
Structural and Biophysical Characterization of Membrane-Coating Proteins from the Nuclear Pore and the Primary Cilium

by

Kevin E. Knockenhauer

B.S. Biochemistry
University of New York at Stony Brook, 2010

SUBMITTED TO THