when shed [48] and (iv) constituting a mechanism for escaping the host's primary humoral response, ensuring long-term survival [92]. When metacyclic trypanosomes are first transmitted during the bloodmeal of the insect vector, they are promptly faced with the full pressure of the

hosts immune system. To ensure their initial survival, the cells possess a surface protein coat comprised of metacyclic VSGs (mVSG) [91]. Although structurally highly similar to the mature BSF VSGs, the number of mVSG encoding genes is severely limited [91]. To avoid elimination as the surface recognition by the host improves, the parasite needs to prevent exhaustion of this repertoire. Differentiation into the long slender BSF stage provides access to the much wider range of bloodstream form VSGs [91].

Most VSGs are homodimeric and contain three functional units: the signal peptide for cell surface targeting (cleaved off after translocation), a large N-terminal domain (NTD) and a smaller C-terminal domain (CTD) [93]. The mature VSGs are approximately 100 to 120 kDa in size with a membrane distal NTD spanning 300-400 residues and the CTD, connected to the NTD via an unstructured linker region, spanning 80 to 120 amino acids [93]. Each VSG protomer is attached to the cell surface via a glycosylphosphatidylinositol (GPI) anchor in the CTD [93]. The NTD is based on an approximately 100 Å long, three-helix bundle scaffold, oriented perpendicular to the parasite membrane, with globular top and bottom lobes at opposite ends of the coiled-coil [93]. The CTD is comprised of either one or two small structured regions with a compact fold, featuring two short α-helices, a short anti-parallel β-sheet, and an optional C-terminal α-helix (Figure 5) [93].
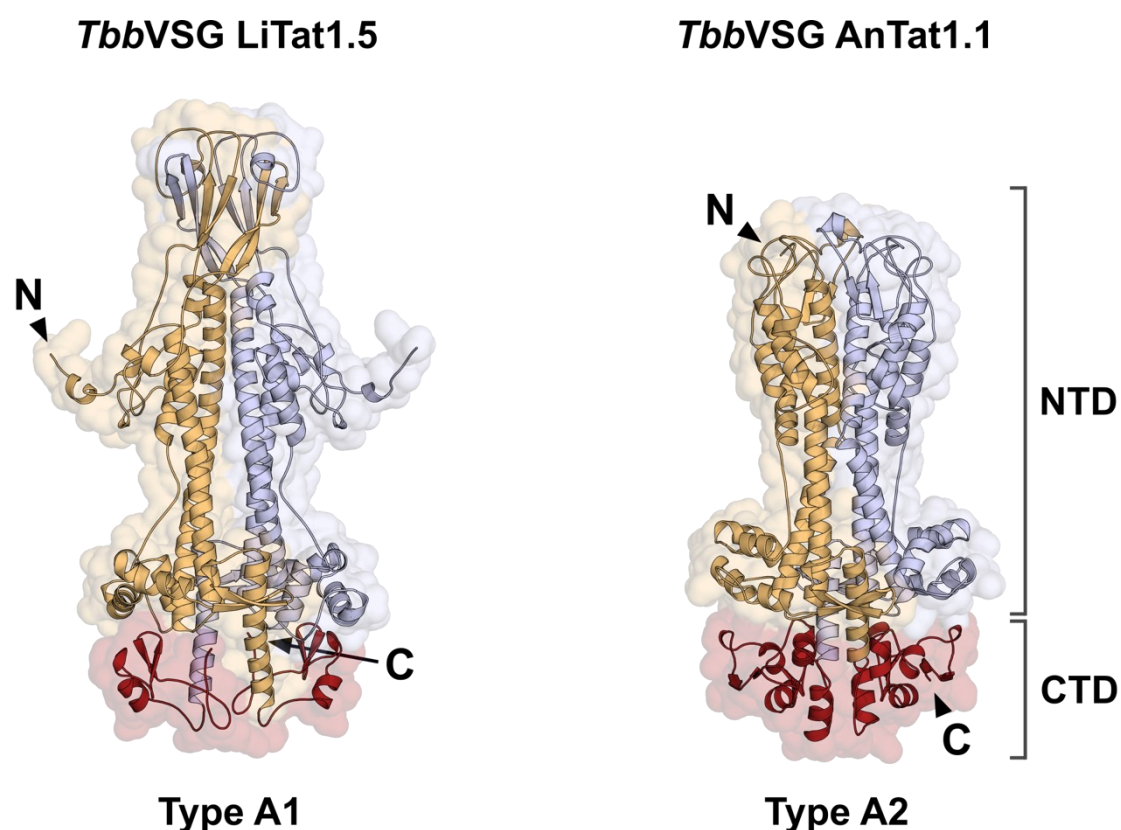
Figure 5 – AlphaFold2 models of type A *T. b. brucei* VSGs.
AlphaFold2 models of the mature homodimers of *T. brucei brucei* VSGs LiTat1.5 (left) and AnTat1.1 (right). The N-terminal domains of individual protomers are coloured in light blue and yellow, the C-terminal domains of both protomers are coloured in red. The position of the N- and C-termini of the same protomer have been marked. This representation was adapted from [94].

Historically, NTDs of VSGs have been categorised into two classes (type A and B) based on the number and position of conserved cysteine residues [93]. However, with an increase in experimental models and recent advances in protein structure prediction, a novel, structure-based classification [93] has been suggested instead: Two structural 'superfamilies' of VSGs are distinguished by the configuration of the NTD bottom lobe, class A and class B. Class A VSGs form homodimers while class B can transition from monomeric to trimeric arrangements in a concentration dependent manner, *in vitro* [95]. Based on the assumed arrangement of VSGs on the parasites surface, it has been postulated that the trimeric arrangement is also found *in vivo* [93], however, this has yet to be confirmed experimentally. Additionally, the top lobe of class B NTDs is subject to O-linked glycosylation, a post-translational modification associated with an increase in pathogenicity of the parasite [96]. The class A superfamily may be divided further into subclasses A1 and A2 based on the

distribution of disulphide bridges, a large, prominent β-sheet in the top lobe of the longer class A1 VSGs, the relative positions of the N-termini and localisation of N-linked glycans (Figure 5) [93]. Subclass A1 can be further subdivided into three additional groups based on size, conformation and twist of the top-lobe β-sheet and the gap size between the three-helix bundles in the dimer [93]. Although no major structural differences occur between the class B NTDs, minor differences allow further categorisation into two subgroups [93].

During early stages of infection, when specific antibody concentrations are low to moderate, VSG endocytosis and recycling is an effective way to remove surface bound antibodies. Studies have shown that *T. brucei* can endocytose and recycle the entire plasma membrane within approximately 12 minutes [97]. Even though this rapid endocytic turnover is highly efficient, it was demonstrated that trypanosomes can remove a complete layer of surface-bound IgG within 120s, greatly exceeding the reported endocytosis rates [98]. This phenomenon has been attributed to hydrodynamic forces created by the cell's motility. Bound immunoglobulins act as 'molecular sails', inducing drag when the parasites are in motion [98]. The VSG-antibody complexes are therefore preferentially pulled towards the posteriorly located flagellar pocket, where they are rapidly endocytosed [98].

The VSG coat is highly immunogenic, inducing a robust, VSG-specific antibody response by the mammalian host, resulting in efficient opsonisation of the parasite cell [92]. With increasing parasitaemia and antibody titres, rapid endocytosis and recycling of Ig-bound VSGs becomes insufficient to remove the opsonising immunoglobulins [98]. Subsequently, trypanosomes are recognised and killed by the primary humoral response. In an effort to evade these antibodies, trypanosomes employ a mechanism called antigenic variation [99]. By continuously switching expression between different VSGs from a repertoire of over 2000 genes, including pseudo- and mosaic genes formed by gene conversion, antibodies recognising a previous variant are rendered obsolete [99]. This dynamic leads to characteristic waves of parasitaemia in the hosts bloodstream [100]. VSGs are transcribed from one of 15 to 20 specialised, subtelomeric bloodstream expression sites (ES) [65]. These sites exhibit a conserved structural pattern, encompassing an RNA polymerase I promoter [99], a series of expression site-associated genes (ESAG1-13) [101], and a solitary VSG gene situated near the telomeric repeats [99]. To ensure strictly monoallelic expression, only a single ES is active at any given time. Antigenic switching can occur via two distinct

mechanisms: transcriptional switching and homologous recombination [64]. In transcriptional switching, the expression of an active ES is turned off and a previously silenced ES is activated *in situ*. Homologous recombination, on the contrary, requires replacement of the active VSG gene within the same ES. This can occur either via gene conversion or reciprocal exchange. The timing of VSG expression during infection appears to depend on the location of VSG genes: While telomeric VSGs are more likely to be expressed early on, recombinant mosaic variants tend to appear in later stages of infection [102, 103].

In addition to the parasite's evasion of the primary humoral immune response of the mammalian host via antigenic switching, the immune response against *T. brucei* is further complicated by the parasites ability to induce a B-cell depletion pathology, characterised by rapid depletion of immature B-cells from the bone marrow, disappearance of transitional and IgM+ marginal zone (MZ) B-cells from the spleen and subsequent, gradual depletion of follicular B-cells (FoB) [49]. The loss of various B-cell populations consequently inhibits the hosts ability to mount a protective antibody response against newly arising VSG variants and disables the development of an effective B-cell memory against previously encountered serotypes [104]. The molecular mechanisms underlying this pathological B-cell depletion have been associated with VSG induced, non-specific, polyclonal B-cell activation leading to clonal exhaustion, IFN-γ-mediated inflammation, natural killer cell (NK) and direct cell-cell contact-mediated B-cell death [2]. Previous studies with animal models showed that trypanosome infection even resulted in the loss of vaccine induced protective immunity to a number of non-trypanosome related pathogens [104].

I.4 Invariant surface glycoproteins

Although the vast majority of the BSF surface coat is dedicated to immune evasion via antigenic variation (*see* I.3.3), embedded in between the VSGs are the highly expressed, yet poorly understood, members of the protein superfamily of invariant surface glycoproteins (ISGs) [105]. ISGs are exclusively expressed by the BSF of African trypanosomes and constitute the second most abundant surface antigen [106]. They are transcribed from multi-gene arrays containing various closely related paralogues that are almost identical in sequence [107]. Despite some

architectural similarities shared with other characterised surface antigens, ISGs also exhibit a set of distinct features.

In addition to evading detection by the hosts immune system, parasite growth and survival also depend on the uptake of a multitude of abundant host substrates such as glucose [108] or iron [109]. Invariant receptors mediating nutrient uptake, such as the TbHpHbR [59, 110] or transferrin receptor [109, 111], are confined to the flagellar pocket to minimize exposure to the host's immune system [36]. On the contrary, ISGs are distributed across the whole cell surface [112]. Instead of being tethered to the plasma membrane via a GPI-anchor, ISGs are type-I transmembrane proteins with a large, extracellular N-terminal domain and a short, cytoplasmic C-terminus [112]. The C-terminus is subject to ubiquitylation and mediates rapid endocytosis, targeting the protein either for endosomal recycling back to the cell surface or lysosomal degradation [113-115]. Noticeably, ISG turnover is significantly faster than VSG recycling (40-fold difference), resulting in a considerable presence of ISGs in the endosomal system [116]. Collectively, these characteristics suggest a functional role in the immunobiology of BSFs, either as part of the immune defence or as nutrient receptors.

The most abundant ISGs are ISG65 and ISG75 with 70.000 and 50.000 copies, respectively [112]. From the several multi-gene ISG families that have been identified, only these two have been extensively studied as their high abundance implies a functional significance. Their invariant nature and relatively high abundance across the cell surface have made ISG65 and ISG75 prime candidates for vaccination trials [33, 117]. They have been shown to be highly immunogenic and trigger a humoral immune response detectable by presence of antibodies in blood [118, 119]. However, subsequent studies demonstrated that while ISGs initially induce very high ISG-specific antibody titres, vaccinations fail to confer protection in memory-recall experiments [33]. Thirty years after its initial discovery in 1992, ISG65 has been identified as a complement receptor for human complement component C3b (*see* 4.2) [85, 88]. While a recent study has demonstrated that ISG75 binds and subsequently internalises the trypanocidal compound suramin, experimental evidence for its structure and exact biological function remains elusive [106].

Part II – Human adenovirus C5

II.1   Introduction to *Adenoviridae*

II.1.1 Human adenoviruses and their diseases

The *Adenoviridae* family is comprised of 6 genera, totalling 87 distinguished species.   The genus of mammalian-infective mastadenoviruses include the seven human infective species termed A-G, which span over 110 different adenovirus (AdV) types [120, 121]. Conditions and the commonly associated causative human adenovirus (HAdV) species include, but are not limited to, meningoencephalitis (HAdV-A, -B, and -D), cystitis (HAdV-B), hepatitis (HAdV-C), pneumonia (HAdV-B,-C, and -E),   keratoconjunctivitis (HAdV-B and -D) and gastrointestinal infections (HAdV-A, -F, -G) [122]. HAdV infections are usually mild and self-limiting in immunocompetent individuals, however, immunocompromised patients often suffer from severe and recurrent pulmonary HAdV infections as well as disseminated adenoviral disease (DAD), which has been attributed with a high mortality rate [122]. The majority of primary HAdV infections are contracted at a young age, owing to the lack of a developed humoral immunity. The incubation period after transmission ranges from two days to two weeks, depending on virus type and mode of transmission [122]. Infections are commonly contracted through respiratory droplet transmission, direct conjunctival inoculation, or faecal-oral transmission.

II.1.3 Structure of the human adenovirus capsid

HAdVs are non-enveloped, pseudo T= 25 icosahedral viruses with a diameter of approximately 920Å (Figure 6) [123]. Proteins constituting the capsid are divided into three 'major' proteins: hexon, penton base and the penton base-associated fibre; as well as four 'minor' proteins (IIIa, VI, VIII and IX) [124]. The icosahedral shape results in 20 quasi-identical, triangular facets. Each facet's surface is comprised of 20 copies of the trimeric hexon capsomer and 3 penton capsomers, interlaced with 12 copies of minor protein IX (Figure 6) [125]. The remaining minor proteins IIIa, VI and VIII are located on the inner surface of the virus capsid. Their arrangement and

function are beyond the scope of this work and have been reviewed extensively elsewhere [125].
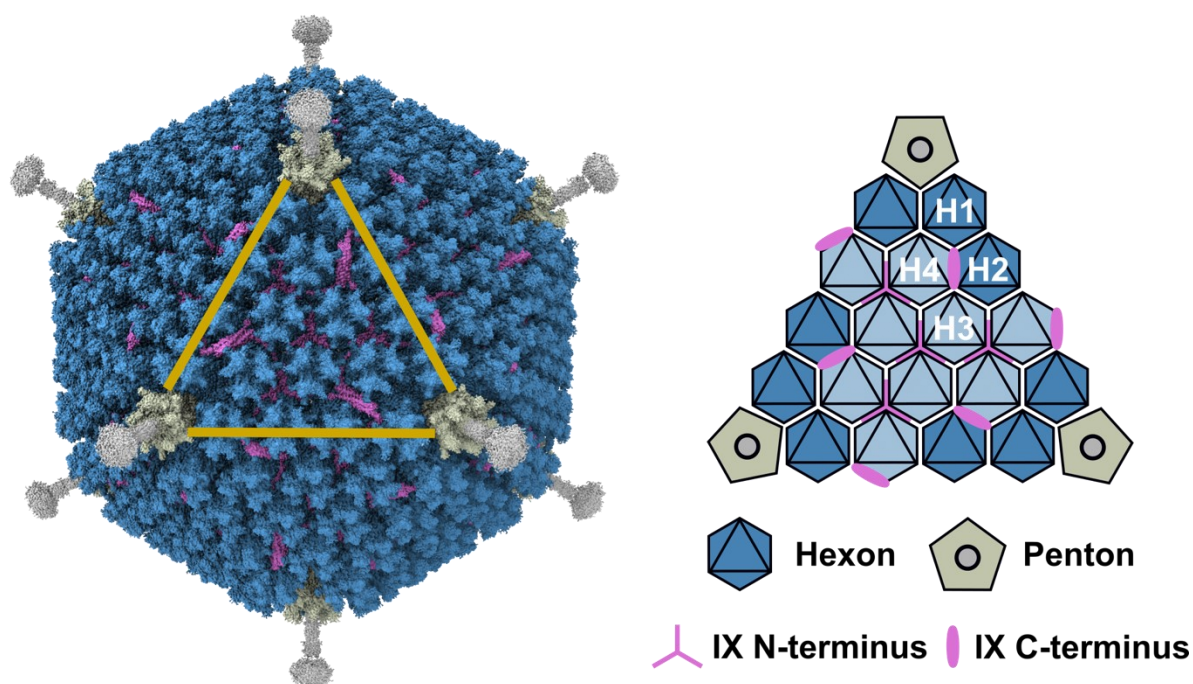

Figure 6 – Structure and organisation of the adenovirus capsid.

Single-particle cryo-electron microscopy (cryo-EM) reconstruction of a complete HAdV-C5 virus (EMD-7034) (left) showing prominent densities for major capsid proteins hexon (blue), penton base (green) and penton associated fibre (grey) as well as minor protein IX (magenta). One of the 20 facets of the icosahedral capsid is outlined in orange. A schematic diagram of the facet illustrates the proteins arrangement within (right). The individual components are not to scale. Hexons constituting the 'group of nines' are coloured in light blue. One example of each of the four types of hexon has been labelled based on its environment within the capsid. Arrangement of the minor proteins on the inside of the capsid is not shown. The schematic illustration of the facet was adapted from [125].

II.1.3.1 Hexon

With 720 monomers constituting 240 trimeric capsomers, hexon is the most prominent capsid protein. As a major structural component, both the interactions among individual hexons as well as the interactions between hexons and the associated minor capsid proteins are integral to capsid assembly and stability [124]. Based on their environment within the capsid, four types of hexon, designated H1 to H4, are differentiated. At the capsid's 12 apices, a total of 60 H1 hexons, also referred to as peripentonal hexons, associate with the penton capsomer bases. The remaining

hexons form so-called 'groups of nine' (GONs), a structurally robust element on each facet of the icosahedron [126]. Hexons located on the two- and threefold symmetry axes are aptly termed H2 and H3, whereas the remaining hexons are designated as H4 (Figure 6) [126].

The assembled trimer forms a pseudo-hexagonal base, created by threefold repetition of two β-barrels at the base of the molecule, a characteristic topology of viral coat proteins termed 'jelly roll' [127, 128]. The capsid distal end of the molecule does not follow the pseudo-hexagonal structure but adapts a triangular shape instead, prominently defined by three protruding 'turrets'. The barrel-shaped hexon capsomer is hollow, forming a large, central cavity (Figure 7). Directly underneath the turrets a so-called beta-constriction forms a 'hydrophobic plug', sealing off the cavity [129].
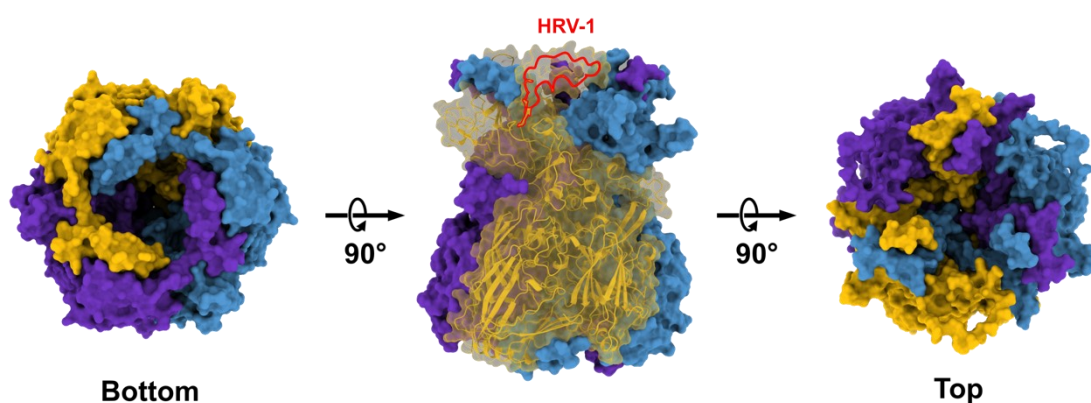

Figure 7 – Structure of the HAdV-C5 hexon.
Surface representation of the trimeric HAdV-C5 hexon capsomer. The three contributing protomers have been coloured in purple, blue and yellow, respectively. Bottom (from the capsid proximal end), side and top (from the capsid distal end) views of the protein are shown (left to right). In the side view, the surface of the yellow-coloured monomer is transparent, showing the underlying model in cartoon representation, prominently featuring the two jellyroll motifs that form the base of the hexon. The HVR-1 loop of the yellow protomer, constituting the turret on top of the hexon, has been outlined in red in the side view. The displayed full-length model is a computationally augmented model based on a HAdV-C5 hexon cryo-EM reconstruction (PDB: 8Q7C). Model and figure have been adapted from [130].

The hexon turrets are created by extended loops intercalated between the β-sheet strands of the N-terminal jelly roll. These loops also house the hyper variable regions (HVRs) [129]. The HVRs vary in length and number between serotypes, HAdV-C5 hexons contain 7 HVRs and feature an atypical abundance of acidic residues in HVR1

[131]. The hexon HVRs are a primary target of neutralising antibodies generated by the host [132]. While pre-existing HAdV-C5 immunity has been shown to decrease immunogenicity of HAdV-C5 based vaccine vectors [133], exchange of the HVRs to those of other serotypes successfully restored immunogenicity [132]. HAdV-C5 HVRs have also been demonstrated to mediate the recruitment of systemic coagulation factor X, facilitating infection and controlling tissue tropism by targeting the virus to the liver [134, 135]. Similarly, it has been shown that species C HAdVs utilise human lactoferrin (hLF) and its antimicrobial cleavage product lactoferricin (Lfcin) for enhanced infection of epithelial cells *(see* II.2.1.3) [136, 137]. Structural insights into the interaction were obtained recently, demonstrating that hLF engages the HVR-1 via the N-terminal Lfcin *(see* 4.4) [130].

II.1.3.2 Penton

The penton capsomer protrudes from each of the 12 vertices of the icosahedron (Figure 6). It comprises a noncovalent complex of two major capsid proteins: the homopentameric penton base and the homotrimeric penton base-associated fibre (from here on also referred to only as 'fibre'). The fibre binds the penton base with high affinity (K_D = 200 nM for HAdV2) and forms a stable complex that is only disrupted by addition of chaotropic agents at elevated temperature [138]. The penton base monomer has an elongated shape, spanning approximately 100 Å across its longest axis. Each monomer is composed of two domains: a capsid proximal domain containing the jellyroll motif and the solvent facing insertion domain [139]. In the pentameric assembly, a ring of jellyroll domains creates a compact base while the insertion domains form a wide top lined with α-helices, creating a complementary shape to the reverse geometry of the hexon (Figure 7, 8). The insertion domain consists of two long insertions intercalated between the jellyroll β-strands. The first insertion, spanning residues 129-436 (in HAdV-C2 and HAdV-C5) contains the glycine and alanine rich, hypervariable RGD loop (res. 301-374). This loop, named after its conserved arginine-glycine-aspartic acid tripeptide sequence motif, exhibits large variability in length and sequence amongst different AdV serotypes [139]. After initial host cell binding of the virion via the penton fibre, the highly conserved RGD motif of

the penton base binds to $\alpha_v\beta_3$ or $\alpha_v\beta_5$ integrins on the cell surface, promoting internalisation into clathrin-coated pits *(see* II.2.1.1) [139].

The penton base-associated fibre can be divided into three distinct regions: the tail, the shaft, and the head (or knob) (Figure 8). Upon assembly of the penton base, a predominantly hydrophobic pore, approximately 15-20Å in diameter, forms on the 5-fold axis of the pentamer [139]. The resulting hydrophobic ring serves as the attachment point for the N-terminal tail of the trimeric fibre [140]. In addition to the hydrophobic interaction, the fibre is also anchored to the penton base via so called 'stay-cables' [140]. The N-terminal tail of the fibre contains a FNPVYPY motif, a sequence highly conserved across adenovirus serotypes. This fibre peptide rests in a groove on the surface of the penton base, formed by two adjacent monomers (Figure 8) [139]. Due to the symmetry mismatch between the trimeric fibre and pentameric base, it has been postulated that the three fibre tails occupy two neighbouring sites and one separated by 144°, as this arrangement would be energetically favourable in comparison to occupation of three adjacent sites [139, 140]. The shaft of the fibre is composed of 15-residues pseudo-repeats, forming three intertwined strands that constitute a rigid and stable shaft (Figure 8). While HAdV-C2 and HAdV-C5 both have 22 pseudo-repeats per strand, the length of the shaft varies amongst other serotypes [140].

Lastly, another ~180 residues form the C-terminal, globular head domain [124]. Each protomer folds into an eight-stranded, anti-parallel β-sandwich. Upon trimerization of the fibre, the knob forms a three-bladed propeller motif (Figure 8) [141]. The head domain of the fibre is responsible for mediating the initial binding to host cell surface receptors, such as the coxsackie and adenovirus receptor (CAR) for serotype C HAdVs [142].
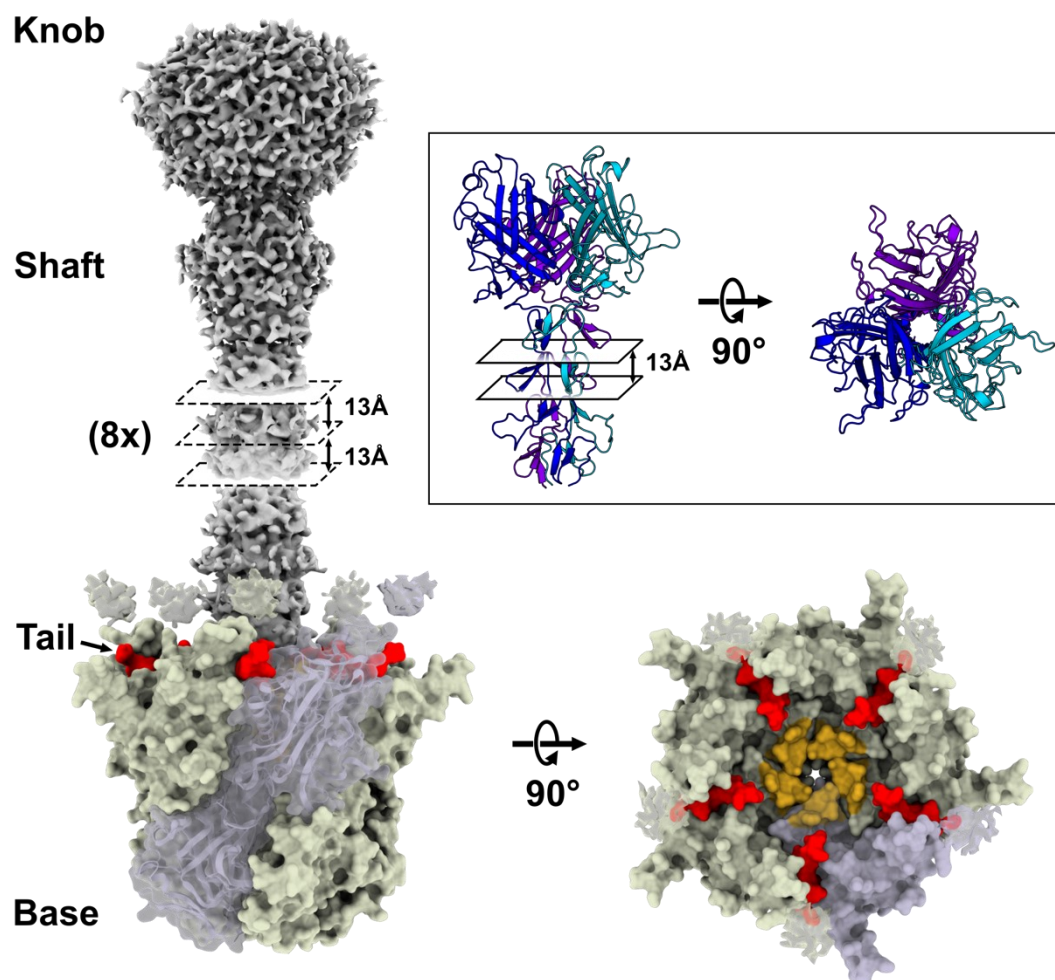
Figure 8 – Illustrative representation of the HAdV penton capsomer.

Due to the flexibility of the penton base-associated fibre, no experimental structure of the complete penton capsomer is available. The illustrative model depicted here has been constructed based on a crystal structure of the HAdV-C2 penton base (PDB 1X9T) (~99% identity to the HAdV-C5 penton base) and a single-particle cryo-EM reconstruction of the HAdV-C5 virus (EMD-7034). Four of five protomers of the penton base are depicted in green surface representation, one protomer is shown in cartoon representation with a semi-transparent surface (light purple). For the penton base-associated fibre, the cryo-EM density is shown (grey). The semi-transparent blobs above the penton base indicate the position of the hypervariable RGD loop. Due to its flexible nature, the precise location of the loop is not resolved in the symmetry averaged cryo-EM reconstruction. Notably, only the first 8 N-terminal pseudo-repeats of the fibre are visible in the cryo-EM reconstruction due to flexibility and masking during data processing. In absence of density for the head domain, the globular density caused by flexibility in the kink of the shaft beyond the visible repeats has been used as an illustrative placeholder (of similar size) for the knob here. The 16 remaining pseudo-repeats missing in the EM reconstruction are representatively illustrated as an 8-fold repetition of a 2-fold pseudo-repeat motif. The individual pseudo-repeats are spaced 13Å apart. In the top view of the penton base (right), the fibre density has been omitted. The hydrophobic ring which serves as the attachment point for the N-terminal tail of the fibre is coloured yellow. The conserved FNPVYPY peptide of the fibres 'stay-cable' (or 'tail') are shown as red volumes in the grooves created by adjacent monomers. Due to the trimeric nature of the fibre, only three of these five illustrated sites are occupied in the assembled capsomer. The inset shows the crystal structure of the HAdV-C2 head with part of the fibre shaft (PDB: 1QIU). The three chains constituting the trimeric

assembly are individually coloured. The 13Å spacing of the pseudo-repeats is shown. The rotated view from the top of the fibre shows the three-bladed propeller motif formed by the fibre knob.

II.1.3.3 Protein IX

With only 14.3 kDa in size, protein IX (pIX) is the smallest of the minor capsid proteins. While strongly conserved amongst the primate mastadenoviruses, it is absent in the other adenovirus genera [125, 143]. On each facet of the icosahedral capsid, twelve copies of protein IX bind nine hexon capsomeres to form the stable GON assemblies (Figure 6) [125]. About halfway down the length of the hexon, protein IX is inlaid into the capsid surface (Figure 6, Figure 9) where it forms a network that lines the hexon boundaries and firmly binds them together [123].

While pIX is not paramount for viral replication, it fundamentally increases thermostability of the virus [143], lending it the commonly used description as a 'cement' protein. Protein IX can be divided into three domains: an N-terminal domain (~50 Å long), a C-terminal helix-bundle domain with a single 12-turn long α-helix (~65 Å) and a rope domain (~70 Å) in between (Figure 9) [123]. The N-terminal domain contains a pair of hydrophobic residues, Y14 and L15, which constitute the so-called 'N-joint' region. The N-termini of three pIX monomers form a hydrophobic core, creating a central trimeric joint (Figure 9). The three contributing protomers radiate from this joint (Figure 6) [123]. The C-termini of pIX form three, ~65 Å long and unusual four helix-bundles per facet [123]. For each bundle, the rope domains of three copies of pIX from within the same facet traverse the length of two hexon edges before merging into the base of the assembled bundle in parallel. The fourth helix however originates from a pIX copy situated in a neighbouring facet. The rope domain traverses the same distance along the hexons, but the helix joins the tip of the bundle and continues to the base in an antiparallel orientation (Figure 9) [123]. While the C-terminal four helix-bundle is essential for the self-association of protein IX, only the N-terminal domain is necessary to increase the thermal stability of the virus [143, 144]. It has been postulated previously that the C-terminal domain may serve as a transcriptional activator of several TATA-containing promoters [144], but the relevance of these findings has been contested [145].
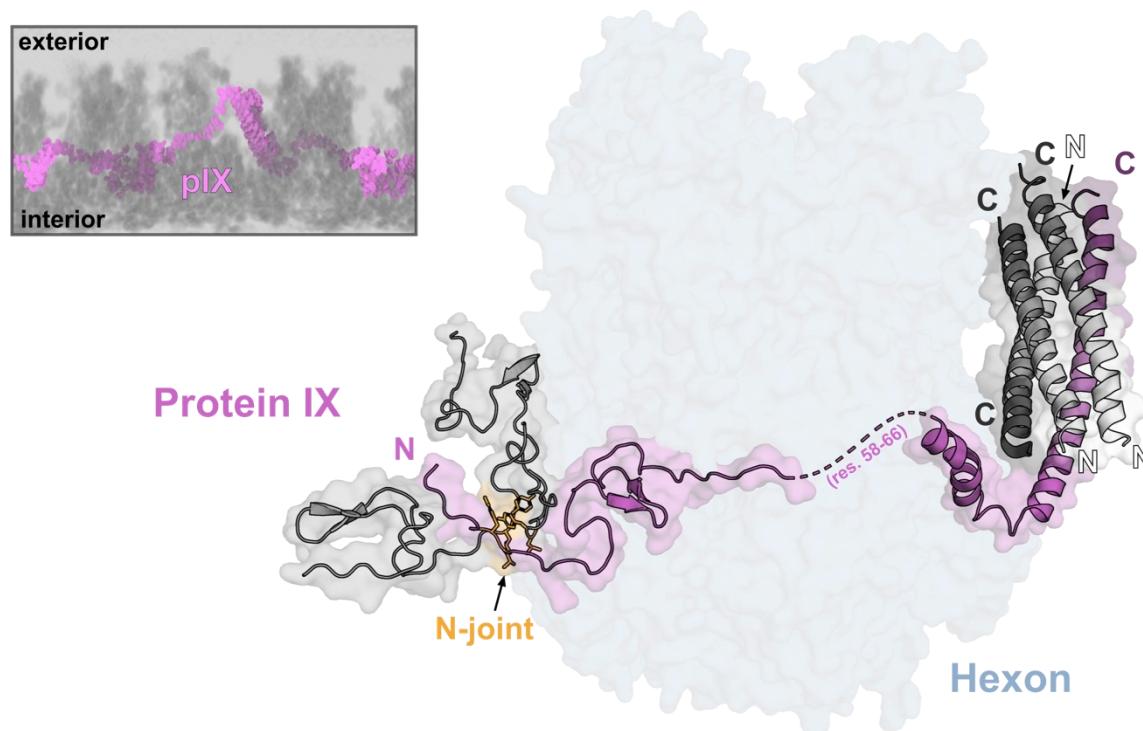
Figure 9 – Structure of HAdV minor capsid protein IX.

The inset depicts a slice through the cryo-EM reconstruction of an HAdV-C5 virion (EMD-24881), showing that pIX, illustrated as magenta spheres, is located on the outside of the viral capsid, buried half-way down the surrounding hexons. The protein IX monomer (PDB: 7S78), shown as magenta coloured cartoon (right), can be divided into three domains: the N-terminal domain forming a trimeric joint (orange) with the N-terminal domains of two additional pIX monomers (grey), the C-terminal four helix-bundle and the connecting 'rope' domain. Residues missing from the atomic model (res. 58-66) are depicted as a dashed line. The three helices of the four helix-bundle provided by other pIX monomers, which are not illustrated here, are coloured using a grayscale ranging from the N- (white) to the C-terminus (dark grey), highlighting the antiparallel orientation of one of the helices. For orientation purposes, the hexon alongside which the pIX monomer is travelling, is shown in a transparent, blue surface representation.

II.2   Life cycle of the human adenovirus C5

Viruses are commonly referred to as intracellular, obligate parasites [146]. As such, completion of their replication cycle, often generalised to 'life cycle' [147], necessitates the entry of host cells where they hijack the replication machinery, propagate, and subsequently exit the cell [148]. Entry of the host cell requires attachment, endocytosis, uncoating and intracellular trafficking to the nucleus. After gene expression and replication of the viral genome, new virions are assembled and released from the host cell [147].

II.2.1 Cell entry

Infection of the host cell starts with binding of the virus to the primary cell surface receptor. While a wide range of molecules has been shown to enable or facilitate HAdV binding, the primary receptor usage varies between different serotypes. The coxsackievirus and adenovirus receptor (CAR) was identified as a HAdV-C5 primary receptor first [149, 150].

II.2.1.1 CAR mediated cell entry

The human coxsackievirus and adenovirus receptor is a 365 amino acid long, glycosylated transmembrane protein and member of the junctional adhesion molecule (JAM) family [142]. The CAR is comprises an intracellular C-terminal domain, a transmembrane helix and two globular, extracellular Ig domains [149]. The N-terminal Ig domain, termed CAR-D1, binds to one of the protomers forming the three-bladed propeller motif in the adenovirus penton base-associated fibre knob [151].
After initial adsorption of the virion to the cell via CAR, secondary interactions with integrins trigger cell entry. Integrins are heterodimeric, single-transmembrane proteins associated with cell signalling [142]. In vertebrates, various combinations of the two extracellular domains, termed α- and β-chains, can produce 24 integrins (from 18 different α- and eight different β-chains) [142]. HAdV binding of $\alpha_V\beta_3$ and $\alpha_V\beta_5$ integrins via the RGD motif of the penton base protein triggers integrin clustering, signalling for endocytosis of the virion [142, 152]. While this model of the HAdV cell entry pathway has been well established in immortalised cell lines, both CAR and integrins are

predominantly buried deep within the tight junctions on the basolateral side of the polarised epithelium, inaccessible to virus particles on the apical side [121, 153]. This apparent contradiction has been partially alleviated by findings demonstrating that invading HAdV-C5 is taken up by resident airway macrophages, triggering a pro-inflammatory response and causing the secretion of interleukin-8 (IL-8) [153]. IL-8 secretion results in increased expression of the apical CAR^{Ex8} isoform as well as apical migration of basolateral $\alpha v \beta_3$ integrin, providing an entry for HAdV-C5 into the polarized epithelium [153, 154]. In addition, a range of CAR independent cell entry pathways of HAdV-C5 have been identified [155, 156].

II.2.1.2 Other cell surface receptors

Human melanoma and breast cancer cells, which express little to no CAR, can be efficiently infected with HAdV-C5 independently of the penton fibre, challenging the idea that HAdV-C5 transduction necessitates binding to a primary receptor before triggering endocytosis via secondary receptors. Very high affinity binding to $\alpha v \beta_5$ integrin is mediated exclusively by the penton base RGD motif instead [157]. Furthermore, HAdV-C5 has been shown to utilise two class A scavenger receptors (SR) on macrophages as an alternative cell entry pathway [158, 159]. Scavenger receptors constitute a large family of structurally and biologically diverse pattern recognition receptors (PRRs), divided into 12 classes [160]. Although well established as key players in innate immunity for their role in mediating phagocytosis of bacterial pathogens, little is known about their involvement in viral pathogenesis [161]. Targeted PEGylation and hepatocyte transduction experiments suggested that HAdV-C5 binds to SR-A1.1 (previously known as SR-AII [160]) via the hexon capsomer HVRs 1, 2, 5 and 7 [162]. Similarly, SR-A1.1 related scavenger receptor MARCO (macrophage receptor with collagenous structure) has been demonstrated to facilitate HAdV-C5 transduction in alveolar macrophages (AM) and AM-like macrophages (designated Max Planck Institute [MPI] cells [163]) [159]. A later study proposed that HAdV-C5 binding to MARCO on murine MPI-2 cells is mediated by the hexon's HVR-1, although a likely interaction with additional hexon regions or other capsid-associated proteins was acknowledged by the authors [164].

Vascular cell adhesion molecule-1 (VCAM-1), an adhesion receptor for leukocytes expressed on the endothelium of atherosclerotic vessels, has been shown to facilitate HAdV-C5 binding and infection *in vitro* [165]. Atherosclerosis is a multifactorial, chronic inflammatory vascular disease of the artery wall for which sustained treatment with current remedies is associated with liver injury and high risk for bleeding [166]. Since VCAM-1 is not expressed on normal, healthy endothelium, atherosclerosis presents a prime candidate for the application of adenoviral gene therapy [166]. Like CAR, the primary cell surface receptor for species C adenoviruses, VCAM-1 is a member of the immunoglobulin gene superfamily [151, 165], characterised by the presence of a variable number of Ig-like domains [167]. While the homology between the receptors would suggest a role for the penton base-associated fibre in HAdV-C5 binding to VCAM-1, the adenoviral protein mediating the interaction is yet to be identified [165]. In addition to the CAR, the $\alpha2$ domain of the major histocompatibility complex-1 (MHC-1) also has been demonstrated to bind the HAdV-C5 fibre [168]. The relevance of this interaction for viral infection *in vivo*, however, remains to be elucidated [156].

In addition to binding of a wide variety of cell-surface receptors other than CAR, HAdVs have also been shown to recruit several 'adaptor molecules' [156]. Human lactoferrin, for example, a host defence protein (HDP) of the innate immune system (*see* II.2.1.3), can facilitate HAdV-C5 (as well as -B35 and -D26) uptake in human dendritic cells by inducing a Toll-like receptor 4 (TLR-4) mediated internalisation in the absence of CAR [169]. When intravascularly administrating HAdV-C5 for clinical applications (*see* II.3), it is predominantly sequestered by hepatocytes in a CAR independent manner [135]. Instead, transduction is facilitated through a heparin-binding site on soluble coagulation factor X (FX) [135].

II.2.1.3 Soluble components for enhanced infection

Coagulation factor X is the zymogen of 'activated coagulation factor X', also commonly referred to as Factor Xa, a vitamin K-dependent serine protease. The soluble 55 kDa protein circulates in human plasma at concentrations of 8-10 µg/mL [135]. Structurally, it can be divided into four distinct domains: the N-terminal, vitamin K-dependent γ-carboxyglutamic acid-rich (Gla) domain, the epidermal growth factor (EGF)-like domain 1, EGF-like domain 2 and the C-terminal serine protease [135]. The

factor X Gla-EGF-like 1 domains bind to the HAdV-C5 hexon with picomolar affinity in a calcium-dependent manner [134, 135]. Despite their trimeric arrangement, each hexon capsomer binds only one FX molecule [134, 135]. Cryo-electron microscopy (cryo-EM) analysis revealed that the N-terminal Gla domain occupies the central depression in between the turrets of the hexon, binding to one of three potential binding sites and thereby obstructing the other two [134, 135]. Replacement of HAdV-C5 HVRs with those from HAdV-D48 resulted in a loss of FX-mediated hepatocyte transduction, confirming the HVRs as the FX binding site on the hexon [135]. The C-terminal serine protease domain of FX contains a heparin-binding site, mediating cell surface binding via heparan sulphate proteoglycans (HSPGs). Binding of FX to the HAdV-C5 hexon and its putative cell surface receptor bridges the interaction between virus and the cellular receptor [135].

In addition to coagulation factor X, human lactoferrin and its proteolytic cleavage product Lfcin have been shown to enhance HAdV-C5 infection of epithelial cells [136, 137]. Lactoferrin, a member of the transferrin gene family, is a glycosylated 80 kDa protein comprising two globular domains, each harbouring a nearly identical metal binding site, primarily coordinating Fe^{2+} and Fe^{3+} (Figure 10) [170, 171]. The protein is secreted by epithelial cells of the mucosa and is found in high concentrations in biological fluids like breast milk, saliva, and tear fluid where it exerts its function as component of the innate immune system [171, 172]. Both full-length hLF and the highly charged, 49-residue long lactoferricin peptide at the N-terminus (Figure 10) exhibit broad antimicrobial activity [173, 174]. In addition to antibacterial and antifungal properties, lactoferrin has been shown to possess antiviral activity against a wide range of both RNA- and DNA viruses, including hepatitis C virus, Rotavirus, Poliovirus, human immunodeficiency virus, cytomegalovirus and herpes simplex virus 1 and 2. Instead of inhibiting viral replication, lactoferrin prevents infection of the host cell either by binding to the virus directly or by binding to viral cell surface receptors, thereby obstructing cell surface attachment [171]. Surprisingly, lactoferrin exerts the adverse effect on species C adenoviruses, promoting their binding and infection of epithelial cells independently of CAR [136, 137]. The biggest improvement of infection efficiency was observed in viral transduction assays when hLF was added to concentrations ranging between 0.1 and 1 mg/mL. A concentration of 10 mg/mL diminished the effect and even reduced binding to a level below that in absence of hLF [136].

Based on the finding that only free but not cell-associated hLF promotes binding, it was proposed that species C adenoviruses 'hijack' lactoferrin to bridge the gap between virus and cell surface receptor [136], as observed for coagulation factor X on hepatocytes [135]. A subsequent study revealed that HAdV-C5 binding and transduction can be promoted by the N-terminal lactoferricin alone. The highly positively charged peptide binds to the hexon capsomer via the predominantly acidic HVR-1 loop, a feature unique to serotype C adenoviruses [137]. In a recent study, detailed structural insights into this interaction were elucidated (*see* 4.4) [130].
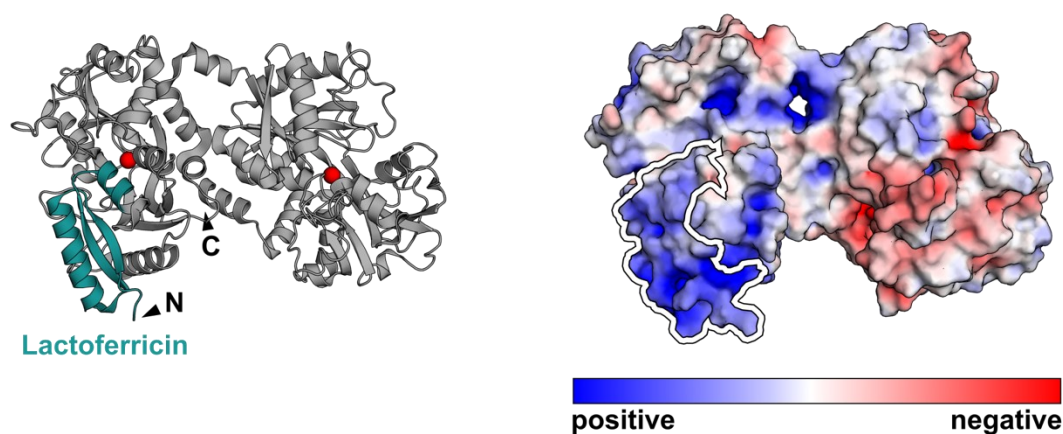

Figure 10 – Structure of human lactoferrin.
The N-terminal lactoferricin peptide of diferric human lactoferrin (grey) (PDB: 1LFG) is coloured in teal (left). The coordinated Fe^{3+} ions are depicted as red spheres. The surface representation of hLF (right) coloured by the electrostatic potential (APBS plugin, PyMol 2.5.4, Schrödinger) reveals the highly positively charged nature of lactoferricin (highlighted with a white outline).

II.2.1.4 Endocytosis and intracellular trafficking

After attachment to the host cell surface, the virion-receptor complexes are internalised via clathrin-mediated endocytosis [175]. Internalisation of species C Adenoviruses requires dynamin, a GTPase enabling the constriction and budding off of clathrin-coated pits into coated vesicles [176], and phosphatidylinositol-3-OH kinase (PI3K), which induces reorganisation of actin filaments [175]. Binding of the HAdV penton base to αvβ3 and αvβ5 integrins triggers integrin clustering, signalling for local formation of clathrin-coated pits and endocytosis of the virion [142, 152]. Micropinocytosis, while not essential for viral uptake into epithelial cells, has been proposed as an additional internalisation pathway [177].

Upon cell surface attachment, sequential disassembly and proteolysis of the virion constituents is initiated to facilitate viral DNA delivery to the nucleus. The first protein to dissociate is the penton base-associated fibre, followed by the release of the penton base protein [175]. Due to their close association with the penton capsomer, the peripentonal hexons and inner capsid proteins IIIa and VI are released simultaneously [178]. Cement protein IX and protein VIII partially dissociate from the capsid as well [179]. As endocytic vesicles bud from the membrane, the vacuolar H^+-ATPase acidifies the endosomal lumen. As the virus-containing endosome matures to early (sorting) endosomes and the pH decreases to 6, endosomal escape of species C adenoviruses is initiated [176]. The pH optimum for viral escape from the endosomal compartment differs amongst the different HAdV species; serotype B HAdV-7 for example localises in the significantly more acidic late endosome [180]. While the disassembly of the capsomeres on the vertices of the icosahedral capsid can occur in the absence of acidification, uncoating is facilitated by a decrease in endosomal pH [175, 181]. Although the precise mechanism of the endosomal escape by HAdVs is still poorly understood, protein VI has been recognised as the lytic component [179, 182]. In addition to the release of protein VI, endosomal escape requires activity of the adenovirus L3 protease and close association of the penton base protein with the endosomal membrane [183]. In spite of being the first capsid component to dissociate, the penton base-associated fibre has been shown to significantly influence the timing of the endosomal escape [176].

After entering the cytosol, the virus particles are rapidly transported towards the nucleus by dynein, a molecular motor which translocates unidirectionally alongside microtubules towards the microtubule organising centre [183, 184]. Upon reaching the perinuclear envelope, the remaining capsid binds to the nuclear pore complex, initiating proteolytic degradation of proteins VI and VIII and thereby detaching the viral DNA core from the capsid. Lastly, dissociation of cement protein IX destabilises the hexon facets, fully disintegrating the remaining capsid and thus liberating the DNA [175, 176]. Translocation of the viral genome into the nucleus is mediated by interaction of the HAdV protein VII with multiple nuclear import receptors [185].

II.2.2 Late gene expression, viral assembly, and release

Adenoviral transcription is divided into two stages: early- and late-stage transcription. The details of early-stage gene expression are well beyond the scope of this work and have been reviewed extensively in literature. Briefly summarised, early AdV gene expression has three fundamental goals [175]. Firstly, early translation products stimulate host cell entry into the S phase of the cell cycle, optimising the cellular environment for viral replication [186]. Secondly, HAdVs need to establish systems for suppression of the host's antiviral immune responses to avoid elimination of infected cells [187]. Lastly, the transcriptional machinery for viral DNA replication has to be synthesised [188]. The transcriptional activators to achieve these initial goals are encoded by the E1A gene [175].

Late-stage expression of viral genes begins simultaneously with viral DNA replication. A complex of the viral proteins E1B-55k and E4orf6 facilitates the translocation and cytoplasmic accumulation of viral mRNAs while simultaneously inhibiting the accumulation of host mRNAs in the cytoplasm [175]. Association of mRNA to the protein complex is mediated by intrinsic RNA binding activity of the E1B-55k component whereas the nuclear export is facilitated by a nuclear export sequence in the E4orf6 protein [175]. It remains unclear what governs the selective export of viral mRNA. Discrimination from host cell mRNA is thought to be caused simply by the cellular localisation of the E1B-55k/E4orf6 complex rather than differences in the nature of the mRNA [189]. Relocating the nuclear export factor to viral transcription centres facilitates viral mRNA transport while simultaneously inhibiting host cell mRNA transport [175]. In addition to restricted host cell mRNA export from the nucleus, translation of cellular mRNAs is also inhibited in infected cells. As a result of viral replication, double-stranded RNA accumulates in the cell, activating the cellular protein kinase R (PKR) which in turn phosphorylates and thereby deactivates the eukaryotic initiation factor 2α (eIF-2α), effectively blocking translation of host cell mRNA [175].

Replication of viral DNA and production of capsid proteins form the base for assembly of progeny virions. Following HAdV hexon protomer synthesis, late viral protein L4-100k, a protein also involved in stimulation of viral translation, nucleates rapid trimerisation. Penton bases and fibres oligomerise independently before merging into the fully assembled capsomer [175]. After their assembly in the cytosol, the capsomers

are translocated into the nucleus to construct the virion. Binding of inner capsid protein VI to the hexon capsomer is a prerequisite for its nuclear import [189]. The formation of a new viral capsid is initiated by association of the capsomers with viral DNA [175]. Virions are assembled concomitantly (assembly of capsid and genome together) rather than sequentially (formation of an empty capsid which is then 'loaded' with DNA) [189]. The newly assembled, immature virions are incapable of uncoating and therefore lack infectivity. Maturation occurs by proteolytic cleavage of capsid and core precursor proteins by the viral protease (AVP), destabilising the capsid and thereby preparing the virus for uncoating [190]. Prior to maturation, polypeptide VIII (pVIII) forms a 'molecular stitch' between the peripentonal and central hexons to stabilise the capsid during assembly. AVP cleavage of pVIII subsequently destabilises the capsid, facilitating vertex release during infection [190]. Analogously, cleavage of polypeptide VI is necessary to reduce its association to the hexon, facilitating the release of mature protein VI during endosomal escape [190]. Lastly, maturation of viral core protein VIII and the terminal protein results in a reduction of viral core condensation [190].

The mechanisms mediating the release of the matured progeny virions are not fully elucidated. In addition to the disruption of intermediate filaments of the cytoskeleton, the viral E3-11.6k protein, also referred to as 'adenovirus death protein', induces host cell death as it accumulates during late-stage infection. The molecular mechanism of E3-11.6k induced cell death remains unknown to this date [175, 189].

II.3 Medical uses

II.3.1 Gene therapy

Gene therapy is a medical technique that involves the transfer of genetic material into patient cells for the treatment of a disease. The therapeutic effect may be achieved either by expression of a transferred gene or by inhibition or modification of an endogenous gene in the target cell [191]. Insertion of foreign genetic material into a cell *in vivo* requires a vehicle for delivery. The nature of these vectors can be divided into viral and non-viral vectors. Non-viral vectors are accompanied by a broad spectrum of advantages and limitations for their use gene therapy, such as reduced immunogenicity at the cost of limited gene delivery capacity [192]. While they have

found successful application in clinical use [191], the following section will focus solely on viral vectors, specifically the use of adenoviruses in gene therapy.

Retroviruses, such as the human immunodeficiency virus, integrate their genetic material into the genome of infected cells. While their use as gene therapy vectors enables continuous expression of the transgene even in mitotic daughter cells, it harbours a significant risk of unpredictable genome integration, potentially disrupting off-target genes at the site of insertion [193]. Following delivery by non-integrating vectors such as adenoviruses, the transgene persist as an episome in the nucleus, reducing the risk of genotoxicity and enabling long-term expression in postmitotic tissue [194]. While this presents a valuable increase of safety in clinical applications, replication-deficient, non-integrating vectors will be progressively lost in proliferating cells. Besides their epichromosomal persistence in the target cell, adenovirus-based vectors have high transduction efficiency in both dividing and non-dividing cells, a broad tissue tropism, scalable production, high expression rates, a high payload capacity and provide the ability to engineer replication-competent and -deficient vectors [193, 195-197].

Since the inception of gene therapy and the use of adenoviral vectors, engineering efforts have focused on improving their capacity, efficacy, longevity of transgene expression and, most importantly, safety [195]. In an attempt to reduce the immunogenicity associated with wild-type adenoviruses, the first generation of adenoviral vectors were created by replacement of the E1A gene with an expression cassette containing the transgene and a high activity promoter, typically the 'cytomegalovirus immediate early' promoter, driving the expression [198]. While E1A deletion results in replication-deficient vectors (*see* II.2.2), strong immunogenicity of these vectors rendered them unsuitable for gene therapy applications [194]. In subsequent efforts to improve upon the safety of the vectors, additional non-essential viral genes were removed. Second-generation adenoviral vectors have deletions of the E1, E2 and/or E4 genes [196]. However, it was a second-generation adenovirus vector that led to the tragic death of 18-year-old Jesse Gelsinger on the 17th of September 1999. Jesse participated in a phase I gene therapy trial for ornithine transcarbamylase (OTC) deficiency when the administration of a high dose of the adenoviral vector based therapeutic (3.8×10^{13} virus particles) induced a cytokine storm, leading to his untimely demise and the first reported fatality in a gene therapy trial [196, 199-201]. The development of the third and latest generation of adenoviral

vectors marked a significant stride in minimising immunogenicity while simultaneously expanding the therapeutical potential. By deleting all viral sequences besides the inverted terminal repeat (ITR) and packing sequences, 'gutless' vectors capable of accommodating DNA cargo of up to 36 kilobase (kb) were created, coining the term 'high-capacity adenoviral vectors' (HCAds) [195, 197]. Naturally, gutless vectors are replication-deficient and 'helper-dependent', necessitating a complementary adenoviral helper virus for replication [202]. While HCAds exhibit reduced immunogenicity, larger capacity and prolonged transgene expression in comparison to earlier generations [194], elimination of potential contaminations by the helper virus are essential for the safety of these vectors in clinical applications [195].

Despite continuous improvements of adenoviral vectors, significant challenges continue to limit their use in gene therapy. Adenovirus infections are widespread and commonly caused by serotype C HAdVs 1, 2, 5 and 6 [203]. Over 60% of adults have neutralising antibodies against HAdV-C5, the most widely used adenoviral vector [204]. Prior infections with the virus generate antigen-specific T-cells which confer lifelong immunity, diminishing the vectors efficacy while inducing strong immunological responses when administered [195, 198]. Although the development of third generation vectors has significantly reduced immunogenicity over previous generations, gutless HCAds still elicit a potent immune response [193]. The use of adenoviral vectors for gene therapy is further complicated by the broad tissue tropism exhibited by the wild-type virus (*see* II.1.1). Using vectors with native adenoviral capsids leads to sequestration at potential off-target sites (*see* II.2.1.2), reducing transduction efficiency in the target tissue [197].

As discussed previously (*see* II.1.3.1), the HVRs of the hexon capsomer pose the primary antigen on adenoviruses and are the target for the majority of the neutralising antibodies [132]. Additional evidence also suggests that antigen-specific cytotoxic T lymphocytes (CTL) against the hexon protein confer protection against HAdV infection [198]. A variety of approaches are being explored to mitigate these problems for the use of HAdVs in gene therapy. To evade pre-existing immunity or to reduce FX mediated sequestration of the HAdV-C5 vector in the liver, chimeric vectors with modifications or replacements in the HVR loops may be generated [197]. While species C adenoviral vectors with lower natural seroprevalence in the human population (HAdV-2, -26 and -35) have been explored, transduction efficiency was significantly lower than that of HAdV-C5 [195]. A seemingly better approach to avoid

problems associated with pre-existing immunity is the use of non-human adenoviral species, as neutralising antibody titres against these vectors are significantly lower. The most frequently used non-human vectors in clinical trials are chimpanzee-derived adenoviral vectors [195].

As much as gene therapy vectors require minimal immunogenicity for efficient transgene delivery and long-term expression, a controlled, localised immune response against vector-antigens in a tumour could also prime an immune response against tumour antigens [197]. Congruently, HAdV-C5 has emerged as a promising candidate for oncolytic virus therapy [193, 205]. In addition to carrying a tumoricidal payload, the oncolytic vector promotes a proinflammatory environment, inducing recruitment of cytotoxic immune cells to eliminate virus-infected cells, aiding in elimination of the cancer [203]. Although HAdV-C5 appears much better suited for application in anti-cancer treatments than gene therapy, pre-existing immunity against frequent serotype C adenoviruses limits infection efficacy also in oncolytic virus therapy. Additionally, cancer cells frequently express little to no CAR, the primary cell surface receptor for adenoviruses, thwarting transduction efficacy in target cells [193].

II.3.2 Vaccines

While the strong immunogenicity of species C HAdVs has significantly limited their use as vectors in gene therapy, it has greatly benefitted the development of adenovirus-based vaccines. The inherent immunogenicity of the vector serves a function similar to that of an adjuvant, inducing a proinflammatory environment and facilitating a strong humoral and antigen-specific T-cell response to the transgene [193, 195, 198]. The primary drawback of adenoviral vaccines is that, while a strong immune response induced by the vector is generally beneficial for vaccine applications, it has the potential to impede the specific immune response targeted towards the desired antigen, effectively 'masking' the intended epitope [197]. Despite this limitation, the advantages of adenovirus-based vaccines were showcased during the 2019 SARS-CoV-2 pandemic (COVID-19), a global health crisis with over 770 million infections. Rapid engineering of genome and capsid, ease of production and induction of robust immune responses with lasting immunity facilitated the rapid

development of four clinically approved vaccines, enabling administration of 13 billion vaccine doses worldwide [193].

Chapter 2

Aim of study

The three-dimensional structure of a protein determines its function in biological processes. Understanding and elucidating these structures is paramount to advancing our knowledge of biology, disease, and, ultimately, the development of novel therapeutics.

Part I – African trypanosomes

The invariant surface glycoproteins ISG65 and ISG75 on the surface of the blood-dwelling, human-infective parasite *T. brucei gambiense* have been extensively studied since their discovery in the 1990s. Their invariant nature and relatively high copy number across the cell surface have made ISG65 and ISG75 prime candidates for vaccination trials. Although the large abundance implied a functional significance for the bloodstream form of the parasite, no biological function had been identified for either protein. The primary research objective of this study was to determine the structure of *T. b. gambiense* ISG65 and gain insights into its biological function. To do so, the study was initially partitioned into three subsidiary objectives:

1. Design, recombinant expression and structural characterisation of a minimal protein construct of *T. b. gambiense* ISG65.

2. Investigation of ISG65s potential as a surface receptor in the human blood stream and its associated biological function.

3. Exploration of the extent to which the findings surrounding ISG65 could be extrapolated to other invariant surface glycoproteins.

Part II – Human adenovirus C5

While respiratory infections by species C HAdVs are common, the underlying mechanisms are still poorly understood. Although cell entry pathways have been studied extensively in immortalised cell lines *in vitro*, inaccessibility of the primary cell surface receptors on the polarised epithelium of the respiratory tract point towards alternative entry pathways *in vivo*. Recent findings demonstrated that HAdV-C5 hijacks lactoferrin, an antimicrobial immunomodulator of the innate immune system, to facilitate cell entry. Although the HVR-1 loop of the hexon capsomer has been suggested as the lactoferrin binding site, a detailed understanding of this mechanism is lacking. Elucidating the structural basis of this immune evasion strategy could aid in the development of novel antivirals or adjuvants for improved transduction efficacy in gene therapy applications. To obtain a mechanistic understanding of this interaction, two subsidiary objectives were set:

1. Structural characterisation of the HAdV-C5 hexon / lactoferrin complex.

2. Biophysical characterisation and validation of the proposed interaction.

Chapter 3

Overview of methods

The following chapter provides an overview of selected methods and techniques that were used throughout course of this study. Complementary methods not included here, work performed by collaborators, as well as more detailed descriptions of materials and methods outlined in this chapter can be found in the respective publications (*see* Chapter 6).

3.1 Recombinant protein expression and purification

3.1.1 *Escherichia coli*

Recombinant expression and purification of all constructs used in this study (*T. b. gambiense* ISG65, ISG43 and ISG64; human C3d; as well as their truncated, mutated or Avi-tagged derivatives) were performed as described previously [85, 206, 207]. Briefly summarised, following the isopropyl-β-D-thiogalactopyranoside (IPTG) induced expression in *Escherichia coli* (*E. coli*) T7 shuffle cells (New England Biolabs), bacteria were harvested, lysed and the N-terminally His-tagged protein of interest was purified from the soluble protein fraction via immobilised metal affinity chromatography (IMAC) and subsequent size exclusion chromatography (SEC).

3.1.2 Insect cells

Recombinant expression of *T. b. gambiense* ISG65 constructs in stable *Drosophila melanogaster* S2 cell lines and subsequent protein purification were performed as described previously [206]. In short, the C-terminally C-tagged protein of interest was purified from cell culture supernatant via affinity chromatography using CaptureSelect C-tag affinity resin (Thermo Fisher Scientific), followed by SEC.

3.2 Protein purification from native source

3.2.1 Human complement factor C3, C3b and C3MA

Native human complement factor C3 was purified from normal human serum as described previously [85]. In brief, a protein fraction containing C3 is isolated by stepwise polyethylene glycol (PEG) precipitation. After resuspension, the protein fraction is passed over an L-lysine affinity chromatography column before being subjected to ion exchange chromatography (IEX) using a Source 15Q anion exchanger. C3 is eluted by application of a NaCl gradient and subsequently purified by SEC. Purified protein was flash-frozen in liquid nitrogen and stored at -80°C until further use. Generation and purification of C3b and C3-methylamine (C3MA) were performed as described [208]. Validation of the purified components by N-terminal sequencing, Western-Blotting, intact mass spectrometry and IEX have been described in detail previously [85].

3.2.2 *Tbg*VSG LiTat3.1

The LiTat3.1 VSG was purified from *T. b. gambiense DAL972* cell pellets as described previously [209]. Briefly summarised, release of GPI-anchored proteins by endogenous GPI-specific phospholipase C was induced by osmotic shock. The solubilised VSGs were purified by anion-exchange chromatography and subsequent SEC.

3.3 Surface plasmon resonance

3.3.1 In vitro biotinylation of ligands

Surface plasmon resonance (SPR) measurements were preferentially performed with ligands immobilised via site specific biotinylation. Biotinylation of Avi-tagged ligands via *E. coli* biotin ligase (BirA) *in vitro* was performed as described [210]. In brief, Avi-tagged protein was incubated with recombinantly produced BirA, $MgCl_2$, adenosine triphosphate (ATP) and D-biotin at 30°C. After the reaction, excess

biotin was removed by dialysis or SEC. Successful biotinylation was confirmed by streptavidin gel-shift [210].

3.3.2 SPR measurements

All SPR measurements were performed on Biacore T200 (GE Healthcare) or Biacore S1+ systems (Cytiva). Wherever possible, biotinylated ligands were immobilised to a Series S CAP sensor chip using the CAPture kit (Cytiva). Alternatively, ligands were immobilised to CM5 sensorchips (Cytiva) via amine coupling. Buffers and injection times were varied as required [85]. Kinetic and steady-state affinity parameters were evaluated using the Biacore T200 evaluation software (GE Healthcare).

3.4 Haemolysis experiments

Inhibition of alternative and classical complement pathway activity in NHS and complement factor depleted serum was assessed using the AP50 and CH50 test kits (HaemoScan) as described previously [85]. Complement activity was determined from the extent of lysis of rabbit (AP) or sensitised sheep erythrocytes (CH50) by measuring the absorbance of free haemoglobin at 415 nm. Absorbance measurements were performed on Infinite 200 PRO plate readers (Tecan).

3.5 Single-particle cryo-EM

Detailed descriptions of the sample preparation, data collection parameters and data processing workflows are outlined in the respective publications [85, 130]. Only brief summaries for the individual samples are outlined here.

3.5.1 ISG65:C3

The ISG65:C3 complex was prepared by concentration of a 2-fold molar excess of ISG65 with C3 and subsequent SEC on a Superdex 200 Increase 10/300 GL column (GE Healthcare). The freshly purified complex was applied to newly glow-discharged

copper C/Flat 1.2/1.3 300 mesh grids (Protochips) and vitrified using a Vitrobot Mark IV (Thermo Fisher Scientific). Grids were imaged at the SOLARIS cryo-EM facility (Krakow, Poland) on a Titan Krios G3i microscope (Thermo Fisher Scientific), operated at 300 kV, and equipped with a K3 detector (GATAN Inc.). 10,380 movies were recorded at 25° stage tilt with a pixel size of 0.86 Å. Data were collected using the automated data collection software EPU (Thermo Fisher Scientific). Data processing was performed using the cryoSPARC (Structura Biotechnology) [211] and Relion 3.1 [212] software suites. The crystal structure of C3 (PDB 2A73) and an ISG65 model predicted by AlphaFold2 [213] were docked into the cryo-EM reconstruction using Phenix [214]. Both models were independently and iteratively refined via automated real-space refinements using Phenix [214] and REFMAC5 [215] as well as manual refinements in Coot [216]. After obtaining satisfactory validation metrics (as determined by MolProbity [217]), the models were manually and automatically refined in tandem using Phenix and Coot.

3.5.2 ISG65:C3b

ISG65 and C3b were combined in a 2:1 molar ratio and crosslinked by addition of a 100-fold molar excess of DSA (Di-(N-succinimidyl)adipate). After quenching of the reaction with ethanolamine, the crosslinked complex was gel-filtered on a Superdex 200 Increase 10/300 GL column (GE Healthcare). For data collection at 25° stage tilt the ISG65:C3b complex was applied to freshly glow-discharged copper C/Flat, 1.2/1.3 and 2/2 300 mesh grids (Protochips). For data collections at a 0° stage tilt, gold R1.2/1.3 300 mesh grids (Protochips) were coated with a graphene monolayer. ISG65:C3b was applied to the grids after glow-discharge. All grids were vitrified using a Vitrobot Mark IV (Thermo Fisher Scientific). Grids were imaged at the CEITEC cryo-EM facility (Brno, Czech Republic) using a Titan Krios, equipped with a K2 detector (GATAN Inc.), and operated at 300 kV. A total of 18,062 movies (at 25° stage tilt) and 8960 movies (at 0° stage tilt) were recorded with a pixel size of 0.818 Å. Data were collected using the automated data collection software EPU (Thermo Fisher Scientific). Processing of all collected datasets was performed independently first before merging selected particles with previous data collections. Particle picking on images acquired at a 25° stage tilt was performed using the convolutional neural

network-based particle picker, crYOLO [218]. All other data processing was performed using the cryoSPARC software suite (Structura Biotechnology) [211]. The crystal structure of C3b (PDB 2I07) and the atomic model of ISG65 obtained from the cryo-EM reconstruction of the ISG65:C3 complex were docked in the obtained density map using Phenix [214]. Both models were simultaneously and iteratively refined via automated real-space refinements using Phenix and REFMAC5 [215] as well as manual refinements in Coot [216]. Model validation was performed using MolProbity [217].

3.5 Small-angle x-ray scattering

Detailed descriptions of sample preparation, data collection and data processing methodology for small-angle x-ray scattering (SAXS) experiments are provided in the respective publications [85, 207]. This section briefly summarises individual data collections and subsequent data processing workflows.

3.5.1 Data collection

SEC-SAXS data for human C3, C3b, C3MA, C3d, *T. b. gambiense* ISG65 and ISG75 as well as an ISG65:C3d complex were collected at 20 °C on a Pilatus 2M detector (DECTRIS) at the BioSAXS beamline BM29 at the ESRF (Grenoble, France). SEC-SAXS data were collected in HPLC mode using a Superdex 200 3.2/300 column (GE Healthcare) [85].
SEC-SAXS data for *T. b. gambiense* ISG43, ISG64, VSG LiTat3.1 and *T. b. brucei* VSG LiTat1.5 were collected on an Eiger X4 M detector (DECTRIS) at 20 °C at the BioSAXS beamline SWING at SOLEIL (Gif-sur-Yvette, France). SEC-SAXS data were collected in HPLC mode using a KW404-4 column (Shodex) [207].

3.5.2 Data processing

The CHROMIXS programme [219] from the ATSAS software package [220] was used to normalise the data to the intensity of the incident beam, average the

recorded frames and subtract buffer scattering for all SEC-SAXS data collected over the course of this study [85, 207].

Structural parameters, namely radius of gyration (Rg), maximum particle dimension (D_{max}), and distance distribution function (p(r)) from data collected at BM29 were derived using the PRIMUS programme from the ATSAS package [220]. Conformational flexibility of the samples was assessed using the ensemble optimisation method (EOM 2.0) via ATSAS online [221] as described [85].

Data collected at SWING was analysed using the PRIMUS programme from the ATSAS [220] package and BioXTas RAW [222]. Theoretical scattering curves of the molecular models generated with AlphaFold2 [213] (TbgISG43, TbgISG65, TbgISG75) and AlphaFold-Multimer [223] (TbbVSG LiTat1.5, TbgVSG LiTat3.1) and their respective fits to the experimental data were calculated with FoXS [224, 225]. The conformational flexibility of the proteins was assessed via SAXS-based ensemble modelling with BILBOMD [226] as described [207].

Chapter 4

Results and discussion

The results of this dissertation are thematically divided into two parts. Part I presents the findings of the study on the structural basis of immune evasion by African Trypanosoma which were published across three articles in international, peer-reviewed journals.

Publication #1:

<u>Sülzen, H.</u>, Votrubová, J., Dhillon, A., Zoll, S., A multifaceted strategy to improve recombinant expression and structural characterisation of a *Trypanosoma* invariant surface protein. *Sci Rep* 12, 12706 (2022).

Publication #2:

<u>Sülzen, H</u>., Beg, J., Dhillon, A., Kereïche, S., Pompach, P., Votrubová, J., Zahedifard, F., Šubrtova, A., Šafner, M., Hubalek, M., Thompson, M., Zoltner, M., Zoll, S., Cryo-EM structures of *Trypanosoma brucei gambiense* ISG65 with human complement C3 and C3b and their roles in alternative pathway restriction. *Nat Commun* 14, 2403 (2023).

Publication #3:

<u>Sülzen, H</u>., Volkov, A., Geens, R., Zahedifard, F., Stijlemans, B., Zoltner, M., Magez, S., Sterckx, Y*., Zoll, S.*, Beyond the VSG Layer: Exploring the Role of Intrinsic Disorder in the Invariant Surface Glycoproteins of African Trypanosomes. *BioRxiv Preprint* (2024)**.

These authors contributed equally to this paper and should be considered joint corresponding authors.

**At the time of writing, the manuscript has been conditionally accepted for publication with minor revisions in PLoS Pathogens. The revised manuscript has been re-submitted on 19.03.2024.*

Although *T. b. gambiense* ISG65 has been the subject of many studies since its discovery 30 years ago, its molecular structure had not yet been determined. While

ISG65 could readily be expressed recombinantly in *E. coli* [119, 227], the commonly used construct did not seem amendable to structural characterisation. Publication #1 presents a multifaceted strategy for improved construct design and expression of challenging proteins, employing hydrogen-deuterium exchange mass spectrometry (HDX-MS) guided construct design, truncation scanning and targeted expression host switching. Publication #2 reports an atomic model of *T. b. gambiense* ISG65, describes the function as a receptor for human C3b on the surface of the parasite and elucidates its role as an inhibitor of the alternative pathway of the human complement system, providing a plausible mechanism as to how the protein may exert its function in the context of the VSG layer. Nucleated by the findings evolving around ISG65, the question whether other members of the ISG superfamily may possess similar functionality arose. Publication #3 shows that several other *T. b. gambiense* ISGs of different sizes may adapt distinct conformational states, suggesting that flexibility of intrinsically disordered linkers tethering the proteins to the parasite membrane are a functional feature of trypanosome receptors of the ISG super-family.

Part II presents the results of investigating the structural aspects of HAdV-C5's ability to bind the human defence protein lactoferrin to facilitate cell entry into epithelial cells. The findings were summarised in an article published in the 'Journal of Virology'.

Publication #4:
Dhillon, A., Persson, D., Volkov, A., **Sülzen, H**., Kádek, A., Pompach, P., Kereïche, S., Lepšik, M., Danskog, K., Uetrecht, C., Arnberg, N., Zoll, S., Structural insights into the interaction between adenovirus C5 hexon and human lactoferrin. *J. Virol.* Feb 7:e0157623 (2024).

A growing body of evidence suggests that instead of the CAR mediated cell entry pathway of species C HAdVs, which has been well characterised *in vitro*, alternative uptake routes facilitate the initial infection of epithelial cells *in vivo* [121, 135, 153, 154, 157]. Human lactoferrin, a component of the non-specific immune system with antiviral activity against a multitude of virus species [171, 172], has been shown to exert the adverse effect on HAdV-C5, facilitating uptake into epithelial cells. Publication #4 presents the structural mechanism underlying the binding of lactoferrin to the viral hexon capsomer, shedding more light onto alternative cell entry pathways

and thereby potentially paving the way for advancements in vaccine or gene therapy applications and the development of novel antivirals.

Part I – African trypanosomes

4.1 A multifaceted strategy to improve recombinant expression and structural characterisation of a *Trypanosoma* invariant surface glycoprotein.

4.1.1 Author's contributions

My contribution to the work included (1) joint conceptualisation of the study, experimental design, and data analysis with the corresponding author; (2) expression and purification of all bacterial expression constructs used in the study; (3) purification of ISG65$_{18\text{-}363}$ and ISG65$_{32\text{-}306}$ from S2 cell culture supernatant; (4) joint preparation of figures with the corresponding author and (5) review and editing of the manuscript.

4.1.2 Results

Secondary structure prediction is likely to be the most commonly used method for the design of recombinant expression constructs. While modern prediction algorithms can achieve >90% accuracy for proteins with homologues of known structure [228], secondary structure prediction of *T. b. gambiense* ISG65 resulted in low confidence scores across the entire sequence with especially low scores in the C-terminus. An initial expression trial of a construct comprising the entire extracellular domain (ISG65$_{18\text{-}386}$), truncated after the N-terminal signal peptide and directly before the transmembrane domain (TMD), failed to express. A revised construct with a C-terminal truncation placed further away from the TMD (ISG65$_{18\text{-}363}$) did express and yield soluble protein [206]. To rationalise the identification of the C-terminal domain boundaries for design of a minimal construct amendable to structural characterisation, HDX-MS was employed to localise low complexity regions in the protein. In comparison to well-structured domains, regions of low complexity exhibit faster deuteration rates [229]. Proteolytic digest of ISG65 with a Nepenthesin-2 column

(AffiPro) resulted in maximum sequence coverage of 97% with a high degree of short, overlapping peptides, providing the foundation for high resolution analysis by HDX-MS. While the N-terminal domain boundaries could be readily inferred from the deuteration plot, the C-terminus exhibited a steady increase in deuteration, starting from residue 260. The deuteration rates plateaued at residue 300 and remained high for the remaining sequence with exception of a local minimum around residue 330 [206]. Truncation of the construct at residue 270, the beginning of the low-complexity region as determined by HDX-MS, resulted in failure to express [206]. To determine the C-terminal domain boundaries, truncation scanning was performed. Starting from the local deuteration minimum at residue 330, a series of constructs with the newly identified N-terminus and sequentially shortened C-termini were generated by introduction of stop codons every two residues. Normalised yields of purified protein increased until the C-terminus was truncated to residue 322, successive shortening increasingly diminished yields. The shortest construct that could be expressed successfully, albeit with exceedingly low yields (< 0.1 mg per litre of culture after two purification steps), was ISG65$_{32-306}$ [206].

With decreasing construct length, the amount of soluble protein aggregates eluting in the void volume of the size exclusion chromatography column increased, indicating that the *E. coli* folding machinery was failing to produce correctly folded protein. To mitigate the problem, the recombinant expression host was switched to an eukaryotic expression system. *Drosophila melanogaster* S2 cells were chosen for their easy handling and relatively low associated costs. To further facilitate folding in the non-reducing environment of the serum-free culture medium, a signal sequence for the secretory pathway was added. Based on a report showing that the use of a C-tag for affinity purification improved expression of a Plasmodium surface protein [230], the N-terminal hexa-His-tag of the bacterial expression construct was replaced with a C-terminal C-tag for the expression in insect cells.

Expression of ISG65$_{32-363}$ and ISG65$_{32-306}$ in S2 cells resulted in yields exceeding the results from expression in *E. coli* 40 to 250-fold [206]. Noticeably, the minimal construct expressed to levels identical to the full-length protein and, based on SEC, neither of the two constructs produced soluble aggregates. Circular dichroism spectroscopy (CDS) revealed that the fraction of random coils in the minimal construct was reduced by over 10% in comparison to the full-length construct. Furthermore, identical melting temperatures were observed for both ISG65$_{18-363}$ produced in *E. coli*

and ISG65$_{32-306}$ produced in S2 cells, indicating that removal of the terminal regions did not compromise the structural integrity of the overall fold [206].

4.1.3 Discussion

By combining well established and readily available technologies in an innovative and targeted fashion, the strategy for recombinant production of complex eukaryotic proteins outlined in this study may serve as a template for expression of other challenging proteins. Using the structurally uncharacterised *T. b. gambiense* ISG65, the presented findings demonstrate that soluble expression in *E. coli* does not guarantee high protein yields and that weak expression is not necessarily a sign for incorrectly chosen domain boundaries [206]. Whether the presented approach bares general validity and may readily be applied to other proteins without homologous refence structures remains to be elucidated. For most proteins, the limiting step of the proposed strategy is likely to be the proteolytic digest for HDX-MS analysis. Insufficient sequence coverage and a lack of small peptides with sufficient overlap will lead to decreased resolution, potentially rendering the identification of domain boundaries even more difficult than in the case of the ISG65 C-terminus.

While the design, expression, and purification of a minimal ISG65 construct did not translate into obtaining an experimental structure, identification of the domain boundaries resulted in significant improvements in computational protein structure prediction. Prediction of the protein structure for ISG65$_{18-363}$ via the popular prediction programme iTASSER [231] did not yield a model with any resemblance to other trypanosome proteins of known structure. Removal of the experimentally determined low complexity region produced a model exhibiting a fold similar to the canonical three-helix bundle commonly found in structurally characterised Trypanosoma surface receptors. While the study was under review, the now widely adapted AlphaFold2 [213] became readily accessible via the fast and easy-to-use ColabFold (published 30th of May 2022) [232]. AlphaFold2 exhibits unprecedented accuracy in protein structure prediction, greatly outperforming almost all other structure prediction tools [233]. Alongside the predicted structural model, AlphaFold2 produces a predicted local distance difference test (pLDDT) score based on the lDDT-Cα metric, reflecting a per-residue estimate of the confidence in the predicted model [234]. Regions with a

pLDDT score above 90 are expected to have correctly predicted backbone and side chains, residues with a score between 70 and 90 are likely to have a good backbone prediction and scores between 50 and 70 are considered low confidence and should be interpreted cautiously. A pLDDT score below 50 indicates a strong propensity for disorder in that region of the protein. Structure prediction of full-length *T. b. gambiense* ISG65$_{18-436}$ via AlphaFold2, only omitting the N-terminal signal peptide, yields a model featuring the canonical three-helix bundle of the extracellular domain, largely predicted with a pLDDT above 90, inspiring confidence in the predicted structure (Figure 11). While the TMD exhibits pLDDT scores above 70, the remainder of the long C-terminal helix is predicted only with low confidence, similarly to the disordered loops in the N-terminal head region of the protein. After the third and last helix of the three-helix bundle, AlphaFold2 predicts a short, 14-residue long α-helix, connected to the long C-terminal helix via a long disordered (pLDDT < 50) linker (Figure 11). The experimentally determined C-terminal boundary for the longest expression construct, ISG65$_{18-363,}$ therefore could have been inferred from the predicted structure. On the contrary, the computational model does not provide any insights into the local decrease of deuteration rates around residue 330 observed during the HDX-MS analysis (Figure 11). Stunningly, a gradual decline of the pLDDT score can be observed around residue 322, the C-terminus of the bacterial expression construct with the highest protein yields. While residue 319 is predicted with low confidence (pLDDT score of 52) and residue 320 only marginally falls into the range residues predicted to be disordered (pLDDT score of 48), residues 322 and 323 show a large decrease in confidence, dropping to pLDDT scores of 41 and 32 respectively, indicating the beginning of a low complexity region (Figure 11). The same observation can be made for N-terminal residues 31 to 33. These predictions coincide well with the observations made using the prokaryotic expression system. While Alphafold2 does predict that residue 306, the C-terminal residue of the shortest expression construct used (ISG65$_{32-306}$), lies at the beginning of a short α-helix, the gradual decline of successive truncations from residues 322 to 306 and failure to express when truncated even further could not have been inferred from the predicted structure (Figure 11). HDX-MS analysis showed a steady increase of deuteration rates, starting with residues approximately half-way down the third α-helix of the three-helix bundle, indicating increased flexibility in solution. Although not necessarily beneficial in

determination of the C-terminal domain boundaries, the high confidence in the prediction of this region in the protein suggests a rigid fold not supported by the experimental data.
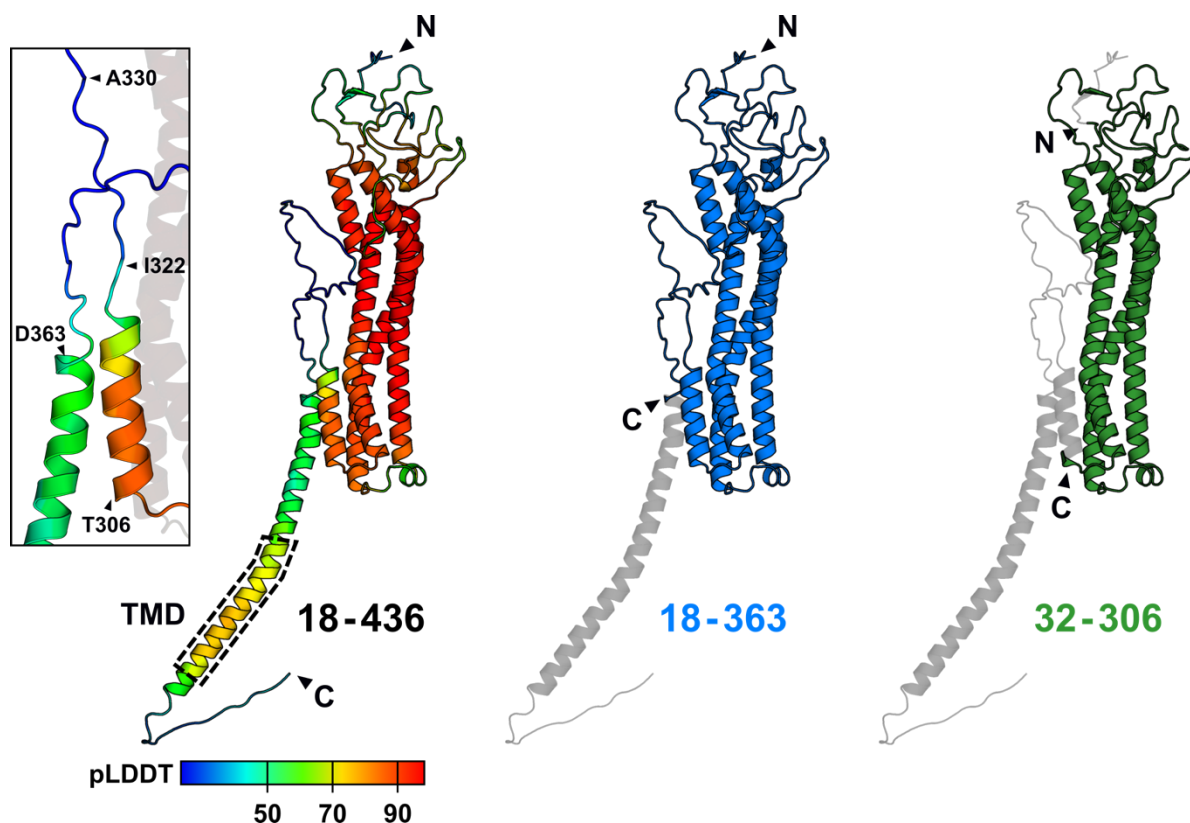

Figure 11 – AlphaFold2 model of *Trypanosoma brucei gambiense* ISG65.
A complete atomic model of *T. b. gambiense* ISG65$_{18-436}$ (lacking only the N-terminal signalling peptide), as predicted by AlphaFold2, is illustrated in cartoon representation (left) and coloured based on the predicted local distance difference test (pLDDT) score, reflecting (local) model quality as indicated by the legend: pLDDT ≥ 90: correct backbone and side chain; pLDDT > 70: correct backbone; pLDDT > 50: low confidence; pLDDT < 50: likely disordered. The region corresponding to the TMD has been highlighted by a dashed box. The inset shows a close-up view on key residues of the study. Using the same model, the predicted structure of the longest (ISG65$_{18-363}$) and shortest (ISG65$_{32-306}$) expression constructs used in the study have been illustrated. Residues contained in the expression constructs have been coloured in blue and green, respectively. Truncated residues are shown in grey. The N- and C-termini are annotated for all models.

In the case of *T. b. gambiense* ISG65, the information AlphaFold2 has provided could have been used to reliably infer the N-terminal domain boundary at Lys32 and the C-terminal domain boundary of the longest expression construct, Asp363 (Figure 11). While it may have provided an approximation of the C-terminal boundary of the minimal construct, the predicted structure incentivises truncation of the construct beyond residue 306, which, at least using prokaryotic expression systems, did not

yield soluble protein in practice. In a recent study with the closely related *T. b. gambiense* ISG75, we could show that protein regions predicted even with very high pLDDT scores do not necessarily accurately describe the behaviour of the protein in solution (*see* 4.3) [207] . While AlphaFold2 undoubtedly adds a highly valuable and versatile tool to the structural biology toolkit, supplementing and improving experimental strategies like the one proposed in this study, it does not serve as a replacement for the latter.

Despite the successful generation of a highly expressing minimal construct, no protein crystals for determination of the structure by x-ray crystallography could be produced. As outlined previously, ISG65 exhibits a disordered N-terminal head region and considerable flexibility in the last α-helix of the bundle, both features likely to interfere with the formation of an orderly crystal. The addition of various combinations of crystallisation chaperones, including three different nanobodies and an antigen-binding fragment (Fab) generated against ISG65, did not generate crystals either (data not shown). Although unsuccessful, the validity of the approach was later corroborated by publication of a crystal structure of *Trypanosoma brucei brucei* ISG65 in complex with human C3d (PDB 7PI6) [88]. While each ISG65 molecule contacted three copies of C3d within the crystal lattice, no crystal contacts in between ISG65 molecules could be observed [88]. The electron density for N-terminal residues 1-26, C-terminal residues beyond residue 316 or the loops of the disordered head region could not be resolved [88].

Upon identification of *Trypanosoma brucei gambiense* ISG65 as a receptor for human complement factor C3b, single-particle cryo-EM was deemed a more appropriate method of structure determination due to the size of the protein complex (*see* 4.2) [85].

4.2 Cryo-EM structures of *Trypanosoma brucei gambiense* ISG65 with human complement C3 and C3b and their roles in alternative pathway restriction.

4.2.1 Author's contributions

My contribution to the work included (1) joint conceptualisation of the study, experimental design and data analysis with the corresponding author; (2) recombinant

protein expression and purification; (3) purification of native C3 from human serum, generation of C3b and C3MA; (4) fluorescent labelling of antibodies and C3b; (5) joint SPR sample preparation; (6) SPR measurements (jointly for proteolytic fragments of C3, solely for validation of interface and AP pro-convertase formation); (6) design, execution and analysis of haemolysis assays; (7) sample preparation for cryo-EM analysis; (8) cryo-EM data processing, model-building and molecular modelling; (9) joint analysis of SAXS data with the corresponding author; (10) joint preparation of the manuscript with the corresponding author; (11) preparation and/or revision of all figures and (12) review and editing of the manuscript.

4.2.2 Results

Based on the high abundance of ISG65 on the surface of blood-stream form *Trypanosoma brucei gambiense*, the search for a potential interaction partner began in the natural habitat of the parasite: the human bloodstream. Using the extracellular domain of ISG65 (ISG65$_{18-363}$) as bait in pulldown experiments, subsequent SDS-PAGE analysis, in-gel digestion, and LC-MS/MS analysis revealed specific enrichment of human complement factor C3. SEC with purified ISG65 and human C3b confirmed the interaction and demonstrated formation of a stable complex [85]. SDS-PAGE analysis after crosslinking with homobifunctional crosslinker Di-(N-succinimidyl)adipate (DSA) indicated equimolar stoichiometry. Fluorescence-activated cell sorting (FACS) with Alexa Fluor 594 (AF$_{594}$) labelled C3b showed efficient internalisation in *T. b. gambiense* cells. Saturation of C3b$_{AF594}$ with ISG65 (by addition of a 1.5 molar excess of the latter) led to a 60% reduction of C3b surface binding and a 40% reduction of AF594 positive cells after 5min of uptake. The addition of *T. b. gambiense* ISG75 had no effect on C3b surface binding or internalisation, confirming the suspected role of ISG65 as a C3 receptor [85].

Previous research showed that *T. b. gambiense* specifically activates the alternative complement pathway without subsequently succumbing to cell lysis [87]. Using immunofluorescence staining of *T. b. gambiense* cells incubated in a variety of human sera depleted of pathway-specific complement factors, the present study revealed that factor Bb, a component of the AP convertases, but not C4b, a core component of the CP and LP convertases, is deposited onto the parasite surface. By demonstrating factor Bb deposition in complement factor C2-depleted serum, activation of the

amplification loop via minor deposition of CP/LP C3 convertase could be ruled out. Congruently, C3b deposition could be observed in C2-depleted, but not fB-depleted serum, confirming specific activation of the alternative pathway by *T. b. gambiense* [85]. The absence of a flagellar pocket signal for C5b, the first component of the terminal pathway, indicated that the inhibition of the complement cascade occurs as a consequence of alternative pathway restriction, upstream of terminal pathway initiation [85]. Addition of ISG65 as an inhibitor in functional haemolytic assays resulted in almost complete inhibition of rabbit erythrocyte lysis upon activation of the AP (AP50 test, Haemoscan), but had no discernible effect on lysis of sensitised sheep erythrocytes following induction of the classical pathway (CH50 test, Haemoscan). Addition of ISG75 in either assay had no measurable impact, confirming that ISG65 specifically restricts alternative pathway activity [85].

Based on SPR experiments demonstrating successful AP C3 pro-convertase formation and computational modelling suggesting unimpaired activity of the mature AP C3 convertase in presence of ISG65, ISG65 mediated inhibition of the AP C5 convertase was assessed using an adapted AP50 assay (Haemoscan). Addition of ISG65 to erythrocytes coated with pre-assembled AP C5 convertase decreased haemolysis by up to 40%, indicating that ISG65 binding to the AP C5 convertase can, at least partially, abrogate C5 substrate cleavage *in vitro* [85]. Addition of ISG75 had no measurable effect on haemolysis.

Throughout the progression of the complement cascade, complement factor C3 undergoes a range of proteolytic cleavages, resulting in a variety of fragments, each with specific functions within the pathway (*see* I.3.2.1). SPR analysis revealed that native C3 has the lowest affinity for ISG65 (K_D = 130 nM), followed by C3b (K_D = 81 nM) and C3-methylamine (C3MA), a mimic for $C3(H_2O)$ (K_D = 18 nM). Noticeably, the increase in the measured affinity correlates well with the decrease in abundance of the respective fragments in human serum [76, 78, 235]. The highest affinity binding was observed for C3d (K_D = 7 nM), the proteolytically liberated thioester domain (TED) of C3. On the contrary, no binding could be observed for C3c, indicating that C3d is the minimally interacting domain [85].

To obtain further mechanistic insight into the restriction of alternative pathway by *T. b. gambiense*, the structures of ISG65 in complex with C3 and C3b were determined by single-particle cryo-EM. The ISG65:C3 complex could be reconstructed to a Fourier shell correlation ($FSC_{0.143}$) of 3.58 Å, resulting in a near complete model

of C3 exhibiting high overall similarity (RMSD = 2 Å, 11.878 atoms) to the previously published structure of C3 (PDB 2A73) [236], and an atomic model of *T. b. gambiense* ISG65, comprising residues Thr35-Ala145, Thr203-Val226 and Glu256-Ala312 (~55% of the ISG65$_{18-363}$ construct used) [85]. N-terminal residues Leu18-Leu34, Arg146-Ser202 and Lys227-Met255, constitute intrinsically disordered loops and could not be resolved. Only poor density was obtained for residues following Ala312 and the disordered C-terminal linker could not be resolved in its entirety. While ISG65 adopts the canonical three-helix bundle fold commonly observed in trypanosome surface receptors of known structure, it exhibits an unusual 20° curvature along its longest axis, facilitating an almost perfect complementary fit to the convex shape of the C3 TED. Binding of ISG65 to the TED is predominantly mediated via hydrophobic contacts and few hydrogen bonds [85]. Residues in the binding interface thought to contribute to the binding were unequivocally confirmed via SPR analysis; individual residues on ISG65 or recombinantly expressed C3d were mutated to alanine, the binding affinity determined and compared to the wild-type proteins. Furthermore, HDX-MS placed all residues modelled to form the interface in regions of low deuteration, thereby confirming their correct identification. In addition to the primary interface with the TED, ISG65 forms a second, smaller interface with the ANA domain (referred to as C3a when proteolytically liberated) of C3 [85]. SPR analysis confirmed binding of recombinant C3a to ISG65. Despite poor side chain densities, two potentially interacting residues could be identified on ISG65 and their identity was confirmed by a reduction in binding affinity upon mutation to alanine.

The complex of ISG65:C3b was reconstructed to a reported FSC$_{0.143}$ of 3.59 Å [85]. The local resolution however varied significantly across the reconstruction, reaching only 5-10 Å in the interface between C3d and ISG65. While secondary structure features like α-helices in ISG65 and β-sheets of the CUB-domain in C3b could still be identified, lack of higher resolution did not permit atomic modelling of ISG65 or the interface. Superposition of the ISG65 model obtained from the ISG65:C3 reconstruction indicated an identical primary interface in both complexes. The majority of C3b could be modelled, producing a model closely resembling the previously published structure of C3b (PDB 2I07) (RSMD = 2 Å, 11.068 atoms) [237]. Minor rearrangements in CUB and TED could be observed and may be caused by binding of ISG65. Coulomb potential corresponding to the disordered head domain of ISG65 suggested a secondary interface with the CUB domain of C3b [85].

To elucidate how ISG65 may exert its function as a receptor for C3b in the context of the VSG surface coat, conformational flexibility of the C-terminal linker was assessed using SAXS and the ensemble optimisation method (EOM) [238]. Consequently, two distinct conformational pools were identified. While the majority of models in the selected ensemble were compact (~100-115 Å), a small subset adapted an extended conformation (~170 Å) [85]. Using AlphaFold2-assisted molecular modelling, a complete model of full-length ISG65$_{18-436}$ was generated. Predicted models of the cytoplasmic domain and the membrane-embedded C-terminal helix were appended to the models of the two distinct conformational states. A model of the disordered N-terminal head domain was generated using AlphaFold2-based, template-guided structure prediction, applying real-space constraints of the cryo-EM reconstruction and MS-based disulphide mapping as selection criteria for the models. The final hybrid model was deposited in the PDBDev database (PDBDEV_ 00000201). Based on the reported findings, a model for the interaction between ISG65 and C3b on the surface of the parasite was proposed [85]. In the compact conformation, ISG65 resides within the VSG coat, exposing only the disordered head domain to the hosts immune system, concealing other epitopes under the protective VSG umbrella. When fully extended, ISG65 protrudes the VSG coat by ~60 Å, thereby enabling the interaction with C3b.

4.2.3 Discussion

In this study, the invariant, immunogenic yet abundant surface protein ISG65 from the blood-stage form of the human parasite *T. b. gambiense* [117, 119] has been identified as a receptor for human complement C3b. The reported findings indicate that ISG65 facilitates innate immune escape by specifically restricting the alternative complement pathway at the level of the C5 convertase, while exerting no control over the classical or lectin pathway [85].

The study demonstrated that binding of ISG65 to native C3 does not prevent formation of the AP C3 proconvertase and computational modelling suggests that assembly of the enzyme-substrate complex (C3bBb:C3b) or subsequent conversion of the substrate are not inhibited in the presence of ISG65, either. On the contrary, addition of ISG65 to the AP C5 convertase, pre-assembled on erythrocytes, significantly reduced subsequent cell lysis. Addition of erythrocytes into serum pre-incubated with

ISG65 abrogated haemolysis almost completely, indicating that *in vitro*, complete inhibition of the AP C5 convertase requires ISG65 interference with both, convertase formation and substrate cleavage [85]. The underlying molecular mechanisms of the inhibition are yet to be elucidated as no structural information about the architecture of the AP C5 convertase is available to this date.

While this study was under review, the crystal structure of ISG65 from the animal infective *Trypanosoma brucei brucei* (*Tbb*) in complex with human C3d (PDB 7PI6) was published [88]. Alignment of *T. b. brucei* ISG65:C3d to the presented cryo-EM structures reveals an overall high similarity of the assemblies ($RMSD_{ISG65:C3 \text{ vs } 7PI6}$: 1.72 Å (1791 atoms); $RMSD_{ISG65:C3b \text{ vs } 7PI6}$: 1.03 Å (1954 atoms)). Congruently, most interface residues occupy similar positions amongst the models. Although positioned in a flexible loop region, the ISG65:C3 cryo-EM structure shows a well resolved side-chain density for ISG65 Trp211, constituting two hydrophobic interactions with the TED. The importance of these interactions was confirmed by HDX-MS and affinity measurements with an alanine mutant, resulting in a significant reduction of the binding affinity [85]. Surprisingly, Trp211 does not engage with C3d in the *Tbb*ISG65:C3d complex [88].

Noticeably, SPR data reported by the authors of the aforementioned study show results that markedly diverge from those reported here. Based on their affinity measurements, the authors proclaim the identification of a secondary binding site within C3c [88]. For their experiments, the authors reportedly immobilised ISG65, site-specifically biotinylated at the C-terminus, onto streptavidin coated SPR chips until a response of 800 response units (RU) was achieved. Injection of 500 nM C3c resulted in a maximum response (R_{max}) of 122 RU. Using a 1:1 binding model to perform a kinetic fit, a K_D of 750 nM is reported. Neither the low curvature of the sensorgrams, the reported K_D well outside of the range of measured concentrations, nor the insufficient injection times used during the measurement instil confidence in the conclusions drawn by the authors. Reliable determination of binding kinetics necessitates an accurate estimation of the R_{max}. While complete saturation of the ligand is not required in order to do so, a wide concentration range for the analyte, typically 0.1x to 10x the K_D, should be used and ideally at least one concentration should reach equilibrium [239]. Importantly, it should be noted that the kinetic parameters reported in the supplementary information of the aforementioned study do not reflect the kinetic parameters reported in the main text and figures, casting further

doubt on the reported findings [88]. The proteins used by the authors for the presented affinity studies were reportedly purchased from CompTech (Texas, USA) but no data validating the identity or heterogeneity of the product are presented [88]. In early attempts of determining the structure of *T. b. gambiense* ISG65 in complex with C3 via cryo-EM during this study, human C3 protein was purchased from CompTech as well. During assessment of the products homogeneity, a heavy contamination with C3b was determined (data not shown) and the product deemed unsuitable. While merely speculative, it cannot be excluded that the commercially available C3c used in the study contains traces of a fragment with binding affinity to ISG65, such as C3, C3b or C3(H_2O). The presence of a small fraction of an ISG65-binding contaminant may also explain the comparatively little response upon injection of C3c. While the data presented by the authors unfortunately precludes a detailed analysis, it should be noted that injection of 500 nM of C3b resulted in a maximum (recorded) response of 832 RU [88]. Although saturation of the ligand has not been achieved, as indicated by the continuously steep slope of the sensorgram throughout the injection, it may illustrate the inconsistencies within the binding affinity data presented for C3c. The highest concentration of C3 used exceeds a 10-fold molar excess of the reported K_D [88], theoretically a concentration sufficient to induce ~90% of R_{max} [239]. Hypothetically, the expected response for C3c at concentrations approximately equal to the reported K_D should elicit ~50% of R_{max} [239]. Based on the recorded maximum response for C3b and the difference in molecular weight between the two C3 fragments, assuming similar association kinetics (9×10^4 $M^{-1}s^{-1}$ and 2×10^4 $M^{-1}s^{-1}$, [88]), injection of 500 nM C3c should have resulted in a response of 614 RU by the end of the injection whereas only 122 RU were observed by the authors [88]. While this comparison is only illustrative and necessitates many assumptions, it supports the suspicion of the presence of an ISG65-binding contaminant in the C3c sample used in the aforementioned study. In addition to the poor quality of the SPR data underlying the authors claim for ISG65 binding to C3c, the lack of evidence for C3c binding in the cryo-EM structures and the contradictory SPR data presented in this study [85] indicates that ISG65 does indeed not bind to C3c.

Although the reported affinity of *T. b. brucei* ISG65 for human C3b (40 nM) [88] is comparable to the K_D determined here for *T. b. gambiense* ISG65 (81 nM) [85], it should be noted that most methodological errors outlined in the context of C3c are also applicable to the other SPR experiments presented in the study [88].

Conspicuously, analysis of the binding kinetics of *T. b. brucei* ISG65 to commercially obtained C3b (CompTech) necessitated the use of a 2-state model [88]. *T. b. gambiense* ISG65 binding to purified C3b, whose identity and homogeneity had been confirmed via intact mass-spectrometry, N-terminal sequencing, Western blotting and IEX, could be analysed using the Langmuir model [85, 240].

In a subsequent study, the authors present a cryo-EM structure of *T. b. brucei* ISG65 in complex with human C3b [241], confirming the secondary interface between ISG65 and the CUB domain predicted here [85]. Based on a higher local resolution of the reconstruction, the authors attribute the secondary interface to the loop surrounding residues Tyr190 on ISG65 and residue Arg954 on C3b [241]. While plausible, the related PDB and EMDB entries were not publicly accessible at the time of analysis (accessed 13.02.2024) and the claims could not be independently verified. Notably, the authors attribute the presence of the aforementioned secondary interface to both, preferential binding of *T. b. brucei* ISG65 to C3b over native C3, as well as the previously observed 2-state binding [88, 241]. Strikingly, neither do the authors provide experimental evidence, such as binding affinity measurements using alanine mutations, to support their hypothesis, nor do they acknowledge that ISG65 also forms a secondary interface with the ANA domain in native C3 [85].

Contrary to C3b, the reported affinity for *T. b. brucei* ISG65 binding to C3d (K_D = 602 nM) [88] stands in stark contrast to the findings reported here (K_D = 7 nM) [85]. It should be noted that *T. b. brucei* is not human infective and the affinity of *Tbb*ISG65 to human C3 and its fragments could conceivably be lower than that of ISG65 from the biologically relevant parasite *T. b. gambiense*. However, a structural comparison of the complex interfaces in *Tbb*ISG65:C3d and *Tbg*ISG65:TED places most interacting residues, with exception of Trp211, in near identical positions [85]. While mutation of Trp211 to alanine decreased the binding affinity of *T. b. gambiense* ISG65 to C3d to 67 nM [85], the absence of this interaction alone does not appear to account for the marked deviation. Although the binding affinity measurements of *Tbb*ISG65 and C3d are reported alongside with and compared to those of C3b and C3c, they were performed using an entirely different experimental setup [88].

Whereas *Tbb*ISG65 had previously been site specifically biotinylated at its C-terminus and immobilised via binding to streptavidin, thus mimicking the likely orientation of the protein on the parasite surface, it was immobilised via non-specific amine coupling for binding affinity measurements of C3d [88]. Reportedly, 345 RU of ISG65 were

immobilised [88], resulting in a theoretical R_{max} of ~300 RU for the marginally smaller C3d fragment. Based on the presented K_D of 602 nM [88], injection of the highest concentration of C3d (4 µM) should, theoretically, elicit a maximum response of at least 150 RU [239]. Despite reaching equilibrium, an injection of 4 µM C3d resulted in a maximum response of 54 RU [88], indicating that as a result of the chosen immobilisation technique, the vast majority of ligand molecules were inaccessible for binding of C3d. While the fraction of inactive ligand molecules does not influence the binding kinetics, previous studies have demonstrated that the presence of free reactive amine groups close to the binding site can reduce the affinity for the analyte [242]. The authors did not disclose the reason for choosing a different immobilisation technique [88]. Surprisingly, the binding kinetics were fitted using the Langmuir model although the goodness of fit to the dissociation rate appears rather low [88]. In light of the fast dissociation kinetics reported and based on the concentration range presented in the study [88], a bulk shift, for example caused by unspecific binding, cannot be distinguished from fast kinetics of a biomolecular interaction [239]. To instil confidence in the presented measurement, saturation of the ligand at high analyte concentrations should have been demonstrated. From their analysis, the authors conclude that the observed binding for C3b is best described by a rapid, initial binding of ISG65 to the TED, as (supposedly) observed for binding to C3d, followed by a secondary binding event which increases the overall binding affinity [88]. This conclusion directly contradicts the cryo-EM structures and binding affinity data presented here [85]. Using the same experimental setup for measurement of binding affinity data of *Tbg*ISG65 to all C3 fragments, data obtained for C3b, C3(MA) and C3d could be fitted using the Langmuir model [85]. On the contrary, native C3, which was demonstrated to form a secondary binding interface with ISG65 via the ANA domain, had to be fitted with a two-state model [85]. Based on the recent structural insights on the nature of the proposed secondary interface of *Tbb*ISG65 and the CUB domain on C3b [241], is seems highly unlikely that the coordination of a single residue constitutes a secondary binding event that markedly stabilises the interaction over 49 bonded and non-bonded contacts found in the primary complex interface [85].

To elucidate the biological role of ISG65 binding to C3b, the authors of the aforementioned study performed mouse infection experiments with a *T. b. brucei* ISG65 knock-out strain [88]. Based on the observation that initial parasite burden in $C3^{+/+}$ and $C3^{-/-}$ mice increased at an equivalent rate, the authors concluded that neither

the alternative nor the lectin pathway of the complement system control infection. Instead, based on the observation of a decrease in parasitaemia progression of the ISG65 K.O. strain, temporally coinciding with the induction of infection-specific IgGs, the authors propose the CP as the likely target of ISG65 [88].

While the alternative pathway activation via 'tickover' formation of C3(H_2O) and assembly of a liquid-phase C3 convertase has been well established, the precise rate of activation and relevance of the AP as a surveillance system for pathogens are still controversial [243]. Regardless, it is widely accepted that formation of C3(H_2O) is a weak initiator of the complement cascade in comparison to initiation of CP or LP [243]. Studies of *T. b. rhodesiense* infections in humans have estimated that, on average, a subcutaneous dose of 300-450 metacyclic parasites might be sufficient to establish a systemic infection [244]. Intraperitoneal injection of 1×10^6 BSF parasites, as performed in the aforementioned study [88], does not only omit events associated with the natural infection pathway [17], but also exceeds the biologically relevant infective dose by more than 3 orders of magnitude [244]. In conjunction, these findings suggest that administration of the reported parasite dose [88] instantaneously overwhelmed the protective capacity of the AP, not permitting any conclusion about activation of the latter. In addition to an excessive initial parasite burden, the first assessment of trypanosome growth was conducted only one day after infection [88], effectively precluding the detection of the AP's prompt actions, which are known to unfold within minutes of exposure to pathogens [245].

Despite acknowledging earlier studies demonstrating that *Trypanosoma brucei gambiense* specifically activate the AP [87], the authors do not comment on the fact that the same study also demonstrates that the CP is not activated [88]. No additional evidence supporting the proposed role of ISG65 as an inhibitor of the classical pathway of the complement is presented [88].

On the contrary, immunofluorescence experiments in the study presented here are in good agreement with earlier studies [87], indicating that the CP/LP does not progress to the formation of the CP/LP C3 convertase C4bC2a [85]. Whether the CP/LP is not triggered or the lack of C4b on the parasite surface is a result of inhibition of the CP/LP prior to C4b deposition was not investigated [85]. Identification of AP convertase components on the surface of *T. b. gambiense* and a lack of C5b, the first component of the terminal pathway, suggested inhibition of the alternative pathway upstream of terminal pathway initiation. Congruently, abrogation of human serum induced

haemolysis by soluble IGS65 was only observed following initiation of the AP but not CP [85]. While the findings reported in this study strongly suggest a role of ISG65 as an inhibitor of the alternative complement pathway *in vitro* [85], further studies will be needed to elucidate the structural mechanism of ISG65 mediated inhibition of the AP C5 convertase and the significance of these findings *in vivo*.

4.3 Beyond the VSG Layer: Exploring the Role of Intrinsic Disorder in the Invariant Surface Glycoproteins of African Trypanosomes.

4.3.1 Author's contributions

My contribution to the work included (1) joint recombinant protein expression and purification; (2) purification of *T. b. gambiense* VSG LiTat3.1 from native source; (3) joint data processing and analysis of SAXS data; (4) joint analysis of HDX-MS data; (5) joint SAXS driven ensemble modelling and subsequent analysis with a corresponding author; (10) joint preparation of the manuscript with the corresponding authors; (11) preparation and/or revision of all figures and (12) review and editing of the manuscript.

4.3.2 Results

Sparked by novel insights into the conformational flexibility of *T. b. gambiense* ISG65 and its role as a receptor for C3b in the context of the VSG coat [85], this study investigated whether prototypical invariant surface glycoproteins *Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75 may exhibit similar conformational dynamics.

All atom models of *Tbb*VSG LiTat1.5, *Tbg*VSG LiTat3.1 and the aforementioned *Tbg*ISGs were predicted by AlphaFold2 with high confidence [207]. For the VSGs, especially the N-terminal domains and the dimer interfaces appear to be modelled with high accuracy as indicated by low predicted aligned error (PAE), good pDockQ scores with very high positive predictive values (PPV) and good interface predicted template modelling (ipTM) scores. Noticeably, AlphaFold2 predicted an unusually long C-terminal helix for *Tbg*VSG LiTat3.1. While good pLDDT scores suggest reliable prediction of that protein region, this architecture has not yet been described for any

other VSG. Structural alignment classified *Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1 as Class A1 and Class A2 VSGs, respectively (*see* I.3.3) [93]. Similarly, the N-terminal domains of the ISGs were predicted with high confidence and are in good agreement with recently published experimental structures of *Tbg*ISG65 [85, 88]. Although significantly lower pLDDT scores reported for the C-terminal domains of *Tbg*ISG43 and *Tbg*ISG64 suggest intrinsic disorder [246], the C-terminus of *Tbg*ISG75 is predicted with reasonably high confidence to consist of two long, intertwined, antiparallel α-helices [207].

In agreement with the predicted structures, circular dichroism spectroscopy (CDS) confirmed high α-helical content of the *Tbg*ISGs. Contrastingly, secondary structure analysis of the CD spectra also revealed a notable degree of disorder unaccounted for by the AlphaFold2 models [207]. Congruently, Kratky analysis of small angle x-ray scattering experiments demonstrated significant flexibility of the proteins, suggesting that the solution behaviour would be best approximated by a conformational ensemble. Predictably, theoretical scattering curves derived from the single-conformer AlphaFold2 models were insufficient to explain the observed solution behaviour of the proteins under study [207].

Using the molecular dynamics based BILBOMD protocol for generation and selection of conformational ensembles against experimental SAXS data [226], the solution behaviour of the studied VSGs was best approximated with a rigid N-terminal domain and flexible CTD, contrasting the single-state conformers predicted by AlphaFold2 [207]. The experimental scattering data for *Tbb*VSG LiTat1.5 was best explained by an ensemble composed of five models, 40% of which adapted a compact conformation (D_{max} = 155 Å) and 60% displayed various degrees of extension (D_{max} = 196 Å for the most extended conformer) [207]. Similarly, the ensemble with the best fit to the experimental data for *Tbg*VSG LiTat3.1 was also obtained when granting full conformational flexibility to the CTD, revealing a primarily compact conformation (D_{max} = 167 Å) and only 21.5% extended conformers (D_{max} = 211 Å) [207]. These findings were in good agreement with an earlier study investigating the structural flexibility of *Tbb*VSG MiTat1.1 and *Tbb*VSG IITat1.24 via a SAXS-based, rigid body modelling approach [209], thereby validating the molecular dynamics-based protocol utilised here.

To assess whether the low pLDDT scores (< 50) in the C-terminal regions of the AlphaFold2 predictions for *Tbg*ISG43 and *Tbg*ISG64 correlate with intrinsic disorder

in the protein, HDX-MS was employed. As exemplified on *Tbg*ISG65, protein regions of lower complexity typically exhibit faster deuteration rates than their well-folded counterparts [229]. Congruently, the canonical three-helical bundle predicted for *Tbg*ISG43, *Tbg*ISG64 and *Tbg*ISG75 with high confidence exhibited significantly lower deuteration rates than the remaining residues. Residues predicted to form the disordered loops of the membrane-distal head domains as well as the C-terminal domains of *Tbg*ISG43 and *Tbg*ISG64 exhibit large degrees of deuteration already after 20s, supporting the AlphaFold2 prediction [207]. Underlining the predictive power of AlphaFold2, residues predicted to form C-terminal helices in *Tbg*ISG43 and *Tbg*ISG64 showed markedly reduced initial deuteration rates, suggesting the presence of secondary structure elements in solution [207].

While deuteration rates of *Tbg*ISG43 and *Tbg*ISG64 aligned well with predicted structural features, the experimental data for *Tbg*ISG75 does not support the existence of the long, C-terminal pair of intertwined α-helices in solution. However, in contrast to the other *Tbg*ISGs studied, the CTD of *Tbg*ISG75 did not appear strictly disordered either. Residues Arg272-Arg317 and Lys339-Ala370 deuterated markedly slower than truly disordered parts of the protein but showed increased deuteration rates in comparison to residues of the predicted three-helical bundle [207]. A significant increase in deuteration rates around residue Lys255 suggested similar domain boundaries of the three-helical bundle as observed for the other *Tbg*ISGs. In sum, these findings indicated that while the C-terminus of *Tbg*ISG75 is largely disordered, it is likely to undergo structural transitions and therefore may occur partially folded in solution. Corroborated by the findings that the single-state AlphaFold2 models inadequately approximated the experimental scattering data, it was concluded that the C-terminal regions of *Tbg*ISG43, *Tbg*ISG64 and *Tbg*ISG75 are subject to a significant degree of intrinsic disorder [207].

To further elucidate the solution behaviour of *Tbg*ISGs, insights into protein flexibility inferred from the HDX-MS data were used as restraints in SAXS-based ensemble modelling via the BILBOMD protocol. In general, N-terminal domains were treated as rigid bodies whereas the C-terminal regions were considered disordered. For all ISGs under study, a good fit to the experimental data was obtained with an ensemble composed of one compact and one highly extended conformer each [207]. The largest deviation between chosen ensemble and the experimental data was observed for *Tbg*ISG43 (χ^2 = 3.28) whereas the *Tbg*ISG64 ensemble could be fitted with a

χ^2 = 1.51. Based on the apparent structural promiscuity of the C-terminal region in HDX-MS, *Tbg*ISG75 was modelled more extensively. A range of ensembles was generated by restricting conformational sampling in either or both of the protein regions characterised by lower deuteration rates, thus preserving the α-helical fold for these residues. Alternatively, no restrictions outside the NTD were applied, accounting for a fully disordered C-terminus. While a two-conformer ensemble with good fit to the experimental data (χ^2 = 1.08 to 1.66) was obtained regardless of the applied restraints, the nature of the ensemble changed markedly. When the C-terminus was granted full flexibility, the ensemble was composed of nearly equal fractions of compact and fully extended conformers, spanning a maximum distance of 90 and 276 Å, respectively [207]. While the D_{max} of the respective conformers did not change drastically when retaining the α-helical fold in residues Arg272-Arg317 (D_{max} = 112 and 308 Å), the fraction of fully extended conformers decreased to only 11%. Despite the noticeable difference, the theoretical scattering curves of both ensembles approximated the experimental SAXS data nearly to the same degree (χ^2 = 1.08 vs χ^2 = 1.1) [207]. Although similar observations were made when the helical fold of residues Lys339-Ala370 was retained, the resulting ensemble showed the least optimal fit to the experimental data (χ^2 = 1.66) [207]. Simultaneously restraining conformational sampling in both residue regions resulted in an ensemble equally representative of the experimental data as a completely disordered C-terminus (χ^2 = 1.08). 86% of conformers were modelled compact, spanning a maximum distance of 103 Å, whereas the remaining 14% stretched to 308 Å [207].

4.3.3 Discussion

The ability of AlphaFold2 to rapidly generate highly accurate, all-atom models without any experimental input has undoubtedly revolutionised the field of structural biology [213]. Despite its relatively recent inception, AlphaFold2 has already proven a valuable tool in a wide variety of applications [247]. However, by generating only a single conformer, the prediction does not account for the behaviour of proteins in solution, severely limiting insights into their potential biological function. This limitation is further exacerbated by intrinsically disordered proteins (IDPs) or proteins with intrinsically disordered regions (IDRs). In such cases, an accurate description of the

conformational flexibility of a protein in solution may be ascertained by combination of AlphaFold2 with SAXS-based conformational sampling [248, 249]. Prompted by recent work demonstrating the conformational flexibility of *Tbg*ISG65 [85], the molecular dynamics driven ensemble modelling approach BILBOMD [226] was used to investigate the structural flexibility of other *Tbg*ISGs and interpret the findings in the context of the VSG coat.

The validity of the applied protocol was established by analysis of the solution behaviour of *Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1 and comparison of the findings to earlier reports for *Tbb*VSG MiTat1.1 and *Tbb*VSG IITat1.24 [209]. In both studies, the solution behaviour of the VSGs was best described by conformational ensembles with rigid N-terminal domains and flexible CTDs. Using a SAXS-based rigid body modelling approach, Bartossek et al. observed 2.5-fold smaller differences in D_{max} between the conformational states than observed when using the BILBOMD protocol [207, 209]. Similarly, the D_{max} of the most extended conformation of the VSGs studied here exceeded the previously reported values by approximately 50 Å. Although the VSGs from both studies are members of the same protein family and were obtained through identical purifications from native source, ultimately, they are different proteins and thus could exhibit different conformational flexibility in solution. Moreover, it seems conceivable that the molecular dynamics-based approach samples a larger conformational space than the rigid-body modelling approach utilised in the previous study. Overall, the ensembles generated using the BILBOMD protocol were found to be in good agreement with those reported in the earlier study [209].

Recently published findings demonstrated that *Tbg*ISG65 is tethered to the parasite membrane via a disordered, C-terminal linker, facilitating extension beyond the boundaries of the VSG layer, enabling interaction with its ligand, human complement C3b [85]. To determine whether this is a structural feature conserved amongst other *Tbg*ISGs, HDX-MS disorder mapping was employed for *Tbg*ISG43, *Tbg*ISG64 and *Tbg*ISG75. The results demonstrated that the *Tbg*ISGs under investigation also possess a membrane-proximal IDR, anchoring the well-structured NTDs to the parasite surface. While AlphaFold2 accurately determined the intrinsically disordered nature of the C-terminal linker for *Tbg*ISG43 and *Tbg*ISG64, the C-terminus of *Tbg*ISG75 was predicted with high confidence scores to form a long pair of intertwined, antiparallel α-helices. Coincidentally, a recent study has found that AlphaFold2 can readily identify conditionally folded IDRs and IDPs [250]. The study demonstrated that

(primarily α-helical) secondary structure predictions with high and very high pLDDT scores in protein regions identified as IDRs via traditional, sequence-based disorder predictors are likely to resemble a conditionally folded conformation of the protein. Although HDX-MS analysis suggested a largely disordered C-terminal linker in *Tbg*ISG75, two regions in particular exhibited reduced deuteration rates, indicating a partial fold in solution [207]. Based on its copy number on the parasite surface, ISG75 has long been suspected to serve a function as an abundant receptor for a yet to be discovered ligand. Considered collectively, the presence of transient secondary structure in the largely disordered C-terminus, the suspected role as a receptor, and the apparent discrepancy between the experimentally determined conformational ensemble and the AlphaFold2 prediction suggests that the *Tbg*ISG75 C-terminal linker may be a conditionally folding IDR. A similar contradiction between the experimental data and the highly confident AlphaFold2 prediction was observed for the unusually long, C-terminal α-helix in *Tbg*VSG LiTat3.1.

*Tbg*ISG65 and *Tbg*ISG75 are highly immunogenic and elicit a robust immune response detectable through the presence of antibodies in patient blood [117]. In line with these findings, immuno-affinity chromatography with *T. b. gambiense* infection IgG could selectively enrich the most abundant invariant surface glycoproteins ISG64, ISG65 and ISG75 [119]. Contrary to expectations, immunisation with ISG75 was unable to confer any protection against repeated infection in memory recall experiments [33]. These observations are consistent with earlier studies where ISG65 on *T. brucei* has been proven generally inaccessible to antibodies whereas only minimal binding to ISG75 could be observed in fixed but not live cells [117]. These findings contradict the prevailing hypothesis of the VSG umbrella as a static, protective barrier, effectively shielding the underlying invariant surface proteins from recognition by the host's immune system. While permissive to penetration by small molecules, larger molecule such as antibodies have been demonstrated to be excluded [105, 251-253]. Consequently, exposure of the ISGs to the hosts immune system must therefore occur either through increased permissiveness of the VSG coat or shedding of the antigens. In addition to the findings presented here, recent studies demonstrating conformational flexibility of *Tbg*ISG65 [85] and metacyclic invariant surface proteins (MISP) [254] have favoured the former hypothesis. In addition to flow-mediated antibody clearance facilitated by fast endocytosis and recycling of surface antigens [98], telescopic extension and retraction of ISGs may enable

biological function beyond the VSG coat while simultaneously minimising exposure to the hosts immune system, thereby effectively reducing their immunogenic footprint.

In synthesis with reports from earlier studies [85, 209, 254], the data gathered here have prompted the formulation of an updated model of the trypanosome surface coat. The revised model acknowledges the apparent conformational flexibility conserved across various invariant surface glycoproteins of *T. b. gambiense* that was previously unaccounted for. Prior studies have estimated the VSG coat to extend approximately 140 Å to 160 Å from the parasite membrane [209]. With maximum distances for the compact conformations approximated to 90 Å (TbgISG64, TbgISG75) and 129 Å (TbgISG43), all of the *Tbg*ISGs under study appear deeply embedded within. It should be acknowledged that, due to technical limitations of the applied methodology, the minimum dimensions inferred from the solution scattering data are likely to be underestimated. In absence of a transmembrane domain anchoring the protein to a lipid bilayer, the intrinsically disordered C-terminal linker may adopt conformations in solution that are not permitted on the surface of the parasite. While applicable to all ISGs studied here, this limitation is best exemplified by the compact conformer of the ensemble calculated for *Tbg*ISG64 where molecular-dynamics driven modelling placed the C-terminus in vicinity of the membrane-distal head domain [207]. Regardless of this limitation, when retracted, the majority of the *Tbg*ISG N-terminal domain is likely to be localised approximately 70 Å off the parasite surface, thereby co-residing with the C-terminal domain of the surrounding VSGs. Earlier studies have demonstrated that the latter is inaccessible to host antibodies [255], suggesting that the majority of invariant epitopes are shielded from the hosts immune system in the compact conformation, potentially exposing only the disordered, N-terminal head domains.

Considering even the largest distances observed for *Tbg*VSG LiTat3.1 (which exceed earlier reports for other VSGs significantly [209]), both *Tbg*ISG43 and *Tbg*ISG75 could extend well beyond the boundaries of the VSG surface coat when fully extended. Based on a comparatively low D_{max} of 234 Å, *Tbg*ISG64 seems to extend at least 40 Å less than the other *Tbg*ISGs under study. The NTD could extend well beyond the CTD of the VSG coat regardless, thereby enabling access to the host's bloodstream and any potential ligands therein. This hypothesis is supported by earlier reports demonstrating that *Tbg*ISG65, a cell surface receptor for complement C3b, extends only to ~170 Å [85].

Although the data presented here provide mechanistical insights into the conformational flexibility of *Trypanosoma brucei* surface coat constituents in solution, it remains to be elucidated whether these findings translate to the biological assembly *in vivo*. While an inherent ramping mechanism has been suggested for a multitude of variant and invariant surface proteins individually [85, 207, 209, 254], the molecular mechanisms choreographing that flexibility into a cohesive biological function are yet to be determined. It has been previously hypothesised that conformational switching may be a result of spatial constraints imposed by in- or decreasing surface protein density [209] but the existence of other thermodynamic or molecular triggers is conceivable. A clue towards the latter is provided by the potential function of the *Tbg*ISG75 C-terminal linker as a conditionally folding IDR. The identification of biologically relevant ligands for ISGs other than ISG65 [85] might shed more light on the matter.

Part II – Human adenovirus C5

4.4 Structural insights into the interaction between adenovirus C5 hexon and human lactoferrin.

4.4.1 Author's contributions

My contribution to the work included (1) joint cryo-EM data processing and data curation for subsequent molecular modelling; (2) SPR measurements[‡]; (3) joint preparation of figures with the first author and (4) writing, review and editing of the manuscript. [‡]*Due to technical limitations, SPR was unsuited for binding affinity determination and was subsequently not included in the published manuscript (see 4.4.3).*

4.4.2 Results

While CAR mediated cell entry of adenoviruses into epithelial cells has been well established *in vitro*, a growing body of evidence suggests the existence of alternative pathways for initial infection of epithelial cells *in vivo* [121, 135,

153, 154, 157]. Results of earlier studies have suggested that lactoferricin, the antimicrobial, N-terminal peptide of the human host defence protein lactoferrin, facilitates transduction of HAdV-C5 in epithelial cells via binding of the HVR-1 loop of the hexon capsomer [137]. The results of the study presented here elucidate the structural basis of this interaction, thereby providing further insights into the alternatives routes of species C adenovirus infection.

Based on observations made in viral transduction assays, association of hLF with the HAdV-C5 hexon was suggested to occur via binding of the alkaline Lfcin peptide to the highly acidic HVR-1 of the viral capsomer [137]. In agreement with these findings, a stable complex of the proteins could only be observed in SEC when the salt concentration of the mobile phase did not exceed 5 mM. Already 10 mM sodium chloride were sufficient to abrogate binding almost entirely [130]. The electrostatic nature of the interaction was unequivocally confirmed by native MS. In 20 mM ammonium acetate (AA) buffer, the trimeric hexon capsomer could be observed in an almost exclusively unbound state, only a small fraction of the total hexon was bound to a single hLF molecule. Reduction of the buffers ionic strength by dilution to approximately 6 mM AA resulted predominantly in formation of a complex with 3:1 stoichiometry (hexon:hLF) as well as small fractions of unbound hexon and a complex containing an additional hLF molecule [130].

Using single-particle cryo-EM, the structure of the HAdV-C5 hexon capsomer could be reconstructed to a $FSC_{0.143}$ of 2.9 Å. With the exception of hypervariable loop residues 137-164, 188-190, 252-256, 271-278 and 431-436, the trimeric assembly could be modelled in its entirety and bears a close resemblance (RMSD = 0.9 Å) to the previously published crystal structure (PDB 3TG7) [130].

While a set of good 2D class averages for 'top' and 'bottom' views of the hexon:hLF complex were obtained, strong preferential orientation, irreproducible grid preparation and a large degree of dissociation of the complex upon vitrification precluded a three-dimensional reconstruction. A previously published modelling protocol [256] utilising the XPLOR-NIH program [257] for refinement of all atom models against experimental 2D projections obtained by cryo-EM was adapted here to generate an atomic model of the hexon:hLF complex.

Distance constraints for initial rigid body docking were generated by crosslinking MS. Incubation of the hexon:hLF complex with homobifunctional crosslinker DSBU (disuccinimidyl dibutyric urea) resulted in crosslinks between hexon residues located

in the protruding turrets (Ser174, Ser446) (*see* II.1.3.1) and hLF residues in the N-lobe (Lys282) as well as the junction of N- and C-lobe (Lys676) (*see* II.2.1.3) [130]. The homotrimeric arrangement of the hexon capsomer hypothetically accommodates crosslinking of hLF to residues on different protomers. Orientation of Lfcin towards the intermolecular interface however was permitted only when both crosslinking pairs were attributed to the same protomer. Based on previous reports demonstrating the importance of Lfcin for this interaction [137], the latter arrangement was assumed to be most likely.

To append the residues missing from the experimental structure, a full-length HAdV-C5 hexon model was generated with AlphaFold2 [213], residues 137-164, 188-190, 251-256, 271-278 and 431-436 were excised from the predicted structure and grafted onto the experimental model. The arrangement of the resulting model deviated noticeably from the quaternary structure inferred from the experimental 2D projections. The class averages obtained for the hexon:hLF complex suggested the presence of a pivoting motion between the protruding turrets, resulting in a significant change in their relative positions [130]. To obtain a more accurate starting model for subsequent modelling steps, a molecular dynamics simulation of the hexon was performed. To accelerate computation, the rigid base of the model was removed and only residues 129-315 and 415-460 of each protomer were used in the simulation. A conformation closely resembling the arrangement inferred from the 2D class averages was selected towards the end of the simulation, post-convergence of the RMSD [130]. The structure of diferric hLF (PDB: 1LFG) [170] and the hexon model selected from the MD simulation were used for rigid body docking and energy minimisation in XPLOR-NIH [257]. Out of 100 generated models, the ten lowest energy solutions without violations of the distance constraints imposed by crosslinking MS were used in subsequent refinement against the experimental 2D projections. From the ten lowest energy solutions generated in the refinement step, the model with the highest cross-correlation score to the experimental data [256] was chosen for subsequent analysis.

In addition to the predicted interaction between Lfcin and the HVR-1 loop of the hexon, the selected model also exhibited a secondary interface between the hLF C-lobe and an adjacent hexon protomer. This finding placed the crosslink previously postulated to be located on the same protomer as the interacting HVR-1 onto a different subunit of the hexon trimer, contradicting the earlier hypothesis [130].

To instil confidence in the interactions inferred from the computationally generated model, binding affinity of hLF to fluorescently labelled HAdV-C5 hexon was assessed via microscale thermophoresis (MST). Charge-reducing point mutations were introduced in residues suspected to facilitate protein binding. Single point mutation of Lfcin residues in direct vicinity to the hexon HVR-1 loop (R24S, R27A, K28A, and R30A) reduced binding affinity of all respective constructs in comparison to the wild-type protein, albeit to various degrees (6 to 53% reduction) [130]. Congruently, a quadruple mutation encompassing all four residues diminished the binding affinity to only 19% of that of the wild-type protein, amounting to the largest reduction amongst all mutants tested [130]. Due to their close proximity to Lfcin and the HVR-1 loop, the residues adjoining Lys282, one of the two hLF residues found to form crosslinks with the HAdV-C5 hexon, were also suspected to facilitate binding. While this region of hLF has not been associated with binding to the viral capsomer previously, a triple mutation (K277S/K280A/K285S) reduced binding affinity by 80% in comparison to the wild-type protein, thereby suggesting a significant role for these residues in complex formation [130]. For validation of the secondary interface between hexon and the C-lobe of lactoferrin, two double mutants (N673A/S678A and E687A/K691A) were introduced in the residues surrounding the second crosslinking residue of hLF, Lys676. Although less pronounced than the observations made for the mutations introduced in the N-lobe, the double mutants exhibited a reduction in binding affinity to 58% and 32%, respectively, confirming a considerable contribution of the C-lobe to the overall binding of HAdV5-C5 hexon binding [130].

4.4.3 Discussion

In addition to other antimicrobial activities [173, 174], human lactoferrin acts as an antiviral against a wide range of both RNA- and DNA viruses [171]. Binding of hLF to invading viruses or their cell surface receptors prevents viral attachment to the host cell surface, thereby inhibiting infection [171]. In striking disparity, previous studies demonstrated that binding of both full-length lactoferrin and its N-terminal lactoferricin peptide to the hexon capsomer of species C adenoviruses exerts the adverse effect, promoting infection of epithelial cells instead [136, 137]. Based on a series of viral transduction assays, the authors hypothesised that the binding is facilitated by

electrostatic interactions between the highly acidic HVR-1 loop of the hexon and the predominantly basic N-terminus of lactoferrin [137]. The biochemical and structural findings outlined in this study provide novel insights into CAR-independent infection pathways and possibly shed light onto an immune evasion strategy employed specifically by species C adenoviruses.

Formation of the hLF:hexon complex in SEC and native MS was observed only in buffers of low ionic strength, supporting the suspected charge-driven nature of the interaction. Already the presence of 10mM NaCl or 20 mM ammonium acetate abrogated formation of the complex almost entirely [130]. These findings appear to contradict earlier transduction experiments which were performed in Dulbecco's modified Eagle's medium (DMEM) [137]. Alongside sodium chloride concentrations exceeding 100 mM, DMEM contains a plethora of other inorganic salts like calcium chloride, magnesium sulphate, potassium chloride, a multitude of amino acids and a wide range of vitamins, resulting in far greater ionic strength than the buffers used in the present study [130]. The conspicuous discrepancy between these findings may be explained by the presence of a secondary cell surface receptor, facilitating attachment to the epithelial cells used in the transduction assay. It is also conceivable that the use of complete virions and the resulting avidity of 240 capsomeres operating in conjunction may compensate for the susceptibility of purified components to the higher ionic strength buffers.

Species C HAdVs exhibit strong tropism for the eye and upper respiratory tract [122], both of which are covered by a thin layer of fluid or mucus with a tightly regulated ion balance [258, 259]. Although electrolyte concentration in the periciliary fluid of the respiratory epithelium is unknow, NaCl abundance in sputum and tear fluid have been estimated to 100 mM (± 27 nM) [259] and 140 mM [260, 261], respectively. While the relevance of hLF mediated enhancement of HAdV-C5 transduction has yet to be demonstrated *in vivo*, average NaCl concentrations reported for tissues specifically targeted by the virus and the earlier results of viral transduction assays in DMEM [137] indicate that the requirement for low ionic strength buffers for complex formation may be caused by the use of purified components.

Efforts to elucidate the structural details governing the binding of hLF to the viral capsomer via single particle cryo-EM were unsuccessful due to inconsistent sample preparation and strong preferential orientation of the hLF:hexon sample upon vitrification. An integrative modelling approach, combining experimental distance

constraints obtained via crosslinking MS, the generation of starting models using protein structure prediction software and molecular dynamics simulations, rigid-body docking and subsequent refinement against 2D projections was used to generate a plausible model of the HAdV-C5 hexon capsomer in complex with human lactoferrin [130]. While providing a good approximation of the complexes overall arrangement, detailed information about residues constituting the two identified protein-protein interfaces could not be inferred from the model [130].

Site-specific mutagenesis of hLF in conjunction with MST confirmed correct prediction of the interfaces and established key residues facilitating the interaction. Previous findings showed that C-terminally truncated Lfcin$_{1-18}$ does not enhance viral transduction efficacy [137]. Congruently, the biggest reduction of binding affinity was observed when mutating positively charged residues R24S, R27A, K28A and R30A in the N-terminal Lfcin peptide [130]. A similar decrease in binding affinity could be observed when mutating solvent accessible residues in a protein region adjacent to the Lfcin peptide. While not previously implicated in hexon binding, charge reducing mutations also revealed a significant contribution to the overall binding affinity [130].

Analysis of crosslinking-MS data indicated the presence of a secondary interface between hexon and the C-lobe of lactoferrin. Mutational analysis of interface residues demonstrated a notable decrease in binding affinity, albeit to a lesser degree than observed for residues of the N-lobe [130]. While these findings are in good agreement with the experimental 2D class averages obtained for the hexon:hLF complex, their significance remain to be elucidated *in vivo*. A secondary interface with little contribution to the overall stability of the complex could facilitate correct orientation of lactoferrin for subsequent binding to a cell surface receptor. Viral transduction assays previously demonstrated that full-length hLF increased transduction efficacy 4-fold whereas equimolar concentrations of Lfcin induced a 6-fold increase in transduction [137], indicating that the hLF C-lobe is not required to bridge the interaction between virus and epithelial cell.

It is conceivable however that correct orientation of full-length lactoferrin is required for binding of a surface receptor on cell types other than epithelial cells. A multitude of lactoferrin receptors have been reported across a wide range of different cell types [262]. Previous reports showed that lactoferrin induces TLR-4 mediated uptake of HAdV species B, C and D in dendritic cells [169]. While Lfcin marginally increases HAdV-B35 transduction into epithelial cells, the HVR-1 region of species D

adenoviruses, such as HAdV-D37, carries a positive net-charge similar to lactoferricin, likely precluding binding of the latter [137]. In synthesis, these observations suggest that adenoviruses may have developed several distinct mechanisms to hijack lactoferrin via the hexon capsomer, figuratively turning it into 'general key', providing access to a variety of different cell types.

Despite the homotrimeric arrangement of the hexon, no capsomers harbouring three copies of lactoferrin were observed in native MS experiments. The predominant form of the complex contained a single lactoferrin molecule and only a minor fraction containing two hLF molecules was identified. The cryo-EM 2D class averages suggested a prominent conformational change in the protruding turrets of the hexon upon lactoferrin binding [130]. Structural alterations during binding of the first lactoferrin copy could diminish the affinity for subsequent binding whereas binding of a second ligand may inhibit the binding of a third entirely. The proposed hypothesis may also explain why full-length hLF increased HAdV-C5 transduction efficacy in epithelial cells 4-fold while equimolar concentrations of Lfcin induced a 6-fold increase [137]. The 49 residue lactoferricin peptide is unlikely to encounter steric inhibition, enabling binding of an additional molecule compared to the much larger full-length protein. This could potentially account for the observed increase in transduction efficacy [137].

To assess the binding dynamics governing the interaction between lactoferrin and hexon, SPR measurements were performed. While an earlier study investigated the binding affinity of hexon protein from various adenovirus serotypes to lactoferrin and lactoferricin via SPR, depiction of any sensorgrams is omitted from the publication, precluding analysis of the underlying dynamics or stoichiometry of the interaction [137]. The authors reportedly immobilised the hexon covalently to a CM5 sensor chip via amine coupling and performed all binding affinity measurements in HBS-EP+ (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% P20, pH 7.4) buffer [137]. Using this experimental setup, the alleged binding affinity of the HAdV-C5 hexon to full-length lactoferrin was reportedly determined to 29 nM, featuring moderate association and slow dissociation rates [137]. These reports strongly contradict the findings shown here. Using a buffer of similar ionic strength (20 mM HEPES, 150 mM NaCl, 0.02% P20, pH 7.2), MST experiments revealed a binding affinity of hLF to HAdV-C5 hexon of approximately 2 µM, a difference of two orders of magnitude [130]. The latter is in good agreement with observations made in SEC and native MS. An argument

could be made that a sensor chip densely coated with immobilised hexon might better approximate the virion surface than the use of purified, soluble components in MST, thereby potentially overcoming the higher ionic strength of the buffer through increased avidity. Adopting the widely accepted approximation that 1 RU is equivalent to a surface concentration of 1 pg/mm^2 on the sensor chip [263], the authors reportedly immobilised hexon to a surface concentration of approximately 1000 pg/mm^2 [137]. On each facet of the virus capsid, 18 hexon capsomers cover a crudely estimated surface area of 68.000 Å^2, totalling approximately 14.242 pg/mm^2, exceeding the surface concentration used in the SPR experiment more than 14-fold. Confidence in the previously reported binding affinity was diminished further by the inability to reproduce the findings. After successful immobilisation of HAdV-C5 hexon to approximately 900 RU onto a CM5 sensor chip (Cytiva) via amine coupling (Figure 12A), injection of 200 nM hLF in low ionic strength buffer (20 mM HEPES, 5 mM NaCl, 0.005% P20, pH 7.2) did not result in any specific binding (Figure 12B).
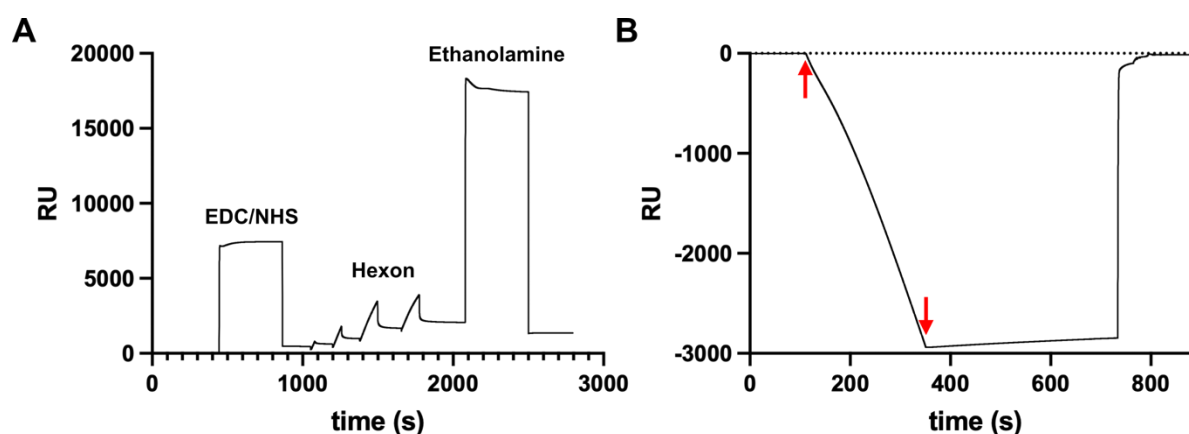
Figure 12 – SPR sensorgrams of HAdV-C5 hexon immobilisation and hLF injection.
A Immobilisation of hexon to a CM5 sensor chip. Activation of the sensor surface with EDC/NHS, four injections of ligand and subsequent quenching with ethanolamine are shown and marked accordingly. After quenching, 900 RU of hexon remained immobilised. **B** Reference subtracted sensorgram of lactoferrin binding on immobilised hexon. Start and end of the injection are highlighted with red arrows. The negative slope of the curve is caused by an increased amount of unspecific background binding in the reference channel. Injection of 1.5 M NaCl removed non-specifically bound protein, effectively regenerating the surface.

The immobilisation step was performed in 10 mM sodium acetate buffer at pH 4.5. In buffers with a pH over 3.5 the dextran matrix of the used CM5 sensor chip is negatively charged. While immobilisation of the hexon in an orientation precluding binding to

lactoferrin cannot be excluded, the highly acidic nature of the hexon turrets renders this hypothesis unlikely. Congruently, immobilisation of site-specifically biotinylated lactoferrin to a CAP sensor chip (Cytiva) and subsequent injection of 200 nM HAdV-C5 hexon, using the same low ionic strength buffer, did not result in any specific binding either (data not shown).

To elucidate the binding mechanism of human lactoferrin to the adenoviral hexon capsomer in detail and to determine whether its interaction may differ across various viral serotypes, the identification of the corresponding cell surface receptors will be crucial. The findings presented here may serve as a basis for development of Lfcin based antivirals against species C HAdVs or pave the way for novel adjuvants in gene therapy applications to facilitate viral transduction efficiency in the lungs.

An increasing number of complex biological structures are being determined via integrative modelling approaches, combining experimental data and computational methodologies [264]. While protein structure prediction via AlphaFold2 has become a versatile tool to support interpretation of cryo-EM reconstructions [265], not all samples are readily amendable to three-dimensional reconstruction. The combination of crosslinking data, rigid body docking and subsequent refinement against a limited number of unique experimental 2D class averages, as described here, produced a likely model of the studied interaction. While general validity of this approach remains to be elucidated, it could serve as a template for generation of models of other specimen which elude structural characterisation by experimental means alone.

Chapter 5

Conclusion

The primary research objective of this study was to elucidate the structural basis of immune evasion mechanisms employed by the human-infective pathogens *Trypanosoma brucei gambiense* and human adenovirus C5. Acquiring a detailed understanding of the molecular events driving such strategies is essential to advance our knowledge of the disease and thus indispensable for the development of novel therapeutics.

Part I – African trypanosomes

Despite a big surge in the number of annually published protein structures determined by cryo-EM following the so-called 'resolution-revolution' [266], x-ray crystallography remains the most commonly used method for structure determination at the time of writing (www.rcsb.org, accessed March 2024) [267]. Although tremendous advances in protein x-ray crystallography have been made since its inception in the late 1950s, the design and production of protein minimal constructs amendable to crystallisation pose a major bottleneck to this date [268]. This problem is further exacerbated if the protein exhibits an especially complex fold and no homologues of known structure are available. The advent of highly reliable protein structure prediction via AlphaFold2 [213] has been hugely beneficial to overcome the latter obstacle, but identification of protein domain boundaries based on computational models alone has proven difficult (*see* 4.1.3).

Using the example of ISG65, a prominent member of a structurally mostly uncharacterised protein family from the human-infective parasite *Trypanosoma brucei gambiense*, a multifaceted strategy for improved construct design and recombinant expression of challenging proteins was developed. Through strategic integration of well-established and widely accessible laboratory techniques such as hydrogen-deuterium exchange mass spectrometry (HDX-MS), truncation scanning and targeted expression host switching, a minimal construct that can be expressed with high yields at relatively low cost was generated [206].

Despite being one of the most abundant invariant surface proteins of the bloodstream form of *T. b. gambiense* and the subject of numerous previous studies, the biological function of ISG65 has remained elusive for the past 30 years. The work published within the scope of this dissertation identified ISG65 as a receptor for the human complement component C3b and demonstrated its role in the selective inhibition of the alternative pathway of the complement system [85]. *In vitro* experiments demonstrated that by interfering with both the assembly and function of the AP C5 convertase, ISG65 prevents progression of the complement cascade to the terminal pathway, thereby thwarting complement-mediated lytic activity. In good agreement with earlier reports [87], no evidence for activation of the classic or lectin pathway by the parasite was found. Congruently, ISG65 had no noticeable effect on classical pathway activity in haemolytic assays.

Additionally, this work presented the cryo-EM structures of native C3 and the proteolytically activated C3b in complex with ISG65, revealing two distinct modes of interaction [85]. In combination with data generated in SAXS experiments, these findings led to the proposal of a likely model for the receptor-ligand interaction in context of the VSG surface coat of the blood-stage parasite and how it may facilitate their escape from the host's innate immune system. *T. b. gambiense* ISG65 is tethered to the parasite membrane via a disordered, C-terminal linker. Conformational flexibility of this membrane-proximal region could grant the ability to adopt both compact and elongated conformations. While the well-structured N-terminal domain of ISG65 resides within the protective VSG layer in the compact conformer, upon extension the protein could reach beyond the boundaries of the protective VSG umbrella, enabling interaction with its ligand [85].

To explore whether conformational flexibility conferred by a disordered C-terminal linker may be a common feature amongst members of the *Trypanosoma brucei gambiense* ISG super-family, three prototypical ISGs of various sizes were studied. The dynamic solution behaviour of *Tbg*ISG43, *Tbg*ISG64 and *Tbg*ISG75 was assessed via an integrative structural biology approach combining AlphaFold2 based structure prediction, HDX-MS and molecular dynamics based conformational ensemble-modelling against experimental SAXS data. Similarly to ISG65, all ISGs under study possess a membrane-proximal IDR, anchoring the well-structured NTDs to the parasite surface. While single-state conformers predicted by AlphaFold2 were insufficient to explain the scattering data, the solution behaviour of all three proteins

was approximated well by a conformational ensemble composed of one compact and one highly extended conformer each. In good agreement with findings of earlier studies [209], application of the same SAXS based ensemble modelling approach to VSGs from both *T. b. brucei* and *T. b. gambiense* (*Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1) revealed a similar conformational dynamic in solution.

In synthesis with reports from earlier studies [85, 209, 254], the data accumulated here have prompted the formulation of a revised model of the trypanosome surface coat, accounting for large-scale conformational changes of surface antigens from the ISG super-family. Flexible extension of ISGs facilitated by intrinsically disordered, membrane-proximal linkers could conceivably serve as a functional feature to overcome the spatial constraints imposed by the surrounding VSG coat. Besides clear significance for the reduced immunogenicity of invariant surface proteins and the potential implications for the parasite's evasion of the host immune system, these findings could also indicate that not only ISG65, but other members of the same protein family may serve as abundant receptors for yet to be discovered ligands.

Part II – Human adenovirus C5

Although respiratory infection by species C HAdVs are widespread and over 60% of adults possess neutralising antibodies against HAdV-C5 [204], the precise molecular interactions governing initial cell entry *in vivo* remain elusive. Using immortalised cell lines, it has been firmly established that viral attachment to the cell surface is facilitated by the coxsackie and adenovirus receptor, followed by integrin mediated endocytosis [142, 149, 150, 152]. However, both CAR and integrins are predominantly localised deep within the tight junctions on the basolateral side of the polarised respiratory epithelium, inaccessible to virus particles on the apical side [121, 153]. A continuously growing body of evidence suggests that alternative, CAR independent pathways facilitate initial HAdV-C5 cell entry instead [135, 153, 157, 169]. Astonishingly, recent studies have demonstrated that human lactoferrin, a soluble component of the human innate immune response with antiviral properties [171, 172], facilitates HAdV-C5 transduction in epithelial cells in a CAR independent manner [136]. In a subsequent study the HVR-1 loop of the viral hexon capsomer was

proposed as the lactoferrin/lactoferricin binding site but a detailed understanding of the mechanism was lacking [137].

In good agreement with previous reports, the work presented here confirmed the suspected electrostatic nature of the interaction via SEC and native MS. In addition to a cryo-EM reconstruction of the HAdV-C5 hexon to a $FSC_{0.143}$ of 2.9 Å, the highest resolution structure obtained for the viral capsomer by electron microscopy yet (www.rcsb.org, accessed March 2024) [267], an integrative modelling approach was employed to construct a plausible hybrid model of the hexon:lactoferrin complex. These structures confirmed a primarily charge driven interaction between lactoferrin's N-terminal lactoferricin domain and the hexon's hypervariable region 1. Targeted introduction of charge-reducing point mutations and subsequent binding affinity measurements using MST revealed critical Lfcin contacts and identified a previously unsuspected, secondary interface between the C-lobe of hLF and an adjacent hexon protomer. In summary, the findings presented in this study provide a structural understanding of the molecular mechanism by which HAdV-C5 exploits a soluble component of the human immune system to facilitate infection of the respiratory epithelium. These insights could provide a solid foundation for the development of novel antivirals or adjuvants for improved transduction efficacy in gene therapy applications.

Chapter 6

Associated publications

6.1 A multifaceted strategy to improve recombinant expression and structural characterisation of a Trypanosoma invariant surface glycoprotein.

scientific reports

Check for updates

A multifaceted strategy to improve recombinant expression and structural characterisation of a *Trypanosoma* invariant surface protein

Hagen Sülzen[1,2], Jitka Votrubova[1], Arun Dhillon[1] & Sebastian Zoll[1✉]

Identification of a protein minimal fragment amenable to crystallisation can be time- and labour intensive especially if large amounts are required and the protein has a complex fold and functionally important post-translational modifications. In addition, a lack of homologues and structural information can further complicate the design of a minimal expression construct. Recombinant expression in *E. coli* promises high yields, low costs and fast turnover times, but falls short for many extracellular, eukaryotic proteins. Eukaryotic expression systems provide an alternative but are costly, slow and require special handling and equipment. Using a member of a structurally uncharacterized, eukaryotic receptor family as an example we employ hydrogen–deuterium exchange mass spectrometry (HDX-MS) guided construct design in conjunction with truncation scanning and targeted expression host switching to identify a minimal expression construct that can be produced with high yields and moderate costs.

Recombinant protein expression in a heterologous system highly depends on precise prediction of secondary structure boundaries to yield soluble protein. Soluble expression however is no guarantor for successful crystallisation of a protein as the presence of flexible, low-complexity regions at the termini can prevent the formation of a highly ordered crystal lattice necessary for high-resolution diffraction. The need to identify a minimal domain amenable to crystallisation therefore calls for mapping the secondary structure landscape of a target protein with the highest possible precision.

Secondary structure prediction has been extensively studied. It has seen many improvements since its inception[1–3] which renders it still the most commonly used method for the design of recombinant expression constructs. Prediction accuracy hereby depends on the chemo-physical properties of amino acids in a sequence as well as the evolutionary relationships of a protein, such as homology to proteins of known structure. Together with ever growing archives (e.g. Protein Data Bank, PDB) of experimentally determined structures, current algorithms reach an accuracy of 80–90%[4,5]. Lack of significant homology to other structurally characterised protein families however has an adverse effect, resulting in low confidence of the prediction. Recently AlphaFold, an AI system developed by DeepMind and EMBL-EBI has entered a developmental stage which does not only allow prediction of secondary but also tertiary structures of proteins with hitherto unprecedented accuracy, using a novel deep learning network[6]. This has also been reported for proteins adopting novel folds, i.e. in absence of structural homologies that could otherwise aid prediction. Examples of accurate novel fold predictions, such as proteins from extremophiles, are however still scarce and therefore require additional validation. Furthermore, prediction accuracy for secondary structure boundaries is lower than for their internal regions as the number of possible secondary structure elements is more conserved than their length[5]. Consequently, secondary structure prediction of proteins from isolated families with no structural homologues and functional assignment has to be aided by experimental data to achieve the highest possible accuracy.

The production of soluble protein alone, however, is not the only concern in recombinant protein expression. The protein of interest also has to be functional and to be produced in quantities sufficient for downstream

[1]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo namesti 2, 16610 Prague 6, Czech Republic. [2]Faculty of Science, Charles University, Albertov 6, 12800 Prague 2, Czech Republic. ✉email: sebastian.zoll@uochb.cas.cz

applications such as functional studies (e.g. Isothermal titration calorimetry), animal immunisations for antibody generation (e.g. Llamas), or structural studies (e.g. NMR and crystallisation)—some of which require considerable amounts of highly pure protein.

Traditionally, *Escherichia coli* is used for high yield expression, but due to its basic procaryotic folding machinery, its usefulness is often limited to intracellular proteins with less complex folds. Lack of glycosylation and improperly formed or absent disulphide bonds can impair or abrogate the function of extracellular proteins of eukaryotic origin when expressed in *E. coli*. Mammalian and insect cells are the most common expression systems of choice for such proteins. An alternative is the *Leishmania* expression system (Lexsy). In recent years many such systems have been commercialised as kits, allowing for scalable and reproducible production of even challenging proteins. Testing dozens of constructs however, particularly in mammalian cells, is, even in the case of transient expressions, expensive (media, transfection reagents, plastics) and time consuming (growth time, contaminations). Additionally, laboratories often lack the necessary infrastructure and expertise. Due to shorter doubling times Lexsy represents an alternative to insect cells, but it does not offer the same spectrum and variety of strains, transfection systems and expression media (e.g. serum-free media) as insect cell systems.

To overcome the described challenges, we propose a hybrid approach for the identification of a minimal expression construct that lacks extended low-complexity regions at the termini. We employ a combination of hydrogen deuterium exchange mass spectrometry (HDX-MS) and truncation scanning together with initial expression screening in *E. coli* and final high-yield production of the identified domain in insect cells.

We use ISG65 (Invariant surface glycoproteins 65) from the eukaryotic parasite *Trypanosoma brucei* as an example. Invariant surface glycoproteins, type I transmembrane proteins, have been discovered about 30 years ago[7] and since then have been the subject of numerous studies[8–13]. Nevertheless, despite many efforts, no structure of an ISG has been determined yet. This might in part be related to difficulties in construct design. While the expression constructs of crystallised GPI(glycosylphosphatidylinositol)-anchored trypanosome surface proteins terminated at the position of the omega site (GPI-anchor attachment site)[14–18], the same straightforward approach for determination of the C-terminal domain boundaries appears not to be applicable to the group of surface proteins harbouring single pass transmembrane domains (TMDs), as shown by ISG65 as an example.

We believe that the strategy outlined here could serve as a template for expression and structure determination of challenging proteins by using established technologies and combining them in a novel, targeted manner that effectively avoids time-consuming trial and error approaches for construct screening.

Results

HDX-MS as a tool to identify low-complexity regions in proteins. No homologues are known for the family of invariant surface glycoproteins. Secondary structure prediction for ISG65 using PsiPred[19,20] shows a low confidence score across the whole sequence and, particularly, in the C-terminal region (Fig. 1A). In a first approach, an initial expression construct comprising the whole extracellular domain (ECD), truncated at the position of the TMD and lacking the N-terminal signal-peptide (ISG65$_{18-386}$), was chosen for expression in *E. coli* T7 shuffle cells to promote disulphide bond formation in the cytoplasm. This construct failed to express, likely due to inaccuracies in the prediction of the transmembrane domain (TMD). A second construct with the new C-terminus further away from the TMD (ISG65$_{18-363}$) could however be produced successfully. To rationalise and accelerate the identification of a minimal construct optimised for crystallisation, we set out to guide our efforts by experimental data using HDX-MS that can be used to identify regions of low complexity that have faster rates of deuteration compared to well-structured regions[21]. Prior to MS analysis, proteolytic digestion of ISG65 was optimised to achieve maximum coverage of the protein sequence with many overlapping peptides (97%, Supplementary Figure 1). This resulted in high resolution mapping of deuterated regions.

The deuteration plot shows a steady increase in deuteration rate starting from residue 260 of ISG65 (Fig. 1B,C; Supplementary Figure 2). Uptake rates plateau at residue 300 and remain high towards the C-terminus of the protein with the exception of a trough approximately around residue 330. This points to the presence of lower secondary structure content in the C-terminal, the membrane-proximal part of the protein.

At the N-terminus deuteration declines rapidly starting from the end of the signal peptide at residue 20 until it reaches almost zero between residues 30–40, indicative for the beginning of a well-structured region which is in line with the secondary structure prediction and position of the signal peptide. Using HDX-MS analysis to approximate a well-expressing, minimal construct amenable to crystallization, we could readily establish the new N-terminus. The precise C-terminal end of the minimal expression construct however was less obvious due to a slower transition to higher deuteration rates in comparison to the N-terminal region and the presence of a local minimum (residues 320–330).

Truncation scanning established the C-terminal domain boundaries. An initial expression trial with a construct truncated at residue 270 at the end of a well-structured region as determined by HDX-MS did not result in soluble expression. Pursuing our aim to generate a minimal construct for crystallisation we therefore decided to use residue 330 within the local minimum of a less structured region as a starting point for a series of constructs with the same N-terminus and sequentially shortened C-termini (Fig. 1B). For truncation scanning a stop codon was introduced every 2 residues. All constructs were expressed in *E. coli* and purified by IMAC using the same methods as for the longer variants. The final protein yields were further normalized to the cell mass before lysis.

The total yields of expressed and purified protein generally declined with successive shorting of the C-terminus (Fig. 2). Purifications also contained aggregates of the target protein that eluted in the void volume of the size exclusion column. This further diminished already low yields, but also pointed to general problems of the expression host producing correctly folded protein. The shortest construct resulting in expression of soluble
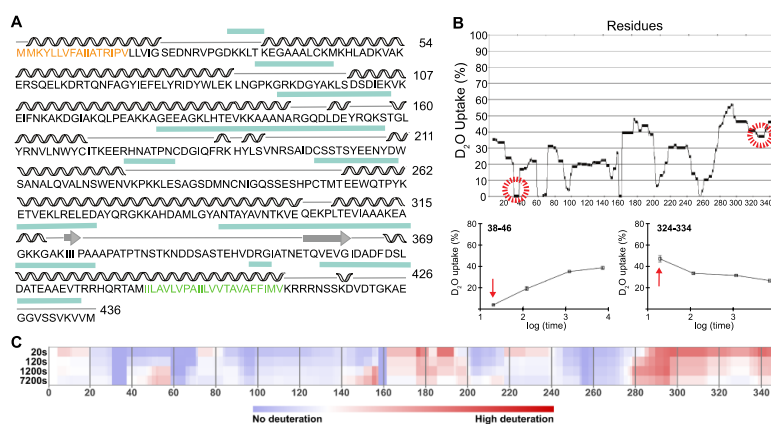
Figure 1. Mapping of ISG65 low-complexity regions. (**A**) Secondary structure prediction of ISG65 adapted from PSIPred[19,20]. Helices are depicted as wavy lines, beta strands as arrows and disordered sections as straight lines. Regions with a low confidence score are marked with a turquoise box above the sequence. Positions of signal-peptide and transmembrane domain are depicted in orange and green, respectively. (**B**) Uptake plot showing the incorporation of deuterium across the sequence of ISG65 20 s after start of the deuteration reaction (main panel). Selected peptides showing the rates of deuterium uptake over time representative for an ordered (bottom left) or disordered (bottom right) region. Red arrows mark the 20 s timepoint. The HDX-MS experiment was performed in triplicates. (**C**) Chiclet plot illustrating the relative deuteration across the sequence of ISG65 at 20, 120, 1200 and 7200 s of deuteration, adapted from MSTools[28]. The amount of relative deuteration is gradually colour coded from blue (no deuteration) to red (high deuteration).

protein was ISG65$_{32-306}$. Expression could no longer be observed in even shorter constructs, suggesting that the C-terminal domain boundary had been identified.

Expression, however, was weak, resulting in exceedingly low yields of soluble protein (< 0.1 mg per litre of culture) after 2 purification steps (Fig. 2). Furthermore, the purified protein still contained significant amounts of contaminations originating from expression in the *E. coli* cytoplasm that would necessitate additional purification steps to establish > 95% purity.

Expression host switching boosts yield of minimal construct 250-fold. Reasoning that resolute shorting of the construct might have further challenged the *E. coli* folding machinery and thus resulted in lower expression rates we switched to a eukaryotic system for expression of the minimal construct. Furthermore, we decided to include a signal sequence for the secretory pathway in order to facilitate folding in the non-reducing environment of the culture medium. The expression system of choice were insect cells because of their easy handling and lack of requirements for expensive equipment (such as CO_2 incubators and shakers).

Protein purity was further increased through usage of serum-free media containing less contaminating proteins in combination with utilisation of a C-tag for affinity purification which was previously reported to improve expression yields for a *Plasmodium* surface protein[22] and is short enough to be unlikely to impair crystallisation.

Expression in insect cells, for both the longest (ISG65$_{32-363}$) as well as the shortest (ISG65$_{32-306}$) construct, yielded approx. 30 mg of highly pure protein per litre of culture after 2 purification steps, exceeding the yields achieved in *E. coli* by approx. 40–250-fold (Fig. 3, Supplementary Figure 3). It is noteworthy that unlike in *E. coli*, the minimal construct showed the same expression level as the full-length protein in insect cells. Moreover, no aggregates of either protein could be detected in SEC. The small shift in retention times compared to the same proteins expressed in *E. coli* can be attributed to the presence of posttranslational modifications that alter the hydrodynamic radius (Fig. 3).

Experimental identification of domain boundaries facilitates structure prediction. In summary we present an integrated strategy for HDX-MS assisted domain identification and high-yield expression of an extracellular eukaryotic protein that can easily be adapted for other challenging proteins. Figure 4A provides an overview of the workflow. While HDX-MS is useful to identify the general region of disorder within a protein, especially at the termini, the method is limited in resolution by the lengths of the obtained peptides (Supple-
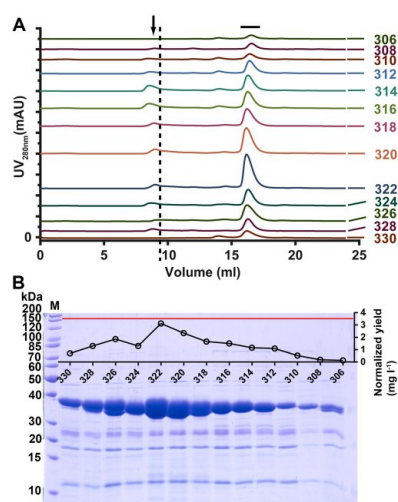
Figure 2. Yields of *E. coli* expressed ISG65 decrease with construct length. (**A**) Stacked size exclusion chromatograms of all ISG65 constructs expressed in *E. coli*. Major tick marks correspond to 300 mAU. Numbers indicate the position of the C-terminal truncation. Peaks corresponding to correctly folded ISG65 are indicated by a black horizontal bar, aggregates by an arrow. The void volume of the column is marked by a dashed, vertical line. (**B**) SDS-PAGE analysis of SEC peak fractions. Total protein yields were calculated for each peak by UV_{280nm} absorption and were normalized to the cell mass in 1 L of culture. All edges of the gel, except the left side, are digitally cropped for clarity. Displayed is the whole separating gel as indicated by the lowest and the highest molecular weight bands of the marker (M).
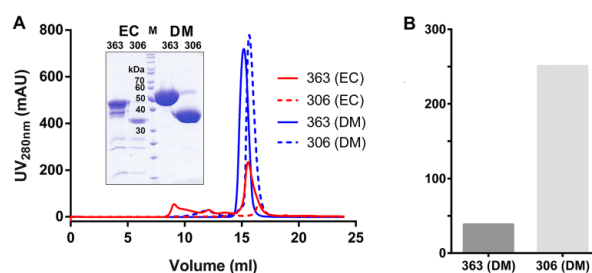
Figure 3. Comparison of recombinant expression yields in *E. coli* and insect cells. Expression in S2 cells results in higher yields, purity and shows no correlation with construct length. (**A**) Size exclusion chromatograms and SDS-gel (inset) of shortest (dashed line) and longest (full line) ISG65 constructs expressed in *E. coli* and insect cells. Bacterial culture volumes were tenfold higher. The protein gel was digitally cropped for clarity. The complete gel is shown in Supplementary Figure 3. (**B**) Expression yields normalised to 1 L of cell culture. Normalised yields were 38-fold (longest construct, dark grey), and, respectively, 250-fold higher (shortest construct, light grey) for the same constructs expressed in insect cells. EC, *E. coli*; DM *Drosophila melanogaster*; M, Marker.
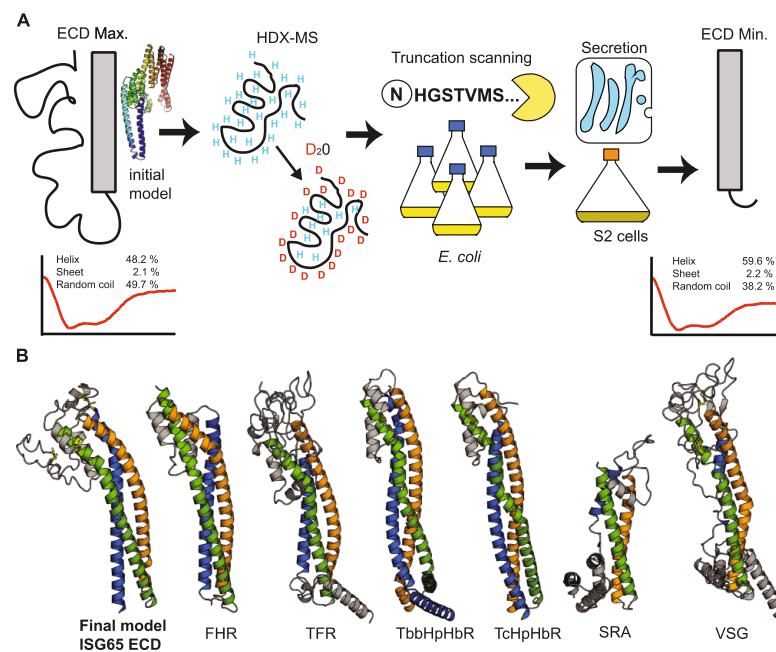
Figure 4. Flow-chart for rational construct design and guided model generation. (**A**) Simplified workflow for identification of a minimal construct that can be produced with high yields and purity. The longest ISG65 construct has a high random coil content (CD measurement, representative curve) which results in generation of an unrealistic model. The identified minimal construct of ISG65 has > 10% lower random coil content according to CD analysis (representative curve below). (**B**) Omission of the identified low-complexity region results in generation of a more realistic ISG65 model (far left). This model shows the same three-helix bundle that is canonical for structurally characterised, GPI-anchored trypanosoma surface proteins ([left to right] PDB IDs: 6XZ6, 6SOY, 4X0J, 4E40, 6ELU, 1VSG). ISG65 ECD, invariant surface glycoprotein 65 extracellular domain; FHR, factor H receptor; TFR, transferrin receptor; TbbHpHbR, *Trypanosoma brucei brucei* haptoglobin-hemoglobin receptor; Tc, *Trypanosoma congolense*; SRA, serum-resistance associated protein; VSG, Variant surface glycoprotein.

mentary Figure 1). HDX-MS does, however, provide a starting point for truncation scanning to determine the domain boundaries more precisely. This step is very time consuming and therefore recommended to be carried out in *E. coli*. Once a minimal construct has been determined, expression can be switched over to insect cells, a moderately expensive and fast eukaryotic system for high yield and high purity production of the target protein (Fig. 4A).

A precise and careful experimental determination of domain boundaries within a protein can compensate for the lack of structural templates in fold prediction and help to generate a more realistic model. Using iTASSER[23], the final model of ISG65 was calculated omitting the long, low-complexity region at the C-terminus of the sequence. This model shows well-defined secondary structure and is, in contrast to the initial model, in agreement with the general fold of other trypanosome surface proteins despite having < 20% sequence identity to the closest structural homologue (Fig. 4B).

The presence of a low-complexity region in the C-terminal domain of ISG65, as identified by HDX-MS, is corroborated by circular dichroism (CD) spectroscopy performed on both the longest as well as the shortest ISG65 construct. CD analysis revealed that the random coil contribution to the overall secondary structure content is

reduced by more than 10% in the minimal construct of ISG65 (Fig. 4A) while the melting temperature remained unchanged (Supplementary Figure 4).

Discussion

The multifaceted strategy outlined in this study combines the advantages of bacterial and eukaryotic expression systems for recombinant production of complex eukaryotic proteins.

Particularly proteins with glycans and multiple disulphide bonds do not present an easy target for heterologous expression in *E. coli*. The use of T7 shuffle cells or similar BL21 derivatives that promote disulphide formation can alleviate the limitations of cytoplasmic expression, but reduced growth rates as compared to standard expression cells diminish yields. Additionally, we consistently found a certain fraction of the soluble, IMAC-purified protein to form aggregates which eluted in the void volume of the gel filtration column. This further decreased the overall amount of functional protein and might be the consequence of incomplete disulphide formation which has been reported previously[24].

Generally, it was possible to produce the target protein in *E. coli*, but correct folding in conjunction with high yields seemed to pose a challenge for the basic prokaryotic folding machinery. In *E. coli*, unlike in insect cells, the expression levels differed drastically between the longest and shortest construct. More remarkably, when comparing the overall yields only for the shortest construct we observed a several 100-fold increase in S2 cells over *E. coli*. This suggests that the truncation of an already complex fold has further challenged the bacterial folding machinery to the point where hardly any protein can be produced anymore. Production of the minimal construct in insect cells on the other hand was not affected by truncation, resulting in yields comparable to the longer construct. Furthermore, both *E. coli* produced ISG65$_{18-363}$ and S2 cell produced ISG65$_{32-306}$ show the same melting temperatures. This implies that the structural integrity of the overall fold was not compromised by removal of the terminal regions and that the differences in expression levels are likely to originate in in the differential efficiency of the two systems to produce a complex eukaryotic protein. Our study, using a structurally uncharacterised surface receptor of a eukaryotic parasite as an example, shows that soluble expression in *E. coli* is no guarantor for high protein yields and that weak expression in a bacterial system is not necessarily an indicator for incorrectly chosen domain boundaries.

The initial model calculated for the extracellular domain of ISG65, using the popular prediction program iTASSER, did not resemble the fold of any known trypanosoma surface protein. Especially the C-terminal domain was modelled rather unrealistic without any connection to the rest of the molecule and an unexpectedly high helical content due to a low confidence score in this region. Experimentally identifying low-complexity regions and excluding them from model building could serve to focus the prediction on the structurally conserved core-region of the protein, thus resulting in a more realistic model. In our case this strategy revealed a fold that is highly similar to the canonical 3-helix bundle fold of structurally known surface receptors[14–18,25]. This structural homology was unexpected and could have neither been inferred from the sequence alone nor from the initial model. This increases confidence in the overall correctness of the predicted model which, for example, could serve as a molecular replacement search model in crystal structure determination. Interestingly, a new AlphaFold2-based resource adapted for *Trypanosoma* and *Leishmania* proteins[26] yielded an initial model that already showed a 3-helix bundle fold, albeit with an unusual arrangement of alpha-helices and long disordered sections (Supplementary Figure 5) in the C-terminal region. However, more reference structures of *Trypanosoma* transmembrane proteins are needed to judge the general validity of our approach as well as the accuracy of predicted models without homologies.

Material and methods

Recombinant expression and purification of ISG65. *Escherichia coli.* A gene fragment coding for the extracellular domain of ISG65 (*Trypanosoma brucei gambiense* DAL972, gene ID 23,858,558) comprising residues 18-363 was cloned into the bacterial expression vector pET15b (EMD Millipore). Further constructs were truncated N-terminally (residue 32) as well as C-terminally (residues 330–303). All constructs contained a N-terminal hexa-histidine tag. C-terminal truncations were carried out in a 2-residue window using the Q5 site-directed mutagenesis kit (New England Biolabs). Expression constructs were introduced in *E. coli* T7 shuffle cells (New England Biolabs) by heat-shock transformation. Protein expression was induced by addition of 1 mM isopropyl-β-D-thiogalacto-pyranoside (IPTG) and carried out for 16 h at 22 °C. Cells were sedimented at 6,000 g for 15 min and pellets weighed for normalisation of protein expression levels. The cell pellets were resuspended in ice-cold buffer A (20 mM Tris pH8, 150 mM NaCl, 10 mM imidazole) and lysed by sonication. Temperature during sonication was monitored and did not exceed 20 °C. Following cell lysis, supernatant was cleared by centrifugation (50,000 g, 1 h, 4 °C). Soluble protein was purified by immobilized metal affinity chromatography (IMAC) using a 1 ml HisTrap HP on ÄKTA start (GE). Immobilized, His-tagged protein was washed with buffer B (20 mM Tris pH8, 150 mM NaCl, 20 mM imidazole) and eluted from the column with buffer C (20 mM Tris pH8, 150 mM NaCl, 500 mM imidazole). Eluate was dialysed against buffer D (20 mM HEPES pH7.5, 150 mM NaCl) for 16 h at 4 °C. The dialysed protein solution was concentrated by ultrafiltration (Amicon Ultra, EMD Millipore, molecular weight cut-off, 10,000 Da) and further purified by size exclusion chromatography (SEC) in buffer D using a Superdex 200 Increase 10/300 GL on ÄKTA pure (GE). SEC peak fractions were analysed for presence and purity of the target protein using SDS-PAGE.

Insect cells. For expression in *Drosophila melanogaster* S2 cells, ISG65 gene fragments encoding residues 18-363 and 32-306 were cloned into the pExpress2.1 vector (ExpreS2ion Biotechnologies). Recombinant ISG65 comprising an N-terminal secretion signal and a C-terminal C-tag was produced following manufacturer's recommendations (ExpreS2ion Biotechnologies). Briefly, 5 ml of 2×10^6 cells/ml, cultured in Ex-cell 420 serum free

6

medium (14420C, SAFC Biosciences), were transfected with 0.25 ml of 1×lipofectamine-based reagent (Expre-S2ion Biotechnologies) and 12.5 µg plasmid DNA. After 4 h, 12.5% FBS was added and cells were incubated in T25 flasks overnight. After 24 h, selection was initiated by addition of 2 mg/ml zeocin to generate a stable cell line. After 3–4 weeks of selection, cells were transferred to shaker flasks and production scaled up. Cells were harvested after 4 days of expression by centrifugation (1000 *g*, 15 °C, 15 min). ISG65-containing supernatant was subsequently cleared from debris by centrifugation (13,000*g* 15 °C, 10 min.) and filtration (0.22 µm). Using tangential flow filtration, the cleared supernatant was concentrated and exchanged to buffer E (20 mM Tris pH7.5, 100 mM NaCl). The protein concentrate was loaded onto a CaptureSelect C-tag affinity column (ThermoFischer Sci.) using an ÄKTA start. After washing with buffer E, C-tagged ISG65 was specifically eluted with buffer E supplemented with 2 M MgCl$_2$.

Circular dichroism spectroscopy. For ISG65, far-UV CD experiments were carried out on a Jasco J-1500 spectropolarimeter with a 0.2 mm path cell. Protein was measured at a concentration of 0.4 mg ml^{-1} in buffer F (20 mM HEPES pH 7.5, 150 mM NaF). Spectra were recorded between 195 and 260 nm wavelength at an acquisition speed of 10 nm min^{-1} and corrected for buffer absorption. For calculation of melting curves at 222 nm, spectra were recorded every 5° between 5 ° and 80C with a slope of 0.16 °C min^{-1}. During measurements the temperature was kept constant.

In both cases, the raw CD data (ellipticity θ in mdeg) were normalized for the protein concentration and the number of residues according to the equation below, yielding the mean residue ellipticity ([θ] in deg cm^2 mol^{-1}), where *MM*, *n*, *C*, and *l* denote the molecular mass (Da), the number of amino acids, the concentration (mg mL^{-1}), and the cuvette path length (cm), respectively.

$$[\theta] = \frac{\theta \cdot MM}{n \cdot C \cdot l}$$

Secondary structure content was calculated using the DichroWeb server[37].

Hydrogen–deuterium exchange mass spectrometry. *Peptide mapping.* 300 pmol of ISG65 were mixed in 1:1 (v/v) ratio with 1 M glycine, 400 mM Tris(2-carboxyethyl)phosphine (TCEP) at pH 2.3 and injected onto a Nepenthesin-2 column. Generated peptides were trapped and desalted using a micro-trap column (Luna Omega 5 µm Polar C18 100 Å Micro Trap 20×0.3 mm) for 3 min at a flow rate 400µL min^{-1} using an isocratic pump delivering 0.4% formic acid in water. Both protease column and trap column were placed in an icebox. After 3 min, peptides were separated on a C18 reversed-phase column (Luna Omega 1.6 µm Polar C18 100 Å, 100×1.0 mm) with a linear gradient of 5–35% B over 26 min, where solvent A was 2% acetonitrile/0.4% formic acid in water and solvent B 95% acetonitrile/5% water/0.4% formic acid. The analytical column was placed in an icebox. A 15 T solariX XR FT-ICR mass spectrometer (Bruker Daltonics) operating in positive MS/MS mode was used for detection of peptides. Data was processed by DataAnalysis 4.2 software (Bruker Daltonics). MASCOT search engine was used for identification of peptides using a sequence of ISG65.

HDX. Hydrogen deuterium exchange was initiated by ten-fold dilution of ISG65 into a deuterated buffer. 50 µl aliquots (100 pmol) were taken after 20 s, 120 s, 1200 s and 7200 s of incubation in deuterated buffer, the reaction was quenched by addition of 50 µl of 1 M glycine, 400 mM TCEP at pH 2.3 and immediate freezing in liquid nitrogen. Aliquots were quickly thawed and analysed using the same system as described above. Peptides were separated over a linear gradient of 10–30% B over 18 min. The mass spectrometer was operated in positive MS mode. Spectra of partially deuterated peptides were processed via Data Analysis 4.2 (Bruker Daltonics, Billerica, MA) and the in-house program DeutEx (unpublished).

Model prediction. Structure prediction was performed using the I-TASSER (Iterative Threading ASSEmbly Refinement) online server with default settings[23].

Data availability
The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References
1. Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **37**(4), 205–211 (1951).
2. Pauling, L. & Corey, R. B. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. U. S. A.* **37**(11), 729–740 (1951).
3. Rost, B. Review: protein secondary structure prediction continues to rise. *J. Struct. Biol.* **134**(2–3), 204–218 (2001).
4. Rost, B., Sander, C. & Schneider, R. Redefining the goals of protein secondary structure prediction. *J. Mol. Biol.* **235**(1), 13–26 (1994).
5. Yang, Y. *et al.* Sixty-five years of the long march in protein secondary structure prediction: the final stretch?. *Brief Bioinform.* **19**(3), 482–494 (2018).
6. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**(7792), 706–710 (2020).
7. Ziegelbauer, K. & Overath, P. Identification of invariant surface glycoproteins in the bloodstream stage of *Trypanosoma brucei. J. Biol. Chem.* **267**(15), 10791–10796 (1992).

8. Koumandou, V. L. *et al.* Evidence for recycling of invariant surface transmembrane domain proteins in African trypanosomes. *Eukaryot. Cell* **12**(2), 330–342 (2013).
9. Leung, K. F. *et al.* Ubiquitylation and developmental regulation of invariant surface protein expression in trypanosomes. *Eukaryot. Cell* **10**(7), 916–931 (2011).
10. Zoltner, M. *et al.* Modulation of the surface proteome through multiple ubiquitylation pathways in African trypanosomes. *PLoS Pathog.* **11**(10), e1005236 (2015).
11. Sullivan, L. *et al.* Proteomic selection of immunodiagnostic antigens for human African trypanosomiasis and generation of a prototype lateral flow immunodiagnostic device. *PLoS Negl. Trop. Dis.* **7**(2), e2087 (2013).
12. Nzou, S. M. *et al.* Development of multiplex serological assay for the detection of human African trypanosomiasis. *Parasitol. Int.* **65**(2), 121–127 (2016).
13. Umaer, K. & Bangs, J. D. Late ESCRT machinery mediates the recycling and rescue of invariant surface glycoprotein 65 in Trypanosoma brucei. *Cell Microbiol.* **22**(11), e13244 (2020).
14. Trevor, C. E. *et al.* Structure of the trypanosome transferrin receptor reveals mechanisms of ligand recognition and immune evasion. *Nat. Microbiol.* **4**(12), 2074–2081 (2019).
15. Higgins, M. K. *et al.* Structure of the trypanosome haptoglobin-hemoglobin receptor and implications for nutrient uptake and innate immunity. *Proc. Natl. Acad. Sci. U. S. A.* **110**(5), 1905–1910 (2013).
16. Macleod, O. J. S. *et al.* A receptor for the complement regulator factor H increases transmission of trypanosomes to tsetse flies. *Nat. Commun.* **11**(1), 1326 (2020).
17. Freymann, D. *et al.* 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. *J. Mol. Biol.* **216**(1), 141–60 (1990).
18. Lane-Serff, H. *et al.* Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor. *Elife* **3**, e05553 (2014).
19. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**(2), 195–202 (1999).
20. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucl. Acids Res.* **47**(W1), W402-w407 (2019).
21. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**(7), 595–602 (2019).
22. Jin, J. *et al.* Accelerating the clinical development of protein-based vaccines for malaria by efficient purification using a four amino acid C-terminal "C-tag". *Int. J. Parasitol.* **47**(7), 435–446 (2017).
23. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**(4), 725–738 (2010).
24. Pleiner, T. *et al.* Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *Elife* **4**, e11349 (2015).
25. Zoll, S. *et al.* The structure of serum resistance-associated protein and its implications for human African trypanosomiasis. *Nat. Microbiol.* **3**(3), 295–301 (2018).
26. Wheeler, R. J. A resource for improved predictions of Trypanosoma and Leishmania protein three-dimensional structure. *PLoS ONE* **16**(11), e0259871 (2021).
27. Whitmore, L. & Wallace, B. A. Dichroweb: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucl. Acids Res.* **32**(Web Server issue), W668–W673 (2004).
28. Kavan, D. M. P. MSTools—Web based application for visualization and presentation of HXMS data. *Int. J. Mass Spectrom.* **302**(1–3), 53 (2011).

Author contributions

Experimental design, H.S. and S.Z.; Data analysis, H.S. and S.Z.; Cloning and generation of stable cell lines, J.V.; Bacterial protein expression and purification, H.S. and S.Z.; Protein production in insect cells, A.D.; Writing–original draft preparation, S.Z.; Writing–review and editing, A.D., H.S. and S.Z.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-16958-x.

Correspondence and requests for materials should be addressed to S.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

6.2 Cryo-EM structures of *Trypanosoma brucei gambiense* ISG65 with human complement C3 and C3b and their roles in alternative pathway restriction.

Cryo-EM structures of *Trypanosoma brucei gambiense* ISG65 with human complement C3 and C3b and their roles in alternative pathway restriction

Hagen Sülzen [1,2], Jakub Began [1,7], Arun Dhillon[1], Sami Kereïche[1,3], Petr Pompach[4], Jitka Votrubova[1], Farnaz Zahedifard[5], Adriana Šubrtova [1], Marie Šafner[1], Martin Hubalek[1], Maaike Thompson [1,6,8], Martin Zoltner [5] & Sebastian Zoll[1]✉

African Trypanosomes have developed elaborate mechanisms to escape the adaptive immune response, but little is known about complement evasion particularly at the early stage of infection. Here we show that ISG65 of the human-infective parasite *Trypanosoma brucei gambiense* is a receptor for human complement factor C3 and its activation fragments and that it takes over a role in selective inhibition of the alternative pathway C5 convertase and thus abrogation of the terminal pathway. No deposition of C4b, as part of the classical and lectin pathway convertases, was detected on trypanosomes. We present the cryo-electron microscopy (EM) structures of native C3 and C3b in complex with ISG65 which reveal a set of modes of complement interaction. Based on these findings, we propose a model for receptor-ligand interactions as they occur at the plasma membrane of blood-stage trypanosomes and may facilitate innate immune escape of the parasite.

The ability of pathogens to survive and proliferate in the host's hostile environment is inextricably linked to their capacity to evade and manipulate its immune responses. African trypanosomes are extreme examples of such competence. The blood-stage parasites are exposed to the full pressure of both branches of the immune system and must employ a plethora of countermeasures to avoid clearance[1]. Many of the molecular mechanisms underlying these elaborate immune evasion strategies have been identified and, particularly with regard to the adaptive immune response, described in detail: variant surface glyco-proteins (VSGs) constituting a protective surface layer, antigenic variation of exposed VSG epitopes, VSG O-glycosylation and flow-mediated clearance of surface-bound antibodies[2–6]. In contrast, defence mechanisms against the innate immune response have so far mostly been described by the example of trypanolytic factors, toxin-bearing higher-order protein complexes that are unique to humans and some primates[7,8]. *T. b. gambiense*-specific glycoprotein TgsGP (together with haptoglobin-haemoglobin receptor HpHbR) in *T. b. gambiense*[9–12] and serum-resistance-associated protein SRA in *T. b. rhodesiense*[13,14] have evolved as adaptations to overcome the action of these factors.

In comparison, our understanding of other innate immune defence mechanisms in human-infective African trypanosomes, parti-cularly during the early stage of infection, is incomplete.

[1]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo namesti 542/2, 16000 Prague, Czech Republic. [2]Faculty of Science, Charles University, Albertov 6, 12800 Prague 2, Czech Republic. [3]First Faculty of Medicine, Charles University, Albertov 4, 12800 Prague, Czech Republic. [4]Institute of Biotechnology of the Czech Academy of Sciences, 25250 Vestec, Czech Republic. [5]Department of Parasitology, Faculty of Science, Charles University Prague, Biocev, 25250 Vestec, Czech Republic. [6]University of Antwerp, Antwerp, Belgium. [7]Present address: Department of Immuno-biology, University of Lausanne, Chemin des Boveresses 155, 1066 Epalinges, Switzerland. [8]Present address: Agidens, Industrial Machinery Manufacturing, Zwijndrecht, Antwerp, Belgium. ✉e-mail: sebastian.zoll@uochb.cas.cz

The complement system is an essential part of the host's innate immune response against invading pathogens. It involves 3 distinct pathways, the classical (CP), the lectin (LP) and the alternative pathway (AP)[15] (Fig. 1). While CP and LP are triggered by recognition of specific structures on the pathogen surface, typically proteins and glycans, the AP can also be initiated spontaneously. Each pathway comprises a series of proteolytic events, resulting in the generation of convertases, multi-component enzyme complexes that serve as central hubs within the cascades. The first convertase of the CP and LP is formed at the junction of both pathways by the covalent binding of C4b to the plasma membrane and subsequent interaction with the serine protease C2a. The emerging CP/LP C3 convertase C4bC2a cleaves native C3 to C3b, which can then associate with C4bC2a to create the CP/LP C5 convertase C4bC2aC3b, the last enzyme of the pathway. Cleavage of C5 initiates the terminal pathway (TP), the point of convergence of CP/LP and AP, during which C5b sequentially binds C6, C7, C8 and C9. C9 polymerises and subsequently forms a lytic pore in the membrane of the invading pathogen, the membrane attack complex (MAC).

While it mainly functions as an amplification pathway of the CP and LP, shortly after an infection, the AP can also act as a first line of defence against invading pathogens, before an antibody response is mounted[16,17]. According to the prevailing hypothesis, it is in a constant state of low-level activation (tick-over), resulting in the formation of the fluid-phase C3 convertase $C3(H_2O)Bb$[17-19]. In the presence of foreign microbial patterns, this can lead to swift activation of the AP cascade. Analogue to the CP and LP, central events of the AP are the formation of C3 and C5 convertases with identical substrate specificities as their counterparts, but different structures and compositions. The AP C3 convertase C3bBb comprises C3b and Bb, a cleavage product of factor B, and the AP C5 convertase C3bBbC3b is most likely established following the binding of a second molecule of C3b. The role of the second C3b molecule has, however, not been fully clarified.

Recently, it was also suggested that surface-bound C3b only primes C5 for cleavage by C3bBb rather than being part of a trimolecular complex[20].

In humans, sleeping sickness or HAT (Human African trypanosomiasis) is caused by only two *Trypanosoma* species, *T. b. rhodesiense* and *T. b. gambiense* of which the latter is responsible for 95% of all cases. Earlier, it was shown that *T. b. gambiense* does not succumb to cell lysis in human serum despite specifically activating the AP, pointing to a dedicated mechanism of complement inactivation at this stage. No complement-binding receptor was, however, identified[21,22].

In this work, we report the identification and characterisation of invariant surface glycoprotein 65 (ISG65) from *T. b. gambiense* as a receptor for C3 and its activation products in human-infective bloodstream forms. We show that only convertases of the AP but not of the CP/LP can form on the surface of *T. b. gambiense* and that ISG65 is capable of specifically inhibiting the activity of the AP C5 convertase in vitro. This prevents initiation of the TP and, thereby, cell lysis. Moreover, we present the structures of *T. b. gambiense* ISG65 in complex with native C3 as well as C3b determined by cryo-EM, describe the differences in the structural basis of complement binding and propose a plausible model for receptor-ligand interactions as they occur in the context of the VSG layer and how they likely facilitate the innate immune escape of the parasite.

Results

ISG65 is a C3b receptor in bloodstream trypanosomes

Initially, the interaction between *T. b. gambiense* ISG65 and C3 was identified using pull-down assays with the extracellular domain of ISG65 as bait and human serum (Supplementary Fig. 1a). To probe the stability of the complex, an excess of ISG65 was incubated with C3b and subjected to size-exclusion chromatography. Two distinct peaks corresponding to the ISG65:C3b complex and free ISG65 were
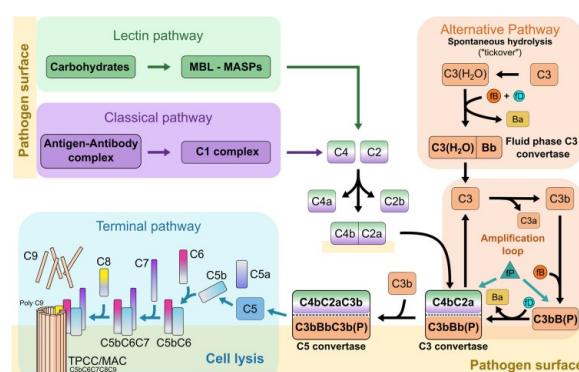
Fig. 1 | Overview of the three complement pathways and their components. The classical (CP) and the lectin pathway (LP) are triggered by specific antigens on pathogen surfaces. The alternative pathway (AP) can start spontaneously by a nucleophilic attack of water on metastable native C3, resulting in the generation of the fluid-phase C3 convertase, but it also serves as an amplification pathway after CP and LP. CP and LP share the same C3 and C5 convertases CP/LP C4bC2a and CP/LP C4bC2aC3b, respectively. Analogue to C3b of the AP C3 convertase, CP/LP C3 convertase formation is initiated by the covalent binding of thioester-containing C4b to the pathogen membrane. Both CP/LP convertases are structurally distinct from the AP C3 and C5 convertases AP C3bBb(P) and AP C3bBbC3b(P), but have the same substrate specificities. Properdin stabilises the AP convertases. Cleavage of C5 initiates the terminal pathway. C5b sequentially binds C6, C7, C8 and poly-C9 and inserts into the plasma membrane where it forms a lytic pore, the membrane attack complex. MBL Mannan-binding lectin, MASP MBL-associated serine proteases, fB factor B, fD factor D, Ba factor B fragment a, Bb factor B fragment b, fP factor P (properdin), MAC membrane attack complex, TPCC terminal pathway complete complex.

Article
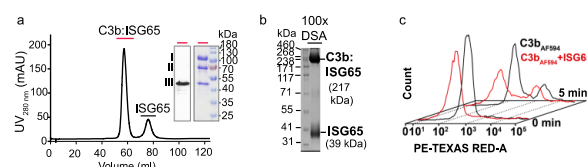
Fig. 2 | ISG65 is a C3b receptor in *T. b. gambiense*. a Chromatogram showing purification of the ISG65:C3b complex (217 kDa) using size-exclusion chromatography. Prior to injection, ISG65 was incubated with C3b in threefold molar excess. The complex formation of ISG65 with C3b was confirmed by analysing fractions underneath the main peak (indicated by a red bar) using reducing SDS-PAGE and Western blotting. Detection with an anti-His antibody confirmed the presence of His-tagged ISG65 in the complex peak; I, a-chain C3b, II, b-chain C3b, III, ISG65.

Displayed Western blot and SDS-gel are representatives of three independent experiments. **b** SDS-PAGE analysis of cross-linked ISG65:C3b complex (100x excess of Di-(N-succinimidyl)adipate (DSA)). The displayed SDS-gel is representative of three independent experiments. **c** Representative scatter plots from FACS analysis of AF_{594} labelled C3b surface binding and uptake by *T. b. gambiense* in the presence (red) and absence (black) of ISG65. Measurements were done in triplicates at two different timepoints. Source data are provided as a Source Data file.

observed. Analysis of peak fractions by SDS-PAGE and Western blotting revealed three bands that corresponded to ISG65 as well as to the α- and β-chains of C3b, demonstrating that a stable complex was formed (Fig. 2a). To further characterise the interaction, the complex was cross-linked. SDS-PAGE analysis showed two bands migrating at approximately 217 and 39 kDa, consistent with the expected sizes of cross-linked complex (at 1:1 stoichiometric ratio) and free ISG65 (Fig. 2b).

No complete ISG65 knock-out cell line was available during the time of this study. To investigate if ISG65 functions as a C3 receptor in *T. b. gambiense*, we compared surface binding and uptake of free C3b to ISG65-saturated C3b using fluorescence-activated cell sorting (FACS). FACS revealed that Alexa Fluor 594 (AF_{594}) labelled human C3b is efficiently internalised via endocytosis in *T. b. gambiense* cells and delivered to the lysosome, similar to AF_{594} labelled human transferrin (Supplementary Fig. 2). The addition of recombinant ISG65 at a 1.5-fold molar excess led to 60% reduction of C3b surface binding and a 40% reduction of AF_{594} positive cells after 5 min of uptake, as compared to treatment with C3b alone (Fig. 2c, Supplementary Fig. 3, and Supplementary Data 1).

Incubation of labelled C3b with another member of the *T. b. gambiense* invariant surface glycoprotein family, ISG75, had no significant effect on surface binding and uptake (Supplementary Fig. 3a). These findings confirmed the suspected role of ISG65 as a C3 receptor. Residual uptake of ISG65-saturated C3b can either be attributed to fluid-phase endocytosis, partial disintegration of the complex or the presence of another C3 receptor.

ISG65 restricts the alternative but not the classical and the lectin pathway

Previous results indicated that *T. b. gambiense* activates the alternative but not the classical pathway, and that a specific molecular mechanism is likely to exist that blocks its progression and prevents cell lysis[21]. Using pathway-specific, fluorescently-labelled antibodies and depleted human sera, we could show that only Bb, a component of AP convertases, but not C4b, an essential component of both CP and LP convertases, is accumulated in the flagellar pocket (Fig. 3a, b and Supplementary Fig. 4a, c) as a consequence of deposition on the surface of *T. b. gambiense*. In contrast, we could readily detect C4b on the surface of sensitised sheep erythrocytes, which were incubated in C6-depleted human serum to avoid premature cell lysis (Fig. 3b). Using C2-depleted serum, we could also confirm that Bb is deposited as a consequence of AP activation and not via the amplification loop through a minor deposition of the CP/LP C3 convertase which contains C2. In agreement with this observation, C3b could only be detected on cells that were incubated in C2-depleted, but not fB-depleted serum (Fig. 3c, d and Supplementary Fig. 4b, d). Since trypanosomes are not affected

by complement-mediated lysis, we investigated whether the TP is initiated at all by testing for binding of C5b, the first complement factor of the TP, using a fluorescently-labelled anti-C5b antibody.

No flagellar pocket signal could be detected on trypanosomes while sheep erythrocytes, when incubated in C7-depleted human serum (allowing membrane-binding of C5b via C6[23], but preventing premature lysis) bound C5b efficiently and presented a robust fluorescent signal (Fig. 3e). This strongly suggests that the TP is not initiated on *T. b. gambiense* and that restriction of the complement cascade occurs upstream.

To ascertain that ISG65 binding to C3 is responsible for the observed restriction of the alternative pathway, we carried out functional haemolytic assays using ISG65 as an inhibitor. The addition of soluble ISG65 to human serum resulted in an almost complete abrogation of rabbit erythrocyte lysis (up to 90% decrease) following activation of the AP (AP50 test, Haemoscan) (Fig. 4a). In agreement with the prior observation that no components of CP/LP convertases could be detected on the parasite surface, we also could not observe any ISG65-mediated decrease in lysis of sensitised sheep erythrocytes after induction of the CP in vitro (CH50 test, Haemoscan) (Fig. 4b). In both assays ISG75 had no measurable impact on erythrocyte lysis, confirming the specificity of the observed effect for ISG65.

Modelling suggests that ISG65 does not interfere with the assembly or activity of the AP C3 convertase (Supplementary Fig. 5a). Additionally, AP C3 pro-convertase formation in the presence of ISG65 was shown by SPR (Supplementary Fig. 5b). Therefore, inhibition via binding to C3b must take place at the stage of the AP C5 convertase, either by preventing its assembly or by diminishing the ability of the assembled convertase to cleave C5 substrate. To investigate how exactly ISG65 interferes with AP C5 convertase activity in vitro, we used a modified AP50 assay (Haemoscan) to separate convertase deposition on the cell surface from the assembly of the membrane attack complex (MAC) in the TP, leading to haemolysis. Erythrocytes were hereby incubated with dilutions of human complement C5-depleted serum, which allowed for C5 convertase assembly on the cell surface, but stalled the progression of the cascade due to a lack of C5 substrate. After a washing step, ISG65 and ISG75 (negative control) were added together with dilutions of human complement factor C3-depleted serum, providing the C5 substrate without introducing competing C3.

The modified AP50 assay was validated by efficient lysis of erythrocytes when both depleted sera were used consecutively (reconstituted NHS). Exchange of either serum with buffer prevented lysis. The addition of ISG65 to the assembled AP C5 convertase revealed a significant effect on TP progression. In comparison to reconstituted NHS, haemolysis was decreased by up to 40% in the presence of ISG65 (Fig. 4c). This demonstrated that binding of ISG65 to pre-assembled AP
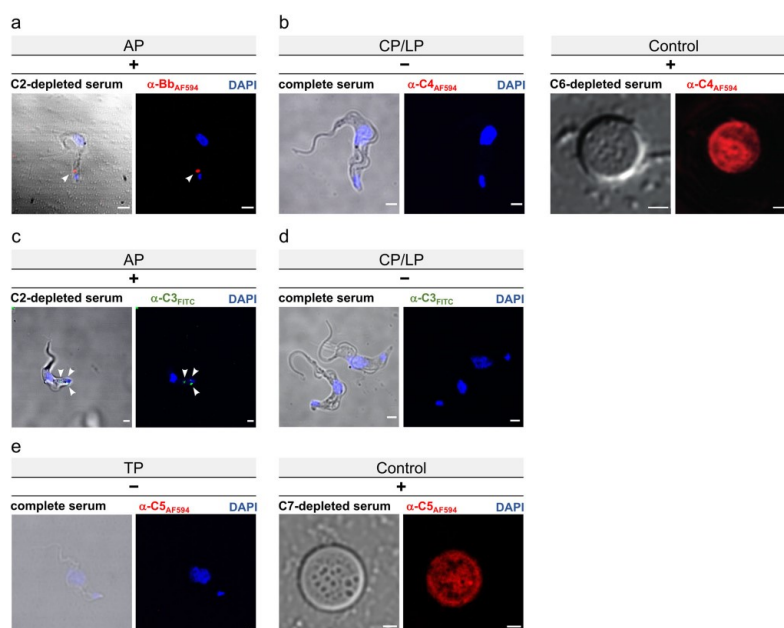
Fig. 3 | Bb and C3b are deposited on the surface of *T. brucei gambiense* in a C2-independent manner, while C4b and C5b are not detected by immunofluorescence. Representative fluorescence micrographs of trypanosomes incubated in depleted or complete human serum tested for deposition of complement factors using fluorescently-labelled antibodies as indicated (red = AF594, green = FITC; blue = DAPI stain; left images are merged with brightfield channel; scale bar = 2 μm). Independent experiments were performed at least in triplicate. Note, that even at low temperatures, molecules bound to the trypanosome surface are rapidly internalised, accumulating in the flagellar pocket before delivery to the early endosome. Flagellar pocket, early endosome and lysosome are highlighted by white arrows on selected images. **a** *T. b. gambiense* in C2-depleted serum (no CP/LP) stained with anti-Bb. Bb was robustly detected at the flagellar pocket. **b** *T. b. gambiense* (left) in complete serum and sensitised sheep erythrocytes (right, positive control) in C6-depleted serum (no MAC assembly) stained with anti-C4. No C4b-specific signal was detected on trypanosomes. **c** *T. b. gambiense* in C2-depleted serum (no CP/LP) stained with anti-C3. C3b was robustly detected at the flagellar pocket. **d** *T. b. gambiense* in fB-depleted serum (no AP) stained with anti-C3. No C3b-specific signal was detected. **e** *T. b. gambiense* (left) in complete serum and sensitised sheep erythrocytes (right, positive control) in C7-depleted serum (no MAC assembly, but C5b-6 membrane deposition) stained with anti-C5. No C5b-specific signal was detected on trypanosomes. For additional experiments, replicates and controls, see Supplementary Fig. 4.

C5 convertase can at least partially abrogate cleavage of C5 substrate in vitro.

ISG65 binds C3 fragments with varying affinities

During the activation of the AP, C3 undergoes multiple proteolytic cleavages that result in an array of fragments, each with a specific role within the AP (Fig. 1). To understand whether ISG65 can distinguish between these naturally occurring fragments, we determined dissociation constants (K_D) using surface plasmon resonance (SPR) (Fig. 5). The identity of all C3 species was confirmed by intact mass spectrometry (MS), N-terminal sequencing, Western blotting and ion-exchange chromatography (Supplementary Fig. 6). ISG65 has clear preferences for certain C3 fragments. Native C3 ($K_D = 130$ nM), the most abundant C3 species in human blood, has the lowest binding affinity, followed by C3b ($K_D = 81$ nM) and C3MA, a mimic for C3(H_2O)[24] ($K_D = 18$ nM). The interaction with C3d (the proteolytically liberated thioester domain of C3, TED) takes place at an even higher affinity ($K_D = 7$ nM), likely due to unimpeded movement in the absence of the remaining C3 scaffold (i.e. C3c).

Dissociation constants for C3 and C3b are in line with the high serum concentrations of both fragments (C3: 5–11 μM, C3b: 152–888 nM)[25]. With reported concentrations ranging from 3–100 nM[18,26,27] for C3(H_2O), the measured affinity of ISG65 to C3MA also appears to be in a physiologically relevant range. Notably, the amount of C3b depends on the complement activation status. Together with a short half-life of C3b, actual concentrations may be lower than reported values. No binding of C3c to ISG65 could be observed, indicating that TED is the main interacting domain of C3. Lack of steric interference from the C3 scaffold could also be the reason for the several-fold higher affinity to C3MA compared to C3b and native C3. In C3MA, the TED has been suggested to be more accessible or more flexibly tethered to the C3 scaffold than in the other fragments[28,29].
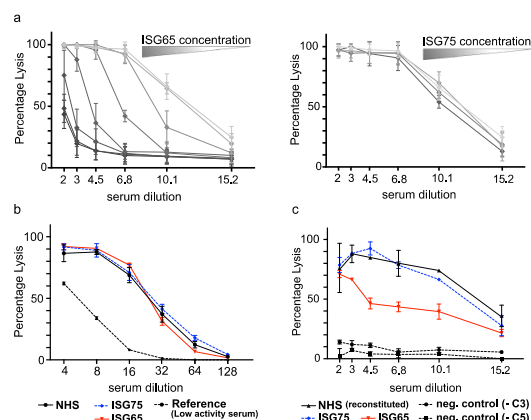
Fig. 4 | ISG65 prevents haemolysis following activation of the alternative pathway, but not the classical pathway. a Haemolysis assay (AP50 test) measuring the lysis of rabbit erythrocytes in human serum in the presence of ISG65 (left) and ISG75 (right) following activation of the alternative complement pathway (*n* = 3 technical replicates; data points are presented as mean values ± SD). ISG65, but not ISG75, considerably reduced haemolysis in a concentration-dependent manner. Colour gradients indicate protein dilutions ranging from 7.5 to 0 µM. **b** Haemolysis assay (CH50 test) measuring the lysis of sensitised sheep erythrocytes in human serum in the presence of ISG65 (red solid line), ISG75 (dashed blue line) and without added inhibitors (NHS, normal human serum; solid black line) following activation of the classical complement pathway (*n* = 4 technical replicates; data points are presented as mean values ± SD). At 24 µM, neither ISG65 nor

ISG75 showed notable inhibition of erythrocyte lysis. Low complement activity reference (*n* = 3 technical replicates; data points are presented as mean values ± SD) is depicted with a dashed black line. **c** Adapted AP50 haemolysis assay using C5- and C3-depleted human sera. Binding of ISG65 to assembled AP C5 convertase C3bBbC3b reduced haemolysis (red solid line). No reduction in haemolysis could be detected for ISG75 (dashed blue line). Human sera depleted for either C5 (-C5) or C3 (-C3) did not cause haemolysis (dashed black lines). Incubation of erythrocytes with C5-depleted serum followed by incubation with C3-depleted serum, reconstituted haemolytic activity (NHS (reconstituted), solid back line). (*n* = 3 technical replicates; data points are presented as mean values ± SD). Source data are provided as a Source Data file.

Cryo-EM structures of ISG65 in complex with C3b and native C3

To gain structural insights into the interaction, we employed single-particle cryo-EM and determined the structures of ISG65:C3 and ISG65:C3b. The ISG65:C3b complex could be reconstructed to a reported FSC$_{0.143}$ of 3.58 Å, while ISG65:C3b was reconstructed to FSC$_{0.143}$ of 3.59 Å (Supplementary Table 1). However, the local resolution range across the reconstructions varies significantly, especially for ISG65:C3b. The resolution of the reconstruction corresponding to the rather rigid MG-ring is in the near-atomic range of 3.5 Å. The local resolution in the CUB and TED region varies between 4 and 5.5 Å, while at the interface and in the remainder of ISG65, it ranges from 5 to beyond 10 Å (Supplementary Fig. 7c). Secondary structure features like β-sheets in the CUB domain and α-helices in ISG65, however, could still be identified (Fig. 6a). For ISG65:C3, the local resolution predominantly ranges from 3 to 5 Å, with lower resolutions in the disordered head region of ISG65, the flexible C345c domain, and glycans in the two glycosylation sites (Supplementary Fig. 8c). The higher variance in local resolution across the ISG65:C3b reconstruction is likely caused by increased flexibility of the extended domains CUB and TED relative to the MG-ring. In contrast, C3 adopts a more compact conformation with TED wedged between MG2, MG8 and CUB. This results in reduced flexibility and thus variance within the molecule, which is reflected by the sloped shape of the FSC of ISG65:C3b compared to a rather vertical drop of the FSC in the ISG65:C3 reconstruction (Supplementary Figs. 7, 8).

C3b could be modelled except for the flexible regions Asn93-Gly101, Gln312-Leu314, Pro665-Ala667, Ser749-Leu751, Glu1372-

Ala1380, and Ser1523-Asp1524 (Fig. 7). C3b resembles the previously published structure of C3b[30] (PDB 2I07), to which ISG65-bound C3b can be aligned with an RMSD = 2.0 (11,068 atoms). Minor rearrangements of the TED and the CUB domain, especially in residues Gln1161-Pro1287, can be observed in C3b and may be caused by ISG65 binding. The lack of near-atomic resolution in ISG65 of the ISG65:C3b reconstruction prevented us from drawing conclusions about the details of the binding interface. However, the superposition of the atomic model of ISG65 with its counterpart in the C3 complex suggests an identical interface with TED (Supplementary Fig. 9). In addition, the electron density corresponding to the head domain of ISG65 also suggests contacts with the CUB domain of C3b (Fig. 7). Should this interaction occur in a physiological context, the second attachment point could restrict the free movement of the TED.

ISG65 adopts a three-helix bundle fold (Helix 1, Lys36-Lys84; Helix 2, Asp100-Asn144; Helix 4, Thr259-Glu300) with a 20° curvature along its longest axis (Figs. 6, 7). The disordered N-terminal residues Leu18-Leu34, as well as residues Arg146-Ser202 and Lys227-Met255, constitute intrinsically disordered loops and thus were not resolved. Instead of distinct conformations, a cloud of electron density was observed, describing the average volume occupied by the loops (Supplementary Fig. 10d). Using MS-based disulphide mapping, we identified three disulphide bridges in the disordered head domain of ISG65 (Cys43-Cys200, Cys168-Cys181 and Cys240-Cys251) (Fig. 8a and Supplementary Fig. 10). The atomic model of the disordered regions was predicted using template-guided modelling in AlphaFold2 using real-space constraints from the reconstruction and disulphide
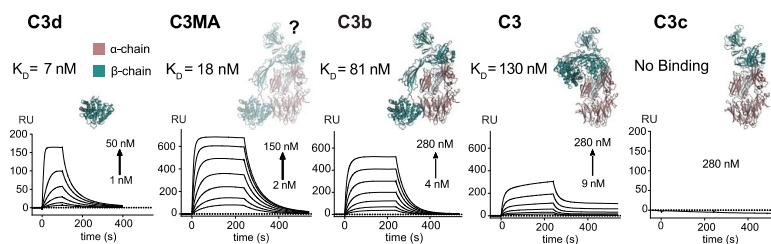
Fig. 5 | ISG65 binds proteolytic fragments of C3 with different affinities. Surface plasmon resonance sensorgrams showing the kinetics of ISG65 binding to C3d, C3MA, C3b, native C3, and C3c (from left to right). Increasing concentrations of C3 analytes were injected into the sensor chip, to which C-terminally biotinylated ISG65 (ligand) was immobilised. Structures of C3 fragments are shown above the corresponding sensorgrams. The atomic structure of C3MA (a mimic of C3(H$_2$O)) is unknown, but its overall architecture has been suggested to be similar to C3b[28], which is shown here instead, indicated with a question mark. Source data are provided as a Source Data file.

bonding as selection criteria for the models (Supplementary Fig. 10). Although resolved at a lower local resolution, the loop connecting helix 1 and helix 2 (Lys85-Asp100) close to the C-terminus could be modelled. Due to a well-resolved electron density, we could confidently place the short helix 3 (residues Gln217 to Val226) and the preceding loop (Asp210-Leu216) near the head domain. Residues Thr3-Asp210 were not resolved to atomic resolution but appeared to have a 3$_{10}$ helical conformation (Figs. 7, 8). Near the C-terminus of the model, Gln301-Leu305 form a loop connecting helix 4 with the short helix 5 (Thr306-Ala312), which points towards the N-terminus of ISG65, away from the parasite surface (Fig. 8b). Beyond these residues, only poorly resolved electron density was observed which precluded further model building. With the exception of the disordered regions Ala666-Ser672, Arg740-Leu751 and Glu1634-Asn1642, C3 was well-resolved and could be modelled in its entirety (Fig. 7). It has high overall similarity (RMSD = 2.0 Å, 11,878 atoms) with the previously published structure of C3[31] (PDB 2A73) (Supplementary Fig. 9c). Positioned above the loop (Gly684-Lys692) connecting the two N-terminal helices of the ANA domain and the N-terminal side of the longest helix in the domain (Ala719-His738) (Fig. 7), binding of ISG65 to C3 is predominantly mediated via hydrophobic contacts and four hydrogen bonds with the TED. Three hydrogen bonds are mediated by Tyr293 (Tyr293$_{ISG65}$-Glu1110$_{C3}$, Tyr293$_{ISG65}$-Glu1111$_{C3}$ and Tyr293$_{ISG65}$-Glu1119$_{C3}$) and one by Arg77 (Arg77$_{ISG65}$-Leu1109$_{C3}$) (Supplementary Table 2). The interaction between the proteins is further facilitated by the almost perfect complementary fit between the convex tip of TED and the concave 3-helix bundle of ISG65.

Validation of the ISG65:TED interface
Based on the higher-resolution structure of ISG65:C3, contacting interface residues (Supplementary Fig. 11a, b), were individually mutated to alanines. The binding affinities of the recombinantly expressed C3d and ISG65 mutants were determined by SPR and compared to the wild-type proteins (Fig. 7). Correct protein folding was assessed via circular dichroism spectroscopy (Supplementary Fig. 12). ISG65 residues across the entire interface contributed to the binding as indicated by the decrease in dissociation constants (Fig. 7 and Supplementary Table 3). Mutation of Tyr293, which is responsible for almost half of all contacts, completely abrogated ISG65 binding to C3d. Notably, mutation of Trp211 caused a decrease in K$_D$ to 42%. This is in agreement with the well-resolved electron density of the adjacent loop (Asp210-Leu216) as well as the side chain of Trp211. Although only contributing two hydrophobic contacts, this interaction might be crucial for the correct alignment of C3d at the upper end of the interface. Similarly, mutation of two of the four hydrogen bond-forming residues in C3d, Leu1109, and Glu1110 also leads to a decrease in binding affinity. This can also be attributed to the fact that the side chains of Leu1109 and Glu1110 are located at the core of the interface and engage in multiple hydrophobic contacts (Fig. 7). Mutation of Arg77 of ISG65 was not included in the analysis due to misfolding of the mutated protein. Further validation of the main interface was carried out by hydrogen-deuterium exchange mass spectrometry (HDX-MS) and small-angle X-ray scattering (SAXS) using C3d instead of C3 (Supplementary Fig. 11c, d and Supplementary Table 4). Both methods unequivocally confirmed our identification of the interface as described here. All interface residues are located in areas of high protection from deuterium uptake and the model of ISG65:C3d is in good agreement with solution scattering data of the same complex. Surprisingly, despite its localisation outside the interface, Arg1134 of C3d was found in a 'protection hotspot', indicating limited solvent accessibility (Supplementary Fig. 11c and Supplementary Data 2). In line with this finding, the mutation did indeed cause a decrease in K$_D$ to 30%. Considering that Arg1134 is located within a long (26 aa) loop close to the interface, it seems conceivable that a transient interaction with ISG65 during binding may occur.

ISG65 interacts with ANA/C3a
A second, smaller interface is formed between loops in the membrane-proximal region of ISG65 and the ANA domain in C3 (Fig. 7). Within the interface, we could identify two possible hydrogen bonds, Lys97$_{ISG65}$ – Glu689$_{C3}$ and Glu302$_{ISG65}$ – Gly717$_{C3}$ (Fig. 7). Although the reconstruction in this flexible region suffers from poor side chains density, the identified residues are well within bonding distance (Supplementary Table 2). To investigate whether the interface between ANA and ISG65 may be physiologically relevant, we tested the binding of ISG65 to recombinant C3a (the proteolytically liberated ANA domain), which revealed a K$_D$ of 18 µM (Fig. 7 and Supplementary Table 3). In agreement with the lack of clear electron density, kinetic profiles showed high on- and off-rates, indicative of a rather short-lived interaction. Specificity of the interaction was further confirmed by alanine mutations in ISG65, each resulting in reduced affinity with dissociation constants of 80 µM for Lys97Ala and 78 µM for Glu302Ala, respectively (Fig. 7, Supplementary Fig. 11b, and Supplementary Table 3).

Mechanisms of C3b binding on the parasite surface
Unlike other *Trypanosoma* receptors described so far, ISG65 is a type-I transmembrane protein that is spread over the entire cell surface where it is embedded into the VSG layer[32]. The C3-binding domain of ISG65 is hereby tethered to its membrane anchor via a 46 amino acid long, disordered linker. In an effort to understand how this linker could
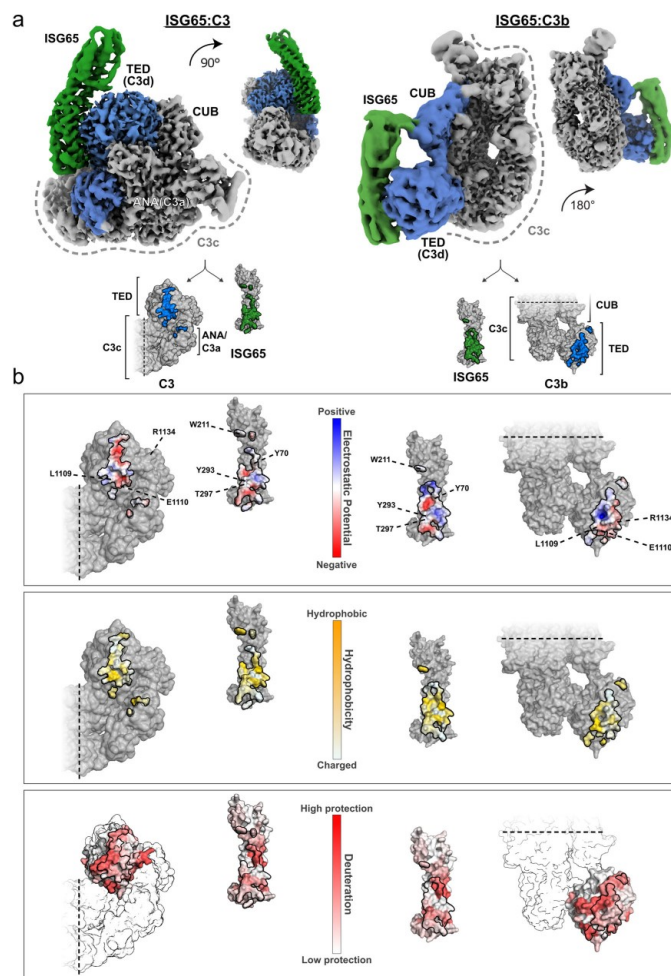
Article https://doi.org/10.1038/s41467-023-37988-7

Fig. 6 | Cryo-EM structures of ISG65 in complex with native C3 and C3b. a Cryo-EM density maps showing side views of ISG65:C3 (left), and ISG65:C3b (right) at two different angles. ISG65 is represented by map regions coloured in green. Interacting domains in C3 and C3b are depicted in blue. In both C3 conformations, TED provides the primary interface. ISG65 and TED show a high degree of shape complementarity. Smaller, secondary interfaces are located in ANA (native C3) and CUB (C3b). The remaining scaffold (C3c, grey) shows no additional contact points

b Open book representations of the interaction interface coloured according to electrostatic potential, hydrophobicity and protection from deuterium uptake as determined by HDX-MS (from top to bottom). HDX-MS measurements were performed with ISG65 and C3d. Positions of contacting residues that contribute to binding as determined by alanine mutagenesis are indicated. Interfaces are delimited by a black line.
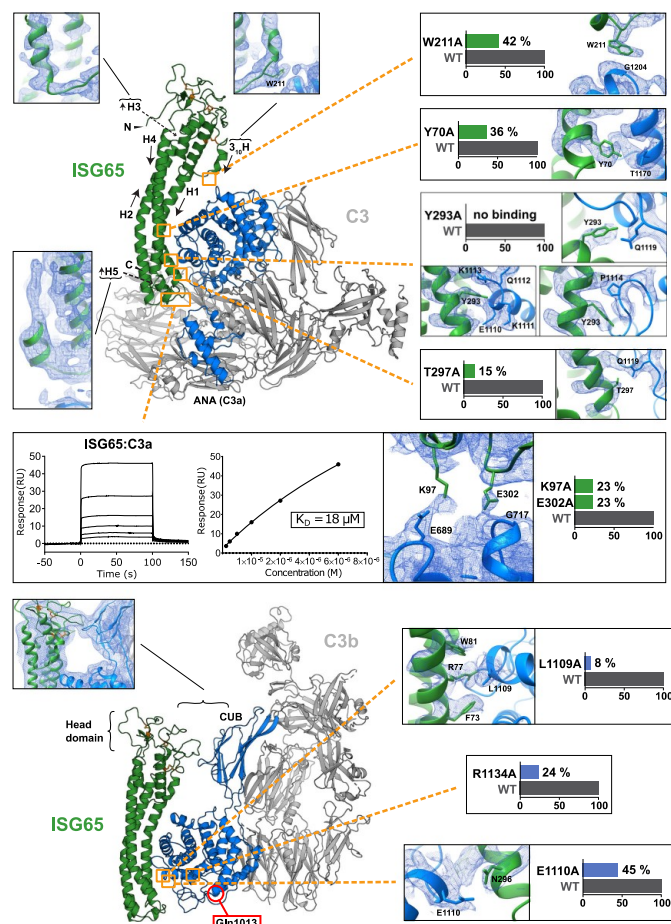
Fig. 7 | Close-up views of different regions in ISG65 and C3/C3b. Models of ISG65:C3 and ISG65:C3b are shown in cartoon representation, the directionality of helices in ISG65 is indicated by arrows. Gln1013 of the broken thioester bond is marked by a red circle. Orange rectangles indicate the positions of ISG65 and TED interface residues that were mutated. Presented bar diagrams show the relative effects of alanine mutations in ISG65 and TED on the overall binding affinity. Close-ups of relevant areas and interacting residues in the ISG65:C3 complex are shown in the unsharpened density map. Arg1134 of TED is not an interacting residue in the presented structures but was found to be protected from deuterium uptake in HDX-MS. Sensorgrams of the ISG65:C3a interaction are shown next to the respective close-up. Source data are provided as a Source Data file. The head domain of ISG65 was modelled using AlphaFold2 guided by experimental constraints (PDBDev: PDBDEV_00000201).

a
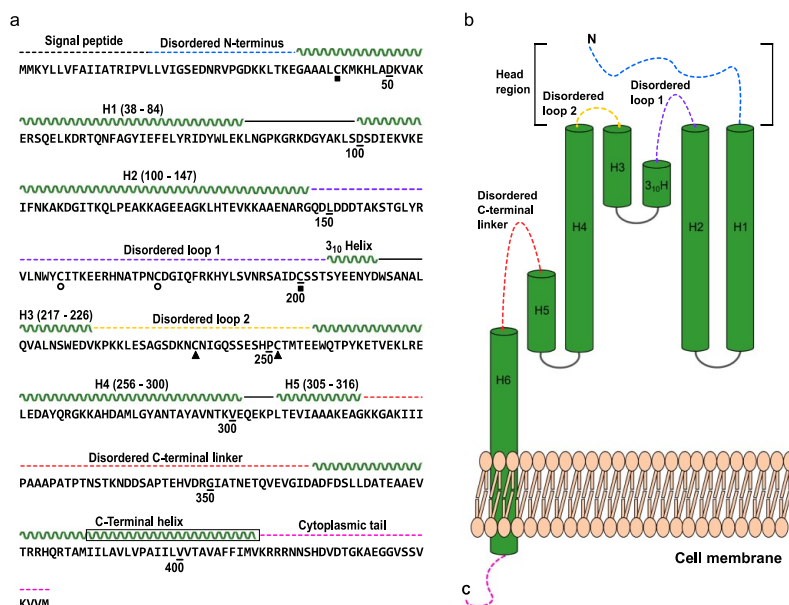
Fig. 8 | Schematic illustration of ISG65 primary and secondary structure.
a Amino acid sequence of full-length ISG65. Secondary structure elements are indicated above the sequence. Helices are displayed as wavy lines; loops present in the experimental model are displayed with solid black lines and loops absent from the experimental model are displayed as dashed lines. The colour scheme corresponds to panel **b**. Disulphide-bonded cysteine residues are indicated by squares, circles and triangles, paired residues are marked with the same geometric shape.

The transmembrane region of the C-terminal helix is highlighted by a black box around the stylised secondary structure. **b** Overview of the connectivity and arrangement of ISG65 secondary structure elements. Disordered regions are indicated as dashed lines. These regions did not give rise to traceable electron density and were amended to the final model based on an experimentally constrained AlphaFold2 model. The colouring scheme is the same as in panel a.

possibly affect the orientation of ISG65 on the cell surface, we determined its likely conformations in solution using SAXS and ensemble optimisation (EOM)[33]. EOM allows for the calculation of several co-existing conformational states of a molecule in solution, taking inter-domain flexibility into account. For ISG65, two distinct conformational pools could be identified (Supplementary Fig. 13). These constitute the ensemble with the most likely conformations of the disordered C-terminal linker that connects helix 5 with the partially membrane-embedded helix 6. While most of the models in the ensemble were compact (-100–115 Å), a small subset adopted an extended conformation of -170 Å (Supplementary Fig. 13e).

Using AlphaFold-assisted molecular modelling[34], we subsequently predicted a model of C-terminal helix 6 of ISG65 and the cytoplasmic domain that we appended to the models of the two most-populated conformational states of ISG65 (Supplementary Fig. 13d). The extra-cellular part of helix 6 spans -30 Å in length. On the trypanosome surface, the compact conformation of ISG65 would therefore extend approx. 130–145 Å from the plasma membrane of the parasite, while in the extended conformation, it could span more than 200 Å, thereby protruding beyond the boundaries of the VSG layer. Based on these findings, we suggest a plausible model for the interaction of ISG65 with

human complement C3b on the cell surface. It has previously been reported that, in *T. b. brucei*, the VSG coat extends about 140–155 Å from the cell membrane[35]. We propose that in its compact conformation, the 'down' position, ISG65 sits flush with the VSG coat, exposing only the disordered head domain to the outside, while other epitopes are hidden underneath the protective VSG umbrella (Fig. 9). Upon extension into its 'up' position, ISG65 protrudes the VSG coat by about 60 Å, enabling it to interact with C3b which is either near or bound to the VSG layer (Fig. 9). While ISG65-binding likely serves to keep C3b away from the plasma membrane, it cannot be ruled out that some fraction of complement escapes the interception, penetrates the VSG layer and covalently attaches to the plasma membrane where it might form convertases. Although the exact conformation of the covalently attached complement remains unknown[36] it is conceivable that ISG65 in its compact conformation might be able to interact due to the same relative position within the VSG layer (Fig. 9).

Discussion
Here we have presented the identification and structure of an invariant, immunogenic[37,38] yet abundant surface protein[18] of the human parasite *T. b. gambiense* that escapes neutralising antibody detection
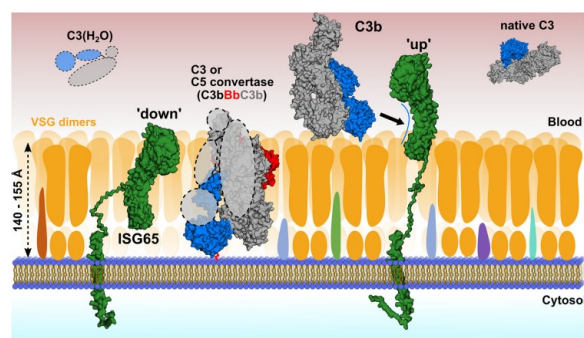
Fig. 9 | Model of complement binding on the surface of *T. b. gambiense*. In its compact 'down' position, ISG65 would remain embedded within the VSG coat, with its C3-binding epitopes being concealed and only the disordered head domain exposed to the outside. In this state, ISG65 may bind and inactivate AP convertases via C3b that may pass through the VSG coat, where it could covalently bind to the plasma membrane via Gln1013 of the former thioester (red connecting line). In its extended 'up' position, ISG65 would be to able intercept C3b outside the VSG coat and transport it to the flagellar pocket for uptake and subsequent lysosomal degradation. C3b would hereby either be present in close proximity to the coat or attached covalently or non-covalently to VSGs. C3(H₂O) and native C3 are depicted in the solution, as it is currently unknown whether they bind to the trypanosome surface. Surface receptors embedded within the VSG coat are depicted in different colours. The atomic structures of C3(H₂O) and the AP C5 convertase have not been determined and are therefore shown as dashed lines. Their overall architectures are presented as inferred earlier[5,28,68].

and likely takes over a role in facilitating innate immune escape through its interaction with human complement C3b. We show that ISG65 is competent to restrict the alternative complement pathway via C5 convertase, but that it does not play a role in the control of the classical and lectin pathway.

ISG65 is evolutionary related to VSGs and characterised by a similar, membrane-distal head domain. Intriguingly, unlike in the structures of GPI-anchored trypanosome receptors, the head domain has not been 'repurposed' for ligand binding[39–41]. It likely acts as an immuno-dominant decoy analogous to the same region in VSGs[42–44]. In ISG65, the ligand binding site is located at the membrane-proximal side of the helix bundle, near the C-terminus. The TED of C3 is the main site of interaction with ISG65, whereby its helix bundle follows the curvature of the globular TED. This strictly complementary conformation likely accounts for the high affinity in the absence of many hydrogen bonds or salt bridges. Additional contact points are present between the membrane-proximal end of ISG65 and ANA in native C3 as well as the head domain of ISG65 and CUB in C3b. Both interactions occur between less structured regions and may only be of transient nature, possibly contributing to correctly position ISG65 in different C3 conformations. For ANA, this is corroborated by the fast on- and off-rates that were measured between ISG65 and C3a.

While this study was under review, the crystal structure of ISG65 from the animal parasite *T. b. brucei* in complex with human C3d was published[45]. *Tbb* ISG65:C3d (PDB 7PI6) and *Tbg* ISG65:C3(TED) can be aligned with an RMSD = 1.7 Å (1791 atoms), confirming high similarity of the overall assembly (Supplementary Fig. 9a). Interface residues are present in similar positions in both structures with the exception of Arg1134 in TED, which is located in a long, flexible loop, and Trp211 in ISG65. Although likewise positioned in a loop region, the electron density for *Tbg* Trp211 is well-resolved and the region is protected from deuterium exchange. The importance of Trp211 for the interaction is further corroborated by affinity measurements using an alanine mutant, resulting in significantly reduced K_D. In contrast, in *Tbb* ISG65:C3d, Trp211 is not engaged with any C3d residues.

Using SPR, the authors of the aforementioned study[45] reported a comparatively low affinity for C3d and additionally claimed to have identified a second binding site located within human C3c (C3b lacking CUB and TED). This contradicts our own findings. *Tbg* ISG65, the receptor that encounters human C3 under natural infection conditions, showed a two magnitude higher affinity for C3d and revealed no interaction with C3c. This is in agreement with our cryo-EM structures which show no evidence for additional contact points with C3c.

As blood-dwelling parasites, African trypanosomes are in constant contact with toxic serum proteins that need to be swiftly removed from the cell surface before they can exert their damaging effects. The underlying mechanism has been described for VSGs and immunoglobulins. Motion-induced drag forces transport VSG:IgG complexes to the flagellar pocket at the posterior of the cell, where they are endocytosed[5]. Analogous to VSGs, ISG65 could mediate immune evasion by serving as a specific, high-affinity attachment point for C3b above the VSG layer, efficiently delivering a toxic cargo first to the flagellar pocket and then to the lysosome for degradation. This is in agreement with the high recycling rate, similar to VSGs, that has been shown for ISG65[46]. Efficient surface removal could thus help in preventing a critically high density of C3b on the surface that has been associated with C5 convertase formation[47]. While we did not demonstrate the surface binding of native C3 and C3(H₂O), it is possible that ISG65 can interact with these fragments in solution after being released from the surface. ISG65-shedding was reported to occur during the differentiation of *T. brucei*[38].

Although ISG65 is the likely site for C3b binding on the parasite surface, we cannot rule out that a fraction of C3b covalently attaches to the cell membrane, where it could serve as a nidus for convertase formation. Therefore, any capacity of ISG65 to block pathway progression, besides surface removal of C3b, could display an additional mechanism to avoid complement-mediated lysis.

Intriguingly, abrogation of haemolysis by soluble IGS65 and thus TP inhibition was only observed following initiation of the AP, but not of the CP. In the CP/LP, C3b is only part of the C5 convertase

C4bC2aC3b. Therefore, if ISG65 is binding to the CP/LP C5 convertase via C3b, it does not affect the cleavage of the C5 substrate and thus could not serve to restrict pathway progression.

Moreover, our immunofluorescence experiments (Fig. 3) indicate that in *T. b. gambiense* the CP/LP does not even progress to the formation of the CP/LP C3 convertase C4bC2a as no C4b was detected in the flagellar pocket. It is, however, unclear if the CP/LP is not triggered, as suggested earlier[21], or if it is inhibited upstream of C4b deposition. In contrast, we could identify Bb, a convertase component of the AP, but not C5b, the first factor of the TP, on the parasite surface, suggesting that the complement cascade is restricted upstream of TP initiation. These results are in agreement with an early study that showed deposition of AP convertases on the surface of *T. b. gambiense*, but found no evidence for activation of the CP[21]. They do, however, contradict the aforementioned recent study[45], which proposes the CP as the likely target of ISG65 based on mouse infection experiments which showed a decrease in parasitaemia progression of an ISG65 knock-out strain, temporally coinciding with the induction of infection-specific IgGs.

Here we provide detailed molecular insight, demonstrating that an effect of ISG65 on the CP/LP via binding to C3b is indeed unlikely and postulate the AP to be the most probable target.

However, how exactly does ISG65 restrict the AP via binding to C3? ISG65 binding to C3b and native C3 does neither prevent assembly of the C3 convertase (C3bBb) and the enzyme-substrate complex (C3bBb:C3), nor does binding to C3 substrate prevent its cleavage by C3bBb (Supplementary Fig. 5). It is, therefore, conceivable that inhibition of the cascade takes place at the level of the AP C5 convertase C3bBbC3b. Here, we could show that binding of soluble ISG65 to C3bBbC3b, pre-assembled on erythrocytes surfaces, significantly reduces cell lysis. This could be due to binding of ISG65 to either one or both C3b molecules of the convertase, thereby possibly reducing C5 substrate affinity and turnover. When ISG65 is additionally allowed to interfere with the assembly of the convertase, as we demonstrate by adding erythrocytes to serum preincubated with ISG65 (Fig. 4), haemolysis is almost fully abrogated. Therefore, in vitro, full inhibition of the AP C5 convertase appears to require ISG65 interference with both, convertase assembly as well as substrate cleavage. Due to the lack of structural information on the AP C5 convertase architecture, it remains to be elucidated which specific C3b interaction of ISG65 could hereby be rate-limiting for C5 turnover. Additional studies will thus be needed to obtain this crucial insight, but also to unravel which further molecular countermeasures trypanosomes employ to withstand other innate immune responses or prevent their activation entirely.

Methods
Cloning, expression and purification
ISG65. A DNA fragment (Genewiz) encoding amino acids (aa) 18–363 of ISG65 from *T. b. gambiense* (Tbg.972.2.1600) was codon optimised for bacterial expression and cloned into the pET15b plasmid using the Gibson assembly method (New England Biolabs). For recombinant expression in *E. coli* T7 shuffle cells (New England Biolabs), a plasmid encoding an N-terminal hexahistidine-tag (His-ISG65$_{18-363}$) and another with an additional C-terminal Avi-tag (His-ISG65$_{18-363}$-Avi) were generated. Amino acids substitutions were introduced into the His-ISG65$_{18-363}$ construct using the Q5 site-directed mutagenesis kit (New England Biolabs). For all ISG65 variants, the same expression and purification protocol was employed. Proteins were produced overnight at 22 °C after induction of protein expression with 1 mM isopropyl-β-D-thiogalacto-pyranoside (IPTG).

Cells were harvested by centrifugation at 6000 × *g*, 4 °C for 30 min. Cell pellets were resuspended in Buffer A (20 mM Tris pH 8.0, 500 mM NaCl, and 10 mM imidazole) and phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 1 mM immediately before cell lysis using an EmulsiFlex-C3 (AVESTIN Europe) with 1000 to 1100 bar lysis pressure. The supernatant was cleared from cell debris by centrifugation at 40.000 × *g*, 4 °C for 1 h.

Soluble protein was subsequently purified from the cleared lysate by immobilised metal-affinity chromatography (IMAC) using nickel-nitrilotriacetic acid (NTA) beads (Qiagen), pre-equilibrated with Buffer A, and gravity-flow columns (Bio-Rad).

After application of the lysate, the beads were washed twice with 10 column volumes (CV) of Buffer A before eluting the immobilised proteins with 10 CV Buffer A, supplemented with 400 mM imidazole. The eluate was fractionated and collected. Fractions containing the protein of interest, as identified by SDS-PAGE, were pooled and dialysed overnight into 20 mM Tris (pH 8.0), 150 mM NaCl at 4 °C, using SnakeSkin Dialysis Tubing (Thermo Fisher Scientific). After dialysis, the purified proteins were concentrated using Amicon Ultra centrifugal filters (Merck Millipore) and subjected to size-exclusion chromatography using a Superdex 200 Increase 10/300 GL (GE Healthcare) equilibrated with 20 mM HEPES (pH 7.5) and 150 mM NaCl. 0.5 mL fractions were collected throughout, and fractions containing the protein of interest, as identified by SDS-PAGE, were pooled and flash-frozen in liquid nitrogen.

ISG75. The gene fragment encoding the extracellular domain of ISG75 (*T. b. gambiense* LiTat 1.3, accession number DQ200220.1) comprising residues 29–462 of the full-length protein was cloned into the mammalian expression vector pHLsec[48]. It encodes an N-terminal secretion sequence and a thrombin-cleavable, C-terminal hexahistidine-tag. Transfection of Expi293F cells (Thermo Fisher) was carried out according to manufacturer instructions. Briefly, 100 ml of exponentially growing cells were transfected with 100 μg of the plasmid. About 5 μM Kifunensine was added to the cell suspension immediately after the transfection to produce a homogenous glycosylation pattern. Cells were harvested 2 days after transfection by centrifugation (10 min, 2000 × *g*, 20 °C). The supernatant was filtered using a 0.45 μm bottle top filter unit and dialysed twice for at least 4 h against 20 volumes of IMAC equilibration buffer (20 mM Tris, 500 mM NaCl, 10 mM Imidazole, and pH 8.0). The dialysed cell supernatant was applied to Nickel-NTA agarose beads (Qiagen) using a gravity-flow column (Bio-Rad), washed with 20 column volumes of IMAC wash buffer (20 mM Tris, 500 mM NaCl, 30 mM Imidazole, and pH 8.0) and the bound protein was eluted with five column volumes of IMAC elution buffer (20 mM Tris, 500 mM NaCl, 400 mM Imidazole, and pH 8.0). Cleavage of the His-tag was carried out by the addition of thrombin (GE Healthcare) to the eluate, which was immediately dialysed against IMAC equilibration buffer at 20 °C overnight. The cleaved protein was separated from the uncleaved protein by applying it again to Nickel beads. The flow-through was subsequently concentrated to 500 μL using an Amicon centrifugal ultrafiltration device (Merck Millipore) with a 10 kDa molecular weight cut-off (3300 × *g*, 4 °C) and injected onto a Superdex S200 10/300 increase size-exclusion column (GE Healthcare) previously equilibrated with SEC running buffer (20 mM HEPES, 150 mM NaCl, and pH 7.5). The fractions containing *Tbg*ISG75 were pooled, flash-frozen in liquid N$_2$ and stored at −80 °C. Each purification step was monitored by SDS-PAGE for the absence of contaminations as well as degradation and >95% purity of the target protein in the final step.

Human complement factor 3. Native C3 was purified from normal human serum (Sigma-Aldrich)[49].

In short, 15% PEG4000 in 0.1 M sodium phosphate (pH 7.4) was added to serum to a final concentration of 5% PEG4000. The precipitate was removed by centrifugation (6000 × *g*, 4 °C, 30 min). 26% PEG4000 in 0.1 M sodium phosphate (pH 7.4) was added to the supernatant to a final concentration of 12% PEG4000. After centrifugation (9000 × *g*, 4 °C, 40 min), the supernatant was discarded and the pellet was immediately resuspended in 0.1 M sodium phosphate (pH 7.4). PMSF was added to a final concentration of 0.5 mM.

Using a gravity-flow column (Bio-Rad), the sample was passed over 25 mL of L-lysine-ceramic hyper D F Hydrogel resin (Sigma-Aldrich), pre-equilibrated with 0.1 M sodium phosphate (pH 7.4) and the flow-through was collected. The resin was washed with 100 mL of cold EDTA (5 mM, pH 8) and the wash fraction was collected. Flow-through and wash were combined and diluted to a specific conductance of 2.5 mS/cm with cold EDTA (5 mM, pH 8). 6-aminohexanoate and PMSF were added to a final concentration of 50 mM and 0.5 mM, respectively. The sample was loaded onto a hand-packed Source 15Q (Amersham) anion exchanger column (10 mL bed volume), pre-equilibrated with Buffer A (25 mM sodium-potassium phosphate, 10 mM EDTA, 50 mM 6-aminohexanoate, 0.5 mM PMSF, pH 7.4). After washing with 5 column volumes of Buffer A, a linear gradient to 100% Buffer B (25 mM sodium-potassium phosphate, 10 mM EDTA, 50 mM 6-aminohexanoate, 0.5 mM PMSF, 500 mM NaCl, pH 7.4) was applied over 30 column volumes, 1 mL fractions were collected throughout the elution. Fractions containing C3 were identified using SDS-PAGE, pooled and concentrated using an Amicon centrifugal ultrafiltration device (Merck Millipore) with a 50 kDa molecular weight cut-off (3300 × g, 4 °C) before being subjected to size-exclusion chromatography using a Superdex 200 16/600 column (GE Healthcare), pre-equilibrated with SEC running buffer (20 mM HEPES, 150 mM NaCl, pH 7.5). Fractions containing the protein of interest, as identified by SDS-PAGE, were pooled and flash-frozen in liquid nitrogen.

Purification and generation of the fragments C3b and C3MA were carried out as described[34].

In short, the thioester in C3 was hydrolysed using methylamine (MA) and deactivated by incubation with iodoacetamide. C3c and C3a were purchased from Complement Technology, Inc. C3d (His-C3d) was produced recombinantly in *E. coli* as described[50]. Amino acid substitutions were introduced using the same method as for ISG65. For SPR assays, a C3d construct (His-C3d-Avi) with an N-terminal His-tag and a C-terminal Avi-tag was generated. Cloning of this construct into pET15b was done by the Gibson assembly method. Expression and purification of all C3d constructs were carried out in the same way as described for ISG65.

Pull-down assay
About 400 µL of settled Ni-NTA resin (Qiagen) were incubated with 200 µg of His-tagged ISG65, washed with 2.5 ml IMAC wash buffer (20 mM Tris, 500 mM NaCl, 20 mM Imidazole, pH 8.0) and incubated with 2.5 ml of human serum. Subsequently, the beads were washed again using the same buffer, before the bound protein was eluted in a total volume of 2 ml IMAC elution buffer (20 mM Tris, 500 mM NaCl, 400 mM Imidazole, pH 8.0). Four fractions were collected and analysed by SDS-PAGE and Coomassie-staining. All steps were carried out in gravity-flow columns (Bio-Rad). ISG65 (without serum) and serum (without ISG65 as bait protein) were used as negative controls.

Western blotting
All Western Blots were performed using the iBlot Dry Blotting System (Invitrogen) and iBlot PVDF Mini Gel Transfer Stacks (Invitrogen) as per the manufacturer's instructions. Blotting was conducted using 'Programme 3' with default settings (20 V, 7 min). After the transfer, membranes for Western Blots of His-tagged proteins were blocked (1 h, RT, shaking) using a Blocking buffer (Penta-His HRP Conjugate Kit; Qiagen), prepared according to the manufacturer's instructions. The membrane was washed three times with PBS-T (0.05% Tween 20) for 5 min, before incubating in a working solution (Blocking Buffer, 0,05% Tween 20, Anti-His HRP conjugate [1/2000 dilution]) for 1 h. The membrane was washed three times for 10 min with PBS-T before removing the wash and developing the Western Blot using Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) according to the manufacturer's instructions. Blots were imaged within 30 min of substrate addition.

After Western Blotting of C3a-containing proteins, membranes were blocked in a Blocking solution (25 mg/mL bovine serum albumin (BSA) in PBS-T) (1 h, RT, shaking).

After rinsing with PBS-T, the membrane was incubated in a working solution (12.5 mg/mL BSA in PBS-T, Mouse anti-human C3a (MAB3677, R&D systems) [1/500 dilution]). After washing three times for 10 min with PBS-T, the membrane was incubated with the detection antibody (12.5 mg/mL BSA in PBS-T, Goat Anti-mouse HRP conjugate (A28177, Invitrogen) [1/10000 dilution]) for 1 h. After washing three times for 10 min with PBS-T, the wash was removed and the Western blot was developed using Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) according to the manufacturer's instructions. Blots were imaged within 30 min of substrate addition. All Chemiluminescence and digitised images were acquired on Image-Quant LAS 4000 and ImageQuant LAS 4000 mini camera systems (GE Healthcare).

In-gel digestion and LC-MS/MS analysis
Single gel bands (three in total), as depicted in Supplementary Fig. 1, were cut, chopped into small pieces, reduced with dithiothreitol, alkylated with iodoacetamide and digested with trypsin overnight. Peptides were extracted from each gel piece, lyophilised in speed-vac, dissolved in 0.1% formic acid and 1/5th of the dissolved sample was separated on an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a Mass Spectrometer Orbitrap Fusion Lumos (Thermo Fisher Scientific). No normalisation was carried out since bands were cut manually. The peptides were trapped and desalted with 2% acetonitrile in 0.1% formic acid at a flow rate of 5 µL/min on an Acclaim PepMap100 column (5 µm, 5 mm by 300 µm internal diameter (ID); Thermo Fisher Scientific). Eluted peptides were separated using an Acclaim PepMap100 analytical column (2 µm, 50 cm × 75 µm ID; Thermo Fisher Scientific). Using a constant flow rate of 300 nL/min a 65 min elution gradient was started at 5% B (0.1% formic acid in 99.9% acetonitrile) and 95% A (0.1% formic acid). The gradient reached 30% B at 52 min, 90% B at 53 min, and was then kept constant until 57 min before being reduced to 5% B at 58 min. For the first minute, nanospray was set to 1600 V, 275 °C source temperature, measuring the scans in the range of m/z 350–2000. An orbitrap detector was used for MS with the resolution 120,000, AGC target value was set as custom with a normalised AGC target of 250%. Maximum injection time was set to 50 ms, MSMS was acquired also using orbitrap with resolution 30,000, the data were acquired in a data-dependent manner, ions were fragmented by HCD collision energy set to 30% with dynamic exclusion set to 60 s.

The software Proteome Discoverer 2.3 (Thermo Fisher Scientific) was used for peptide and protein identification using Sequest and MS Amanda as search engines and databases of *Homo sapiens* (downloaded from Uniprot combining reviewed and unreviewed sequences on November 12 2018) and common contaminants (downloaded from Proteome Discoverer software on December 21 2018). The protein sequence of ISG65 was included in a database for the search. Mass tolerance for MS was 10 ppm and MSMS 0.6 Da for Sequest, tolerance in MS 5 ppm and MSMS 0.06 Da in MS Amanda. The fixed modification included carbamidomethylation on cysteine, variable modifications were set to methionine oxidation and deamidation of asparagine and glutamine, two missed cleavages were allowed, FDR target was set to 1%.

Cross-linking
ISG65:C3b complex was cross-linked using homobifunctional cross-linker DSA (Di-(N-succinimidyl)adipate). For cross-linking, proteins were transferred into a buffer containing 20 mM HEPES (pH 7.5) and 150 mM NaCl. ISG65 and C3b were mixed in a 2:1 molar ratio, and freshly prepared DSA was added to the proteins at a hundred molar excess. The reaction mixture was incubated at room temperature for

30 min and quenched with 10 molar excess of ethanolamine. For EM-grid preparation, the cross-linked complex was also gel-filtered on Superdex 200 Increase 10/300 GL (GE Healthcare).

C3b surface binding and uptake analysis

About 4×10^5 *T. brucei gambiense* cells were harvested from a mid-exponential grown culture and resuspended in 200 µl serum-free HMI-11 supplemented with 1% bovine serum albumin. Uptake assays were carried out with AF_{594} labelled C3b at a concentration of 50 µg/ml or AF_{594} labelled human transferrin (T13343, Thermo Fisher Scientific) as a positive control at a concentration of 50 µg/ml, with or without the addition of recombinant ISG65 or ISG75 extracellular fragment at a 1.5-fold molar excess (preincubated with C3b for 10 min). Samples were incubated at 37 °C and harvested at different timepoints. To analyse the contribution of surface binding, samples were pre-cooled on ice for 10 min before mixing with test proteins and harvested immediately. To demonstrate specific, receptor-mediated uptake, we compared $C3b_{AF594}$ surface binding and uptake at a concentration of 10 µg/ml in the presence and absence of non-labelled C3b at a concentration of 40 µg/ml (adding 80 µg/ml ISG65 or ISG75, respectively). Cells were harvested by centrifugation ($1400 \times g$, 4 °C), washed twice with ice-cold PBS and resuspended in 500 µl PBS. Cells were fixed for 30 min at RT after the addition of an equal volume of 4% formaldehyde in PBS, then washed once with PBS and resuspended in a final volume of 300 µl PBS. An aliquot of each sample was mounted on a glass slide with Vectashield Antifade mounting medium with 2-(4-amidinophenyl)−1H-indole-6-carboxamidine (DAPI) for fluorescence microscopy imaging on a Leica TCS SP8 WLL SMD-FLIM inverted confocal microscope. Images were processed in Fiji[51]. Samples were analysed by flow cytometry on an LS Fortessa (BD Biosciences) using an excitation wavelength of 560 nm in combination with a PE-Texas Red filter ($\lambda_{emission} = 610$ nm, bandpass = 20 nm) and counting 10^4 cells. Data were processed using BD FACSDiva software v8.0.1 using a refined gate based on size versus granularity to exclude cell debris. Untreated cells were subjected to the same procedure in parallel as negative control and used to define the gate for AF_{594} positive cells. Flow cytometry standard (FCS) files used in this study are available in the Flow-Repository database[52] under accession code FR-FCM-Z5XC.

Detection of complement factor surface deposition by immunofluorescence

About 1×10^5 *T. brucei gambiense* cells were harvested from a mid-exponential grown culture ($800 \times g$, 8 min, RT), washed once in 1 ml serum-free HMI-11 supplemented with 1% bovine serum albumin and then resuspended in 100 µl human serum (complete serum, C2-depleted serum (C0913, Sigma) or Fb-depleted serum (A506; Quidel)) containing 5 µg [w/v] dilution) AF594 (complement factor Bb monoclonal antibody (10-09; Invitrogen); complement C4 polyclonal (ab47788; Abcam); complement C5 polyclonal antibody (PA596933; Invitrogen) or fluorescein isothiocyanate (FITC) labelled antibody (complement C3 polyclonal, (PA1-28933, Invitrogen)). Unlabelled antibodies were fluorescently labelled using the Alexa Fluor 594 Conjugation Kit (Fast)−Lightning-Link (Abcam).

After 2 min at RT, 10 mM sodium azide was added and cells were sedimented ($800 \times g$, 8 min, 4 °C). Cells were resuspended in 100 µl PBS containing the respective antibody (at the same dilution) and incubated on ice for 2 min before addition of 0.2% formaldehyde, followed by two washes in ice-cold PBS and mounting on microscopy slides. For testing complement deposition on sensitised sheep erythrocytes, 2 µl of concentrated cells (CH50 Test, K002-1; Haemoscan) were diluted in 1 ml diluent (K002-2, Haemoscan), sedimented by centrifugation ($400 \times g$, 5 min, 4 °C) and then resuspended in 100 µl human serum (C6-depleted serum (C1288, Sigma) or C7-depleted serum (C1413, Sigma)). Erythrocytes were incubated for 15 min at 37 °C, then sedimented ($400 \times g$, 5 min, 4 °C) and resuspended in 100 µl cold PBS containing 5 µg (5% [w/v] dilution) labelled antibody. After 15 min incubation under gentle tumbling, formaldehyde was added to 0.2% and erythrocytes were washed twice in ice-cold PBS. An aliquot of each sample was mounted on a glass slide with Vectashield Antifade mounting medium with DAPI for fluorescence microscopy imaging on a Leica TCS SP8 WLL SMD-FLIM inverted confocal microscope. Images were processed in Fiji[51].

Haemolysis assays

Normal human serum. Inhibition of complement haemolytic activity by ISG65 and ISG75 (negative control) was assessed using the AP50 and CH50 test kits (HaemoScan), according to the manufacturer's instructions. Dilutions (2–15.2-fold (AP50) & 4-128-fold (CH50)) of non-heat inactivated human serum (Sigma-Aldrich) were incubated with 24 µM ISG65 and ISG75 (CH50 assay) or serially diluted ISG65 and ISG75 (7.5–0 µM) (AP50 assay).

Complement inhibition was determined from the extent of lysis of rabbit (AP50) or antibody-sensitised sheep erythrocytes (CH50) measured at 415 nm. The concentration of free haemoglobin is hereby directly proportional to complement activity. The measurements were performed in technical triplicates ($n = 3$, AP50) or quadruplets ($n = 4$, CH50). Graphs were created using GraphPad Prism.

Complement factor depleted serum. Inhibition of complement haemolytic activity by ISG65 and ISG75 (negative control) was assessed using the AP50 test kit (HaemoScan), using an adapted version of the manufacturer's instructions. Rabbit erythrocytes were incubated with dilutions (2–15.2-fold) of human complement C5-depleted serum (Sigma-Aldrich). After incubation for 30 min at 37 °C, the erythrocytes were centrifuged ($400 \times g$, 10 min, RT), the supernatant removed, and the erythrocytes washed twice with dilution buffer (HaemoScan). The erythrocytes were resuspended and dilutions (2–15.2-fold) of human complement factor C3-depleted serum (Sigma-Aldrich), were incubated with 7.5 µM ISG65 or ISG75 added. Lysis was performed for 30 min at 37 °C.

The modified protocol was validated by efficient lysis of erythrocytes when both sera were consecutively used (reconstituted NHS). Exchange of either serum with buffer prevented lysis. Complement inhibition was determined as described for normal human serum.

Intact mass measurement

Samples were diluted to 0.2 mg/ml with 0.1% formic acid. About 5 µl were injected into a desalting column (MassPREP desalting, Waters) and desalted using a fast gradient (4 min) of acetonitrile in water supplemented with 0.1% formic acid. The separation was carried out by an LC system (I-class, Waters) that is online coupled to a mass spectrometer (Synapt G2, Waters) to acquire protein masses by electrospray ionisation. Prior to sample measurement, the system was calibrated with sodium formate in the range of 100–1500 m/z. The calibration was checked by measuring myoglobins ions in the same range. For myoglobin ions, the ppm error was less than 1 ppm. For the measurement of samples of interest, the instrument was set up to accept a calibration error of less than 10 ppm. The raw spectrum was processed in MassLynx (Waters). The deconvolution was performed in MaxEnt1 with the mass range 160,000–200,000 m/z, resolution 0.5 Da/channel and uniform Gaussian used to calculate the damage model (Waters). The final spectrum was produced in the mMass software (mmass.org)[53].

N-terminal protein sequencing

Purified proteins were run on reducing SDS-PAGE, electroblotted onto PVDF membrane and stained with Coomassie Brilliant Blue R-250. The N-terminal amino acid sequence of the α-chain of C3b was determined by Edman degradation using Procise 494 cLC Protein Sequencing

System (Applied Biosystems), following the manufacturer's recommendation. Briefly, in each sequencing cycle, the N-terminus of the protein was treated with phenylisothiocyanate, and specifically released phenylthiohydantoin amino acid derivative was analysed by reverse-phase HPLC in 3.5% tetrahydrofurane and 12% n-propanol/acetonitrile mobile phases, respectively.

Surface plasmon resonance

Binding-impaired mutants. Binding-impaired mutants of ISG65 and C3d were subjected to kinetic binding analysis using SPR. For interrogation of ISG65 mutants, His-C3d-Avi was used as a ligand, and for C3d mutants, His-ISG65$_{18-363}$-Avi was the ligand. Site-specific Avi-tag biotinylation for immobilisation on the SPR chip was carried out as previously described[54]. SPR experiments were performed at 25 °C using Series S sensor chip CAP (Cytiva) on a BIAcore T200 system (GE Healthcare). All binding analyses, as well as dilutions, were performed in SPR running buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) TWEEN-20). Biotinylated ligands were coupled to flow path 2 at 10 μl min^{-1} for 120 s. Dilution series (1:1) of both analytes (ISG65, 1.5–100 nM; C3d, 0.78–50 nM) were applied to flow paths 1 and 2 at 30 μl min^{-1} for 120 s, followed by 300 s of dissociation time. The chip surface was regenerated in between titrations with 6 M guanidinium hydrochloride dissolved in 0.25 M sodium hydroxide. The binding data was reference subtracted, and kinetic and steady-state affinity parameters were evaluated using BIAcore T200 evaluation software (GE Healthcare).

C3 fragments. Binding analyses of native C3, C3b, C3MA and C3c to immobilised His-ISG65$_{18-363}$-Avi was carried out analogous to the method described above for the binding-impaired mutants. The binding and dissociation times of analytes were altered depending on the kinetic profiles. For binding analyses of C3a, His-ISG65$_{18-363}$ was covalently immobilised via amine coupling according to the manufacturer's instructions (amine coupling kit, Cytiva). The reference channel was activated and subsequently quenched without immobilisation of a ligand. C3a (8, 4, 2, 1, 0.5, 0.25 μM) was flown over both channels at 30 μl min^{-1}, allowing 100 s for association and 300 s for dissociation. Between kinetic runs, the chip surface was regenerated with 1 M sodium chloride for 30 s at 30 μl min^{-1}.

Disulphide mapping

For detection of disulphide bonds under low pH conditions to avoid disulphide-bridge scrambling, 100 pmol of ISG65 were diluted into 0.5 M glycine, pH 2.3 and online digested using a Nepenthesin-2 column (AffiPro)[55]. Generated peptides were trapped and desalted on a micro-trap column (Luna Omega 5 μm Polar C18 100 Å Micro Trap 20 × 0.3 mm) for 3 min at a 200 μL/min using an isocratic pump delivering 0.4 % formic acid in water. Both the protease column and trap column were placed in an icebox. After 3 min, peptides were separated on a C18 reversed-phase column (Luna Omega 1.6 μm Polar C18 100 Å, 100 × 1.0 mm) with a linear gradient 5–35% B in 26 min, where solvent A was 2% acetonitrile/0.4% formic acid in water and solvent B 95% acetonitrile/5% water/0.4% formic acid. Peptides were measured by a timsToF Pro PASEF mass spectrometer (Bruker Daltonics) operated in a positive data-dependent mode in the m/z range 300–1600. The scheduling target intensity was set at 20,000. The intensity threshold was 2500 and the PASEF charge range was 1–6. The mass spectrometer was externally calibrated using an ESI low-concentration tuning mix (Agilent Technologies). The data were processed by Data Analysis 5.0 software (Bruker Daltonics) and ProteinScape 4.0 (Bruker Daltonics). Non-specific cleavage and dehydro-cysteine were used as variable modifications in the peptide search. The peptide tolerance was 15 ppm and MS/MS tolerance was 0.1 Da. The data were searched using MASCOT against the database containing ISG65 protein. The MASCOT ions score cut-

off was 2, the peptide rank cut-off was set at 3, and the minimal peptide length was 4.

Cryo-electron microscopy
Grid preparation and data collection
ISG65:C3. About 3 μl (0.16 mg/ml) of ISG65:C3 were applied to freshly glow-discharged copper C/Flat 1.2/1.3 300 mesh grids (Protochips) and vitrified by being plunged into liquid ethane using a Thermo Fisher Scientific Vitrobot Mark IV system (4 °C, 100% relative humidity, no wait time, 2 s blotting time). The grids were then transferred to a Titan Krios G3i microscope (Thermo Fisher Scientific) equipped with a K3 detector (GATAN Inc.) and operated at 300 kV. Images were recorded at a magnification of 105,000X while tilting the stage by 25°, yielding a pixel size of 0.86 Å. Multi-frame movies (40 frames with a total dose of 41.10 e$^-$/Å^2 were recorded using a nominal defocus (dF) range of −2.5 to −1.5 μm. Data were collected using the automated data collection software EPU (Thermo Fisher Scientific).

ISG65:C3b. To prepare grids used in the collection of datasets at a 25° tilt, 3.5 μl (0.1 mg/ml) of ISG65:C3b complex was applied to glow-discharged copper C/Flat, 300 mesh 1.2/1.3, and 2/2 TEM grids (Protochips). The grids were vitrified by being plunged into liquid ethane using the Thermo Scientific Vitrobot Mark IV system (4 °C, 100% relative humidity, 30 s waiting time, 4 s blotting time) and then transferred to a Titan Krios microscope (Thermo Fisher Scientific) for data acquisition.

For data collections at a 0° stage tilt, Au, 300 mesh, R1.2/1.3 TEM grids (Protochips) were coated with a graphene monolayer using an in-house developed protocol. About 3.5 μl (0.15 mg/ml) of ISG65:C3b complex was applied to freshly plasma-cleaned TEM grids and vitrified as explained for grids used to obtain the 25° tilt dataset. The grids were subsequently transferred to a Titan Krios microscope (Thermo Fisher Scientific) for data acquisition.

The data were collected at 300 kV using SerialEM[56] software. The data were collected on a K2 direct electron detection camera positioned behind a Gatan Imaging Filter (Bioquantum 967, Gatan). The camera was operated in the electron counting mode, and the data were collected at the calibrated pixel size of 0.818 Å/px. The data from a 5.0 s exposure were split into 40 frames comprising an overall dose of 60 e$^-$/Å^2.

Image processing
ISG65:C3. A total of 10,380 movies were acquired during data collection. Motion correction, as well as CTF correction (CTFFIND4 wrapper), were done using cryoSPARC[57]. About 5,508,854 particles were picked using the reference-free Gaussian Blob picker in cryoSPARC. Particles were extracted with a 400 px box with twofold binning applied, which resulted in a pixel size of 1.72 Å/px. Iterative panning using reference-free 2D classification in cryoSPARC was performed until a subset of 432,224 particles was identified. The particles were re-extracted using a 400 px box with no binning applied. The re-extracted particles were submitted to another round of reference-free 2D classification in Relion 3.1[58]. About 406,545 particles were selected and used for ab initio modelling and subsequently subjected to the 3D classification job in Relion. Particles were classified into five classes, all yielding well-defined reconstructions of the desired complex. All particles selected after 2D classification in Relion were later used for the creation of an ab initio model in cryoSPARC, which was subsequently refined using the non-uniform refinement procedure (cryoSPARC v3.2 and later) and local CTF refinement to yield the final reconstruction.

ISG65:C3b. A total of 18,062 movies were collected at a 25° stage tilt in two separate data collections (for two different grid types), and 8960 movies were collected at a 0° stage tilt on graphene-coated grids. The processing of each dataset was performed independently before

merging selected particles with previous data collections. Motion correction and CTF correction for all 27,022 movies were performed using cryoSPARC's patch motion and patch CTF correction implementations, respectively. On images acquired at a 25° stage tilt, particles were picked using the convolutional neural network-based particle picker, crYOLO[59]. Particles were initially identified using the supplied general model. The resulting particle picks were manually inspected, and reference-free 2D classification was performed in cryoSPARC. Particles constituting the 2D classes which represented the protein (regardless of whether they were the desired complex or not) were selected and used to re-train the outer layers of the general model on a subset of micrographs at varying defocus. crYOLO identified a total of 2,836,445 particles (1,414,868 and 1,421,577). The processing of the two datasets followed the same overall processing pipeline, using cryoSPARC. Particles were extracted using a box with 400 px and twofold binning, resulting in a pixel size of 1.656 Å/px. The iterative, reference-free 2D classification was performed until satisfactory 2D classes were identified. The selected particles were subjected to ab initio modelling using 3 classes. Models representative of the desired ISG65:C3b complex and C3b were chosen for subsequent Heterogenous Refinement. Particles constituting the reconstruction representing ISG65:C3b were selected, re-extracted using a 400 px box (no binning applied resulted in a pixel size of 0.828 Å/px), and subjected to Non-uniform Refinement.

For images acquired at a 0° stage tilt, the Gaussian particle picker in cryoSPARC was utilised to identify particles. The remaining processing was performed as described for data collected at a 25° stage tilt.

For the final reconstruction, 204,946 combined particles were re-subjected to ab initio modelling using 2 classes. The class representative of the desired complex was subjected to Non-uniform Refinement and local CTF refinement. A total of 145,172 particles, of which 54,453 particles came from datasets at 25° tilt (22,266 and 32,187, respectively) and 90,719 particles came from a 0° tilt dataset, contributed to the final reconstruction.

Model building and refinement
ISG65: C3. The initial step in the model building was performed by docking starting models (the crystal structure of C3 (PDB 2A73)) and an ISG65 model predicted by AlphaFold2[34] into the obtained electron density map using Phenix dock_in_map[60]. After an initial real-space refinement using Phenix real_space_refine, both models were independently and iteratively refined using a combination of manual refinements in Coot as well as automated real-space refinements using Phenix and REFMACS[61]. Model validation was performed using MolProbity (http://molprobity.biochem.duke.edu)[62]. After the individual models had reached satisfactory validation metrics and map correlation, they were manually and automatically refined together in Coot[63] and Phenix[60] to model interactions between the two proteins.

ISG65: C3b. Due to large variations in the local resolution of the obtained electron density map, the modelling of ISG65 at atomic resolution was not possible. Therefore, the atomic model obtained from the ISG65:C3 data was docked in the electron density map using Phenix dock_in_map[60]. For C3b, the same was done using the crystal structure of C3b (PDB 2I07). After an initial real-space refinement using Phenix real_space_refine[60], the model was iteratively refined using a combination of manual refinements in Coot[63] as well as automated real-space refinements in Phenix and REFMACS[61]. Model validation was performed using MolProbity[62].

Molecular modelling
The model of the full-length mature ISG65 (aa 18–436), excluding the signal sequence but including the transmembrane domain and the cytoplasmic tail, was predicted by using the cryo-EM structure of ISG65 (aa 35–316) as a template into AlphaFold2. The obtained molecular model was used to complete the experimentally determined structure of the C3-binding domain of ISG65. The model was previously incomplete due to disorder in the N-terminal portion (aa 18–34), in the loop-rich membrane-distal head domain (aa 147–201 and 227–252), and in the C-terminal portion. A model was selected based on the presence of disulphide bonds that were determined experimentally prior to modelling as well by the general fit of their head domains to the experimental electron density that delineated this region. No constraints were imposed on the selection of the C-terminal portion that is composed of a long-disordered C-terminal linker (aa 317–363), a C-terminal alpha-helix with an extracellular (aa 364–387) as well as membrane-embedded part (aa 387–410), and the cytoplasmic tail.

Circular dichroism spectroscopy
For ISG65, far-UV CD experiments were carried out on a Jasco J-1500 spectropolarimeter with a 0.2 mm path cell. ISG65 was dissolved in 10 mM HEPES pH 7.5, 150 mM NaF at a concentration of 0.4 mg/ml. Spectra were recorded between 195 and 260 nm wavelength at an acquisition speed of 10 nm/min and corrected for buffer absorption. For the calculation of melting curves at 222 nm, spectra were recorded every 5 degrees between 5 °C and 80 °C with a slope of 0.16 °C/min. During measurements, the temperature was kept constant. The raw CD data (ellipticity θ in mdeg) were normalised for the protein concentration and for the number of residues, according to the equation below, yielding the mean residue ellipticity ([θ] in deg cm^2 mol^{-1}), where MM, n, C and l denote the molecular mass (Da), the number of amino acids, the concentration (mg/mL), and the cuvette path length (cm), respectively.

$$[\theta] = \frac{\theta \cdot MM}{n \cdot C \cdot l}$$

Hydrogen-deuterium exchange mass spectrometry
Peptide mapping. 300 pmol of ISG65 or C3d were mixed in a 1:1 (v/v) ratio with 1 M glycine, 200 mM Tris(2-carboxyethyl)phosphine (TCEP) at pH 2.3 and injected onto a Nepenthesin-2 column (Affipro). Generated peptides were trapped and desalted using a micro-trap column (Luna Omega 5 µm Polar C18 100 Å Micro Trap 20 × 0.3 mm) for 3 min at a flow rate of 400 µl/min using an isocratic pump delivering 0.4% (v/v) formic acid in water. Both the protease column and the trap column were placed in an icebox. After 3 min, the peptides were separated on a C18 reversed-phase column (Luna Omega 1.6 µm Polar C18 100 Å, 100 × 1.0 mm) with a linear gradient of 5–35% B over 26 min, where solvent A was 2% (v/v) acetonitrile/0.4% (v/v) formic acid in water and solvent B was 95% (v/v) acetonitrile/4.5% (v/v) water/0.4% (v/v) formic acid. The analytical column was placed in an icebox. A 15 T solariX XR FT-ICR mass spectrometer (Bruker Daltonics) operating in positive MS/MS mode was used for the detection of peptides. Data were processed by Data Analysis 4.2 software (Bruker Daltonics) and ProteinScape 4.0 (Bruker Daltonics). The peptide ion tolerance was 5 ppm and MS/MS tolerance was 0.05 Da. No enzyme specificity was set for peptide search. MASCOT ion score cut-off was set at 2, peptide rank cut-off at 3 and minimal peptide length was 4. The MASCOT search engine was used for the identification of peptides, using a database containing the sequence of ISG65.

Hydrogen-deuterium exchange. Hydrogen-deuterium exchange (HDX) was initiated by a tenfold dilution of 160 µM ISG65: C3d complex into a deuterated buffer (20 mM HEPES, 150 mM NaCl, pD 7.5). The hydrogen-deuterium exchange reaction was performed at RT. 50 µl aliquots (100 pmol) were taken after 20, 120, 1200 and 7200 s of incubation in deuterated buffer and quenched by the addition of 50 µl of 1 M glycine, 200 mM TCEP at pD 2.3, followed by immediate freezing

using liquid nitrogen. Aliquots were quickly thawed and analysed as described for peptide mapping. Peptides were separated by a linear gradient of 10–30% B over 18 min. The mass spectrometer was operated in positive MS mode in the mass range 350–2000 m/z. Data were measured with 1 Mb acquisition with two selective accumulations. The capillary voltage was set at 3900 V, the drying gas temperature was 180 °C and the nebuliser gas was 2.0 bar. Spectra of partially deuterated peptides were processed by Data Analysis 4.2 (Bruker Daltonics, Billerica, MA) and the in-house programme DeutEx[64].

In the DeutEx software calculation of the envelope width intensity was set at 25, the error for deuteration was 10 ppm and the error for isotopes was 10 ppm. An intensity filter was set at a value of 25 and a rate filter was set at 5. No back exchange correction was applied, because quantification of the absolute amount of exchange was not desired.

Small-angle X-ray scattering and ensemble optimisation
Small-angle X-ray scattering (SAXS) data were collected at ESRF, Grenoble (France), using SAXS beamline BM29 with a wavelength of 0.99 Å on a Pilatus 2 M detector (DECTRIS) at 20 °C. For SEC-SAXS, 50 μl of His-ISG65$_{18-363}$ at 11 mg/ml were injected onto a Superdex 200 3.2/300 column (equilibrated in 20 mM HEPES pH 7.5, 150 mM NaCl, 3% (v/v) glycerol) at a flow rate of 75 μl/min. Scattering data were acquired as components eluted from the column and passed through the SAXS measuring cell. The ATSAS[65] software package was used to normalise the data to the intensity of the incident beam, to average the frames, and to subtract the scattering contribution from the buffer. In detail, 10 frames corresponding to the void volume of the column were averaged and subtracted from ten averaged frames of the main elution peak. The radius of gyration (R_g), maximum particle dimension (D_{max}), and distance distribution function ($p(r)$) were evaluated using the programme PRIMUS[66] as part of the ATSAS package.

The ensemble optimisation method (EOM 2.0)[33] as part of ATSAS online was used to determine the flexibility and conformational dynamics of the disordered C-terminal linker region (aa 317–aa 366) connecting the structured, C3-binding domain of ISG65 (aa 18–316) to the C-terminal, partially membrane-embedded, alpha-helix that was predicted using AlphaFold2. The EOM 2.0 web-based application was run with default settings using scattering data obtained for ISG65$_{18-363}$ in combination with the cryo-EM structure of ISG65. To generate the final model of the full-length ISG65$_{18-436}$, the C-terminal helix and the cytoplasmic tail (aa 364–436) from the ISG65 AlphaFold model were appended to SAXS models representative of compact and extended conformations within the EOM obtained ensemble.

***T. b. gambiense* culturing**
Bloodstream form *T. b. gambiense* DAL972 was cultured in HMI-11 complete medium (HMI-11 supplemented with 10% (v/v) fetal bovine serum (non-heat-inactivated), 100 units/ml penicillin and 100 units/ml streptomycin)[67] at 37 °C with 5% CO_2 in a humid atmosphere, in culture flasks with vented caps.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Models of the presented complexes have been deposited in the Protein Data Bank under accession codes 7ZGJ (ISG65:C3) and 7ZGK (ISG65:C3b). The associated cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under accession codes EMD-14707 (ISG65:C3) and EMD-14708 (ISG65:C3b). The hybrid model of ISG65 has been deposited in PDBDev under accession code PDBDEV_00000201. Starting models of C3b and C3 used in the modelling of the complexes are deposited in the Protein Data Bank under accession

codes 2I07 (C3b) and 2A73 (C3). The SAXS data generated in this study is available in SASBDB under accession codes SASDP99 (ISG65) and SASDPA9 (ISG65:C3d). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD036611 (C3 and proteolytic fragments) and PXD033606 (Analysis of ISG65 disulfides and HDX-MS). Flow cytometry standard (FCS) files used in this study are available in the FlowRepository database (https://flowrepository.org/) under accession code FR-FCM-Z5XC. Complete Flow cytometry traces are provided in Supplementary Data 1. Peptide mapping, HDX-MS analysis of the ISG65:C3d complex, and HDX-MS summary table are provided in Supplementary Data 2. All other source data are provided in the Source Data file. Source data are provided with this paper.

References
1. Stijlemans, B. et al. Immune evasion strategies of *Trypanosoma brucei* within the mammalian host: progression to pathogenicity. *Front. Immunol.* **7**, 233 (2016).
2. Cross, G. A. M. Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei. Parasitology* **71**, 393–417 (1975).
3. Cross, G. A. Release and purification of *Trypanosoma brucei* variant surface glycoprotein. *J. Cell Biochem.* **24**, 79–90 (1984).
4. Pinger, J. et al. African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat. *Nat. Microbiol.* **3**, 932–938 (2018).
5. Engstler, M. et al. Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131**, 505–515 (2007).
6. Mugnier, M. R., Cross, G. A. & Papavasiliou, F. N. The in vivo dynamics of antigenic variation in *Trypanosoma brucei. Science* **347**, 1470–1473 (2015).
7. Rifkin, M. R. Identification of the trypanocidal factor in normal human serum: high density lipoprotein. *Proc Natl Acad Sci USA* **75**, 3450–3454 (1978).
8. Hawking, F., Ramsden, D. B. & Whytock, S. The trypanocidal action of human serum and of baboon plasma. *Trans. R. Soc. Trop. Med. Hyg.* **67**, 501–516 (1973).
9. DeJesus, E., Kieft, R., Albright, B., Stephens, N. A. & Hajduk, S. L. A single amino acid substitution in the group 1 *Trypanosoma brucei gambiense* haptoglobin-hemoglobin receptor abolishes TLF-1 binding. *PLoS Pathog.* **9**, e1003317 (2013).
10. Uzureau, P. et al. Mechanism of *Trypanosoma brucei gambiense* resistance to human serum. *Nature* **501**, 430–434 (2013).
11. Symula, R. E. et al. *Trypanosoma brucei gambiense* group 1 is distinguished by a unique amino acid substitution in the HpHb receptor implicated in human serum resistance. *PLoS Negl. Trop. Dis.* **6**, e1728 (2012).
12. Higgins, M. K. et al. Structure of the trypanosome haptoglobin-hemoglobin receptor and implications for nutrient uptake and innate immunity. *Proc. Natl Acad. Sci. USA* **110**, 1905–1910 (2013).
13. Vanhamme, L. et al. Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* **422**, 83–87 (2003).
14. Zoll, S. et al. The structure of serum resistance-associated protein and its implications for human African trypanosomiasis. *Nat. Microbiol.* **3**, 295–301 (2018).
15. Merle, N. S., Church, S. E., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement system part I - molecular mechanisms of activation and regulation. *Front. Immunol.* **6**, 262 (2015).
16. Fromell, K. et al. Assessment of the role of C3(H2O) in the alternative pathway. *Front. Immunol.* **11**, 530 (2020).
17. Ricklin, D., Reis, E. S., Mastellos, D. C., Gros, P. & Lambris, J. D. Complement component C3 - the "Swiss Army Knife" of innate immunity and host defense. *Immunol. Rev.* **274**, 33–58 (2016).
18. Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. Formation of the initial C3 convertase of the alternative complement

pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3. *J. Exp. Med.* **154**, 856–867 (1981).

19. Lachmann, P. J., Lay, E. & Seilly, D. J. Experimental confirmation of the C3 tickover hypothesis by studies with an Ab (S77) that inhibits tickover in whole serum. *FASEB J.* **32**, 123–129 (2018).

20. Mannes, M. et al. Complement inhibition at the level of C3 or C5: mechanistic reasons for ongoing terminal pathway activity. *Blood* **137**, 443–455 (2021).

21. Devine, D. V., Falk, R. J. & Balber, A. E. Restriction of the alternative pathway of human complement by intact Trypanosoma brucei subsp. gambiense. *Infect. Immun.* **52**, 223–229 (1986).

22. Ferrante, A. & Allison, A. C. Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. *Parasite Immunol.* **5**, 491–498 (1983).

23. Menny, A. et al. CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. *Nat. Commun.* **9**, 5316 (2018).

24. Sanchez-Corral, P. et al. Separation of active and inactive forms of the third component of human complement, C3, by fast protein liquid chromatography (FPLC). *J. Immunol. Methods* **122**, 105–113 (1989).

25. Liu, X. et al. Changes in complement levels and activity of red blood cells, fresh frozen plasma, and platelet concentrates during storage. *Ind. J. Hematol. Blood Transfus.* **37**, 140–146 (2021).

26. Elvington, M. et al. Development and optimization of an ELISA to quantitate C3(H 2 O) as a marker of human disease. *Front. Immunol.* **10**, 703 (2019).

27. Pangburn, M. K. & Muller-Eberhard, H. J. Initiation of the alternative complement pathway due to spontaneous hydrolysis of the thioester of C3. *Ann. N Y Acad. Sci.* **421**, 291–298 (1983).

28. Chen, Z. A. et al. Structure of complement C3(H2O) revealed by quantitative cross-linking/mass spectrometry and modeling. *Mol. Cell Proteomics* **15**, 2730–2743 (2016).

29. Rodriguez, E., Nan, R., Li, K., Gor, J. & Perkins, S. J. A revised mechanism for the activation of complement C3 to C3b: a molecular explanation of a disease-associated polymorphism. *J. Biol. Chem.* **290**, 2334–2350 (2015).

30. Janssen, B. J., Christodoulidou, A., McCarthy, A., Lambris, J. D. & Gros, P. Structure of C3b reveals conformational changes that underlie complement activity. *Nature* **444**, 213–216 (2006).

31. Janssen, B. J. et al. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505–511 (2005).

32. Field, M. C. & Carrington, M. The trypanosome flagellar pocket. *Nat. Rev. Microbiol.* **7**, 775–786 (2009).

33. Tria, G., Mertens, H. D., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).

34. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

35. Bartossek, T. et al. Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. *Nat. Microbiol.* **2**, 1523–1532 (2017).

36. Berends, E. T. et al. Molecular insights into the surface-specific arrangement of complement C5 convertase enzymes. *BMC Biol.* **13**, 93 (2015).

37. Sullivan, L., Wall, S. J., Carrington, M. & Ferguson, M. A. Proteomic selection of immunodiagnostic antigens for human African trypanosomiasis and generation of a prototype lateral flow immunodiagnostic device. *PLoS Negl. Trop. Dis.* **7**, e2087 (2013).

38. Ziegelbauer, K. & Overath, P. Organization of two invariant surface glycoproteins in the surface coat of Trypanosoma brucei. *Infect. Immun.* **61**, 4540–4545 (1993).

39. Trevor, C. E. et al. Structure of the trypanosome transferrin receptor reveals mechanisms of ligand recognition and immune evasion. *Nat. Microbiol.* **4**, 2074–2081 (2019).

40. Macleod, O. J. S. et al. A receptor for the complement regulator factor H increases transmission of trypanosomes to tsetse flies. *Nat. Commun.* **11**, 1326 (2020).

41. Stodkilde, K., Torvund-Jensen, M., Moestrup, S. K. & Andersen, C. B. Structural basis for trypanosomal haem acquisition and susceptibility to the host innate immune system. *Nat. Commun.* **5**, 5487 (2014).

42. Freymann, D. et al. 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. *J. Mol. Biol.* **216**, 141–160 (1990).

43. Clarke, M. W., Barbet, A. F. & Pearson, T. W. Structural features of antigenic determinants on variant surface glycoproteins from *Trypanosoma brucei*. *Mol. Immunol.* **24**, 707–713 (1987).

44. Miller, E. N., Allan, L. M. & Turner, M. J. Topological analysis of antigenic determinants on a variant surface glycoprotein of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **13**, 67–81 (1984).

45. Macleod, O. J. S. et al. Invariant surface glycoprotein 65 of *Trypanosoma brucei* is a complement C3 receptor. *Nat. Commun.* **13**, 5085 (2022).

46. Koumandou, V. L., Boehm, C., Horder, K. A. & Field, M. C. Evidence for recycling of invariant surface transmembrane domain proteins in African trypanosomes. *Eukaryot Cell* **12**, 330–342 (2013).

47. Zwarthoff, S. A. et al. Functional characterization of alternative and classical pathway C3/C5 convertase activity and inhibition using purified models. *Front. Immunol.* **9**, 1691 (2018).

48. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1243–1250 (2006).

49. O'Rear, L. D. & Ross, G. D. Isolation and purification of C3 from human plasma. *Curr. Protoc. Immunol.* **Chapter 13**, Unit 13 13 (2001).

50. Guthridge, J. M. et al. Structural studies in solution of the recombinant N-terminal pair of short consensus/complement repeat domains of complement receptor type 2 (CR2/CD21) and interactions with its ligand C3dg. *Biochemistry* **40**, 5931–5941 (2001).

51. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).

52. Kotecha, N., Krutzik, P. O. & Irish, J. M. Web-based analysis and publication of flow cytometry experiments. *Curr. Protoc. Cytom.* **Chapter 10**, Unit10 17 (2010).

53. Strohalm, M., Kavan, D., Novak, P., Volny, M. & Havlicek, V. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Anal. Chem.* **82**, 4648–4651 (2010).

54. Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.* **1266**, 171–184 (2015).

55. Yang, M. et al. Recombinant nepenthesin II for hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **87**, 6681–6687 (2015).

56. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

57. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

58. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* https://doi.org/10.7554/eLife.42166 (2018).

59. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).

60. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).

61. Murshudov, G. N. et al. REFMAC5 for the refinement of macro-molecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).

62. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).

63. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

64. Daniel Kavan, P. M. MSTools—Web based application for visualization and presentation of HXMS data. *Int. J. Mass Spectrom.* **302**, 53–58 (2011).

65. Manalastas-Cantos, K. et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355 (2021).

66. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

67. Hirumi, H. & Hirumi, K. Axenic culture of African trypanosome bloodstream forms. *Parasitol Today* **10**, 80–84 (1994).

68. Kinoshita, T. et al. C5 convertase of the alternative complement pathway: covalent linkage between two C3b molecules within the trimolecular complex enzyme. *J. Immunol.* **141**, 3895–3901 (1988).

Author contributions
H.S. and S.Z. conceptualised the study. Initial pull-down experiments were performed by SZ. J.V., M.S. and J.B. cloned the recombinant expression constructs. H.S. and J.B. performed protein expression, purifications and SPR measurements, assisted by A.S. and M.T. Haemolysis assays were performed by H.S. S.K. prepared and subsequently screened for suitable TEM grids for ISG65:C3 data collections. Fluorescence microscopy and FACS measurements were performed by M.Z. and F.Z. P.P. performed HDX-MS, XL-MS and SS-bond mapping experiments. M.H. performed intact MS and LC-MS/MS experiments. Cryo-EM data processing, subsequent model building, and molecular modelling was done by H.S. SAXS measurements were evaluated by H.S. and S.Z. H.S. and S.Z. wrote the manuscript with input from M.Z. A.D., H.S. and S.Z. revised and edited the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-37988-7.

Correspondence and requests for materials should be addressed to Sebastian Zoll.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

6.3 Beyond the VSG Layer: Exploring the Role of Intrinsic Disorder in the Invariant Surface Glycoproteins of African Trypanosomes.

Beyond the VSG Layer: Exploring the Role of Intrinsic Disorder in the Invariant Surface Glycoproteins of African Trypanosomes

Hagen Sülzen[1,2], Alexander N. Volkov[3,4], Rob Geens[3,5], Farnaz Zahedifard[6], Benoit Stijlemans[7,8], Martin Zoltner[6], Stefan Magez[7,9,10], Yann G.-J. Sterckx[5,‡*], Sebastian Zoll[1,‡*]

[1]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo namesti 542/2, 16000, Prague, Czech Republic
[2]Faculty of Science, Charles University, Albertov 6, 12800, Prague 2, Czech Republic
[3]VIB-VUB Center for Structural Biology, Flemish Institute of Biotechnology (VIB), Pleinlaan 2, 1050 Brussels, Belgium
[4]Jean Jeener NMR Centre, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium
[5]Laboratory of Medical Biochemistry (LMB) and the Infla-Med Center of Excellence, Department of Pharmaceutical Sciences, Universiteit of Antwerp, B-2610 Wilrijk, Belgium
[6]Department of Parasitology, Faculty of Science, Charles University in Prague, Biocev, 25250 Vestec, Czech Republic
[7]Brussels Center for Immunology (BCIM), Department of Bioengineering Sciences, Vrije Universiteit Brussel, B-1050 Brussels, Belgium
[8]Myeloid Cell Immunology Laboratory, VIB Center for Inflammation Research, B-1050 Brussels, Belgium
[9]Department of Biochemistry and Microbiology, Ghent University, B-9000 Gent, Belgium
[10]Laboratory for Biomedical Research, Department of Molecular Biotechnology, Environment Technology and Food Technology, Ghent University Global Campus, Incheon 406-840, South Korea

[‡]These authors contributed equally to this paper and should be considered joint senior authors

[*]Corresponding authors: yann.sterckx@uantwerpen.be, sebastian.zoll@uochb.cas.cz

ABSTRACT

In the bloodstream of mammalian hosts, African trypanosomes face the challenge of protecting their invariant surface receptors from immune detection. This crucial role is fulfilled by a dense, glycosylated protein layer composed of variant surface glycoproteins (VSGs), which undergo antigenic variation and provide a physical barrier that shields the underlying invariant surface glycoproteins (ISGs). The protective shield's limited permeability comes at the cost of restricted access to the extracellular host environment, raising questions regarding the specific function of the ISG repertoire. In this study, we employ an integrative structural biology approach to show that intrinsically disordered membrane-proximal regions are a common feature of members of the ISG superfamily, conferring the ability to switch between compact and elongated conformers. While the folded, membrane-distal ectodomain is buried within the VSG layer for compact conformers, their elongated counterparts would enable the extension beyond it. This dynamic behavior enables ISGs to maintain a low immunogenic footprint while still allowing them to engage with the host environment when necessary. Our findings add further evidence to a dynamic molecular organization of trypanosome surface antigens wherein intrinsic disorder underpins the characteristics of a highly flexible ISG proteome to circumvent the constraints imposed by the VSG coat.

INTRODUCTION

As strictly extracellular parasites, African trypanosomes rely on sophisticated defense, communication, and sensing mechanisms to evade host immune responses and initiate developmental transformations. Well-documented examples of the parasite's bloodstream form (BSF, predominant form in the mammalian host) include: i) inhibition of innate immune responses through trypanosome receptor-like adenylate cyclases [1–3]; ii) complement and trypanolytic factor binding proteins [4–7]; iii) antigenic variation of variant surface glycoproteins (VSGs) to counter adaptive immune responses [8,9]; iv) inter-parasite communication through quorum sensing [10,11]; and v) the release of extracellular vesicles to manipulate the host environment, mediate parasite-parasite communication and facilitate specific forms of motility [12,13]. Among these, the most studied (and most notorious) mechanism remains the structure-function relationship of the parasite's VSG coat. The classical model depicts the VSG coat as a dense, protective barrier that shields underlying invariant proteins from recognition by the host's immune system. While small molecules may penetrate this shield, large immune reactive molecules at the size of an antibody have been shown to be excluded [14–17]. This viewpoint was further maintained by an extended model of the VSG coat organization, according to which VSG dimers can adopt two different conformations, generating a flexible topology while still maintaining the shielding function [18]. This, for the first time, linked the plasticity of a trypanosome surface protein to a specific function. However, as trypanosomes reside freely in the bloodstream, they must engage with numerous host ligands in the extracellular environment through dedicated surface receptors. The current VSG umbrella model and the assumed exclusion limit of the VSG coat raise questions as to how such surface receptors would be able to perform their biological function.

Invariant surface glycoproteins (ISGs) are a unique super-family of surface antigens that are exclusively expressed by the BSFs of African trypanosomes. Two distinct features set ISGs apart from other characterized BSF surface antigens. First, while most invariant receptors are confined to the flagellar pocket to minimize immune exposure [19,20], ISGs are embedded within the VSG coat and thus distributed over the whole cell surface [20–22]. Second, instead of being attached to the plasma membrane via GPI-anchors, ISGs are type-I transmembrane proteins with a short cytosolic domain, which is subject to ubiquitylation, therefore mediating rapid recycling [21,23]. Both features suggest a potentially important role in the immunobiology of BSFs, such as capturing/scavenging nutrients or protecting the parasite from host immune factors. While several multi-gene ISG families have been identified, only ISG65 and ISG75 have been studied extensively due to their high abundance, indicating functional significance. Notably, ISG65 has recently been identified as a complement receptor [6,7,24], featuring an adaptation of the canonical three-helical bundle fold with unique characteristics, whereas experimental evidence for structure and biological function of ISG75 remain elusive. Owing to their relative abundance and invariant nature, ISG65 and ISG75 have been considered as potential vaccine candidates [22], as they are immunogenic and elicit an immune response detectable through the presence of antibodies in patient blood. Indeed, studies have shown that ISG75, ISG65, and ISG64 rank among the most abundant surface antigens and can be selectively enriched through immuno-affinity chromatography using *T. b. gambiense* infection IgG [25]. However, immunization with ISGs conferred no or only partial protection against repeated infections [26]. The accessibility of ISGs for the interaction with extracellular host ligands (including antibodies) becomes enigmatic should they strictly conform to the prevailing VSG shielding model [18,27]. Consequently, ISG exposure is likely to occur either through shedding or increased permissiveness of the VSG coat. Recent studies involving ISG65 from BSFs and metacyclic invariant surface proteins (MISP) from metacyclic forms have favored the latter hypothesis [7,28]. For both proteins, intrinsically disordered C-terminal regions (IDRs) were found, which could enable them to operate within as well as beyond the VSG coat. Protrusion beyond the coat's boundaries would allow for the acquisition of large-sized host ligands at the expense of increased exposure to the host's immune system. Despite this

functional importance, the precise molecular mechanisms by which IDRs facilitate these phenomena remain unexplored.

In this study, we investigate the conformational flexibility of prototypical ISGs of different sizes from the human-infective parasite *T. b. gambiense* (*Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75) and describe our findings in the context of the VSG umbrella model. Using an integrative structural modelling approach, combining AlphaFold2-based structure prediction, hydrogen-deuterium exchange mass spectrometry (HDX-MS), and small-angle X-ray scattering (SAXS)-driven conformational sampling, we demonstrate that, similarly to the previously described *Tbg*ISG65 [7], *Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75 all possess disordered C-terminal linkers that are extremely variable in length. We show that these linkers have the capacity to adopt distinct conformational states, enabling the proteins to either reside within the VSG coat or protrude from it, thereby enabling them to interact with large extracellular ligands. Our results suggest that large-scale conformational changes mediated by intrinsically disordered membrane-proximal linkers are a functional feature of trypanosome receptors of the ISG super-family that may have evolved as an adaptation to their presence within the VSG coat.

MATERIAL AND METHODS

Protein production and purification

T. b. gambiense ISGs. ISG75 was cloned, recombinantly produced and purified as previously described in [29]. DNA fragments (Genewiz) encoding amino acids (aa) 22–349 of *Tbg*972.5.200 (from here on referred to as *Tbg*ISG43) and aa 24-365 from *Tbg*ISG64 (*Tbg*972.5.6550) were codon-optimized for bacterial expression and cloned into the pET15b plasmid using the Gibson assembly method (New England Biolabs).

*Tbg*ISG43 and *Tbg*ISG64 were recombinantly produced overnight at 22°C in *E. coli* T7 shuffle cells (New England Biolabs) after induction of protein expression by addition of isopropyl-β-D-thiogalacto-pyranoside (IPTG) to a final concentration of 1 mM. Cells were harvested by centrifugation for 15 min at 6.000xg, 4°C. Cell pellets were resuspended in Buffer A (20 mM Tris, 500 mM NaCl, 10 mM imidazole, pH 8.0), to which phenylmethylsulfonyl fluoride (PMSF) was added at a final concentration of 1 mM immediately before cell lysis. Cells were lysed using an EmulsiFlex-C3 (AVESTIN Europe) with 1000 to 1100 bar lysis pressure at 4°C. Cell debris were removed by centrifugation for 45 min at 20.000xg, 4°C. Soluble protein was purified using immobilized-metal affinity chromatography (IMAC) by application of the cleared supernatant to nickel-nitrilotriacetic acid (NTA) beads (Qiagen) in a gravity flow column (Bio-Rad), pre-equilibrated with Buffer A. The resin was washed with 10 column volumes (CV) Buffer A twice before eluting the bound protein fraction with 10 CV Buffer B (20 mM Tris, 500 mM NaCl, 400 mM imidazole, pH 8.0). The eluate was fractionated, and those fractions containing the protein of interest were identified via SDS-PAGE, pooled, transferred into SnakeSkin Dialysis tubing (Thermo Fisher Scientific) and dialyzed overnight into Buffer C (20 mM Tris, 500 mM NaCl, pH 8.0) at 4°C. The dialyzed eluate was concentrated using Amicon Ultra centrifugal filters (Merck Millipore) with a 10 kDa MWCO before being subjected to size exclusion chromatography (SEC) on a Superdex 200 16/60 column (GE Healthcare) connected to an Äkta FPLC system (GE Healthcare), pre-equilibrated with Buffer D (20mM HEPES, 150 mM NaCl, pH 7.5). 1.5 mL elution fractions were collected throughout the run, fractions containing the protein of interest, as determined by SDS-PAGE, were pooled, flash-frozen in liquid-nitrogen and stored at -80°C until further use.

T. b. brucei and T. b. gambiense VSGs. T. b. brucei LiTat1.5 and *T. b. gambiense* LiTat3.1 (*Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1, respectively) were obtained from native source as described in [30] and [18], respectively.

Hydrogen-deuterium exchange mass spectrometry

Hydrogen deuterium exchange was initiated by 10-fold dilution of *Tbg*ISG43, *Tbg*ISG65 or *Tbg*ISG75 into deuterated buffer (20 mM HEPES, 150 mM NaCl, pD 7.5). 50 µL aliquots (100 pmols) were taken after 20 s, 120 s and 1200 s of incubation in deuterated buffer and quenched by the addition of 50 µL of 4 M urea, 1 M glycine and 200 mM TCEP at pD 2.3, followed by immediate flash freezing in liquid nitrogen. Aliquots were quickly thawed and injected onto a Nepenthesin-2/pepsin column (AffiPro, Czech Republic). Generated peptides were trapped and desalted via a Micro-trap column (Luna Omega 5 µm Polar C18 100 Å Micro Trap, 20 x 0.3 mm) for 3 min at a flow rate of 200 µL/min using an isocratic pump delivering 0.4% (v/v) formic acid in water. Both the protease and the trap column were placed in an icebox. After 3 min, peptides were separated on a C18 reversed phase column (Luna® Omega 1.6 µm Polar C18 100 Å, 100 x 1.0 mm) and analyzed using a timsToF Pro mass spectrometer (Bruker Daltonics, Billerica, MA). Peptides were separated by a linear gradient of 10 - 30% B over 18 min, where solvent A was 2% (v/v) acetonitrile/0.4% (v/v) formic acid in water and solvent B was 95% (v/v) acetonitrile/4.5% (v/v) water/0.4% (v/v) formic acid. The mass spectrometer was operated in positive MS mode. Spectra of partially deuterated peptides were processed by Data Analysis 4.2 (Bruker Daltonics, Billerica, MA) and by in-house program DeutEx [31].

Small angle X-ray scattering and ensemble modelling
All experiments were performed at the BioSAXS beamlines SWING (SOLEIL, Gif-sur-Yvette, France, [32]) and BM29 (ESRF, Grenoble, France, [33]). SEC-SAXS data were collected in HPLC mode using a Shodex KW404-4F column pre-equilibrated with SAXS buffer. Samples were concentrated on site to ~10 mg/mL using a 10-kDa cutoff centrifugal filter (Amicon). Eighty-microliter samples were injected and eluted at a flow rate of 0.2 mL/min, while scattering data were collected with an exposure time of 750 msec and a dead time of 10 msec. The scattering of pure water was used to calibrate the intensity to absolute units [34]. Data were processed using CHROMIXS [35] and analyzed using BioXTas RAW [36] and the ATSAS package [37]. The information on data collection and derived structural parameters is summarized in Supplemental Table S1.
Molecular models were generated with AlphaFold2 [38] (*Tbg*ISG43, *Tbg*ISG65, *Tbg*ISG75) and AlphaFold-Multimer [39] (*Tbb*VSG LiTat1.5, *Tbg*VSG LiTat3.1). Theoretical scattering curves of the AlphaFold2 models and their respective fits to the experimental data were calculated using FoXS [40]. SAXS-based ensemble modelling was carried out using BILBOMD [40–42]. For all runs, 800 conformations were generated per R_g, with minimal and maximal R_g values set at 7% and 35% of the experimentally determined Rg, respectively. Molecular graphics visualization and analysis were performed with UCSF ChimeraX [43] and PyMOL Molecular Graphics System (Schrödinger).

Circular dichroism spectroscopy
Far-UV CD experiments were carried out on a Jasco J-1500 spectropolarimeter with a 0.2 mm path cell. Proteins were dissolved in 20 mM HEPES pH 7.5, 150 mM NaF at a concentration of 0.4 mg/mL. Spectra were recorded between 195 and 260 nm wavelength at an acquisition speed of 10 nm/min and corrected for buffer absorption. During measurements, the temperature was kept constant at 25°C. The raw CD data (ellipticity θ in mdeg) were normalized for the protein concentration and for the number of residues, according to the equation below, yielding the mean residue ellipticity ([θ] in deg cm^2 mol^{-1}), where *MM*, *n*, *C*, and *l* denote the molecular mass (Da), the number of amino acids, the concentration (mg/mL), and the cuvette path length (cm), respectively.

$$[\theta] = \frac{\theta \cdot MM}{n \cdot C \cdot l}$$

Secondary structure analysis was performed with the CD Pro software package.

Mass spectrometry
Single gel bands were cut, chopped into small pieces, reduced with dithiothreitol, alkylated with chloroacetamide and digested with trypsin overnight. Peptides were extracted from each gel piece, lyophilized in a speed-vac, dissolved in 0.1% formic acid and 20% of the dissolved sample was separated on an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). The peptides were trapped and desalted with 2% acetonitrile in 0.1% formic acid at a flow rate of 5 µL/min on an Acclaim PepMap100 column (5 µm, 5 mm by 300 µm internal diameter (ID); Thermo Fisher Scientific). Eluted peptides were separated using an Acclaim PepMap100 analytical column (2 µm, 50 cm × 75 µm ID; Thermo Fisher Scientific). Using a constant flow rate of 300 nL/min, a 65 min elution gradient was started at 5% B (0.1% formic acid in 99.9% acetonitrile) and 95% A (0.1% formic acid). The gradient reached 30% B at 52 min, 90% B at 53 min, and was then kept constant until 57 min before being reduced to 5% B at 58 min. For the first minute, nanospray was set to 1600 V, 350 °C source temperature, measuring the scans in the range of m/z 350–2000. An orbitrap detector was used for MS with the resolution 120,000, the AGC target value was set as custom with a normalized AGC target of 250%. Maximum injection time was set to 50 ms, MSMS was acquired also using orbitrap with

resolution 30,000, the data were acquired in a data-dependent manner, ions were fragmented by HCD collision energy set to 30% with dynamic exclusion set to 60 s.

The resulting raw data files were searched to identify the peptides by PEAKS (PEAKS Studio 10.0) software. PEAKS DB database search was set to search for Carbamidomethylation on Cysteine as fixed modification, Deamidation on N or Q, Oxidation on M and glycosylation specified as follows (Hex(2) HexNAc, HexNAc(3), Hex2 dHex2) was set as the variable modifications. The PEAKS DB was followed by PEAKS PTM and SPIDER.

The data files were processed also by MASCOT (Version: 2.6.2). To address the range of modifications, the error tolerant search was applied.

In all searching steps, the *T. b. gambiense* protein database (downloaded from UniProt on 31st October 2022) was used.

RESULTS

Single-conformer AlphaFold2 models do not accurately represent the in-solution behavior of variant and invariant surface glycoproteins

To explore the conformational dynamics of ISGs and VSGs in solution, this study employed a combination of AlphaFold2-based structure prediction for generating all-atom structural models, and the collection of small angle X-ray scattering (SAXS) data for model validation.
The structures of both VSGs and ISGs could be predicted with relatively high confidence as judged from local pLDDT values and overall zDOPE and pTM scores (Supplementary Figure 1). For both VSGs (*Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1), especially the N-terminal domains (NTDs) and the dimer interfaces are modelled with high accuracy, as evidenced by various validation metrics (PAE, pDockQ, and ipTM). Structural alignment reveals that *Tbb*VSG LiTat1.5 is a Class A1 VSG (top lobe structures reminiscent of VSG13 [44]), while *Tbg*VSG LiTat3.1 can be classified as a Class A2 VSG [45] (N-terminal lobe structure as found in VSG1, Figure 1A). In addition, the occurrence of one C-terminal domain (CTD) is predicted per monomer, albeit with lower confidence scores. Interestingly, the *Tbg*VSG LiTat3.1 AlphaFold2 model displays a relatively long C-terminal helix. While the pLDDT value analysis of this region would suggest a reliable prediction, such structures have not yet been described for VSG C-terminal regions. The structure prediction of the ISGs under study (*Tbg*ISG43, *Tbg*ISG64, *Tbg*ISG75) follows a similar trend. The N-terminal domains are modelled with high accuracy and are in accordance with the recently published, experimental structure of *Tbg*ISG65 (Figure 1A). In contrast, the ISG43 and ISG64 C-terminal domains display significantly lower pLDDT scores (<50; Supplementary Figure 1). For *Tbg*ISG75, AlphaFold2 predicts the C-terminus to consist of two long, intertwined, antiparallel α-helices with reasonably good pLDDT scores (Supplementary Figure 1).
A first insight into the solution behavior of the studied proteins is offered by circular dichroism (CD) spectroscopy (Figure 1B). As expected, the CD spectra show a high α-helical content, but also reveal a significant amount of intrinsic disorder, which is not apparent in the single-conformer AlphaFold2 models. In both VSGs and *Tbg*ISG75, ~50% of the residues can be found in low complexity regions (turns and intrinsic disorder). This observation is further supported by the normalized Kratky plots (indicators for intrinsic disorder), which show that the solution structures of both VSGs and ISGs are characterized by a significant degree of flexibility (Figure 1C). The latter suggests that the VSG and ISG solution structures would be best described by a conformational ensemble. Indeed, as shown in Figure 2, single-conformer AlphaFold2 models are insufficient to explain the solution behavior of the VSG and ISG molecules under study. Theoretical scattering curves derived from the static models differ significantly from the experimental SAXS curves as indicated by the poor fits and high χ^2 values. The largest disagreements between predicted and solution structures were observed for *Tbb*VSG LiTat 1.5 ($\chi^2 = 522$) and *Tbg*ISG75 ($\chi^2 = 150$).
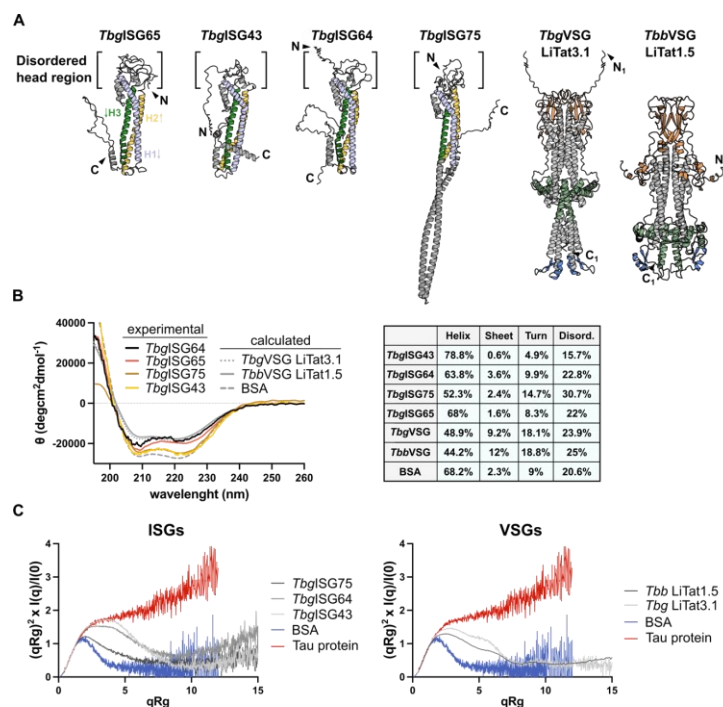
Figure 1. The VSG and ISG solution structures exhibit a high degree of flexibility. (A) The predicted AlphaFold2 models of *Tbg*ISG43, *Tbg*ISG64, *Tbg*ISG75, *Tbg*VSG LiTat3.1 and *Tbb*VSG LiTat1.5 illustrated in cartoon representation, depicted next to the hybrid structure of *Tbg*ISG65$_{18\text{-}363}$ (PDBDEV_00000201). The three helices in the *Tbg*ISGs constituting the canonical three helix bundle have been colored uniformly and labelled on *Tbg*ISG65, arrows indicate the direction of the peptide chain from N- to C-terminus. Bottom and top lobe as well as C-terminal domains of the VSGs are highlighted in orange, green and blue, respectively. For representation purposes, the N-termini of *Tbg*ISG64 and *Tbg*VSG have been shortened, deleted residues are indicated with a line break (full length models are shown in Figure 2). **(B)** CD spectra for *Tbg*ISGs (experimental), VSGs (calculated from AlphaFold2 models) and BSA (calculated from PDB 4F5S). Secondary structure analyses (performed using CONTINLL [46,47]) for the respective spectra are summarized in the table. All CD spectra exhibit minima at 208 and 222 nm, characteristic for a largely alpha-helical fold. Similarly, all proteins have a high content of turns and disordered regions. **(C)** Dimensionless Kratky plots (R$_g$) for *Tbg*ISGs (left) and *Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1 (right). Comparison of the plots for *Tbg* and *Tbb* proteins to reference samples for strictly folded (BSA, blue) and completely disordered proteins (Tau protein, red) demonstrate that all measured proteins contain significant fractions of both ordered and disordered components.
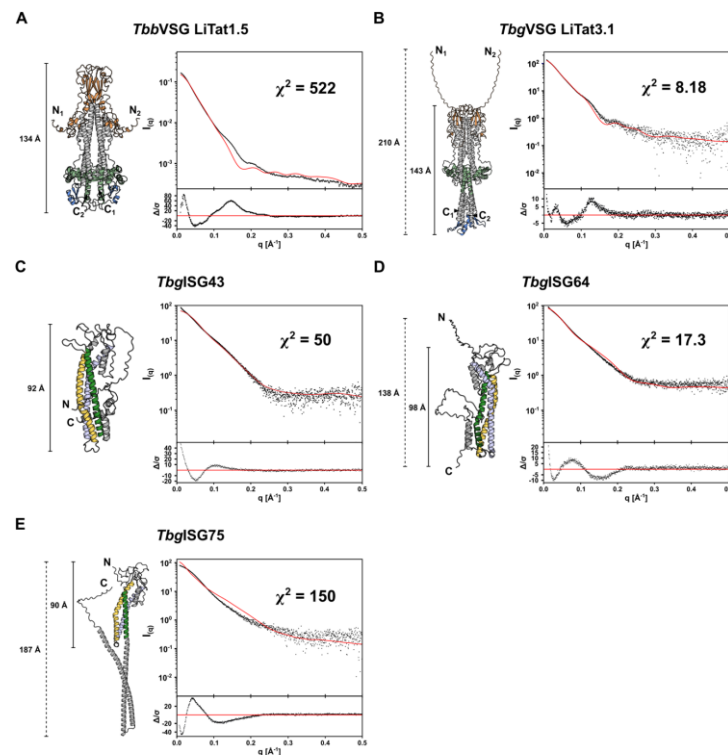
Figure 2. The VSG and ISG single-conformer AlphaFold2 models do not account for the experimentally observed solution behavior. The theoretical scattering curves of the unaltered AlphaFold2 models for *Tbb*VSG LiTat1.5 (A), *Tbg*VSG LiTat3.1 (B), *Tbg*ISG43 (C), *Tbg*ISG64 (D) and *Tbg*ISG75 (E) were calculated (red curves) and compared to the experimental SAXS data (black dots). The fit residuals are displayed as insets below the scattering curves. The respective AlphaFold2 models used for the calculations are shown (colored using the same scheme as in Figure 1).

VSGs adopt distinct structural states in solution that are best described by conformational ensembles

Structural flexibility of VSGs (*Tbb*VSG MiTat1.1 and *Tbb*VSG IlTat1.24) has previously been investigated [18]. Bartossek and co-workers employed a SAXS-based, rigid body modelling approach to demonstrate that the VSG solution behavior is best described by a conformational ensemble, in which the NTDs behave as rigid bodies, while the CTDs are flexible. In line with these findings, we found that the static AlphaFold2 models of *Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1 inadequately describe the solution behavior (Figure 2, panels A and B).

By employing an alternative, molecular dynamics-based method for generating and selecting conformers during SAXS-driven ensemble modeling, we could confirm the initial findings of

Bartossek et al. As shown in Figure 3A, the CTD of VSGs is highly flexible, thereby allowing the molecule to adopt more than one conformational state. For *Tbb*VSG LiTat1.5, the experimental scattering data was best explained by an ensemble ($\chi^2 = 2.00$), 40% of which consists of a compact conformer ($R_g = 42.16$ Å, $D_{max} = 155$ Å), while the remaining 60% comprises conformers with varying degrees of extension ($R_g = 47.6$ Å, $D_{max} = 196$ Å for the most extended conformer). The application of the same protocol to *Tbg*VSG LiTat3.1 reveals that the best fit to the experimental data is provided by a conformational ensemble in which the CTD is also granted full conformational flexibility ($\chi^2 = 1.75$, Figure 3B), contrasting the AlphaFold2 model discussed above (Figure 1A). The majority of the ensemble (78.5%) comprises a compact conformation ($R_g = 43.57$ Å, $D_{max} = 167$ Å), while the remaining 21.5% corresponds to an extended conformation ($R_g = 48.67$ Å, $D_{max} = 211$ Å).

In contrast to the relatively small differences in D_{max} (up to 16 Å) between the 2 conformational states of *Tbb*VSGs reported by Bartossek et al., the differences in D_{max} found for the compact and the most extended conformers of *Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1 are much larger with 41 Å (Figure 3A) and 44 Å (Figure 3B), respectively. Furthermore, the most extended conformers of the VSGs investigated here exceed the D_{max} reported for VSGs by Bartossek et al. by approximately 50 Å.
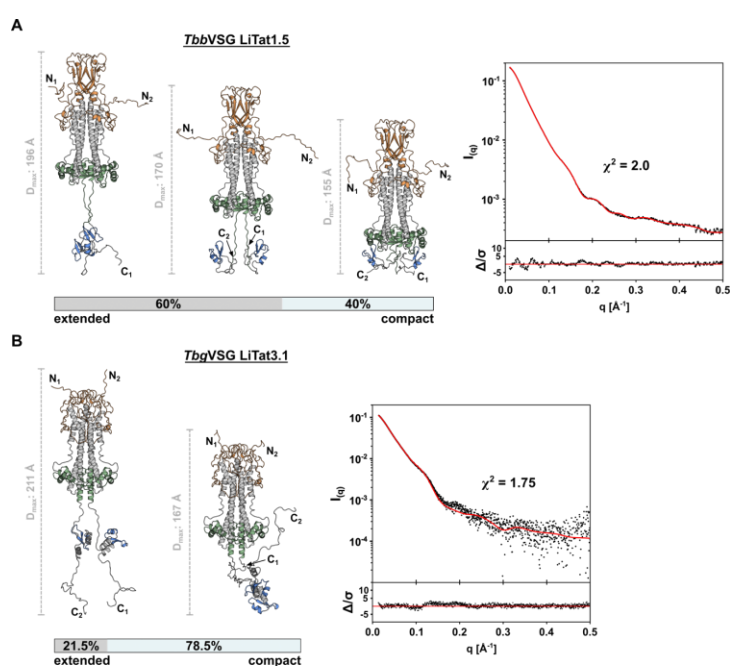
Figure 3. Conformational ensemble modelling reveals structural flexibility of VSGs in solution. (A) The solution scattering of *Tbb*VSG LiTat1.5 is best described by a conformational ensemble composed of 5 models. Three representative conformations are shown, alongside a scale bar indicating the D_{max} of the individual

conformer. **(B)** The SAXS data of *Tbg*VSG LiTat3.1 is best described by a conformational ensemble composed of 2 models. Both conformations are depicted next to a scale bar indicating the D_{max} of the individual conformer. In both panels, the experimental scattering data (black dots) and calculated ensemble scattering curves (red line) are shown. The residuals of the fit are shown below.

The membrane-proximal, C-terminal regions of ISGs are intrinsically disordered

Considering that low pLDDT scores (<50) can serve as an indicator for disorder [48], the AlphaFold2 models presented in this study suggest the presence of intrinsic disorder in the C-terminal regions of *Tbg*ISG43 and *Tbg*ISG64. To investigate this further, we employed hydrogen-deuterium exchange mass spectrometry (HDX-MS) (Figure 4). Differences in relative deuteration rates can provide insight into protein secondary structure, as protein regions with low complexity typically exhibit faster deuteration rates than their well-folded counterparts [49]. This is clearly illustrated by *Tbg*ISG65, which serves as a benchmark here, demonstrating that deuteration rates exhibit a strong correlation with the experimental structure (Figure 4, bottom panel). Similarly, for *Tbg*ISG43, *Tbg*ISG64 and *Tbg*ISG75, residues predicted by AlphaFold2 with high confidence to constitute the canonical three-helical bundle show significantly lower deuteration rates compared to the rest of the protein. Underlining the low pLDDT scores, the HDX-MS analysis also confirms a disordered N-terminus for all proteins analyzed, as indicated by the fast deuteration rates. Notably, *Tbg*ISG43 N-terminal residues Gly2-Leu14 appear to exhibit a lower deuteration rate than residues Val25-Glu39, indicating a stable secondary structure. This however is most likely an artifact caused by technical limitations, such as insufficient proteolytic digest or low ionization efficiency. Gaps in the sequence coverage, caused by the lack of a sufficient quantity of smaller peptides from that part of the protein, may result in a loss of resolution, producing nonrepresentative deuteration rates (Supplementary Figure 2). In all proteins analyzed, the residues predicted to form membrane-distal, disordered head domains undergo a large degree of deuteration already after 20 s. The C-terminal residues of *Tbg*ISG43, *Tbg*ISG64 and *Tbg*ISG65 (predicted to be largely disordered) deuterate to a similar degree as the disordered loops in the head domains, supporting the AlphaFold2 prediction. C-terminal residues *Tbg*ISG43 Val328-Arg348 are predicted to form an α-helix, which appears to be supported by the slow initial deuteration rate over the first 20 s. The same observation can be made for *Tbg*ISG64 residues Ala354-Leu360.

HDX-MS analysis of *Tbg*ISG75 however contradicts the AlphaFold2 model (Figure 1C) as this experimental evidence does not support the existence of the two predicted, continuous C-terminal helices in solution (residues Lys262-Ala330 and Glu336-Gly417). Unlike for the other *Tbg*ISGs studied, the HDX-MS analysis furthermore suggests that the C-terminus of *Tbg*ISG75 is not entirely disordered either. Residues Arg272-Arg317 initially exhibit significantly lower deuteration rates than truly disordered parts of the protein, such as the head domains, but show increased deuteration after longer incubation periods when compared to the predicted helices of the canonical three-helical bundle. A similar trend can be observed for residues Lys339-Ala370, and to a lesser degree for residues Glu371-Glu408. After incubation for 1200 s, residues Arg317-Glu438 show a large degree of deuteration, suggesting a high degree of flexibility. This indicates that parts of the C-terminus of *Tbg*ISG75 undergo structural transitions and therefore may occur as both folded and disordered in solution.

In line with a high χ^2 (and hence a disagreement between the *Tbg*ISG75 Alphafold2 and solution structures), HDX profiles also indicate that the prediction of an unusually long helix 3 is incorrect. High deuterium incorporation shown for residues Lys255-Ser264 strongly suggest that helix 3 terminates around residue Lys255.

Hence, in conclusion, these data show that the membrane-proximal, C-terminal regions of *Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75 are characterized by a significant degree of intrinsic disorder. This discovery aligns with the previously described observation that, similarly to the VSGs, the single-conformer AlphaFold2 models for the *Tbg*ISGs proved inadequate for explaining the solution scattering data (Figure 2, panels C to E).
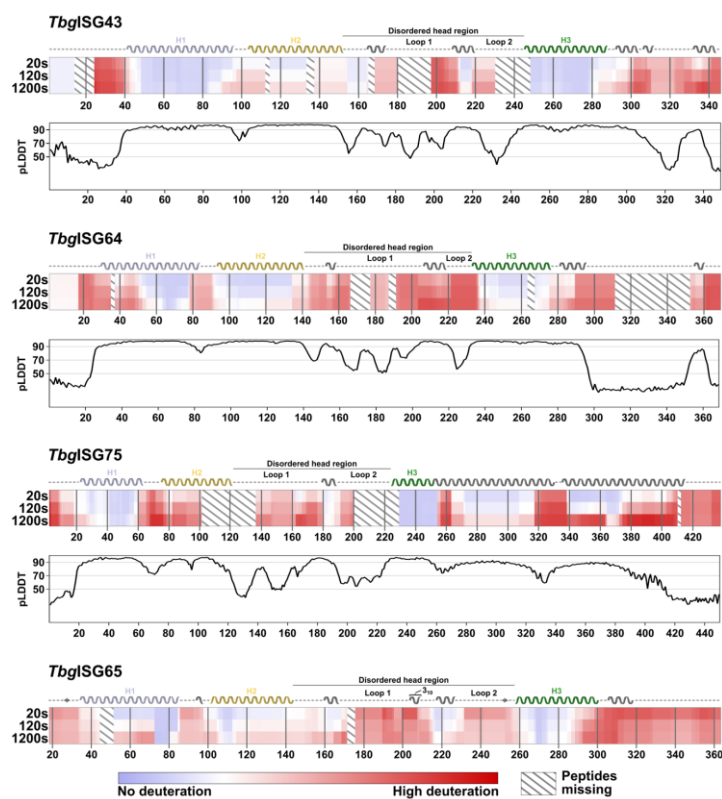
Figure 4. HDX-MS reveals intrinsically disordered regions in TbgISGs. Chiclet plots showing the relative deuteration across the sequences of *Tbg*ISGs at 20, 120 and 1200 s of incubation. The amount of relative deuteration is indicated by color gradients, ranging from blue (no deuteration) to red (high deuteration). Peptides absent in the analysis are represented by diagonal lines. For *Tbg*ISG75, *Tbg*ISG64 and *Tbg*ISG43, the secondary structure, as predicted by AlphaFold2, is indicated above the corresponding chiclet plot. For *Tbg*ISG65, the secondary structure of the integrative *Tbg*ISG65 hybrid structure deposited to the PDBDev (PDBDEV_00000201) is shown [7]. Disordered regions are represented by a dashed line, beta sheets as arrows and alpha helices as helices. The three helices constituting the canonical three-helical bundle are labelled (H1 to H3) and colored as in previous figures. The 3_{10}-helix in ISG65 is indicated with a label. The disordered loops forming the disordered, membrane-distal head region of the ISGs are labeled accordingly. Underneath the chiclet plots of *Tbg*ISG43, *Tbg*ISG64 and *Tbg*ISG75, the per-residue pLDDT scores of the respective AlphaFold2 models are plotted. Thresholds for high (>90), medium (>70) and low (<50) modelling confidence are indicated with grey lines.

Intrinsic disorder enables T. b. gambiense ISGs to switch between compact and elongated conformational states

To improve the fit to the experimental data and, thus, provide structural models that represent the *Tbg*ISG solution behavior best, the HDX-MS data were employed as restraints in the SAXS-based ensemble modelling (Supplementary Table 1). In all three cases, the best results were obtained when the structures are described by a conformational ensemble, in which the membrane-distal, N-terminal domains are treated as rigid bodies and the membrane-proximal, C-terminal regions are disordered. A thorough investigation of the conformational properties revealed that all ensembles yielding a good fit to the experimental data consist of two conformers: a compact and a highly extended conformer (Figure 5).

For *Tbg*ISG43, the selected ensemble provided a $\chi^2 = 3.28$, indicating a reasonable fit of the ensemble's theoretical scattering curve to the experimental data. While 70% of the ensemble consists of a compact conformer ($R_g = 33.47$ Å, $D_{max} = 129$ Å), the remaining 30% corresponds to a highly elongated form ($R_g = 69.03$ Å, $D_{max} = 276$ Å) (Figure 5A). A similar result was obtained for *Tbg*ISG64, for which the selected two-model ensemble (comprising 57% and 43% compact and elongated conformers, respectively) could be fitted to the experimental data with a $\chi^2 = 1.51$.

For *Tbg*ISG75, our modelling attempts were more extensive given that the HDX-MS data indicated that the C-terminal region displays an intriguing structural promiscuity, possibly transitioning between helical and disordered forms. When treating the entire C-terminus of *Tbg*ISG75 as disordered, the best fitting ensemble ($\chi^2 = 1.08$) consisted of 54% compact ($R_g = 28.37$ Å, $D_{max} = 90$ Å) and 46% extended ($R_g = 89.96$ Å, $D_{max} = 323$ Å) conformers. To explore ensembles representing the HDX-MS data more closely, ensembles with α-helical elements in their C-terminal regions corresponding to lower deuteration rates were generated. Among these ensembles, the least optimal fit between theoretical and experimental scattering data was obtained when restraining conformational sampling in residues Arg272-Arg317 ($\chi^2 = 1.66$), conserving the predicted α-helical fold for these residues. Even though the distances spanned by the compact ($R_g = 31.78$ Å, $D_{max} = 112$ Å) and the extended ($R_g = 95.2$ Å, $D_{max} = 308$ Å) conformers did not change dramatically, the fraction of conformers in each state did, with only 11% of the molecules modelled in the extended conformation. When excluding residues Lys339-Ala370 from the conformational sampling, thereby enforcing the predicted α-helical secondary structure in this region, a comparable distribution was observed. Although the fractions of the two conformers contributing to the ensemble did not change much with 16% of elongated and 84% of compact conformers, the calculated distances the two conformations spanned were noticeably different ($D_{max, \, compact} = 117$ Å and $D_{max, \, extended} = 282$ Å). This ensemble appears to describe the experimental data better in comparison with the previous restrains as the χ^2 improved to 1.1. Finally, when applying α-helix restrains in both regions of the C-terminus simultaneously, an ensemble with a similar distribution, 14% compact ($R_g = 31.07$ Å, $D_{max} = 103$ Å) and 86% extended ($R_g = 88.79$ Å, $D_{max} = 308$ Å) conformers was produced. Interestingly, the theoretical scattering could be fitted to the experimental data with a $\chi^2 = 1.08$, identical to the ensemble with no restraints in the C-terminus. These findings correlate well with the HDX-MS data, supporting the hypothesis that both states, entirely disordered and partially folded C-termini, may be present in solution. Noticeably, the presence of one or two helices in the CTD shifted the equilibrium from an equal distribution of compact and extended conformations, as observed in a fully disordered CTD, towards a prevalence of compact states.
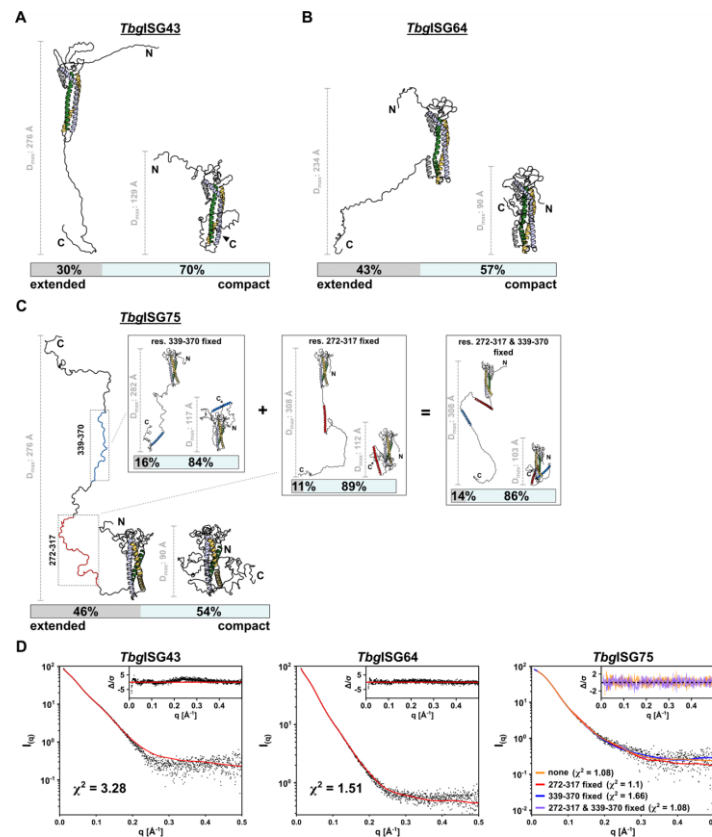
Figure 5. Conformational ensemble modelling reveals structural flexibility of ISGs in solution. Representative ensembles of (A) *Tbg*ISG43, (B) *Tbg*ISG64 and (C) *Tbg*ISG75 for the experimental SAXS data, calculated using BILBOMD. The smallest ensemble producing a reasonable improvement in goodness of fit of the theoretical scattering curve to the experimental data was chosen. Models constituting the selected ensemble are shown with their respective weight. For each model, the D_{max} is displayed with a representative scalebar. For ISG75 in (C) four ensembles are shown, corresponding to ensemble calculation without HDX-MS restraints, restraints applied to residues Arg272-Arg317 (highlighted in red) [left insert], residues Lys339-Ala370 (highlighted in blue) [middle insert] and both [right insert]. (D) Experimental scattering data are shown, overlaid with the theoretical scattering curves of the selected ensembles and the respective χ^2 of the fit. Residuals are shown as insets. For ISG75, the residuals for the ensemble without restraints for modelling of the CTD (orange) and with restraints in both res. 272-317 and res. 339-370 (purple) are shown.
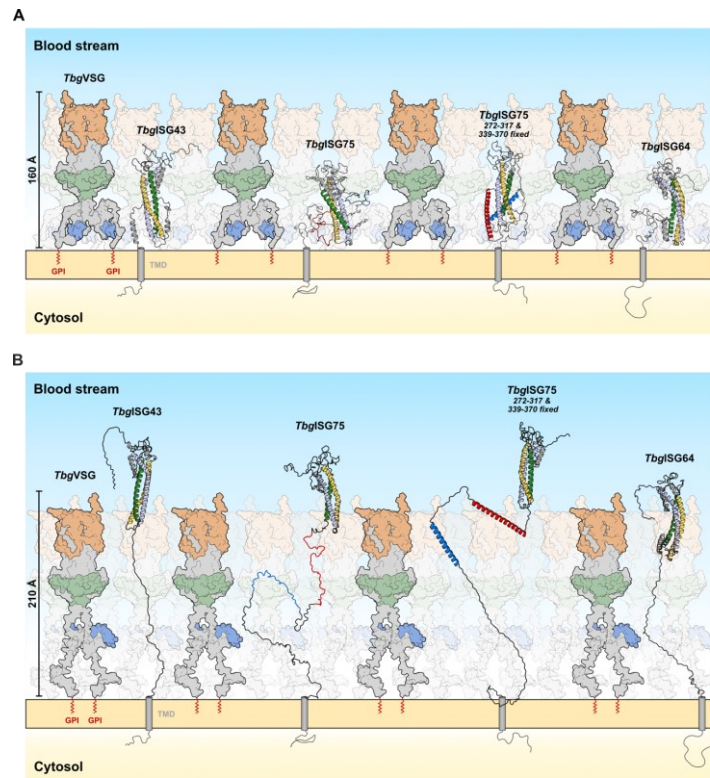
Figure 6. An updated model of the *Trypanosoma* surface coat – Functional implications of ISG conformational flexibility within the VSG umbrella model. (A) In the retracted form, ISGs reside within the VSG coat (represented here by *Tbg*VSG LiTat3.1), leaving only their disordered, loop-rich head domains accessible to molecules in the host's blood stream, while concealing all other epitopes under the protective VSG umbrella. **(B)** In the extended conformation, both ISG43 and ISG75 protrude well beyond the boundaries of the VSG layer, even when the latter is in its extended conformation. Whilst in this model ISG64 would not be able to fully extend beyond the maximally extended VSG umbrella, it still protrudes beyond the VSGs CTD, thereby implying accessibility to potential ligands. The GPI anchors of the VSGs are depicted as a red "zigzag", while transmembrane domains of ISGs are shown as gray cylinders. The N- and C-termini of the protein models obtained from SAXS modeling were manually modified for this figure to allow for membrane anchoring via the C-terminal linker without altering the D_{max} of the models.

DISCUSSION

The advent of AI-based structure prediction programs such as AlphaFold2 [38] and RoseTTAFold [50] has revolutionized and democratized structural biology. These tools can readily provide highly accurate, all-atom structural models without any experimental input. However, the generated structures do not always capture the solution behavior of proteins and their complexes, thus providing a very limited insight into their potential biological functions. This is especially true for intrinsically disordered proteins (IDPs) or proteins containing intrinsically disordered regions (IDRs). In these cases, AlphaFold2 structural models can be employed in conjunction with SAXS-based conformational sampling to provide an accurate description of a protein's conformational ensemble in solution [51,52]. Such an approach was applied here to investigate the structural features of *Tbg*ISGs (*Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75), interpret the results within the context of the *Tbg*VSG layer and provide a comparison with recent work published on *Tbg*ISG65 [7].

We first sought to validate the SAXS-based modeling approach employed here by collecting solution scattering data for *Tbb*VSG LiTat1.5 and *Tbg*VSG LiTat3.1 and comparing these to the results previously reported for *Tbb*VSG MiTat1.1 and *Tbb*VSG IITat1.24 [18]. In their work, Bartossek et al. showed that the *Tbb*VSG umbrella extends approximately 140 Å to 155 Å from the parasite surface. For both VSGs employed in this study, the solution behavior is best described by conformational ensembles with average D_{max} values of approximately 170 Å. Considering that in both studies the proteins were directly obtained from the parasite plasma membrane by GPI-specific phospholipase C cleavage, thereby eliminating all steric restraints imposed by localization on the parasite's membrane (e.g. the presence of neighboring molecules and protomer GPI-anchoring), we found our solution scattering data and resulting ensemble models to be in good agreement with the published *Tbb*VSG results [18]. Notably, the observed maximum dimensions of the extended models within the ensembles were larger in our approach than previously reported. It is conceivable that an MD-based approach can sample a larger conformational space of intrinsic disorder than the rigid body modelling based protocol employed earlier. Additionally, although the VSGs used in these studies belong to the same protein family, they are not identical, thus potentially explaining the observed differences.

After this validation step, the same SAXS-based modeling approach was employed to investigate the solution behavior of *Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75. For all three *Tbg*ISGs, we identified a clear requirement for ensemble-based structural modelling, *i.e.*, while single-conformer AlphaFold2 models could not adequately explain the SAXS data, conformational ensembles accurately describe the solution scattering data. Our findings show that the *Tbg*ISGs possess a C-terminal, membrane-proximal IDR, constituting a flexible tether between the TMD and the well-structured NTD. This configuration allows exploration of a large conformational landscape, encompassing highly compact to maximally extended conformers. For most ISGs the intrinsic disorder of these tails was captured well by AlphaFold2 (these regions have pLDDT scores around 50, which is indicative of intrinsic disorder [48]). However, the C-terminus of *Tbg*ISG75 was predicted to form a long pair of coiled α-helices, with confidence scores ranging from high (pLDDT between 70 to 90) to very high (pLDDT ≥ 90). Interestingly, a recent study has found AlphaFold2 capable of identifying conditionally folded IDRs and IDPs [53]. Alderson et al. have reported that secondary structure predictions with high and very high confidence scores in protein regions identified as IDRs via sequence-based prediction software are likely to resemble conditionally folded conformations, induced either upon post-translational modification or ligand binding. Coincidentally, most conditionally folded IDRs predicted by AlphaFold2 are helical. Considering the long-suspected role of ISG75 as an abundant receptor for a yet to be discovered ligand, the apparent discrepancy between the experimentally determined conformational ensemble and the single-conformer AlphaFold2 model could suggest that the *Tbg*ISG75 C-terminal linkers are conditionally folding IDRs. Likewise, the *Tbg*VSG LiTat3.1 AlphaFold2 model contains a predicted C-terminal α-helix (Val360-Ala410) harboring high to very high pLDDT scores, which is not in accordance with

its solution behavior and may potentially highlight this stretch as a conditionally folding IDR. While the molecular cue for the conditional folding remains speculative, both ligand-binding induced changes and folding in proximity of the parasite membrane are conceivable. In the absence of a stabilizing entity, the integrative structural biology approach adopted here indicates that the *Tbg*ISG75 and *Tbg*VSG LiTat3.1 solution structures are best described by conformational ensembles. For *Tbg*ISG75, the latter consists of a mixture of compact and extended conformers with transiently formed α-helices (Arg272-Arg317 and Lys339-Ala370).

The solution behavior observed for *Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75 is reminiscent of similar findings recently published for *Tbg*ISG65 [7]. Like the *Tbg*ISGs used in this study, *Tbg*ISG65 possesses a C-terminal, disordered linker, enabling the extension beyond the boundaries of the VSG layer and thereby facilitating interaction with its ligands, human complement factor C3 and its proteolytically activated fragments [6,7,24]. In a similar fashion, the IDRs of *Tbg*ISG43, *Tbg*ISG64, and *Tbg*ISG75 enable these proteins to adopt distinct conformational states, which would allow them to either protrude from or reside within the VSG coat (Figure 6). Indeed, with compact conformers spanning 90 Å (*Tbg*ISG64, *Tbg*ISG75) to 129 Å (*Tbg*ISG43), all three *Tbg*ISGs studied here appear to be deeply embedded within the VSG coat (Figure 6A). However, it is important to note that the experimentally determined dimensions of the most compact *Tbg*ISG conformers presented here are likely underestimated. This is because the absence of steric constraints imposed by TMD-anchoring would permit the C-terminal linker to adopt conformations that are not possible within the confines of the membrane context. Nonetheless, in their compact conformation, much of the *Tbg*ISG N-terminal domain would remain buried and be located at a similar distance from the parasite surface as the VSG's C-terminal domain (~70 Å). Given that the latter was postulated to be difficult to target by the host immune system [54,55], such a compact molecular surface organization would effectively shield the *Tbg*ISGs from an immune response, perhaps only leaving their membrane-distal head domains exposed, protruding into the gap between adjacent VSG dimers. Using HDX-MS disorder mapping we could demonstrate that, in addition to *Tbg*ISG65, other members of the *Tbg*ISG super-family also possess these domains. This is a new feature of trypanosomal invariant surface proteins that was so far only described for VSGs where similar structures act as immune-dominant decoys [56–58]. This is in line with previous studies showing that ISG65 is generally inaccessible to antibodies in *T. brucei*, while only minimal antibody binding to ISG75 is observed in fixed but not live cells [22]. In combination with the well-known rapid endocytosis and efficient turnover of surface antigens exhibited by trypanosomes [59], the IDR-mediated retractability of ISGs to reduce surface exposure could indeed be regarded as a highly efficient evasion mechanism for invariant determinants. In contrast, the elongated *Tbg*ISG43 and *Tbg*ISG75 conformers would be capable of reaching beyond the VSG layer boundary, even when compared to the overextended VSG solution state observed for *Tbg*VSG Litat3.1 (Figure 6B). Although *Tbg*ISG64 appears to span the smallest maximum distance of the three *Tbg*ISGs in its extended conformation (D_{max} = 234 Å), the N-terminal domain may still protrude well beyond the CTD of the VSG coat and, therefore, gain access to potential ligands. Although we believe our data to accurately depict the solution behavior of soluble ISGs and VSGs, it is possible that TMDs and GPI anchors as well as steric hindrances imposed by neighboring molecules add additional restraints on the trypanosome surface that could not be incorporated into our analysis. It also remains to be elucidated whether molecular triggers exist that can initiate vertical movement of surface receptors via their C-terminal IDRs and how ramping of variant and invariant proteins is coordinated. Our observation that the conditional presence of more secondary structure in the *Tbg*ISG75 CTD (i.e. fully disordered vs formation of 1 or 2 α-helices) yields predominantly compact conformations may provide a clue to how this could possibly be achieved.

The results presented in this study further expand our understanding of the structure-function relationship of *Trypanosoma* ISGs, which have been implicated in a wide range of functions

(including nutrient uptake, adhesion, environmental sensing, and immune evasion). In agreement with previous reports [7,28], these findings add to a growing body of evidence observed for non-VSG surface antigens in which intrinsic disorder appears to play a prominent, functional role at the host-parasite interface. This enables us to contribute to an updated model of the trypanosome surface coat and its embedded proteins, portraying it as a highly dynamic structure rather than a static barrier.

CONTRIBUTIONS

Y.G.-J.S. and S.Z. conceptualized the study. H.S. and S.Z. performed recombinant protein expression and purification. F.Z. and M.Z. cultured *T. b. gambiense* and provided cell pellets for VSG extraction. H.S. purified *Tbg*VSG LiTat3.1 from native source. B.S. and S.M. purified *Tbb*VSG LiTat1.5 from native source. AlphaFold2 predictions were made by Y.G.-J.S. R.G. and Y.G.-J.S. performed the SAXS measurements. SAXS data processing and analysis was done by H.S., A.N.V., S.Z. and Y.G.-J.S. HDX-MS data analysis was performed by H.S. and S.Z. SAXS driven ensemble modelling and analysis was done by H.S. (*Tbg*ISGs) and Y.G.-J.S. (*Tbb*VSG and *Tbg*VSG). The manuscript was written by H.S., Y.G.-J.S. and SZ.

REFERENCES

1. Salmon D, Vanwalleghem G, Morias Y, Denoeud J, Krumbholz C, Lhommé F, et al. Adenylate cyclases of Trypanosoma brucei inhibit the innate immune response of the host. Science (New York, NY). 2012;337: 463–466. doi:10.1126/science.1222753

2. Salmon D, Bachmaier S, Krumbholz C, Kador M, Gossmann JA, Uzureau P, et al. Cytokinesis of Trypanosoma brucei bloodstream forms depends on expression of adenylyl cyclases of the ESAG4 or ESAG4-like subfamily. 2012;84: 225–242. doi:10.1111/j.1365-2958.2012.08013.x

3. Salmon D. Adenylate Cyclases of Trypanosoma brucei, Environmental Sensors and Controllers of Host Innate Immune Response. Pathogens. 2018;7: 48. doi:10.3390/pathogens7020048

4. SEED JR, SECHELSKI OB, LOOMIS MR. A Survey for a Trypanocidal Factor in Primate Sera. J Protozool. 1990;37: 393–400. doi:10.1111/j.1550-7408.1990.tb01163.x

5. Rifkin MR. Identification of the trypanocidal factor in normal human serum: high density lipoprotein. Proc Natl Acad Sci. 1978;75: 3450–3454. doi:10.1073/pnas.75.7.3450

6. Macleod OJS, Cook AD, Webb H, Crow M, Burns R, Redpath M, et al. Invariant surface glycoprotein 65 of Trypanosoma brucei is a complement C3 receptor. Nat Commun. 2022;13: 5085. doi:10.1038/s41467-022-32728-9

7. Sülzen H, Began J, Dhillon A, Kereïche S, Pompach P, Votrubova J, et al. Cryo-EM structures of Trypanosoma brucei gambiense ISG65 with human complement C3 and C3b and their roles in alternative pathway restriction. Nat Commun. 2023;14: 2403. doi:10.1038/s41467-023-37988-7

8. Bangs JD. Evolution of Antigenic Variation in African Trypanosomes: Variant Surface Glycoprotein Expression, Structure, and Function. Bioessays. 2018;40: 1800181. doi:10.1002/bies.201800181

9. Pinger J, Nešić D, Ali L, Aresta-Branco F, Lilic M, Chowdhury S, et al. African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat. Nature microbiology. 2018;3: 932. doi:10.1038/s41564-018-0187-6

10. Mony BM, MacGregor P, Ivens A, Rojas F, Cowton A, Young J, et al. Genome-wide dissection of the quorum sensing signalling pathway in Trypanosoma brucei. Nature. 2013; 1–17. doi:10.1038/nature12864

11. Rojas F, Silvester E, Young J, Milne R, Tettey M, Houston DR, et al. Oligopeptide Signaling through TbGPR89 Drives Trypanosome Quorum Sensing. Cell. 2019;176: 306-317.e16. doi:10.1016/j.cell.2018.10.041

12. Szempruch AJ, Sykes SE, Kieft R, Dennison L, Becker AC, Gartrell A, et al. Extracellular Vesicles from Trypanosoma brucei Mediate Virulence Factor Transfer and Cause Host Anemia. Cell. 2016;164: 246–257. doi:10.1016/j.cell.2015.11.051

13. Eliaz D, Kannan S, Shaked H, Arvatz G, Tkacz ID, Binder L, et al. Exosome secretion affects social motility in Trypanosoma brucei. Hill KL, editor. PLoS Pathogens. 2017;13: e1006245-38. doi:10.1371/journal.ppat.1006245

14. Hall T, Esser K. Topologic mapping of protective and nonprotective epitopes on the variant surface glycoprotein of the WRATat 1 clone of Trypanosoma brucei rhodesiense. J Immunol. 1984;132: 2059–2063. doi:10.4049/jimmunol.132.4.2059

15. Cross GAM. Structure of the variant glycoproteins and surface coat of Trypanosoma brucei. Philos Trans R Soc Lond B, Biol Sci. 1984;307: 3–12. doi:10.1098/rstb.1984.0104

16. Masterson WJ, Taylor D, Turner MJ. Topologic analysis of the epitopes of a variant surface glycoprotein of Trypanosoma brucei. J Immunol. 1988;140: 3194–3199. doi:10.4049/jimmunol.140.9.3194

17. Overath P, Chaudhri M, Steverding D, Ziegelbauer K. Invariant surface proteins in bloodstream forms of Trypanosoma brucei. Parasitol Today. 1994;10: 53–58. doi:10.1016/0169-4758(94)90393-x

18. Bartossek T, Jones NG, fer CS x000E4, x00107 MC, Glogger M, Mott HR, et al. Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. Nature microbiology. 2017; 1–12. doi:10.1038/s41564-017-0013-6

19. Mussmann R, Engstler M, Gerrits H, Kieft R, Toaldo CB, Onderwater J, et al. Factors Affecting the Level and Localization of the Transferrin Receptor in Trypanosoma brucei *. J Biol Chem. 2004;279: 40690–40698. doi:10.1074/jbc.m404697200

20. Field MC, Carrington M. The trypanosome flagellar pocket. Nature Reviews Microbiology. 2009;7: 775–786. doi:10.1038/nrmicro2221

21. Ziegelbauer K, Overath P. Identification of invariant surface glycoproteins in the bloodstream stage of Trypanosoma brucei. J Biol Chem. 1992;267: 10791–6.

22. Ziegelbauer K, Overath P. Organization of two invariant surface glycoproteins in the surface coat of Trypanosoma brucei. Infect Immun. 1993;61: 4540–4545. doi:10.1128/iai.61.11.4540-4545.1993

23. Chung W, Leung KF, Carrington M, Field MC. Ubiquitylation is Required for Degradation of Transmembrane Surface Proteins in Trypanosomes. Traffic. 2008;9: 1681–1697. doi:10.1111/j.1600-0854.2008.00785.x

24. Cook AD, Carrington M, Higgins MK. Trypanosomes and complement: more than one way to die? Trends Parasitol. 2023;39: 1014–1022. doi:10.1016/j.pt.2023.09.001

25. Sullivan L, Wall SJ, Carrington M, Ferguson MAJ. Proteomic Selection of Immunodiagnostic Antigens for Human African Trypanosomiasis and Generation of a Prototype Lateral Flow Immunodiagnostic Device. Plos Neglect Trop D. 2013;7: e2087. doi:10.1371/journal.pntd.0002087

26. Magez S, Li Z, Nguyen HTT, Torres JEP, Wielendaele PV, Radwanska M, et al. The History of Anti-Trypanosome Vaccine Development Shows That Highly Immunogenic and Exposed Pathogen-Derived Antigens Are Not Necessarily Good Target Candidates: Enolase and ISG75 as Examples. Pathogens. 2021;10: 1050. doi:10.3390/pathogens10081050

27. Cross GAM. Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology. 1975;71: 393–417. doi:10.1017/s003118200004717x

28. Casas-Sanchez A, Ramaswamy R, Perally S, Haines LR, Rose C, Aguilera-Flores M, et al. The Trypanosoma brucei MISP family of invariant proteins is co-expressed with BARP as triple helical bundle structures on the surface of salivary gland forms, but is dispensable for parasite development within the tsetse vector. PLOS Pathog. 2023;19: e1011269. doi:10.1371/journal.ppat.1011269

29. Makarov A, Began J, Mautone IC, Pinto E, Ferguson L, Zoltner M, et al. The role of invariant surface glycoprotein 75 in xenobiotic acquisition by African trypanosomes. Microb Cell. 2023;10: 18. doi:10.15698/mic2023.02.790

30. Moon S, Janssens I, Kim KH, Stijlemans B, Magez S, Radwanska M. Detrimental Effect of Trypanosoma brucei brucei Infection on Memory B Cells and Host Ability to Recall Protective B-cell Responses. J Infect Dis. 2022;226: 528–540. doi:10.1093/infdis/jiac112

31. Kavan D, Man P. MSTools—Web based application for visualization and presentation of HXMS data. Int J Mass Spectrom. 2011;302: 53–58. doi:10.1016/j.ijms.2010.07.030

32. Thureau A, Roblin P, Pérez J. BioSAXS on the SWING beamline at Synchrotron SOLEIL. J Appl Crystallogr. 2021;54: 1698–1710. doi:10.1107/s1600576721008736

33. Tully MD, Kieffer J, Brennich ME, Aberdam RC, Florial JB, Hutin S, et al. BioSAXS at European Synchrotron Radiation Facility – Extremely Brilliant Source: BM29 with an upgraded source, detector, robot, sample environment, data collection and analysis software. J Synchrotron Radiat. 2023;30: 258–266. doi:10.1107/s1600577522011286

34. Orthaber D, Bergmann A, Glatter O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. Journal of Applied Crystallography. 2000;33: 218–225. Available: http://onlinelibrary.wiley.com.ezproxy.vub.ac.be:2048/store/10.1107/S0021889899015216/asset/S0021889899015216.pdf?v=1&t=gpgt3kev&s=1798c05aa7c262675ca49710f8431d86c43fbe46

35. Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. Bioinformatics. 2018;34: 1944–1946. doi:10.1093/bioinformatics/btx846

36. Hopkins JB, Gillilan RE, Skou S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. J Appl Cryst (2017) 50, 1545-1553 [doi:101107/S1600576717011438]. 2017; 1–9. doi:10.1107/s1600576717011438

37. Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV, Molodenskiy DS, et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. J Appl Crystallogr. 2021;54: 343–355. doi:10.1107/s1600576720013412

38. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589. doi:10.1038/s41586-021-03819-2

39. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2022; 2021.10.04.463034. doi:10.1101/2021.10.04.463034

40. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Research. 2010;38: W540-4. doi:10.1093/nar/gkq461

41. Pelikan M, Hura G, Hammel M. Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen Physiol Biophys. 2009;28: 174–189. doi:10.4149/gpb_2009_02_174

42. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. 2016;44: W424–W429. doi:10.1093/nar/gkw389

43. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: Tools for structure building and analysis. Protein Sci. 2023;32: e4792. doi:10.1002/pro.4792

44. Đaković S, Zeelen JP, Gkeka A, Chandra M, Straaten M van, Foti K, et al. A structural classification of the variant surface glycoproteins of the African trypanosome. PLOS Neglected Trop Dis. 2023;17: e0011621. doi:10.1371/journal.pntd.0011621

45. So J, Sudlow S, Sayeed A, Grudda T, Deborggraeve S, Ngoyi DM, et al. VSGs Expressed during Natural T. b. gambiense Infection Exhibit Extensive Sequence Divergence and a Subspecies-Specific Bias towards Type B N-Terminal Domains. Mbio. 2022; e02553-22. doi:10.1128/mbio.02553-22

46. Provencher SW, Gloeckner J. Estimation of globular protein secondary structure from circular dichroism. Biochemistry. 1981;20: 33–37. doi:10.1021/bi00504a006

47. Stokkum IHM van, Spoelder HJW, Bloemendal M, Grondelle R van, Groen FCA. Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal Biochem. 1990;191: 110–118. doi:10.1016/0003-2697(90)90396-q

48. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. Nature. 2021;596: 590–596. doi:10.1038/s41586-021-03828-1

49. Masson GR, Burke JE, Ahn NG, Anand GS, Borchers C, Brier S, et al. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. Nat Methods. 2019;16: 595–602. doi:10.1038/s41592-019-0459-y

50. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373: 871–876. doi:10.1126/science.abj8754

51. Brookes E, Rocco M, Vachette P, Trewhella J. AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank. J Appl Crystallogr. 2023;56: 910–926. doi:10.1107/s1600576723005344

52. Receveur-Bréchot V. AlphaFold, small-angle X-ray scattering and ensemble modelling: a winning combination for intrinsically disordered proteins. J Appl Crystallogr. 2023;56: 1313–1314. doi:10.1107/s1600576723008403

53. Alderson TR, Pritišanac I, Kolarić Đ, Moses AM, Forman-Kay JD. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. Proc Natl Acad Sci. 2023;120: e2304302120. doi:10.1073/pnas.2304302120

54. Schwede A, Jones N, Engstler M, Carrington M. The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Molecular and biochemical parasitology. 2011;175: 201–204. doi:10.1016/j.molbiopara.2010.11.004

55. Hempelmann A, Hartleb L, Straaten M van, Hashemi H, Zeelen JP, Bongers K, et al. Nanobody-mediated macromolecular crowding induces membrane fission and remodeling in the African trypanosome. Cell Reports. 2021;37: 109923. doi:10.1016/j.celrep.2021.109923

56. Miller EN, Allan LM, Turner MJ. Topological analysis of antigenic determinants on a variant surface glycoprotein of Trypanosoma brucei. Mol Biochem Parasitol. 1984;13: 67–81. doi:10.1016/0166-6851(84)90102-6

57. Clarke MW, Barbet AF, Pearson TW. Structural features of antigenic determinants on variant surface glycoproteins from Trypanosoma brucei. Mol Immunol. 1987;24: 707–713. doi:10.1016/0161-5890(87)90052-6

58. Freymann D, Down J, Carrington M, Turner M, Wiley D. 2.9A resolution structure of the N-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. Journal Of Molecular Biology. 1990;216: 141–160. Available: message:%3CCAD0U+0Ce_Q8xzvu0Qao98pUSMiChQ21kmdUGUEypzT3ttz3pfg@mail.gmail.com%3E

59. Engstler M, Pfohl T, Herminghaus S, Boshart M, Wiegertjes G, Heddergott N, et al. Hydrodynamic Flow-Mediated Protein Sorting on the Cell Surface of Trypanosomes. Cell. 2007;131: 505–515. doi:10.1016/j.cell.2007.08.046

6.4 Structural insights into the interaction between adenovirus C5 hexon and human lactoferrin.

Structural insights into the interaction between adenovirus C5 hexon and human lactoferrin

Arun Dhillon,[1] B. David Persson,[2] Alexander N. Volkov,[3,4] Hagen Sülzen,[1,5] Alan Kádek,[6,7] Petr Pompach,[8] Sami Kereïche,[1,9] Martin Lepšík,[1] Katarina Danskog,[10] Charlotte Uetrecht,[11] Niklas Arnberg,[10] Sebastian Zoll[1]

AUTHOR AFFILIATIONS See affiliation list on p. 16.

ABSTRACT Adenovirus (AdV) infection of the respiratory epithelium is common but poorly understood. Human AdV species C types, such as HAdV-C5, utilize the Coxsackie-adenovirus receptor (CAR) for attachment and subsequently integrins for entry. CAR and integrins are however located deep within the tight junctions in the mucosa where they would not be easily accessible. Recently, a model for CAR-independent AdV entry was proposed. In this model, human lactoferrin (hLF), an innate immune protein, aids the viral uptake into epithelial cells by mediating interactions between the major capsid protein, hexon, and yet unknown host cellular receptor(s). However, a detailed understanding of the molecular interactions driving this mechanism is lacking. Here, we present a new cryo-EM structure of HAdV-5C hexon at high resolution alongside a hybrid structure of HAdV-5C hexon complexed with human lactoferrin (hLF). These structures reveal the molecular determinants of the interaction between hLF and HAdV-C5 hexon. hLF engages hexon primarily *via* its N-terminal lactoferricin (Lfcin) region, interacting with hexon's hypervariable region 1 (HVR-1). Mutational analyses pinpoint critical Lfcin contacts and also identify additional regions within hLF that critically contribute to hexon binding. Our study sheds more light on the intricate mechanism by which HAdV-C5 utilizes soluble hLF/Lfcin for cellular entry. These findings hold promise for advancing gene therapy applications and inform vaccine development.

IMPORTANCE Our study delves into the structural aspects of adenovirus (AdV) infections, specifically HAdV-C5 in the respiratory epithelium. It uncovers the molecular details of a novel pathway where human lactoferrin (hLF) interacts with the major capsid protein, hexon, facilitating viral entry, and bypassing traditional receptors such as CAR and integrins. The study's cryo-EM structures reveal how hLF engages hexon, primarily through its N-terminal lactoferricin (Lfcin) region and hexon's hypervariable region 1 (HVR-1). Mutational analyses identify critical Lfcin contacts and other regions within hLF vital for hexon binding. This structural insight sheds light on HAdV-C5's mechanism of utilizing soluble hLF/Lfcin for cellular entry, holding promise for gene therapy and vaccine development advancements in adenovirus research.

KEYWORDS adenovirus, lactoferrin, cryo-EM, immune evasion, viral entry

Adenoviruses are non-enveloped viruses with a capsid of icosahedral symmetry that encapsulates a double-stranded DNA genome (1). Currently, more than 110 human adenovirus (HAdV) types have been identified that are categorized into "species" A through G based on tissue tropism and genetic diversity (2). HAdVs usually cause self-limiting ocular (species B, C, D, and E), respiratory (species B, C, and E), gastrointestinal (species A and F), and urinary tract (species B) infections, which might turn lethal in immuno-compromised individuals and children (3). Species C adenoviruses that include HAdV-5 may also cause persistent infections of the adenoids and tonsils (4, 5)

resulting in viral shedding for months (6). Historically, adenoviruses have played pivotal roles as molecular biology tools in the discovery of eukaryotic biological mechanisms such as RNA splicing and protein folding (7, 8). In addition, the non-integrating nature of adenoviruses has been harnessed in the field of gene therapy (9) where a major obstacle to overcome for the use of HAdVs in gene therapy is the presence of pre-existing immunity (10) and the relatively poor transduction efficiency in the target tissue (11). Nevertheless, after decades of research, a HAdV-based gene therapy was recently approved by regulators in the United States of America (9). In addition, vaccines based on modified adenoviruses against Ebola (12), HIV (13), ZIKV (14), RSV (15), and SARS-CoV2 (16) are currently in various phases of clinical trials or already being used in the clinic.

Species C adenoviruses infect respiratory epithelial cells by interacting with the cellular receptor CAR using the protruding fiber protein, followed by secondary interactions between the penton base and $\alpha_v\beta_3$ integrins that trigger entry (1, 17, 18). This fundamentally simple model is likely used in immortalized cell lines but is highly unlikely in the polarized epithelium where CAR is hidden deep within the tight junctions (19–21). However, additional models and mechanisms have been suggested to overcome this dilemma. HAdV-C5 infection may be facilitated by airway macrophages that take up invading HAdV particles resulting in the secretion of IL-8 as part of their inflammatory response (22). The secretion of IL-8 results in the apical migration of the CAR isoform CAR^{Ex7} and $\alpha_v\beta3$ from the basolateral membrane while enhancing the expression of the CAR^{Ex8} at the apical membrane (22). In addition, several CAR-independent cell entry pathways have been identified for HAdV-C5, such as the use of heparin on hepatocytes (23), $\alpha_v\beta_5$ on melanoma and breast cancer cells (24), and Toll-like receptor 4 on dendritic cells (25). Finally, several different types of HAdVs have been shown to use different soluble components present in body fluids to enhance the infection. For example, HAdV-5C has been shown to recruit and hi-jack coagulation factor X (FX) for infection of hepatocytes (23), which not only enhances infection but also controls the tropism by directing the virus to the liver. During infection, FX binds to protruding loops of the HAdV-hexon on the capsid thus allowing FX to bridge an interaction between the cellular receptor and the virus particle. Hexon is a trimeric, barrel-shaped protein that represents the main building block of the virus capsid. In addition to coagulation factors, species C HAdVs have been shown to use hLF and its proteolytic cleavage product lactoferricin (Lfcin), a highly positively charged, 49 amino acid peptide with antimicrobial activity (26), for enhanced infection of epithelial cells (27, 28). Even though this mechanism is less well described, both hLF and Lfcin have been suggested to bind the HVR-1 of hexon acting as a bridge between the virus capsid and cell surface receptor(s). Interestingly, similar mechanisms have been described for species A HAdVs that bind factor IX for a more effective infection (29, 30) indicating that this mechanism might be an efficient way for HAdVs to bypass initial infection using CAR.

In this study, we present a new cryo-EM structure of HAd-5C hexon at FSC$_{0.143}$ of 2.9 Å together with a detailed hybrid structure of hLF in complex with HAdV-5C hexon providing structural insight into the previously reported interaction between hLF and HAdV-C5 hexon (28). We show that hLF engages the HVR-1 of one hexon protomer *via* its N-terminal Lfcin region with additional stabilizing contacts between the hLF C-terminal lobe and a second hexon protomer. In addition, structure-based mutational studies identified key residues within Lfcin as well as an adjacent helix in the N-lobe. Taken together, our structural and biophysical characterizations shed more light onto the mechanism by which HAdV-C5 utilizes soluble hLF/Lfcin for cellular entry. These detailed insights may eventually aid the development of novel adjuvants for use in HAdV-C5-based gene therapy.

RESULTS

The interaction between HAdV-C5 hexon and hLF is charge driven

Based on a series of viral transduction experiments, it was previously suggested that the interaction between HAdV-C5 hexon and hLF involves the stretch of acidic amino acid residues in the HAdV-C5 hexon HVR-1 and the basic N-terminal region of hLF (28) (Fig. 1A and B). When proteolytically liberated, the N-terminal region of hLF, Lfcin, is equally capable of promoting transduction (26). Here, we provide direct biophysical evidence for a charge-driven interaction. Using size exclusion chromatography (SEC) and native mass spectroscopy (MS), we could show that the stability of the complex is highly dependent on the ionic strength of the buffer used which is typical for charge-driven interactions (Fig. 1C and D). In SEC, the HAdV-C5 hexon:hLF complex stays intact only at a salt concentration of ≤5 mM (Fig. 1C). Any further increase in salt concentration leads to gradual disruption of the complex. At 10 mM NaCl, the complex largely disintegrates with free hLF adsorbing to the matrix of the SEC column. Therefore, at 10 mM NaCl, no additional peak corresponding to free hLF such as in the presence of 150 mM NaCl, can be observed. Excess hLF not bound to hexon could afterward be eluted from the column with a higher salt concentration. To rule out aberrant behavior of the protein due to nonspecific interaction with the stationary phase of the column under low salt conditions, we carried out native MS measurements. Consistent with observations made in SEC, the binding of hLF to hexon could only be observed under low ionic strength conditions (Fig. 1D; Fig. S1). In this case, ammonium acetate (AA) was used as an MS-compatible buffer surrogate. In the presence of 20 mM AA, the hexon:hLF complex was not stable. Compared to free hexon, only a fraction of hexon:hLF at 3:1 (hexon protomers:hLF) stoichiometry was detectable. Upon dilution from 20 mM AA to around 5 mM, predominantly a complex with 3:1 stoichiometry was observed with smaller amounts of a complex with 2 hLF molecules per hexon trimer as well as free hexon.

Cryo-EM structure of HAdV-C5 hexon at 2.9 Å

The structure of the HAdV-C5 hexon was reconstructed to $FSC_{0.143}$ of 2.9 Å (Fig. 2A and B; Table S1). To our knowledge, this is the highest resolution structure of HAdV-C5 hexon determined by cryo-EM. Due to the high stability of the trimer, the resolution is relatively uniform across the whole structure and thus it could be modeled unambiguously with a high average cross-correlation of 0.736 (Fig. 2C). Throughout the length of the hexon, the protomers are in close contact forming large interfaces in addition to intertwined regions at the "turrets" and the base. Each protomer contributes approximately 16,000 $Å^2$ to the interface with the two other protomers. Significantly lower resolution is observed only for the hypervariable loop regions, the "turrets" at the top of the molecule which are largely disordered and could not be built due to a lack of traceable density in the final reconstruction. Specifically, regions 137–164, 188–190, 252–256, 271–278, and 431–436 could not be sufficiently reconstructed. These regions are also lacking in the crystal structure (3TG7) of hexon. The same structure was used as a starting model to build hexon in the cryo-EM reconstruction but no major conformational changes were observed. The RMSD between both structures is 0.9 Å, indicating high overall similarity (Fig. S2).

Integrative modeling of the HAdV-C5 hexon:hLF complex

Despite many attempts, the strong preferential orientation of the hexon:hLF complex on EM grids could not be overcome. Unlike for hexon alone, which also showed a higher number of top and bottom views (Fig. 2B), this phenomenon was exacerbated in the complex and precluded its reconstruction. Standard remedial approaches such as stage tilting and the use of detergents were ineffective. Furthermore, grid preparation was highly irreproducible with only small percentages of intact complex particles being detectable. Likewise, reconstructing hLF bound to whole virus particles was
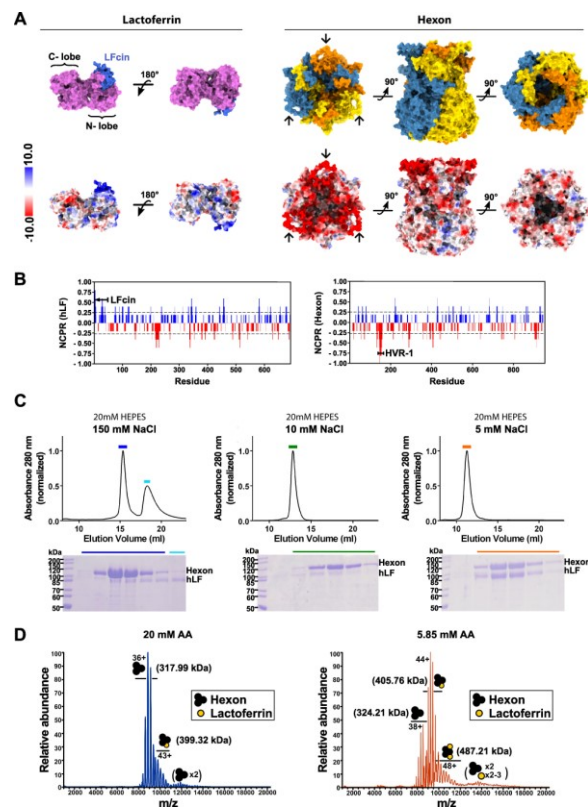
FIG 1 Effect of ionic strength on hexon:hLF complex formation. (A) Surface views of lactoferrin (PDB: 1LFG) (left) and hexon, completed with AlphaFold2 (see methods) (right). The Lfcin peptide on the N-terminal lobe of LF is colored in blue. The three protomers of hexon are shown in different colors: orange, yellow, and blue. The HVR-1 loop of each subunit is marked with an arrow. Surface representations colored according to electrostatic potential are depicted below. Notably, the area on lactoferrin corresponding to the Lfcin peptide exhibits a strong positive charge while the HVR-1 loops of hexon are strongly negatively charged. (B) Net charge per residue (NCPR) distribution along the amino acid sequence of hLF (left) and hexon (right). Blue color indicates positive NCPR and red color indicates negative NCPR. Residues corresponding to the Lfcin and HVR-1 have been highlighted. NCPR distribution was calculated *via* CIDER with the default blob size. (C) Hexon:hLF complex formation at different ionic strengths during size exclusion chromatography (top). Reducing SDS gels (bottom) confirms the co-elution of hexon and hLF. Fractions loaded onto the SDS gel have been highlighted with colored bars in the corresponding chromatogram (D) Native MS of 0.5 µM hexon:hLF complex electro-sprayed from 20 mM (left) and 5.85 mM (right) ammonium acetate (AA) pH 7.0. Hexon is detected exclusively trimeric (black) with up to 2 hLF molecules bound (yellow). The main charge states and corresponding molecular weights are indicated. Small amounts of dimeric hexon trimer are artifacts due to ESI concentration effects. For a more detailed peak assignment, see Fig. S1.
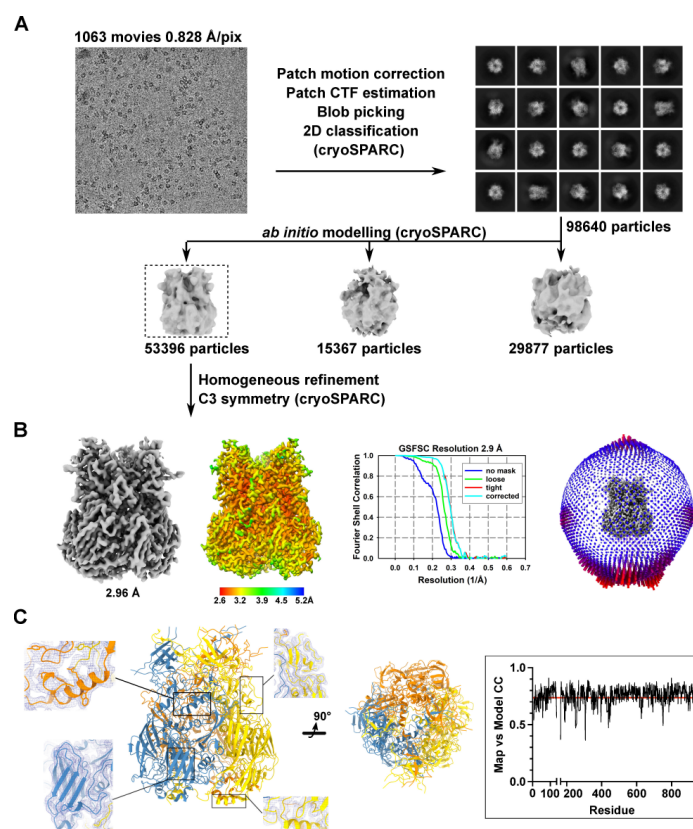
FIG 2 Structure of HAdV-C5 hexon determined by single particle cryo-EM at 2.9 Å. (A) Visual representation of the workflow adopted for data processing using cryoSPARC. A particle set identified by iterative, reference-free 2D classification was used to obtain three *ab initio* models. The model with the most recognizable features was selected (boxed) and homogeneously refined to reconstruct the trimeric hexon. 53,396 particles contributed to the final reconstruction. A representative micrograph (top left) is shown alongside a panel of selected 2D class averages. (B) The final, unsharpened map of the hexon reconstruction is shown in gray and colored according to local resolution. Next to the gold-standard FSC curves is the angular distribution plot of the reconstruction. (C) Side and bottom view of the hexon model built from the cryo-EM reconstruction. The three protomers of the hexon are depicted in the color scheme used previously. Insets show close-up views of the model inside the unsharpened map from different regions of the hexon. The plot of the Map vs Model cross-correlation for one protomer (boxed) shows a high average cross-correlation of 0.736 (red line) at $FSC_{0.143}$: 2.9 Å.

unsuccessful. Here, the primary challenge arose from the different occupational states of the hexon trimer. Data from native MS and single-particle EM have shown that the hexon:hLF complex exhibits stoichiometries of 1:3 and 2:3, in addition to the presence of

unbound hexon. This heterogeneity in coating complicates the use of symmetry operators for virus reconstruction. As a result, achieving high-quality reconstructions would require the collection of very large data sets, given the expected diversity in particle morphology.

Despite the lack of side views of the hexon:hLF complex on EM grids, good 2D class averages were obtained for the top and bottom views of the complex. We therefore set out to use this information in the calculation of a hybrid atomic model of the complex, adapting a protocol by Walti et al (31). Implemented in the program XPLOR-NIH (32), this protocol allows the fitting of protein models to two-dimensional electron density projection images. In the first step of our integrative modeling approach, we performed crosslinking mass spectrometry (XL-MS) with the HAdV-C5 hexon:hLF complex using the homobifunctional crosslinker disuccinimidyl dibutyric urea (DSBU). Two pairs of crosslinks were identified (Fig. 3A; Fig S3). The hexon residues (S174, S446) crosslinked to hLF residues are present in the "turrets," while one of the corresponding hLF residues is located in the N-lobe (K282) and the second (K676) one at the junction of N- and C-lobe. Since, hexon is a homotrimer, in principle, hLF could be crosslinked to residues belonging to the same protomer or residues on different protomers of hexon. However, only when hLF was assumed to be crosslinked to the same hexon protomer did the Lfcin region face hexon. As prior evidence exists for the involvement of Lfcin in hexon binding (28), it was assumed that hLF is crosslinked to residues on the same protomer of hexon (hexon S446: hLF K282; hexon S174: hLF K676). Next, we carried out a molecular dynamics simulation of hexon alone (Fig. 3A) and selected a conformation from the latter part of the trajectory, post-convergence of RMSD, that closely mirrors the observed hexon conformation in experimental 2D classes of the complex (Fig. 3B; Fig. S4). Complex projections suggest a conformational change in the flexible "turret" region of hexon upon binding of hLF.

These steps created the hexon input model and established the principal binding site and distance constraints for docking of the complex in XPLOR-NIH. The docked and energy-minimized complex (Fig. 3A) was then rigid body refined against the 2D projections and models were calculated (Fig. 3B and C). Models without crosslink distance violations were ranked according to their correlation with the experimental projections of the complex (Fig. 3B; Fig S5). The model of HAdV-C5 hexon:hLF complex with the highest C-value was selected as the final model and used for further analysis (Fig. 3C). In the selected model, residues 1–49 of hLF, constituting Lfcin, are positioned facing the hexon HVR-1, while another previously not reported point of close contact is near crosslink 2 (hexon S174-hLF K676) between the C-lobe of hLF and an adjacent protomer within hexon.

Residues on both hLF lobes contribute to HAdV-C5 hexon binding

To validate our model and obtain further molecular insight, the mutational analysis of hLF binding to hexon was carried out. Guided by the 2D class averages and the generated hybrid model, surface exposed and solvent accessible hLF residues in the regions of two apparent contact points were mutated, and binding affinities to fluorescently labeled hexon were measured using microscale thermophoresis (MST) (Fig. 4A). Single, charge-reducing mutations in Lfcin (mut1-4; R24S, R27A, K28A, and R30A) all resulted in reduced binding affinity of the respective hLF mutant. The importance of each of these residues was shown by their additive effect in a quadruple mutant (mut5; R24S/R27A/K28A/R30A) which showed the lowest binding affinity within the set of mutants, retaining less than 20% affinity of the WT protein. Point mutations were also introduced in a region directly adjacent to Lfcin which also harbors crosslink 1 (Hexon S446-hLF K282). Interestingly, this region, which has not been implicated in hexon binding before, also seems to play an important role in complex stability (Fig. 4A and B). A triple mutant (mut6; K277S/K280A/K285S) of hLF was found to be similarly impaired in its binding to hexon. Although not measured, it is plausible that a mutant combining mutations from mut5 and 6 might not interact with hexon anymore.
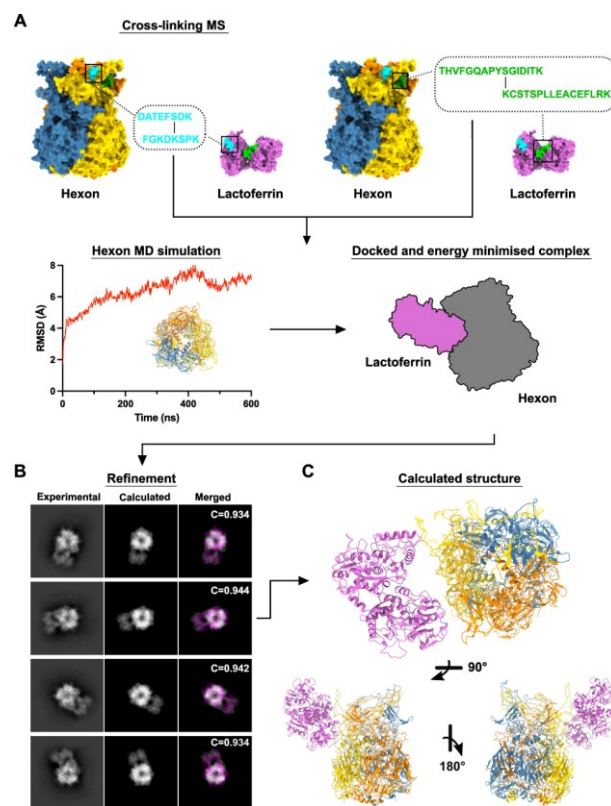
FIG 3 Visual representation of computation workflow for a hybrid model of the hexon:hLF complex. (A) Top, Crosslinking MS revealed two crosslinks, locations of identified peptides are highlighted on surface representations of hLF (PDB: 1LFG) and hexon in cyan and green. Peptide sequences are shown in the corresponding color, crosslinked residues are marked with a black line. Bottom left; Hexon MD simulation—the evolution of backbone RMSD, overlay of several snapshots from the MD trajectory. Bottom right; Identified crosslinks were used as distance restrains in the subsequent docking and energy minimization of the complex in XPLOR-NIH, visualized by representative schematic. (B) Refinement of the docked hexon:hLF complex against experimental 2D class averages of the complex using XPLOR-NIH (see methods). The left column shows the experimentally obtained 2D class averages for the hexon:hLF, the central column shows the calculated back projections for the hexon:hLF complex obtained after docking and energy minimization, and the right column shows the result of the superposition of experimental (green) and calculated (magenta) projections. Cross-correlation (C) of experimental and calculated projections is shown for individual projections. (D) Different views of the final hybrid model (best C-value) of hexon:hLF complex. LF is colored pink, the protomers of hexon are individually colored using the color scheme as before.
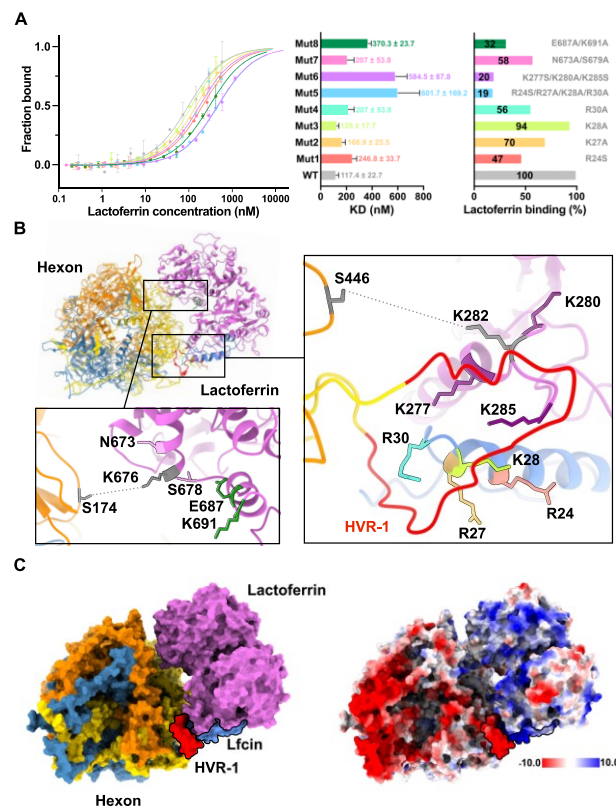
FIG 4 Validation of the hexon:hLF hybrid model. (A) Microscale thermophoresis (MST) was used to determine the effect of mutations in hLF interface residues on the affinity to hexon. The fraction of hLF bound to fluorescently labeled hexon was determined and plotted as a dose–response curve (*n* = 2). The binding curve of WT hLF is shown in gray, tested hLF mutants are shown in color. The colors of dose–response curves are identical to the colors used in bar graphs. The left diagram depicts the measured affinity; the right diagram shows the relative binding capacity of lactoferrin mutants to hexon in comparison to the WT. The percentage of lactoferrin binding is shown inside the bar. Residues mutated in hLF are shown next to their respective bar. (B) Top view of the hexon:hLF hybrid model in cartoon representation. Mutants, with the exception of Mut5, have been colored using the same color scheme as in (A). In addition, HVR-1 on hexon has been colored red, Lfcin on the lactoferrin is colored blue. Insets show close-up views of the mutated residues. Amino acids in identified crosslinks are depicted in gray, the crosslinks are symbolically depicted as dotted lines. (C) Surface representation of the computed hexon:hLF complex in the color scheme used throughout and colored according to the electrostatic potential. HVR-1 and Lfcin have been highlighted on both representations with a black outline.

Double mutants mut7 (N673A/S678A) and mut8 (E687A/K691A) were introduced near the site of crosslink two which appears to be in proximity to hexon in our hybrid model (Fig. 4B and C) as well as 2D class averages (Fig. 3C). Although, to a lesser extent than

mutations in the tip of the N-lobe, both double mutants are also significantly impaired in their binding to hexon. Binding affinities of these mutants are reduced by 40% and 70% for mut7 and mut8, respectively, pointing to a considerable role of the C-lobe in hexon binding (Fig. 4A).

DISCUSSION

The human lung has developed an intricate system for protecting its epithelial lining against invading pathogens (33). For species C HAdVs, the main mechanism for attachment and entry in respiratory epithelial cells uses CAR followed by secondary attachment to integrins that trigger uptake. This mechanism is well established and is likely the main port of infection in various immortalized cell lines but the relevance for the respiratory epithelium is questionable, given the hidden location of CAR deep within cellular junctions sealed off by tight junctions (19–21). Therefore, it is not surprising that alternative mechanisms for initial infections have been identified, such as using soluble components of various body fluids to gain cell entry. One such mechanism is to utilize hLF, and Lfcin, both of which are present in high concentrations within mucosal tissue (34). Naturally, hLF is a protein of the innate protection of the lung and has been shown to interfere with the infection ability of, for example, RSV (35) and HIV (36). In inflamed tissue, neutrophils have been shown to secrete hLF, which apart from its role as an antimicrobial is also regarded as the suppressor of inflammation (37). Therefore, it is very interesting that a human pathogen has evolved to highjack this system for infection. Unfortunately, very limited detailed information is available for this mechanism. In this study, we provide structural and biochemical data for the interaction between hLF and the HAdV-C5 hexon, which fills an important gap in our understanding of how HAdVs engage hLF for a CAR-independent infection of epithelial cells.

A recent study has shown that the interaction with hLF likely involves the HVR-1 of HAdV-C5 hexon, which is the longest HVR of the protein (28). In addition, all species C adenoviruses have a longer HVR-1 containing multiple negatively charged residues. Substituting HVR-1 of HAdV-C5 with a significantly shorter and less anionic HVR of another species has significantly reduced viral transduction *in vitro* (28). It has therefore been suggested that the length and charge of HVR-1 are essential for the ability to utilize lactoferrin for infection. We show here that the interaction between hLF and hexon is highly dependent on the ionic strength of the environment as buffers (20 mM HEPES) with NaCl concentrations above 5 mM disrupt the complex making it impossible to purify it by size exclusion chromatography. This was further verified by native mass spectrometry where a stable complex is present at ammonium acetate of 5–6 mM but not at 20 mM. These observations present a contradiction when compared to previous *in vitro* infection studies with HAdV-C5 (28), which demonstrated successful transduction in cell culture media with higher ionic strength. This discrepancy is not only surprising from a biological standpoint but also raises questions about the biological significance of the interaction observed.

Although the underlying molecular mechanisms facilitating LF-enhanced adenovirus transduction in higher ionic strength conditions (e.g., cell culture) remain speculative, we propose two factors that might contribute to mitigate the pronounced salt sensitivity observed with isolated, purified components. First, as binding occurs on the cell surface, it may be further stabilized by secondary interactions with cellular receptors. Second, infection studies in cell cultures employed entire virus particles, in contrast to our experiments using isolated hexon proteins. The collective binding strength (avidity) provided by 240 hexon trimers on each capsid (38) could partially offset the diminished affinity that was observed in higher ionic strength conditions (Fig. S6).

However, it is entirely possible that viruses also infect cells under low ionic strength conditions which could be utilized by viruses to increase cellular entry. In both the lungs and the eye, tissues targeted by species C adenoviruses, there are local microenvironments present in which the ionic strength is regulated (39). The epithelial cell layer of the respiratory tract, including the trachea and bronchi, is covered by a thin layer of fluid, or

mucus that ensures the proper function of the beating ciliated cells as well as protecting the epithelial cells from invading pathogens (40, 41). In this air-liquid interface, the balance of ions such as sodium, chloride, and bicarbonate is essential to preserve the characteristics of the mucins to ensure a normal mucociliary clearance (42, 43). Like in the conducting airways, the ionic strength in the tear film covering the cornea is tightly regulated to ensure proper function (44). Thus, as ions are pumped in and out of the cells, the local ionic strength is altered to a degree that may, at times, be beneficial for the hLF-mediated infection of the epithelium. Even if the concentration of sodium has been determined in the sputum to be about 100 nM (±27 nM), the concentration in the periciliary fluid is unknown and virtually impossible to determine (40).

The trimeric structure of the HAdV-C5 hexon suggests a potential interaction with an equivalent number of hLF molecules in a 3:3 stoichiometric ratio to maximize avidity. Our native MS data, in agreement with 2D class averages of the complex, however, suggest a 3:1 stoichiometry of hexon to lactoferrin within the complex. Occasionally, a second molecule of lactoferrin was found to be bound, both in native MS (Fig. 1D) and in 2D projections (Fig. S7) while a 3:3 stoichiometry was not observed. A possible explanation could be an allosteric effect of the first binding event that introduces conformational changes in hexon resulting in reduced affinities for the second and third molecules of hLF. Alternatively, with two protomers of the hexon trimer already being occupied, the binding of a third hLF might be sterically impaired.

An increasing amount of hybrid protein structures is currently being determined making use of an integrative approach combining experimental data and computational techniques (45). The structure of HAdV-C5 hexon in complex with human lactoferrin was well suited for this approach. Although we were unable to reconstruct a 3D model from our cryo-EM data alone, 2D class averages clearly showed assembly of the complex and identity of its components. Combining crosslinking data and refinement of the docked complex against the experimental 2D projections resulted in a likely model of the interaction. We believe that the workflow described here could also serve as a template to generate hybrid models of other difficult to obtain structures. While our hybrid model realistically depicts the overall assembly of the complex, it insufficiently describes direct contacts between interface residues in both structures. This is due to the inherent low resolution of the 2D class averages which only allow for rigid body refinement in addition to the high flexibility of hexon's hypervariable loops in the hexon. These are typically largely disordered and therefore cannot be modeled or computationally predicted with accuracy. Furthermore, our hybrid model does not sufficiently take into account conformational changes that occur in both molecules upon binding. These result in small, observable disagreements between experimentally obtained 2D class averages and calculated 2D projections of the complex.

Further molecular insight was therefore obtained through site-directed mutagenesis of selected hLF residues present at its interface with hexon. Hereby, we could confirm that a stretch of basic residues of hLF between amino acid positions 24–30 within the Lfcin region shows the highest contribution to the overall binding affinity. These residues appear to face the HVR-1 of hexon in the proposed model. This finding is in agreement with the earlier observation that an LFcin peptide comprising residues 1–18 did not increase HAdV-C5 transduction (28). Likewise, bovine LF (bLF), which carries substitutions in the mutated human lactoferrin residues, lacked transduction-enhancing activity (28) (Fig. S8).

The described region forms a continuum with an adjacent patch of basic residues outside the Lfcin region which also contribute significantly to the binding affinity, or the lack thereof when mutated to Ala or Ser. It therefore seems that the region in the N-lobe of hLF that interacts with hexon HVR-1 might not be limited to the Lfcin peptide but comprises a larger area.

Interestingly, a possible second attachment point between hLF and hexon was observed in 2D class averages which appeared to involve the region around the junction of N- and C-lobe of hLF. The presence of such a site was corroborated by

the identification of a crosslink connecting hLF K676 (present at the junction of N- and C-lobe) with hexon S174 (present in the "turrets"). Based on the electrostatic potential distribution in both molecules the interaction between the C-lobe of hLF and a second hexon protomer seems to be less charge dependent (Fig. 4E). Mutational analysis also indicated a significant but lesser contribution to overall affinity within the complex. It is therefore possible that this second contact site is less crucial for overall complex stability but rather for proper orientation of LF toward its cellular receptor.

Taken together, our data provide novel structural insight into the interaction between HAdV-C5 and hLF. Combining computational modeling with cryo-EM, XL-MS, and biophysical characterization, our hybrid model agrees well with mutational data, realistically depicting the interaction between HAdV-5C hexon and hLF. In the future, our results may be used for the generation of Lfcin-based antivirals that block the interaction between HVR-1 and hLF. In addition, novel Lfcin-based adjuvants may be developed that improve transduction efficiency in the lung during gene therapy.

MATERIALS AND METHODS

Cloning and protein purification

Human lactoferrin (hLF)

A commercially synthesized DNA fragment (Genewiz) encoding hLF (UniProt: Q19KS2) was cloned into the pHLsec vector (46) using the Gibson assembly method (New England Biolabs). Eight hLF mutants were produced by introducing desired point mutations during the synthesis of DNA fragments (GenScript). The plasmids encoding the wild-type (WT) LF, and its mutants were transfected and transiently expressed in Expi293F cells (ThermoFisher Scientific) following the manufacturer's protocol.

LF and its mutants were purified from the culture supernatant obtained after centrifugation. After dialysis against 20 mM HEPES pH 7.2, 30 mM NaCl (buffer A), the supernatant was loaded onto a Mono S cation exchange chromatography column (GE Healthcare) equilibrated with buffer A using an ÄKTA Pure system (GE Healthcare). LF and its mutants were eluted by a linear NaCl gradient using buffer A and 20 mM HEPES pH 7.2, 1 M NaCl (buffer B) and subsequently analyzed by SDS-PAGE. Desired fractions were incubated with 0.5 mM $FeCl_3$ in the presence of 200 mM $NaHCO_3$ for 1 h at room temperature and further purified SEC on a S200 16/600 column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.2, 150 mM NaCl (buffer C). The purified LF was concentrated using Amicon Ultra centrifugal filters (Merck Millipore) and stored at −80°C until further use.

Hexon

Hexon was purified from virus particles propagated in A549 cells (ATCC CCL-185). A549 cells were maintained in DMEM (Sigma) containing 5% (vol/vol) fetal bovine serum (FBS; Gibco), and 1% (vol/vol) penicillin-streptomycin mix (Gibco) at 37°C under 5% CO_2. Cells from a confluent culture were seeded into 5×175 cm^2 flasks. Upon reaching confluence, the medium was removed and fresh 5 mL medium containing 2% (vol/vol) FBS was added before infecting with HAdV-5C (200 µL). The flasks were incubated at 37°C for 90 min and gently mixed every 15 min. Later, the medium was replaced with fresh 30 mL medium containing 2% (vol/vol) fetal bovine serum and flasks were left undisturbed for 4 days. Cells were harvested by tapping the flasks on the bench followed by centrifugation of the culture medium at 5,000 g for 10 min. Prior to centrifugation, 1 mL culture from a flask was collected to be used as infectious material for the next round of hexon purification. A cell pellet from 5×175 cm^2 was re-suspended in 5 mL 20 mM HEPES, 5 mM NaCl pH 7.2 (buffer D). To disrupt the cells, three freeze-thaw (37°C/−80°C) cycles were done and then 6 mL of Vertrel XF was added (1,1,1,2,3,4,4,5,5,5-decafluoropentane; Sigma-Aldrich). The cell suspension was mixed by gentle shaking of tubes by hands for 3 min followed by centrifugation at 3,000 g for 10 min. The contents of the clear

top layer above the Vertrel phase were collected and separated over a CsCl gradient ultracentrifuged using the SW41 rotor (Beckman Coulter) at 25,000 rpm, 4°C for 90 min. The clear top phase obtained above a white band was collected and separated over a 5 mL HiTrap Q FF column (GE Healthcare) equilibrated with buffer A. Hexon was eluted with a linear NaCl gradient using buffers A and B. The elution fractions were analyzed by SDS-PAGE, concentrated, and further purified by SEC on S200 16/600 column (GE Healthcare) equilibrated with buffer C. The purified hexon was stored at −80°C until further use.

Purification of hexon:hLF complex

Both hexon and hLF were dialyzed against buffer D before the preparation of the complex. A twofold molar excess of LF was mixed with 925 pmoles of hexon and incubated on ice for 5 min followed by SEC on Superose 6 increase 10/300 (GE Health-care) equilibrated with buffer D. Peak fractions were analyzed by SDS-PAGE to confirm the co-elution of hexon and hLF. The fractions containing the complex were concentrated to 1 mg/mL using a Vivaspin centrifugal concentrator (Sartorius).

Loop prediction using AlphaFold

The structure of the hexon was predicted using AlphaFold2 (47). Using PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) and Coot (48), the loop regions 137–164, 188–190, 251–256, 271–278, and 431–436 from the hexon AlphaFold2 model were appended to the cryo-EM model lacking these regions.

Native MS

Prior to native MS measurements, hexon:hLf complex purified in buffer D was buffer exchanged into 20 mM aqueous ammonium acetate solution pH 7.0 (AA; ≥99.99% trace metals basis, Honeywell Research Chemicals) using two cycles of spin gel-filtration (MicroBioSpin P-6, Bio-Rad). The complex was subsequently diluted by 20 mM AA to 0.5 µM (estimated based on 3 + 3 stoichiometry). Alternatively, for low ionic strength conditions, the dilution was performed with LC-MS grade water instead, resulting in a 5.85 mM AA concentration. Reference spectra of isolated hexon and hLF were obtained from 5 µM (monomer) samples in 150 mM AA pH 7.0.

After short equilibration on ice, the samples were introduced into an Orbitrap Q Exactive UHMR mass spectrometer (Thermo Scientific) *via* static nanoelectrospray ionization. The samples were electrosprayed from gold-coated borosilicate glass capillaries Kwik-Fil 1B120F-4 (World Precision Instruments) prepared in-house essentially as described before (49, 50). The mass spectrometer was tuned for best-signal quality and intensity while maintaining minimal ion activation. Specifically, electrospray voltage was kept at 1.25 kV, source desolvation temperature 120°C, in-source trapping −50 V, ion transfer profile "high m/z," analyzer profile "low m/z," analyzer target resolution 1,500 acquiring in a mass range 1,000–20,000 m/z, averaging five microscans. Further desolvation and collisional cooling in an HCD cell were performed with nitrogen at a relative gas pressure setting of 8.0 with gentle collisional activation by 10 V HCD voltage gradient. Raw spectra were averaged over at least 50 scans in Thermo Xcalibur Qual Browser 4.2.47 (Thermo Scientific) and mass deconvoluted in the UniDec 6.0.3 package (51).

Cross-linking MS

Hexon:hLF complex was crosslinked using the homobifunctional crosslinker disuccinimidyl dibutyric urea (DSBU; CF Plus Chemicals). The complex was mixed with the excess of DSBU at a 1:500 molar ratio. The reaction mixture was then incubated at room temperature for 1 h and quenched with Tris at 30× higher concentration than DSBU.

Samples were diluted with 50 µL of 50 mM ammonium bicarbonate. Proteins were reduced by 10 mM TCEP and alkylated by 30 mM iodoacetamide. For deglycosylation of

hLF, PNGase F (New England Biolabs) was added to the sample in a 1:20 (wt/wt) ratio and incubated for 4 h at 37°C. Proteins were digested by trypsin (Promega) overnight at 37°C. After the digestion, samples were off-line desalted using a Microtrap C18 cartridge (Optimize Technologies) and dried using a SpeedVac concentrator. Samples were re-suspended in 40 µL of 2% (vol/vol) acetonitrile, 0.1% (vol/vol) TFA and analyzed using the Vanquish liquid chromatography system (ThermoScientific), connected to the timsToF SCP mass spectrometer equipped with captive spray (Bruker Daltonics). The mass spectrometer was operated in a positive data-dependent mode. One microliter of the peptide mixture was injected by autosampler on the C18 trap column (PepMap Neo C18 5 µm, 300 µm x 5 mm, Thermo Scientific). After 3 min of trapping, peptides were eluted from the trap column and separated on a C18 column (DNV PepMap Neo 75 µm x 150 mm, 2 µm, Thermo Scientific) by a linear 35 min water-acetonitrile gradient from 5% (vol/vol) to 35% (vol/vol) acetonitrile at a flow rate of 350 nL/min. The trap and analytical columns were both heated to 50°C. The TIMS scan range was set between 0.6 and 1.6 V s/ cm^2 with a ramp time of 100 ms. The number of PASEF MS/MS ramps was 4. Precursor ions in the m/z range between 100 and 1,700 with charge states ≥ 1 + and ≤8 + were selected for fragmentation. The active exclusion was enabled for 0.4 min.

The raw data were processed by Data Analysis 5.3 (Bruker Daltonics). The mgf files were uploaded to MeroX 2.0.1.4 software (52). The search parameters were set as follows: enzyme—trypsin (specific), carbamidomethylation as a fixed modification, oxidation of methionine as variable modifications. Precursor precision was set at 10.0 ppm and fragment ion precision was set at 0.1 Da.

Cryo-EM sample preparation and data collection

HAdV-C5 hexon

3.5 µL of Hexon (0.7 mg/mL) was applied to freshly glow discharged Au, 300 mesh, R1.2/1.3 TEM grids (Protochips) coated in-house with a graphene monolayer. The grids were blotted at 100% humidity and 4°C (Vitrobot, ThemoFisher Scientific) and vitrified by plunging into liquid ethane. The grids were imaged using a Titan Krios G3i microscope (ThermoFisher Scientific) operated at 300 kV. Automated data collection was performed using EPU software (ThermoFisher Scientific) software. In all, 1,063 movies (40 frames/ movie; 5-s exposure) were collected using a Quantum K2 LS detector (Gatan) at a magnification of 165,000×, resulting in a pixel size of 0.828 Å. The nominal defocus during data collection ranged from 1.2 µm to 2.5 µm and the total dose used was 55 e⁻/Å^2.

HAdV-C5 hexon:hLF

Grids for cryo-EM were prepared by applying a freshly purified batch of hexon:hLF complex to Au, 300 mesh, R1.2/1.3 TEM grids (Quantifoil). An amount of 3 µL of Hexon-hLF complex (1 mg/mL) was applied to freshly glow discharged grids that were blotted at 100% humidity and 4°C (Vitrobot, ThemoFisher Scientific) and vitrified by plunging into liquid ethane. For imaging, the grids were transferred to a Talos Arctica microscope (ThermoFisher Scientific) equipped with a Falcon 3 detector (ThermoFisher Scientific). In total, 680 movies (60 frames/movies) were automatically collected using EPU software (ThermoFisher Scientific). The resulting pixel size was 1.23 Å and the nominal defocus ranged from 2 to 3.2 µm.

Cryo-EM image processing

HAdV-C5 hexon

The recorded movies were entirely processed using cryoSPARC (version 3.2 and later) (53). Motion correction to obtain averaged micrographs from multi-frame movies followed by CTF estimation was performed using Patch Motion Correction, and CTFFIND4 algorithms integrated in cryoSPARC (53, 54). Hexon particles were picked

using the Gaussian Blob Picker in cryoSPARC. Particles were extracted with a box size of 300 px and subjected to iterative, reference-free 2D classification until no ill-defined class averages were obtained. The final selected subset of 98,640 particles was used to produce three *ab initio* models. The class containing the maximum number of particles (53,996) closely resembled the hexon and was chosen for homogeneous refinement with C3 symmetry imposed to yield the reconstructed hexon.

HAdV-C5 hexon:hLF

The recorded movies were imported into cryoSPARC (version 3.2 and later) and averaged micrographs from multi-frame movies were obtained *via* the Patch Motion Correction algorithm. Micrographs were CTF corrected using the Patch CTF Estimation algorithm. A Blob Picker was employed to pick particles which were extracted with a box size of 230 px. The set of extracted particles was cleared of ill-defined particles by iterative, reference-free 2D classification to obtain the final set of class averages, used for refinement of the HAdV-C5 hexon:hLF hybrid structure.

Model building HAdV-C5 hexon cryo-EM structure

To generate a model for hexon using the cryo-EM reconstruction, the crystal structure of hexon (PDB: 3TG7) was docked into the cryo-EM coloumb potential map using the dock_in_map procedure of Phenix (55). This docked model was subjected to automatic real-space refinement by Phenix. The obtained model was iteratively refined using multiple rounds of manual refinement performed in Coot followed by another round of automatic refinement in Phenix. After each round of refinement, the model was validated using MolProbity (56) until satisfactory parameters were achieved.

Molecular dynamics simulation HAdV-C5 hexon structure

System preparation

The model of hexon was trimmed down for more efficient calculation by discarding the rigid base, that is, only residues 129–315 and 415–460 of all three protomers were retained in the final model. Hydrogens were added using the tLEaP program of the AMBER20 suite (57). Titratable residues were modeled in their standard state at pH = 7.4. Protein force field ff19SB (58) was used along with a truncated octahedral box of SPC/E water molecules (59) extending 12 Å from the solute. Sodium and chloride counter ions were added. Prior to molecular dynamics (MD) simulation, the systems were relaxed, heated to 300 K, and equilibrated at 300 K and 1 atm pressure using the published protocol (60).

Molecular dynamics

The production run was 600 ns long in a NpT ensemble, that is, keeping the number of particles (N), pressure (p), and temperature (T) constant. The SHAKE algorithm (61) was used to restrain all bond vibrations and hydrogen mass repartitioning to 3 Da (62) allowed us to apply a time step of 4 fs. Residues adjacent to those of the discarded base were fixed with harmonic restraint of 500 kcal mol^{-1} Å^{-2} to maintain the conformations compatible with linking to the base. All the analyses were done in the CPPTRAJ program (63).

Docking of HAdV-C5 hexon:hLF complex and refinement of the docked model against 2D class averages

All simulations were performed in XPLOR-NIH v 3.4 (32) starting from the MD structure of Hexon and the structure of hLF (PDB: 1LFG). Protons and missing side chain atoms in both proteins were added in XPLOR-NIH, followed by minimization of the energy function consisting of the standard geometric (bonds, angles, dihedrals, and impropers) and steric (van der Waals) terms. Based on the published analysis (64), the XL-MS data

were converted into pairwise S446 (Hexon)—K282 (hLF) and S174 (Hexon)—K676 (hLF) intermolecular distance restraints, defined as the upper limit bounds of 24 Å between the Cα atoms and 15 Å bounds between the corresponding Ser OG and Lys NZ side chain atoms. The cryo-EM 2D class averages imported into XPLOR-NIH as two-dimensional coloumb potential maps (31, 65) were subsequently used to fit the protein molecules.

Following the computational protocol of Clore and co-workers (31), the refinement of the hexon:hLF complex was carried out in three steps. First, we performed a rigid-body protein-protein docking, driven by the intermolecular distance restraints. To this end, the positions of the hexon atoms were kept fixed, except for the "turret" loops (residues 136–164, 173–217, 268–287, and 429–448 in each subunit of the hexon homo-trimer, which includes the crosslink residues S174 and S446), which were given full torsional degrees of freedom. The incoming hLF molecule was treated as a rigid-body group, except for the crosslink residues K282 and K676, whose side chains were allowed to move. Starting from randomly oriented proteins separated by 100 Å, the docking comprised an initial rigid-body simulated annealing step, followed by the full side-chain energy minimization (32). In addition to the energy terms mentioned above, the total energy function included the intermolecular restraint term and a knowledge-based dihedral angle potential (32). In this first step, 100 structures were calculated, and the 10 lowest-energy solutions—showing no distance restraint violations, imposed by crosslinks—were retained for the subsequent refinement.

In the second step, energy terms for the 2D projections were added sequentially, with the map centering and (initially random) orientation optimized for each image from the cryo-EM 2D class stack. The back-calculated projections were computed as a sum over all heavy atoms, where each atom contributed a 2D quartic density according to its van der Waals radius and weighted by the atom's number of electrons (31). The Gaussian blurring was employed throughout, with Gaussian diameter given by the resolution of the 2D class, and the agreement between the experimental and back-calculated electron density projections was estimated by the cross-correlation C metric (31). Out of 10 slices in the 2D classes image stack, one consistently failed to yield a good fit (C < 0.9) and, thus, was discarded in the subsequent refinement rounds.

In the third and final step, the hexon:hLF complex was refined by allowing full rigid-body and torsional degrees of freedom in the motional setup described above, with simultaneous minimization of all constituent energy terms, including those for the intermolecular restraints and the nine individual electron density projections (31). In this run, 100 structures were calculated and, 10 lowest-energy solutions—showing the best agreement with the experimental cryo-EM data (highest C values) and no distance restraint violations—were retained for the final analysis. Individual back-calculated 2D projection images and their overlays with the experimental 2D class densities were output for visualization.

Microscale thermophoresis

Microscale thermophoresis (MST) analysis was carried out to determine the binding affinity between hexon and hLF using Monolith NT.115 (Nanotemper Technologies). Hexon was labeled with AlexaFluor594 (ThermoFisher Scientific) according to the manufacturer's instructions. Excess dye was removed by SEC on S200 16/600 column (GE Healthcare) equilibrated with buffer C. MST experiments were performed in buffer D containing 0.05% (vol/vol) Tween 20. A range of 16 different concentrations of hLF or hLF mutants was generated by serial dilution (1:1) while the same amount of labeled hexon was added to each dilution. Following 5-min incubation at 25°C, the samples were taken up in capillaries for MST measurements that were performed at 25°C at 50% LED power, and 40% MST power. The MST curves were fitted using MO.Affinity analysis software (Nanotemper Technologies).

ACKNOWLEDGMENTS

AUTHOR AFFILIATIONS

[1]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic
[2]Swedish Veterinary Agency, Uppsala, Sweden
[3]VIB-VUB Center for Structural Biology, Flemish Institute of Biotechnology (VIB), Brussels, Belgium
[4]Jean Jeener NMR Centre, Vrije Universiteit Brussel (VUB), Brussels, Belgium
[5]Faculty of Science, Charles University, Prague, Czech Republic
[6]Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic
[7]Leibniz Institute of Virology (LIV), Hamburg, Germany
[8]Biotechnology and Biomedical Center of the Academy of Sciences and Charles University in Vestec, Vestec, Czech Republic
[9]First Faculty of Medicine, Charles University, Prague, Czech Republic
[10]Department of Clinical Microbiology, Umeå University, Umeå, Sweden
[11]Department of Health Sciences and Biomedicine, Faculty V: School of Life Sciences, CSSB Centre for Structural Systems Biology, Deutsches Elektronen Synchrotron DESY and Leibniz Institute of Virology, Hamburg, University of Siegen, Siegen, Germany

AUTHOR ORCIDs

Sebastian Zoll http://orcid.org/0000-0001-5376-9698

FUNDING

AUTHOR CONTRIBUTIONS

Arun Dhillon, Formal analysis, Methodology, Visualization | B. David Persson, Writing – original draft | Alexander N. Volkov, Formal analysis, Methodology, Writing – review and editing | Hagen Sülzen, Methodology, Validation, Visualization, Writing – review and editing | Alan Kádek, Formal analysis, Methodology, Writing – review and editing | Petr Pompach, Formal analysis, Methodology, Writing – review and editing | Sami Kereïche, Methodology | Martin Lepšík, Methodology, Writing – review and editing | Katarina

Danskog, Methodology | Charlotte Uetrecht, Supervision | Niklas Arnberg, Conceptualization, Writing – review and editing | Sebastian Zoll, Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The model of the HAdV-C5 hexon has been deposited in the Protein Data Bank under accession code 8Q7C (https://www.rcsb.org/structure/8Q7C; https://doi.org/10.2210/pdb8Q7C/pdb). The associated Coulomb potential map has been deposited in the Electron Microscopy Data Bank under accession code EMD-18212 (https://www.ebi.ac.uk/emdb/EMD-18212). The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE (66) partner repository with the data set identifier PXD045900 (https://www.ebi.ac.uk/pride/archive/projects/PXD045900). Thirty snapshots from the MD trajectory used in integrative modeling are available in the Mendeley repository (10.17632/2xrxbm8g6j.1). The native MS data have been deposited in ZENODO (10.5281/zenodo.10534068).

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental material (JVI01576-23-S0001.docx). Figures S1 to S8; Table S1.

REFERENCES

1. Greber UF, Flatt JW. 2019. Adenovirus entry: from infection to immunity. Annu Rev Virol 6:177–197. https://doi.org/10.1146/annurev-virology-092818-015550
2. Benkő M, Aoki K, Arnberg N, Davison AJ, Echavarría M, Hess M, Jones MS, Kaján GL, Kajon AE, Mittal SK, Podgorski II, San Martín C, Wadell G, Watanabe H, Harrach B. 2022. ICTV virus taxonomy profile: adenoviridae 2022. J Gen Virol 103. https://doi.org/10.1099/jgv.0.001721
3. Echavarría M. 2008. Adenoviruses in immunocompromised hosts. Clin Microbiol Rev 21:704–715. https://doi.org/10.1128/CMR.00052-07
4. Proenca-Modena JL, de Souza Cardoso R, Criado MF, Milanez GP, de Souza WM, Parise PL, Bertol JW, de Jesus BLS, Prates MCM, Silva ML, Buzatto GP, Demarco RC, Valera FCP, Tamashiro E, Anselmo-Lima WT, Arruda E. 2019. Human adenovirus replication and persistence in hypertrophic adenoids and palatine tonsils in children. J Med Virol 91:1250–1262. https://doi.org/10.1002/jmv.25441
5. Lion T. 2019. Adenovirus persistence, reactivation, and clinical management. FEBS Lett 593:3571–3582. https://doi.org/10.1002/1873-3468.13576
6. Ohnishi Y, Noda S. 1976. [Studies on the rabbit conjunctival cells *in vitro* (author's transl)]. Nippon Ganka Gakkai Zasshi 80:1255–1263.
7. Berk AJ, Sharp PA. 1978. Structure of the adenovirus 2 early mRNAs. Cell 14:695–711. https://doi.org/10.1016/0092-8674(78)90252-0
8. Cepko CL, Sharp PA. 1982. Assembly of adenovirus major capsid protein is mediated by a nonvirion protein. Cell 31:407–415. https://doi.org/10.1016/0092-8674(82)90134-9
9. Lee A. 2023. Nadofaragene firadenovec: first approval. Drugs 83:353–357. https://doi.org/10.1007/s40265-023-01846-z
10. Sumida SM, Truitt DM, Kishko MG, Arthur JC, Jackson SS, Gorgone DA, Lifton MA, Koudstaal W, Pau MG, Kostense S, Havenga MJE, Goudsmit J, Letvin NL, Barouch DH. 2004. Neutralizing antibodies and CD8+ T lymphocytes both contribute to immunity to adenovirus serotype 5 vaccine vectors. J Virol 78:2666–2673. https://doi.org/10.1128/jvi.78.6.2666-2673.2004
11. Tillman BW, de Gruijl TD, Luykx-de Bakker SA, Scheper RJ, Pinedo HM, Curiel TJ, Gerritsen WR, Curiel DT. 1999. Maturation of dendritic cells accompanies high-efficiency gene transfer by a CD40-targeted adenoviral vector. J Immunol 162:6378–6383.
12. Milligan ID, Gibani MM, Sewell R, Clutterbuck EA, Campbell D, Plested E, Nuthall E, Voysey M, Silva-Reyes L, McElrath MJ, et al. 2016. Safety and immunogenicity of novel adenovirus type 26- and modified vaccinia ankara-vectored Ebola vaccines: a randomized clinical trial. JAMA 315:1610–1623. https://doi.org/10.1001/jama.2016.4218
13. Barouch DH, Tomaka FL, Wegmann F, Stieh DJ, Alter G, Robb ML, Michael NL, Peter L, Nkolola JP, Borducchi EN, et al. 2018. Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-blind, placebo-controlled, phase 1/2A clinical trial (APPROACH) and in rhesus monkeys (NHP 13-19). Lancet 392:232–243. https://doi.org/10.1016/S0140-6736(18)31364-3
14. Larocca RA, Mendes EA, Abbink P, Peterson RL, Martinot AJ, Iampietro MJ, Kang ZH, Aid M, Kirilova M, Jacob-Dolan C, Tostanoski L, Borducchi EN, De La Barrera RA, Barouch DH. 2019. Adenovirus vector-based vaccines confer maternal-fetal protection against zika virus challenge in pregnant IFN-alphabetaR(-/-) mice. Cell Host Microbe 26:591–600. https://doi.org/10.1016/j.chom.2019.04.001
15. Williams K, Bastian AR, Feldman RA, Omoruyi E, de Paepe E, Hendriks J, van Zeeburg H, Godeaux O, Langedijk JPM, Schuitemaker H, Sadoff J, Callendret B. 2020. Phase 1 safety and Immunogenicity study of a respiratory syncytial virus vaccine with an adenovirus 26 vector encoding prefusion F (Ad26.RSV.preF) in adults aged >/=60 years. J Infect Dis 222:979–988. https://doi.org/10.1093/infdis/jiaa193
16. Sadoff J, Gray G, Vandebosch A, Cárdenas V, Shukarev G, Grinsztejn B, Goepfert PA, Truyers C, Fennema H, Spiessens B, et al. 2021. Safety and efficacy of single-dose Ad26.COV2.S vaccine against COVID-19. N Engl J Med 384:2187–2201. https://doi.org/10.1056/NEJMoa2101544
17. Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, Hong JS, Horwitz MS, Crowell RL, Finberg RW. 1997. Isolation of a common receptor for coxsackie B viruses and adenoviruses 2 and 5. Science 275:1320–1323. https://doi.org/10.1126/science.275.5304.1320
18. Storm RJ, Persson BD, Skalman LN, Frangsmyr L, Lindstrom M, Rankin G, Lundmark R, Domellof FP, Arnberg N. 2017. Human adenovirus type 37 uses alpha(V)beta(1) and alpha(3)beta(1) integrins for infection of human corneal cells. J Virol 91
19. Mateo M, Generous A, Sinn PL, Cattaneo R. 2015. Connections matter—how viruses use cell-cell adhesion components. J Cell Sci 128:431–439. https://doi.org/10.1242/jcs.159400

20. Walters RW, Freimuth P, Moninger TO, Ganske I, Zabner J, Welsh MJ. 2002. Adenovirus fiber disrupts CAR-mediated intercellular adhesion allowing virus escape. Cell 110:789–799. https://doi.org/10.1016/s0092-8674(02)00912-1

21. Walters RW, Grunst T, Bergelson JM, Finberg RW, Welsh MJ, Zabner J. 1999. Basolateral localization of fiber receptors limits adenovirus infection from the apical surface of airway epithelia. J Biol Chem 274:10219–10226. https://doi.org/10.1074/jbc.274.15.10219

22. Kotha PLN, Sharma P, Kolawole AO, Yan R, Alghamri MS, Brockman TL, Gomez-Cambronero J, Excoffon K. 2015. Adenovirus entry from the apical surface of polarized epithelia is facilitated by the host innate immune response. PLoS Pathog 11:e1004696. https://doi.org/10.1371/journal.ppat.1004696

23. Waddington SN, McVey JH, Bhella D, Parker AL, Barker K, Atoda H, Pink R, Buckley SMK, Greig JA, Denby L, Custers J, Morita T, Francischetti IMB, Monteiro RQ, Barouch DH, van Rooijen N, Napoli C, Havenga MJE, Nicklin SA, Baker AH. 2008. Adenovirus serotype 5 hexon mediates liver gene transfer. Cell 132:397–409. https://doi.org/10.1016/j.cell.2008.01.016

24. Lyle C, McCormick F. 2010. Integrin alphavbeta5 is a primary receptor for adenovirus in CAR-negative cells. Virol J 7:148. https://doi.org/10.1186/1743-422X-7-148

25. Chéneau C, Eichholz K, Tran TH, Tran TTP, Paris O, Henriquet C, Bajramovic JJ, Pugniere M, Kremer EJ. 2021. Lactoferrin retargets human adenoviruses to TLR4 to induce an abortive NLRP3-associated pyroptotic response in human phagocytes. Front Immunol 12:685218. https://doi.org/10.3389/fimmu.2021.685218

26. Bellamy W, Takase M, Yamauchi K, Wakabayashi H, Kawase K, Tomita M. 1992. Identification of the bactericidal domain of lactoferrin. Biochim Biophys Acta 1121:130–136. https://doi.org/10.1016/0167-4838(92)90346-f

27. Johansson C, Jonsson M, Marttila M, Persson D, Fan X-L, Skog J, Frängsmyr L, Wadell G, Arnberg N. 2007. Adenoviruses use lactoferrin as a bridge for CAR-independent binding to and infection of epithelial cells. J Virol 81:954–963. https://doi.org/10.1128/JVI.01995-06

28. Persson BD, Lenman A, Frängsmyr L, Schmid M, Ahlm C, Plückthun A, Jenssen H, Arnberg N. 2020. Lactoferrin-hexon interactions mediate CAR-independent adenovirus infection of human respiratory cells. J Virol 94:e00542-20. https://doi.org/10.1128/JVI.00542-20

29. Jonsson MI, Lenman AE, Frängsmyr L, Nyberg C, Abdullahi M, Arnberg N. 2009. Coagulation factors IX and X enhance binding and infection of adenovirus types 5 and 31 in human epithelial cells. J Virol 83:3816–3825. https://doi.org/10.1128/JVI.02562-08

30. Lenman A, Müller S, Nygren MI, Frängsmyr L, Stehle T, Arnberg N. 2011. Coagulation factor IX mediates serotype-specific binding of species A adenoviruses to host cells. J Virol 85:13420–13431. https://doi.org/10.1128/JVI.06088-11

31. Wälti MA, Canagarajah B, Schwieters CD, Clore GM. 2021. Visualization of sparsely-populated lower-order oligomeric states of human mitochondrial Hsp60 by cryo-electron microscopy. J Mol Biol 433:167322. https://doi.org/10.1016/j.jmb.2021.167322

32. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. 2003. The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73. https://doi.org/10.1016/s1090-7807(02)00014-9

33. Martin TR, Frevert CW. 2005. Innate immunity in the lungs. Proc Am Thorac Soc 2:403–411. https://doi.org/10.1513/pats.200508-090JS

34. Rosa L, Cutone A, Lepanto MS, Paesano R, Valenti P. 2017. Lactoferrin: a natural glycoprotein involved in iron and inflammatory homeostasis. Int J Mol Sci 18:1985. https://doi.org/10.3390/ijms18091985

35. Sano H, Nagai K, Tsutsumi H, Kuroki Y. 2003. Lactoferrin and surfactant protein A exhibit distinct binding specificity to F protein and differently modulate respiratory syncytial virus infection. Eur J Immunol 33:2894–2902. https://doi.org/10.1002/eji.200324218

36. Berlutti F, Pantanella F, Natalizi T, Frioni A, Paesano R, Polimeni A, Valenti P. 2011. Antiviral properties of lactoferrin–a natural immunity molecule. Molecules 16:6992–7018. https://doi.org/10.3390/molecules16086992

37. Lönnerdal B, Iyer S. 1995. Lactoferrin: molecular structure and biological function. Annu Rev Nutr 15:93–110. https://doi.org/10.1146/annurev.nu.15.070195.000521

38. Gallardo J, Pérez-Illana M, Martín-González N, San Martín C. 2021. Adenovirus structure: what is new. Int J Mol Sci 22:5240. https://doi.org/10.3390/ijms22105240

39. Toczyłowska-Mamińska R, Dołowy K. 2012. Ion transporting proteins of human bronchial epithelium. J Cell Biochem 113:426–432. https://doi.org/10.1002/jcb.23393

40. Luk CK, Dulfano MJ. 1983. Effect of pH, viscosity and ionic-strength changes on ciliary beating frequency of human bronchial explants. Clin Sci (Lond) 64:449–451. https://doi.org/10.1042/cs0640449

41. Tabary O, Muselet C, Miesch MC, Yvin JC, Clément A, Jacquot J. 2003. Reduction of chemokine IL-8 and RANTES expression in human bronchial epithelial cells by a sea-water derived saline through inhibited nuclear factor-kappaB activation. Biochem Biophys Res Commun 309:310–316. https://doi.org/10.1016/j.bbrc.2003.07.006

42. Fajac I, Burgel PR. 2023. Cystic fibrosis. Presse Med 52:104169. https://doi.org/10.1016/j.lpm.2023.104169

43. Saint-Criq V, Gray MA. 2017. Role of CFTR in epithelial physiology. Cell Mol Life Sci 74:93–115. https://doi.org/10.1007/s00018-016-2391-y

44. Hosoya K, Lee VHL, Kim K-J. 2005. Roles of the conjunctiva in ocular drug delivery: a review of conjunctival transport mechanisms and their regulation. Eur J Pharm Biopharm 60:227–240. https://doi.org/10.1016/j.ejpb.2004.12.007

45. Vallat B, Webb B, Fayazi M, Voinea S, Tangmunarunkit H, Ganesan SJ, Lawson CL, Westbrook JD, Kesselman C, Sali A, Berman HM. 2021. New system for archiving integrative structures. Acta Crystallogr D Struct Biol 77:1486–1496. https://doi.org/10.1107/S2059798321010871

46. Aricescu AR, Lu W, Jones EY. 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr 62:1243–1250. https://doi.org/10.1107/S0907444906029799

47. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. Colabfold: making protein folding accessible to all. Nat Methods 19:679–682. https://doi.org/10.1038/s41592-022-01488-1

48. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493

49. Hernández H, Robinson CV. 2007. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat Protoc 2:715–726. https://doi.org/10.1038/nprot.2007.73

50. Krichel B, Bylapudi G, Schmidt C, Blanchet C, Schubert R, Brings L, Koehler M, Zenobi R, Svergun D, Lorenzen K, Madhugiri R, Ziebuhr J, Uetrecht C. 2021. Hallmarks of alpha- and betacoronavirus non-structural protein 7+8 complexes. Sci Adv 7

51. Marty MT, Baldwin AJ, Marklund EG, Hochberg GKA, Benesch JLP, Robinson CV. 2015. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem 87:4370–4376. https://doi.org/10.1021/acs.analchem.5b00140

52. Iacobucci C, Götze M, Ihling CH, Piotrowski C, Arlt C, Schäfer M, Hage C, Schmidt R, Sinz A. 2018. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the merox software for studying protein structures and protein-protein interactions. Nat Protoc 13:2864–2889. https://doi.org/10.1038/s41596-018-0068-8

53. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14:290–296. https://doi.org/10.1038/nmeth.4169

54. Rohou A, Grigorieff N. 2015. Ctffind4: fast and accurate defocus estimation from electron micrographs. J Struct Biol 192:216–221. https://doi.org/10.1016/j.jsb.2015.08.008

55. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD. 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in phenix. Acta Crystallogr D Struct Biol 75:861–877. https://doi.org/10.1107/S2059798319011471

56. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. https://doi.org/10.1107/S0907444909042073

57. D.A. Case HMA, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, Cerutti DS, Cisneros GA, Cruzeiro VWD, Darden TA, Forouzesh N, et al. 2020. Amber 2020. University of California, San Francisco.

58. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C. 2020. Ff19Sb: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J Chem Theory Comput 16:528–552. https://doi.org/10.1021/acs.jctc.9b00591

59. Berendsen HJC, Grigera JR, Straatsma TP. 1987. The missing term in effective pair potentials. J Phys Chem 91:6269–6271. https://doi.org/10.1021/j100308a038

60. Srb P, Svoboda M, Benda L, Lepšík M, Tarábek J, Šícha V, Grüner B, Grantz-Šašková K, Brynda J, Řezáčová P, Konvalinka J, Veverka V. 2019. Capturing a dynamically interacting inhibitor by paramagnetic NMR spectroscopy. Phys Chem Chem Phys 21:5661–5673. https://doi.org/10.1039/c9cp00416e

61. Ryckaert J-P, Ciccotti G, Berendsen HJC. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341. https://doi.org/10.1016/0021-9991(77)90098-5

62. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. 2015. Long-time-step molecular dynamics through hydrogen mass repartitioning. J Chem Theory Comput 11:1864–1874. https://doi.org/10.1021/ct5010406

63. Roe DR, Cheatham TE, 3rd. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 9:3084–3095. https://doi.org/10.1021/ct400341p

64. Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. 2014. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. Protein Sci 23:747–759. https://doi.org/10.1002/pro.2458

65. Schwieters CD, Bermejo GA, Clore GM. 2018. Xplor-NIH for molecular structure determination from NMR and other data sources. Protein Sci 27:26–40. https://doi.org/10.1002/pro.3248

66. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaíno JA. 2022. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res 50:D543–D552. https://doi.org/10.1093/nar/gkab1038

Bibliography

1. Kennedy, P.G.E., *Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness).* The Lancet Neurology, 2013. **12**(2): p. 186-194.

2. Radwanska, M., et al., *Salivarian Trypanosomosis: A Review of Parasites Involved, Their Global Distribution and Their Interaction With the Innate and Adaptive Mammalian Host Immune System.* Frontiers in Immunology, 2018. **9**.

3. Büscher, P., et al., *Human African trypanosomiasis.* The Lancet, 2017. **390**(10110): p. 2397-2409.

4. Steverding, D., *The history of African trypanosomiasis.* Parasites & Vectors, 2008. **1**(3).

5. Cayla, M., et al., *African trypanosomes.* Parasites & Vectors, 2019. **12**(1).

6. Franco, J.R., et al., *Epidemiology of human African trypanosomiasis.* Clin Epidemiol, 2014. **6**: p. 257-275.

7. Dutton, J.E., *Preliminary note upon a trypanosome occurring in the blood of man.* 1902.

8. Stephens, J.W.W., H.B. Fantham, and R. Ross, *On the peculiar morphology of a trypanosome from a case of sleeping sickness and the possibility of its being a new species (T. rhodesiense).* Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character, 1910. **83**(561): p. 28-33.

9. *World Health Organisation*. 2024; Available from: https://www.who.int/data/gho/data/themes/topics/human-african-trypanosomiasis.

10. Franco, J.R., et al., *The elimination of human African trypanosomiasis: Achievements in relation to WHO road map targets for 2020.* PLOS Neglected Tropical Diseases, 2022. **16**(1).

11. Aksoy, S., et al., *Human African trypanosomiasis control: Achievements and challenges.* PLOS Neglected Tropical Diseases, 2017. **11**(4).

12. Meisner, J., et al., *The effect of livestock density on Trypanosoma brucei gambiense and T. b. rhodesiense: A causal inference-based approach.* PLOS Neglected Tropical Diseases, 2022. **16**(8).

13.   Giordani, F., et al., *The animal trypanosomiases and their chemotherapy: a review.* Parasitology, 2016. **143**(14): p. 1862-1889.

14.   Abro, Z., et al., *The potential economic benefits of controlling trypanosomiasis using waterbuck repellent blend in sub-Saharan Africa.* PLOS ONE, 2021. **16**(7).

15.   Rogers, D.J. and S.E. Randolph, *A response to the aim of eradicating tsetse from Africa.* Trends in Parasitology, 2002. **18**(12): p. 534-536.

16.   Murray, M., et al., *Genetic Resistance to African Trypanosomiasis.* The Journal of Infectious Diseases, 1984. **149**(3): p. 311-319.

17.   Alfituri, O.A., et al., *To the Skin and Beyond: The Immune Response to African Trypanosomes as They Enter and Exit the Vertebrate Host.* Frontiers in Immunology, 2020. **11**.

18.   Kennedy, P.G.E. and J. Rodgers, *Clinical and Neuropathogenetic Aspects of Human African Trypanosomiasis.* Frontiers in Immunology, 2019. **10**.

19.   Kennedy, P.G.E., *The continuing problem of human African trypanosomiasis (sleeping sickness).* Annals of Neurology, 2008. **64**(2): p. 116-126.

20.   Odiit, M., F. Kansiime, and J.C. Enyaru, *Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda.* East Afr Med J, 1997. **74**(12): p. 792-5.

21.   Checchi, F., et al., *Estimates of the duration of the early and late stage of gambiense sleeping sickness.* BMC Infectious Diseases, 2008. **8**(1).

22.   MacLean, L.M., et al., *Focus–Specific Clinical Profiles in Human African Trypanosomiasis Caused by Trypanosoma brucei rhodesiense.* PLOS Neglected Tropical Diseases, 2010. **4**(12).

23.   Kato, C.D., et al., *Clinical Profiles, Disease Outcome and Co-Morbidities among T. b. rhodesiense Sleeping Sickness Patients in Uganda.* PLOS ONE, 2015. **10**(2).

24.   Venturelli, A., et al., *Current Treatments to Control African Trypanosomiasis and One Health Perspective.* Microorganisms, 2022. **10**(7).

25.   Brun, R., et al., *Human African trypanosomiasis.* The Lancet, 2010. **375**(9709): p. 148-159.

26.   Jamabo, M., et al., *Tackling Sleeping Sickness: Current and Promising Therapeutics and Treatment Strategies.* International Journal of Molecular Sciences, 2023. **24**(15).

27. Papagni, R., et al., *Human African Trypanosomiasis (sleeping sickness): Current knowledge and future challenges.* Frontiers in Tropical Diseases, 2023. **4**.

28. Babokhov, P., et al., *A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis.* Pathogens and Global Health, 2013. **107**(5): p. 242-252.

29. Lindner, A.K., et al., *New WHO guidelines for treatment of gambiense human African trypanosomiasis including fexinidazole: substantial changes for clinical practice.* The Lancet Infectious Diseases, 2020. **20**(2).

30. Kasozi, K.I., et al., *An Update on African Trypanocide Pharmaceutics and Resistance.* Frontiers in Veterinary Science, 2022. **9**.

31. P. De Koning, H., *The Drugs of Sleeping Sickness: Their Mechanisms of Action and Resistance, and a Brief History.* Tropical Medicine and Infectious Disease, 2020. **5**(1).

32. La Greca, F. and S. Magez, *Vaccination against trypanosomiasis.* Human Vaccines, 2011. **7**(11): p. 1225-1233.

33. Magez, S., et al., *The History of Anti-Trypanosome Vaccine Development Shows That Highly Immunogenic and Exposed Pathogen-Derived Antigens Are Not Necessarily Good Target Candidates: Enolase and ISG75 as Examples.* Pathogens, 2021. **10**(8).

34. Hoare, C.A. and F.G. Wallace, *Developmental Stages of Trypanosomatid Flagellates: a New Terminology.* Nature, 1966. **212**(5068): p. 1385-1386.

35. Ralston, K.S. and K.L. Hill, *The flagellum of Trypanosoma brucei: New tricks from an old dog.* International Journal for Parasitology, 2008. **38**(8): p. 869-884.

36. Field, M.C. and M. Carrington, *The trypanosome flagellar pocket.* Nature Reviews Microbiology, 2009. **7**(11): p. 775-786.

37. Cavalcanti, D.P. and W. de Souza, *The Kinetoplast of Trypanosomatids: From Early Studies of Electron Microscopy to Recent Advances in Atomic Force Microscopy.* Scanning, 2018.

38. Rodrigues, J.C.F., J.L.P. Godinho, and W. de Souza, *Biology of Human Pathogenic Trypanosomatids: Epidemiology, Lifecycle and Ultrastructure*, in *Proteins and Proteomics of Leishmania and Trypanosoma*, A.L.S. Santos, et al., Editors. 2014, Springer Netherlands: Dordrecht. p. 1-42.

39. Michels, P.A.M., et al., *Metabolic functions of glycosomes in trypanosomatids.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2006. **1763**(12): p. 1463-1477.

40. Sommer, J.M., et al., *In vivo import of firefly luciferase into the glycosomes of Trypanosoma brucei and mutational analysis of the C-terminal targeting signal.* Molecular Biology of the Cell, 1992. **3**(7): p. 749-759.

41. Rotureau, B. and J. Van Den Abbeele, *Through the dark continent: African trypanosome development in the tsetse fly.* Frontiers in Cellular and Infection Microbiology, 2013. **3**.

42. Rotureau, B., et al., *A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly.* Development, 2012. **139**(10): p. 1842-1850.

43. Vassella, E., et al., *Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway.* Journal of Cell Science, 1997. **110**(21): p. 2661-2671.

44. Mony, B.M. and K.R. Matthews, *Assembling the components of the quorum sensing pathway in African trypanosomes.* Molecular Microbiology, 2015. **96**(2): p. 220-232.

45. Global Health, D.o.P.D.a.M. 2024; Available from: www.cdc.gov/dpdx.

46. Dyer, N.A., et al., *Flying tryps: survival and maturation of trypanosomes in tsetse flies.* Trends in Parasitology, 2013. **29**(4): p. 188-196.

47. Ponte-Sucre, A., *An Overview of Trypanosoma brucei Infections: An Intense Host–Parasite Interaction.* Frontiers in Microbiology, 2016. **7**.

48. Onyilagha, C. and J.E. Uzonna, *Host Immune Responses and Immune Evasion Strategies in African Trypanosomiasis.* Frontiers in Immunology, 2019. **10**.

49. Stijlemans, B., et al., *Immune Evasion Strategies of Trypanosoma brucei within the Mammalian Host: Progression to Pathogenicity.* Frontiers in Immunology, 2016. **7**.

50. Machado, H., et al., *Trypanosoma brucei triggers a broad immune response in the adipose tissue.* PLOS Pathogens, 2021. **17**(9).

51. Raper, J., V. Nussenzweig, and S. Tomlinson, *The main lytic factor of Trypanosoma brucei brucei in normal human serum is not high density lipoprotein.* Journal of Experimental Medicine, 1996. **183**(3): p. 1023-1029.

52.   Pays, E. and B. Vanhollebeke, *Mutual self-defence: the trypanolytic factor story.* Microbes and Infection, 2008. **10**(9): p. 985-989.

53.   Raper, J., et al., *Characterization of a Novel Trypanosome Lytic Factor from Human Serum.* Infection and Immunity, 1999. **67**(4): p. 1910-1916.

54.   Vanhamme, L., et al., *Apolipoprotein L-I is the trypanosome lytic factor of human serum.* Nature, 2003. **422**(6927): p. 83-87.

55.   Pérez-Morga, D., et al., *Apolipoprotein L-I Promotes Trypanosome Lysis by Forming Pores in Lysosomal Membranes.* Science, 2005. **309**(5733): p. 469-472.

56.   Wheeler, R.J., *The trypanolytic factor–mechanism, impacts and applications.* Trends in Parasitology, 2010. **26**(9): p. 457-464.

57.   del Pilar Molina-Portela, M., et al., *Trypanosome lytic factor, a subclass of high-density lipoprotein, forms cation-selective pores in membranes.* Molecular and Biochemical Parasitology, 2005. **144**(2): p. 218-226.

58.   Thomson, R., M. Samanovic, and J. Raper, *Activity of trypanosome lytic factor: a novel component of innate immunity.* Future Microbiology, 2009. **4**(7): p. 789-796.

59.   Vanhollebeke, B., et al., *A Haptoglobin-Hemoglobin Receptor Conveys Innate Immunity to Trypanosoma brucei in Humans.* Science, 2008. **320**(5876): p. 677-681.

60.   Lane-Serff, H., et al., *Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor.* eLife, 2014. **3**.

61.   Lecordier, L., et al., *Identification of Trypanosoma brucei components involved in trypanolysis by normal human serum.* Molecular Microbiology, 2014. **94**(3): p. 625-636.

62.   Verdi, J., et al., *Inducible Germline IgMs Bridge Trypanosome Lytic Factor Assembly and Parasite Recognition.* Cell Host & Microbe, 2020. **28**(1): p. 79-88.

63.   De Greef, C., et al., *A gene expressed only in serum-resistant variants of Trypanosoma brucei rhodesiense.* Molecular and Biochemical Parasitology, 1989. **36**(2): p. 169-176.

64.   Rudenko, G., *African trypanosomes: the genome and adaptations for immune evasion.* Essays in Biochemistry, 2011. **51**: p. 47-62.

65.   Huang, H., et al., *Trypanosomes* ☆, in *Encyclopedia of Microbiology (Fourth Edition)*, T.M. Schmidt, Editor. 2019, Academic Press: Oxford. p. 495-509.

66.   Pays, E., et al., *The VSG expression sites of Trypanosoma brucei: multipurpose tools for the adaptation of the parasite to mammalian hosts.* Molecular and Biochemical Parasitology, 2001. **114**(1): p. 1-16.

67.   Van Xong, H., et al., *A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in Trypanosoma rhodesiense.* Cell, 1998. **95**(6): p. 839-846.

68.   Pays, E., et al., *The trypanolytic factor of human serum.* Nature Reviews Microbiology, 2006. **4**(6): p. 477-486.

69.   Zoll, S., et al., *The structure of serum resistance-associated protein and its implications for human African trypanosomiasis.* Nature Microbiology, 2018. **3**(3): p. 295-301.

70.   Pays, E., et al., *The molecular arms race between African trypanosomes and humans.* Nature Reviews Microbiology, 2014. **12**(8): p. 575-584.

71.   Uzureau, P., et al., *Mechanism of Trypanosoma brucei gambiense resistance to human serum.* Nature, 2013. **501**(7467): p. 430-434.

72.   Merle, N.S., et al., *Complement System Part I – Molecular Mechanisms of Activation and Regulation.* Frontiers in Immunology, 2015. **6**.

73.   Janeway, C., *Immunobiology 5 : The Immune System in Health and Disease*. 5th ed. 2001, New York: Garland Pub. New York.

74.   Zwarthoff, S.A., et al., *Functional Characterization of Alternative and Classical Pathway C3/C5 Convertase Activity and Inhibition Using Purified Models.* Frontiers in Immunology, 2018. **9**.

75.   Klos, A., et al., *The role of the anaphylatoxins in health and disease.* Molecular Immunology, 2009. **46**(14): p. 2753-2766.

76.   Pangburn, M.K., R.D. Schreiber, and H.J. Müller-Eberhard, *Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3.* Journal of Experimental Medicine, 1981. **154**(3): p. 856-867.

77.   Lachmann, P.J., E. Lay, and D.J. Seilly, *Experimental confirmation of the C3 tickover hypothesis by studies with an Ab (S77) that inhibits tickover in whole serum.* The FASEB Journal, 2018. **32**(1): p. 123-129.

78.   Liu, X., et al., *Changes in Complement Levels and Activity of Red Blood Cells, Fresh Frozen Plasma, and Platelet Concentrates During Storage.* Indian Journal of Hematology and Blood Transfusion, 2021. **37**(1): p. 140-146.

79.   Chen, Z.A., et al., *Structure of Complement C3(H2O) Revealed By Quantitative Cross-Linking/Mass Spectrometry And Modeling.* Molecular & Cellular Proteomics, 2016. **15**(8): p. 2730-2743.

80.   Morgan, H.P., et al., *Structural basis for engagement by complement factor H of C3b on a self surface.* Nature Structural & Molecular Biology, 2011. **18**(4): p. 463-470.

81.   Cunnion, K.M., P.S. Hair, and E.S. Buescher, *Cleavage of Complement C3b to iC3b on the Surface of Staphylococcus aureus Is Mediated by Serum Complement Factor I.* Infection and Immunity, 2004. **72**(5): p. 2858-2863.

82.   Chen, J.Y., C. Cortes, and V.P. Ferreira, *Properdin: A multifaceted molecule involved in inflammation and diseases.* Molecular Immunology, 2018. **102**: p. 58-72.

83.   Fearon, D.T. and K.F. Austen, *Properdin: binding to C3b and stabilization of the C3b-dependent C3 convertase.* Journal of Experimental Medicine, 1975. **142**(4): p. 856-863.

84.   Xie, C.B., D. Jane-Wit, and J.S. Pober, *Complement Membrane Attack Complex: New Roles, Mechanisms of Action, and Therapeutic Targets.* The American Journal of Pathology, 2020. **190**(6): p. 1138-1150.

85.   Sülzen, H., et al., *Cryo-EM structures of Trypanosoma brucei gambiense ISG65 with human complement C3 and C3b and their roles in alternative pathway restriction.* Nature Communications, 2023. **14**(1).

86.   Ferrante, A. and A.C. Allison, *Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat.* Parasite Immunology, 1983. **5**(5): p. 491-498.

87.   Devine, D.V., R.J. Falk, and A.E. Balber, *Restriction of the alternative pathway of human complement by intact Trypanosoma brucei subsp. gambiense.* Infection and Immunity, 1986. **52**(1): p. 223-229.

88.   Macleod, O.J.S., et al., *Invariant surface glycoprotein 65 of Trypanosoma brucei is a complement C3 receptor.* Nature Communications, 2022. **13**(1).

89.   Grünfelder, C.G., et al., *Accumulation of a GPI-Anchored Protein at the Cell Surface Requires Sorting at Multiple Intracellular Levels.* Traffic, 2002. **3**(8): p. 547-559.

90.   Engstler, M., et al., *Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in Trypanosoma brucei.* Journal of Cell Science, 2004. **117**(7): p. 1105-1115.

91.   Magez, S., et al., *Salivarian Trypanosomes Have Adopted Intricate Host-Pathogen Interaction Mechanisms That Ensure Survival in Plain Sight of the Adaptive Immune System.* Pathogens, 2021. **10**(6).

92.   Horn, D., *Antigenic variation in African trypanosomes.* Molecular and Biochemical Parasitology, 2014. **195**(2): p. 123-129.

93.   Đaković, S., et al., *A structural classification of the variant surface glycoproteins of the African trypanosome.* PLOS Neglected Tropical Diseases, 2023. **17**(9).

94.   Schwede, A., et al., *How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins?* PLOS Pathogens, 2016. **11**(12).

95.   Umaer, K., et al., *Dynamic, variable oligomerization and the trafficking of variant surface glycoproteins of Trypanosoma brucei.* Traffic, 2021. **22**(8): p. 274-283.

96.   Pinger, J., et al., *African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat.* Nature Microbiology, 2018. **3**(8): p. 932-938.

97.   Overath, P. and M. Engstler, *Endocytosis, membrane recycling and sorting of GPI-anchored proteins: Trypanosoma brucei as a model system.* Molecular Microbiology, 2004. **53**(3): p. 735-744.

98.   Engstler, M., et al., *Hydrodynamic Flow-Mediated Protein Sorting on the Cell Surface of Trypanosomes.* Cell, 2007. **131**(3): p. 505-515.

99.   Hovel-Miner, G., et al., *A Host–Pathogen Interaction Reduced to First Principles: Antigenic Variation in T. brucei*, in *Pathogen-Host Interactions: Antigenic Variation v. Somatic Adaptations*, E. Hsu and L. Du Pasquier, Editors. 2015, Springer International Publishing: Cham. p. 23-46.

100.   Matthews, K.R., R. McCulloch, and L.J. Morrison, *The within-host dynamics of African trypanosome infections.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2015. **370**(1675).

101. Abbas, A.H., et al., *The Structure of a Conserved Telomeric Region Associated with Variant Antigen Loci in the Blood Parasite Trypanosoma congolense.* Genome Biology and Evolution, 2018. **10**(9): p. 2458-2473.

102. Hall, J.P.J., H. Wang, and J.D. Barry, *Mosaic VSGs and the Scale of Trypanosoma brucei Antigenic Variation.* PLOS Pathogens, 2013. **9**(7).

103. Marcello, L. and J.D. Barry, *From Silent Genes to Noisy Populations—Dialogue Between the Genotype and Phenotypes of Antigenic Variation1.* Journal of Eukaryotic Microbiology, 2007. **54**(1): p. 14-17.

104. Radwanska, M., et al., *Trypanosomiasis-Induced B Cell Apoptosis Results in Loss of Protective Anti-Parasite Antibody Responses and Abolishment of Vaccine-Induced Memory Responses.* PLOS Pathogens, 2008. **4**(5).

105. Overath, P., et al., *Invariant surface proteins in bloodstream forms of Trypanosoma brucei.* Parasitology Today, 1994. **10**(2): p. 53-58.

106. Makarov, A., et al., *The role of invariant surface glycoprotein 75 in xenobiotic acquisition by African trypanosomes.* Microb Cell, 2023. **10**(2): p. 18-35.

107. Ziegelbauer, K., et al., *Genomic organization of an invariant surface glycoprotein gene family of Trypanosoma brucei.* Molecular and Biochemical Parasitology, 1995. **69**(1): p. 53-63.

108. Bringaud, F. and T. Baltz, *Differential Regulation of Two Distinct Families of Glucose Transporter Genes in Trypanosoma brucei.* Molecular and Cellular Biology, 1993. **13**(2): p. 1146-1154.

109. Steverding, D., et al., *Transferrin-binding protein complex is the receptor for transferrin uptake in Trypanosoma brucei.* Journal of Cell Biology, 1995. **131**(5): p. 1173-1182.

110. Higgins, M.K., et al., *Structure of the trypanosome haptoglobin–hemoglobin receptor and implications for nutrient uptake and innate immunity.* Proceedings of the National Academy of Sciences, 2013. **110**(5): p. 1905-1910.

111. Kariuki, C.K., B. Stijlemans, and S. Magez, *The Trypanosomal Transferrin Receptor of Trypanosoma Brucei—A Review.* Tropical Medicine and Infectious Disease, 2019. **4**(4).

112. Ziegelbauer, K., G. Multhaup, and P. Overath, *Molecular characterization of two invariant surface glycoproteins specific for the bloodstream stage of Trypanosoma brucei.* Journal of Biological Chemistry, 1992. **267**(15): p. 10797-10803.

113. Leung Ka, F., et al., *Ubiquitylation and Developmental Regulation of Invariant Surface Protein Expression in Trypanosomes.* Eukaryotic Cell, 2011. **10**(7): p. 916-931.

114. Zoltner, M., et al., *Modulation of the Surface Proteome through Multiple Ubiquitylation Pathways in African Trypanosomes.* PLOS Pathogens, 2015. **11**(10).

115. Umaer, K. and J.D. Bangs, *Late ESCRT machinery mediates the recycling and Rescue of Invariant Surface Glycoprotein 65 in Trypanosoma brucei.* Cellular Microbiology, 2020. **22**(11).

116. Koumandou, V.L., et al., *Evidence for Recycling of Invariant Surface Transmembrane Domain Proteins in African Trypanosomes.* Eukaryotic Cell, 2013. **12**(2): p. 330-342.

117. Ziegelbauer, K. and P. Overath, *Organization of two invariant surface glycoproteins in the surface coat of Trypanosoma brucei.* Infection and Immunity, 1993. **61**(11): p. 4540-4545.

118. Biéler, S., et al., *Evaluation of Antigens for Development of a Serological Test for Human African Trypanosomiasis.* PLOS ONE, 2016. **11**(12).

119. Sullivan, L., et al., *Proteomic Selection of Immunodiagnostic Antigens for Human African Trypanosomiasis and Generation of a Prototype Lateral Flow Immunodiagnostic Device.* PLOS Neglected Tropical Diseases, 2013. **7**(2).

120. Benkő, M., et al., *ICTV Virus Taxonomy Profile: Adenoviridae 2022.* Journal of General Virology, 2022. **103**(3).

121. Greber, U.F. and J.W. Flatt, *Adenovirus Entry: From Infection to Immunity.* Annual Review of Virology, 2019. **6**(Volume 6, 2019): p. 177-197.

122. Lion, T., *Adenovirus Infections in Immunocompetent and Immunocompromised Patients.* Clinical Microbiology Reviews, 2014. **27**(3): p. 441-462.

123. Liu, H., et al., *Atomic Structure of Human Adenovirus by Cryo-EM Reveals Interactions Among Protein Networks.* Science, 2010. **329**(5995): p. 1038-1043.

124. Russell, W.C., *Adenoviruses: update on structure and function.* Journal of General Virology, 2009. **90**(1): p. 1-20.

125. Vellinga, J., S. Van der Heijdt, and R.C. Hoeben, *The adenovirus capsid: major progress in minor proteins.* Journal of General Virology, 2005. **86**(6): p. 1581-1588.

126. Prage, L., et al., *Structural proteins of adenoviruses: IV. Sequential degradation of the adenovirus type 2 virion.* Virology, 1970. **42**(2): p. 341-358.

127. Richardson, J.S., *The Anatomy and Taxonomy of Protein Structure*, in *Advances in Protein Chemistry*, C.B. Anfinsen, J.T. Edsall, and F.M. Richards, Editors. 1981, Academic Press. p. 167-339.

128. Bamford, D.H., R.M. Burnett, and D.I. Stuart, *Evolution of Viral Structure.* Theoretical Population Biology, 2002. **61**(4): p. 461-470.

129. Athappily, F.K., et al., *The Refined Crystal Structure of Hexon, the Major Coat Protein of Adenovirus Type 2, at 2·9 Å Resolution.* Journal of Molecular Biology, 1994. **242**(4): p. 430-455.

130. Dhillon, A., et al., *Structural insights into the interaction between adenovirus C5 hexon and human lactoferrin.* Journal of Virology, 2024. **98**(3).

131. Crawford-Miksza, L. and D.P. Schnurr, *Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues.* Journal of Virology, 1996. **70**(3): p. 1836-1844.

132. Roberts, D.M., et al., *Hexon-chimaeric adenovirus serotype 5 vectors circumvent pre-existing anti-vector immunity.* Nature, 2006. **441**(7090): p. 239-243.

133. Sumida, S.M., et al., *Neutralizing Antibodies to Adenovirus Serotype 5 Vaccine Vectors Are Directed Primarily against the Adenovirus Hexon Protein1.* The Journal of Immunology, 2005. **174**(11): p. 7179-7185.

134. Kalyuzhniy, O., et al., *Adenovirus serotype 5 hexon is critical for virus infection of hepatocytes in vivo.* Proceedings of the National Academy of Sciences, 2008. **105**(14): p. 5483-5488.

135. Waddington, S.N., et al., *Adenovirus Serotype 5 Hexon Mediates Liver Gene Transfer.* Cell, 2008. **132**(3): p. 397-409.

136. Johansson, C., et al., *Adenoviruses Use Lactoferrin as a Bridge for CAR-Independent Binding to and Infection of Epithelial Cells.* Journal of Virology, 2007. **81**(2): p. 954-963.

137. Persson, B.D., et al., *Lactoferrin-Hexon Interactions Mediate CAR-Independent Adenovirus Infection of Human Respiratory Cells.* Journal of Virology, 2020. **94**(14).

138. Boudin, M.-L. and P. Boulanger, *Assembly of adenovirus penton base and fiber.* Virology, 1982. **116**(2): p. 589-604.

139. Zubieta, C., et al., *The Structure of the Human Adenovirus 2 Penton.* Molecular Cell, 2005. **17**(1): p. 121-135.

140. Liu, H., L. Wu, and Z.H. Zhou, *Model of the Trimeric Fiber and Its Interactions with the Pentameric Penton Base of Human Adenovirus by Cryo-electron Microscopy.* Journal of Molecular Biology, 2011. **406**(5): p. 764-774.

141. van Raaij, M.J., et al., *A triple β-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein.* Nature, 1999. **401**(6756): p. 935-938.

142. Stasiak, A.C. and T. Stehle, *Human adenovirus binding to host cell receptors: a structural view.* Medical Microbiology and Immunology, 2020. **209**(3): p. 325-333.

143. Vellinga, J., et al., *The Coiled-Coil Domain of the Adenovirus Type 5 Protein IX Is Dispensable for Capsid Incorporation and Thermostability.* Journal of Virology, 2005. **79**(5): p. 3206-3210.

144. Rosa-Calatrava, M., et al., *Functional Analysis of Adenovirus Protein IX Identifies Domains Involved in Capsid Stability, Transcriptional Activity, and Nuclear Reorganization.* Journal of Virology, 2001. **75**(15): p. 7131-7141.

145. Sargent Kathy, L., A. Meulenbroek Robert, and J. Parks Robin, *Activation of Adenoviral Gene Expression by Protein IX Is Not Required for Efficient Virus Replication.* Journal of Virology, 2004. **78**(10): p. 5032-5037.

146. Summers, W.C., *Virus Infection*, in *Encyclopedia of Microbiology (Third Edition)*, M. Schaechter, Editor. 2009, Academic Press: Oxford. p. 546-552.

147. Ryu, W.-S., *Chapter 3 - Virus Life Cycle*, in *Molecular Virology of Human Pathogenic Viruses*, W.-S. Ryu, Editor. 2017, Academic Press: Boston. p. 31-45.

148. Cann, A.J., *Replication of Viruses*, in *Encyclopedia of Virology (Third Edition)*, B.W.J. Mahy and M.H.V. Van Regenmortel, Editors. 2008, Academic Press: Oxford. p. 406-412.

149. Bergelson, J.M., et al., *Isolation of a Common Receptor for Coxsackie B Viruses and Adenoviruses 2 and 5.* Science, 1997. **275**(5304): p. 1320-1323.

150. Tomko, R.P., R. Xu, and L. Philipson, *HCAR and MCAR: The human and mouse cellular receptors for subgroup C adenoviruses and group*

B coxsackieviruses. Proceedings of the National Academy of Sciences, 1997. **94**(7): p. 3352-3356.

151. Bewley, M.C., et al., *Structural Analysis of the Mechanism of Adenovirus Binding to Its Human Cellular Receptor, CAR.* Science, 1999. **286**(5444): p. 1579-1583.

152. Wickham, T.J., et al., *Integrins αvβ3 and αvβ5 promote adenovirus internalization but not virus attachment.* Cell, 1993. **73**(2): p. 309-319.

153. Kotha, P.L.N., et al., *Adenovirus Entry From the Apical Surface of Polarized Epithelia Is Facilitated by the Host Innate Immune Response.* PLOS Pathogens, 2015. **11**(3).

154. Lütschg, V., et al., *Chemotactic antiviral cytokines promote infectious apical entry of human adenovirus into polarized epithelial cells.* Nature Communications, 2011. **2**(1).

155. Arnberg, N., *Adenovirus receptors: implications for targeting of viral vectors.* Trends in Pharmacological Sciences, 2012. **33**(8): p. 442-448.

156. Hensen, L.C.M., R.C. Hoeben, and S.T.F. Bots, *Adenovirus Receptor Expression in Cancer and Its Multifaceted Role in Oncolytic Adenovirus Therapy.* International Journal of Molecular Sciences, 2020. **21**(18).

157. Lyle, C. and F. McCormick, *Integrin αvβ5 is a primary receptor for adenovirus in CAR-negative cells.* Virology Journal, 2010. **7**(1).

158. Haisma, H.J., et al., *Scavenger Receptor A: A New Route for Adenovirus 5.* Molecular Pharmaceutics, 2009. **6**(2): p. 366-374.

159. Maler, M.D., et al., *Key Role of the Scavenger Receptor MARCO in Mediating Adenovirus Infection and Subsequent Innate Responses of Macrophages.* mBio, 2017. **8**(4).

160. PrabhuDas, M.R., et al., *A Consensus Definitive Classification of Scavenger Receptors and Their Roles in Health and Disease.* The Journal of Immunology, 2017. **198**(10): p. 3775-3789.

161. Alquraini, A. and J. El Khoury, *Scavenger receptors.* Current Biology, 2020. **30**(14): p. R790-R795.

162. Khare, R., et al., *Identification of Adenovirus Serotype 5 Hexon Regions That Interact with Scavenger Receptors.* Journal of Virology, 2012. **86**(4): p. 2293-2301.

163. Fejer, G., et al., *Nontransformed, GM-CSF–dependent macrophage lines are a unique model to study tissue macrophage functions.* Proceedings of the National Academy of Sciences, 2013. **110**(24): p. E2191-E2198.

164. Stichling, N., et al., *Lung macrophage scavenger receptor SR-A6 (MARCO) is an adenovirus type-specific virus entry receptor.* PLOS Pathogens, 2018. **14**(3).

165. Chu, Y., et al., *Vascular Cell Adhesion Molecule-1 Augments Adenovirus-Mediated Gene Transfer.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2001. **21**(2): p. 238-242.

166. Cao, G., et al., *Gene Therapy for Cardiovascular Disease: Basic Research and Clinical Prospects.* Frontiers in Cardiovascular Medicine, 2021. **8**.

167. Adelsman, M.A. and Y. Shimizu, *Adhesion Molecules*, in *Encyclopedia of Immunology (Second Edition)*, P.J. Delves, Editor. 1998, Elsevier: Oxford. p. 26-33.

168. Hong, S.S., et al., *Adenovirus type 5 fiber knob binds to MHC class I α2 domain at the surface of human epithelial and B lymphoblastoid cells.* The EMBO Journal, 1997. **16**(9): p. 2294-2306.

169. Chéneau, C., et al., *Lactoferrin Retargets Human Adenoviruses to TLR4 to Induce an Abortive NLRP3-Associated Pyroptotic Response in Human Phagocytes.* Frontiers in Immunology, 2021. **12**.

170. Haridas, M., B.F. Anderson, and E.N. Baker, *Structure of human diferric lactoferrin refined at 2.2 A resolution.* Acta Crystallographica Section D, 1995. **51**(5): p. 629-646.

171. van der Strate, B.W.A., et al., *Antiviral activities of lactoferrin.* Antiviral Research, 2001. **52**(3): p. 225-239.

172. Farnaud, S. and R.W. Evans, *Lactoferrin—a multifunctional protein with antimicrobial properties.* Molecular Immunology, 2003. **40**(7): p. 395-405.

173. Bellamy, W., et al., *Identification of the bactericidal domain of lactoferrin.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1992. **1121**(1): p. 130-136.

174. Gruden, Š. and N. Poklar Ulrih, *Diverse Mechanisms of Antimicrobial Activities of Lactoferrins, Lactoferricins, and Other Lactoferrin-Derived Peptides.* International Journal of Molecular Sciences, 2021. **22**(20).

175. Shenk, T.E., *Adenoviridae: The Viruses and Their Replication*, in *Fundamental Virology*, D.M. Knipe and P.M. Howley, Editors. 2001, Lippincott Williams & Wilkins. p. 1053-1088.

176. Medina-Kauwe, L.K., *Endocytosis of adenovirus and adenovirus capsid proteins.* Advanced Drug Delivery Reviews, 2003. **55**(11): p. 1485-1496.

177. Meier, O., et al., *Adenovirus triggers macropinocytosis and endosomal leakage together with its clathrin-mediated uptake.* Journal of Cell Biology, 2002. **158**(6): p. 1119-1131.

178. Saban Susan, D., et al., *Visualization of α-Helices in a 6-Ångstrom Resolution Cryoelectron Microscopy Structure of Adenovirus Allows Refinement of Capsid Protein Assignments.* Journal of Virology, 2006. **80**(24): p. 12049-12059.

179. Wiethoff, C.M. and G.R. Nemerow, *Adenovirus membrane penetration: Tickling the tail of a sleeping dragon.* Virology, 2015. **479-480**: p. 591-599.

180. Miyazawa, N., G. Crystal Ronald, and L. Leopold Philip, *Adenovirus Serotype 7 Retention in a Late Endosomal Compartment prior to Cytosol Escape Is Modulated by Fiber Protein.* Journal of Virology, 2001. **75**(3): p. 1387-1400.

181. Maier, O., et al., *An N-terminal domain of adenovirus protein VI fragments membranes by inducing positive membrane curvature.* Virology, 2010. **402**(1): p. 11-19.

182. Hernando-Pérez, M., et al., *Dynamic competition for hexon binding between core protein VII and lytic protein VI promotes adenovirus maturation and entry.* Proceedings of the National Academy of Sciences, 2020. **117**(24): p. 13699-13707.

183. Leopold, P.L. and R.G. Crystal, *Intracellular trafficking of adenovirus: Many means to many ends.* Advanced Drug Delivery Reviews, 2007. **59**(8): p. 810-821.

184. Greber, U.F., *Signalling in viral entry.* Cellular and Molecular Life Sciences CMLS, 2002. **59**(4): p. 608-626.

185. Kulanayake, S. and S.K. Tikoo, *Adenovirus Core Proteins: Structure and Function.* Viruses, 2021. **13**(3).

186. Berk, A.J., *Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus.* Oncogene, 2005. **24**(52): p. 7673-7685.

187. Oliveira, E.R.A. and M. Bouvier, *Immune evasion by adenoviruses: a window into host–virus adaptation.* FEBS Letters, 2019. **593**(24): p. 3496-3503.

188. Hoeben, R.C. and T.G. Uil, *Adenovirus DNA Replication.* Cold Spring Harbor Perspectives in Biology, 2013. **5**(3).

189. Pied, N. and H. Wodrich, *Imaging the adenovirus infection cycle.* FEBS Letters, 2019. **593**(24): p. 3419-3448.

190. Mangel, W.F. and C. San Martín, *Structure, Function and Dynamics in Adenovirus Maturation.* Viruses, 2014. **6**(11): p. 4536-4570.

191. Tang, R. and Z. Xu, *Gene therapy: a double-edged sword with great powers.* Molecular and Cellular Biochemistry, 2020. **474**(1): p. 73-81.

192. Ramamoorth, M. and A. Narvekar, *Non viral vectors in gene therapy- an overview.* J Clin Diagn Res, 2015. **9**(1).

193. Trivedi, P.D., B.J. Byrne, and M. Corti, *Evolving Horizons: Adenovirus Vectors' Timeless Influence on Cancer, Gene Therapy and Vaccines.* Viruses, 2023. **15**(12).

194. Watanabe, M., et al., *Adenovirus Biology, Recombinant Adenovirus, and Adenovirus Usage in Gene Therapy.* Viruses, 2021. **13**(12).

195. Bulcha, J.T., et al., *Viral vector platforms within the gene therapy landscape.* Signal Transduction and Targeted Therapy, 2021. **6**(1).

196. Athanasopoulos, T., M.M. Munye, and R.J. Yáñez-Muñoz, *Nonintegrating Gene Therapy Vectors.* Hematology/Oncology Clinics of North America, 2017. **31**(5): p. 753-770.

197. Sallard, E., et al., *The Adenovirus Vector Platform: Novel Insights into Rational Vector Design and Lessons Learned from the COVID-19 Vaccine.* Viruses, 2023. **15**(1).

198. Wold, S.M.W. and K. Toth, *Adenovirus Vectors for Gene Therapy, Vaccination and Cancer Gene Therapy.* Current Gene Therapy, 2013. **13**(6): p. 421-433.

199. Sibbald, B., *Death but one unintended consequence of gene-therapy trial.* Cmaj, 2001. **164**(11).

200. Somia, N. and I.M. Verma, *Gene therapy: trials and tribulations.* Nature Reviews Genetics, 2000. **1**(2): p. 91-99.

201. Teichler Zallen, D., *US gene therapy in crisis.* Trends in Genetics, 2000. **16**(6): p. 272-275.

202. Shakti, S., K. Rakesh, and A. Babita, *Adenoviral Vector-Based Vaccines and Gene Therapies: Current Status and Future Prospects*, in *Adenoviruses*, D. Yulia, Editor. 2018, IntechOpen: Rijeka.

203. Tan, E.W., et al., *Engineered Oncolytic Adenoviruses: An Emerging Approach for Cancer Therapy.* Pathogens, 2022. **11**(10).

204. Mennechet, F.J.D., et al., *A review of 65 years of human adenovirus seroprevalence.* Expert Review of Vaccines, 2019. **18**(6): p. 597-613.

205. Lawler, S.E., et al., *Oncolytic Viruses in Cancer Treatment: A Review.* JAMA Oncology, 2017. **3**(6): p. 841-849.

206. Sülzen, H., et al., *A multifaceted strategy to improve recombinant expression and structural characterisation of a Trypanosoma invariant surface protein.* Scientific Reports, 2022. **12**(1).

207. Sülzen, H., et al., *Beyond the VSG Layer: Exploring the Role of Intrinsic Disorder in the Invariant Surface Glycoproteins of African Trypanosomes.* bioRxiv, 2023.

208. Sánchez-Corral, P., et al., *Separation of active and inactive forms of the third component of human complement, C3, by fast protein liquid chromatography (FPLC).* Journal of Immunological Methods, 1989. **122**(1): p. 105-113.

209. Bartossek, T., et al., *Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat.* Nature Microbiology, 2017. **2**(11): p. 1523-1532.

210. Fairhead, M. and M. Howarth, *Site-Specific Biotinylation of Purified Proteins Using BirA*, in *Site-Specific Protein Labeling: Methods and Protocols*, A. Gautier and M.J. Hinner, Editors. 2015, Springer New York: New York, NY. p. 171-184.

211. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nature Methods, 2017. **14**(3): p. 290-296.

212. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* eLife, 2018. **7**.

213. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.

214. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallographica Section D, 2019. **75**(10): p. 861-877.

215. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallographica Section D, 2011. **67**(4): p. 355-367.

216. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(12): p. 2126-32.

217.   Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Research, 2007. **35**: p. W375-W383.

218.   Wagner, T., et al., *SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM.* Communications Biology, 2019. **2**(1).

219.   Panjkovich, A. and D.I. Svergun, *CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data.* Bioinformatics, 2018. **34**(11): p. 1944-1946.

220.   Manalastas-Cantos, K., et al., *ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis.* Journal of Applied Crystallography, 2021. **54**(1): p. 343-355.

221.   *ATSAS Online*. 2024; Available from: www.embl-hamburg.de/biosaxs/atsas-online.

222.   Hopkins, J.B., R.E. Gillilan, and S. Skou, *BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis.* Journal of Applied Crystallography, 2017. **50**(5): p. 1545-1553.

223.   Evans, R., et al., *Protein complex prediction with AlphaFold-Multimer*. 2021, bioRxiv.

224.   Schneidman-Duhovny, D., M. Hammel, and A. Sali, *FoXS: a web server for rapid computation and fitting of SAXS profiles.* Nucleic Acids Research, 2010. **38**: p. W540-W544.

225.   Schneidman-Duhovny, D., et al., *FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles.* Nucleic Acids Research, 2016. **44**: p. W424-W429.

226.   Pelikan, M., G.L. Hura, and M. Hammel, *Structure and flexibility within proteins as identified through small angle X-ray scattering.* Gen Physiol Biophys, 2009. **28**(2): p. 174-89.

227.   Nzou, S.M., et al., *Development of multiplex serological assay for the detection of human African trypanosomiasis.* Parasitology International, 2016. **65**(2): p. 121-127.

228.   Yang, Y., et al., *Sixty-five years of the long march in protein secondary structure prediction: the final stretch?* Briefings in Bioinformatics, 2018. **19**(3): p. 482-494.

229.   Masson, G.R., et al., *Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments.* Nature Methods, 2019. **16**(7): p. 595-602.

230.   Jin, J., et al., *Accelerating the clinical development of protein-based vaccines for malaria by efficient purification using a four amino acid C-terminal 'C-tag'.* International Journal for Parasitology, 2017. **47**(7): p. 435-446.

231.   Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nature Protocols, 2010. **5**(4): p. 725-738.

232.   Mirdita, M., et al., *ColabFold: making protein folding accessible to all.* Nature Methods, 2022. **19**(6): p. 679-682.

233.   Schaeffer, R.D., et al., *Assessment of domain interactions in the fourteenth round of the Critical Assessment of Structure Prediction (CASP14).* Proteins: Structure, Function, and Bioinformatics, 2021. **89**(12): p. 1700-1710.

234.   Mariani, V., et al., *IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests.* Bioinformatics, 2013. **29**(21): p. 2722-2728.

235.   Pangburn, M.K. and H.J. Müller-Eberhard, *Initiation of the Alternative Complement Pathway due to Spontaneous Hydrolysis of the Thioester of C3a.* Annals of the New York Academy of Sciences, 1983. **421**(1): p. 291-298.

236.   Janssen, B.J.C., et al., *Structures of complement component C3 provide insights into the function and evolution of immunity.* Nature, 2005. **437**(7058): p. 505-511.

237.   Janssen, B.J.C., et al., *Structure of C3b reveals conformational changes that underlie complement activity.* Nature, 2006. **444**(7116): p. 213-216.

238.   Tria, G., et al., *Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering.* IUCrJ, 2015. **2**(2): p. 207-217.

239.   Marquart, J.A., *Surface Plasmon Resonance and Biomolecular Interaction Analysis: Theory and Practice*. 2008.

240.   Oshannessy, D.J., et al., *Determination of Rate and Equilibrium Binding Constants for Macromolecular Interactions Using Surface Plasmon Resonance: Use of Nonlinear Least Squares Analysis Methods.* Analytical Biochemistry, 1993. **212**(2): p. 457-468.

241.   Cook, A.D., M. Carrington, and M.K. Higgins, *Molecular mechanism of complement inhibition by the trypanosome receptor ISG65.* eLife Sciences Publications, 2023.

242.   Kortt, A.A., et al., *Nonspecific Amine Immobilization of Ligand Can Be a Potential Source of Error in BIAcore Binding Experiments and May Reduce Binding Affinities.* Analytical Biochemistry, 1997. **253**(1): p. 103-111.

243.   Pangburn, M.K., *Initiation of the alternative pathway of complement and the history of "tickover".* Immunological Reviews, 2023. **313**(1): p. 64-70.

244.   Fairbairn, H. and E. Burtt, *The Infectivity to Man of a Strain of Trypanosoma Rhodesiense Transmitted Cyclically by Glossina Morsitans Through Sheep and Antelope: Evidence that Man Requires a Minimum Infective Dose of Metacyclic Trypanosomes.* Annals of Tropical Medicine & Parasitology, 1946. **40**(3-4): p. 270-313.

245.   Caruso, A., et al., *Modeling the activation of the alternative complement pathway and its effects on hemolysis in health and disease.* PLOS Computational Biology, 2020. **16**(10).

246.   Tunyasuvunakool, K., et al., *Highly accurate protein structure prediction for the human proteome.* Nature, 2021. **596**: p. 590-596.

247.   Yang, Z., et al., *AlphaFold2 and its applications in the fields of biology and medicine.* Signal Transduction and Targeted Therapy, 2023. **8**(1).

248.   Brookes, E., et al., *AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank.* Journal of Applied Crystallography, 2023. **56**(4): p. 910-926.

249.   Receveur-Brechot, V., *AlphaFold, small-angle X-ray scattering and ensemble modelling: a winning combination for intrinsically disordered proteins.* Journal of Applied Crystallography, 2023. **56**(5): p. 1313-1314.

250.   Alderson, T.R., et al., *Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2.* Proceedings of the National Academy of Sciences, 2023. **120**(44).

251.   Hall, T. and K. Esser, *Topologic mapping of protective and nonprotective epitopes on the variant surface glycoprotein of the WRATat 1 clone of Trypanosoma brucei rhodesiense.* The Journal of Immunology, 1984. **132**(4): p. 2059-2063.

252.   Cross, G.A.M., S. Cohen, and G.A.M. Cross, *Structure of the variant glycoproteins and surface coat of Trypanosoma brucei.* Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 1997. **307**(1131): p. 3-12.

253.   Masterson, W.J., D. Taylor, and M.J. Turner, *Topologic analysis of the epitopes of a variant surface glycoprotein of Trypanosoma brucei.* The Journal of Immunology, 1988. **140**(9): p. 3194-3199.

254.   Casas-Sanchez, A., et al., *The Trypanosoma brucei MISP family of invariant proteins is co-expressed with BARP as triple helical bundle structures on the surface of salivary gland forms, but is dispensable for parasite development within the tsetse vector.* PLOS Pathogens, 2023. **19**(3).

255.   Schwede, A., et al., *The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes.* Molecular and Biochemical Parasitology, 2011. **175**(2): p. 201-204.

256.   Wälti, M.A., et al., *Visualization of Sparsely-populated Lower-order Oligomeric States of Human Mitochondrial Hsp60 by Cryo-electron Microscopy.* Journal of Molecular Biology, 2021. **433**(24).

257.   Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package.* Journal of Magnetic Resonance, 2003. **160**(1): p. 65-73.

258.   Dartt, D.A., *Regulation of mucin and fluid secretion by conjunctival epithelial cells.* Progress in Retinal and Eye Research, 2002. **21**(6): p. 555-576.

259.   Luk, C.K. and M.J. Dulfano, *Effect of pH, Viscosity and Ionic-Strength Changes on Ciliary Beating Frequency of Human Bronchial Explants.* Clinical Science, 1983. **64**(4): p. 449-451.

260.   Mitsubayashi, K., et al., *Measurement of tear electrolyte concentration and turnover rate using a flexible conductimetric sensor.* Technology and Health Care, 1995. **3**: p. 117-121.

261.   Whikehart, D.R., *Chapter 1 - Water and Ocular Fluids: Physical Chemistry of Ocular Fluids*, in *Biochemistry of the Eye (Second Edition)*, D.R. Whikehart, Editor. 2003, Butterworth-Heinemann: Philadelphia. p. 1-14.

262.   Kell, D.B., E.L. Heyden, and E. Pretorius, *The Biology of Lactoferrin, an Iron-Binding Protein That Can Help Defend Against Viruses and Bacteria.* Frontiers in Immunology, 2020. **11**.

263. Calabuig-Navarro, M.V., et al., *A randomized trial and novel SPR technique identifies altered lipoprotein-LDL receptor binding as a mechanism underlying elevated LDL-cholesterol in APOE4s.* Scientific Reports, 2017. **7**(1).

264. Vallat, B., et al., *New system for archiving integrative structures.* Acta Crystallographica Section D, 2021. **77**(12): p. 1486-1496.

265. Corum, M.R., et al., *Predictive modeling and cryo-EM: A synergistic approach to modeling macromolecular structure.* Biophysical Journal, 2024. **123**(4): p. 435-450.

266. Kühlbrandt, W., *The Resolution Revolution.* Science, 2014. **343**(6178): p. 1443-1444.

267. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

268. Kermani, A.A., S. Aggarwal, and A. Ghanbarpour, *Chapter 11 - Advances in X-ray crystallography methods to study structural dynamics of macromolecules*, in *Advanced Spectroscopic Methods to Study Biomolecular Structure and Dynamics*, P. Saudagar and T. Tripathi, Editors. 2023, Academic Press. p. 309-355.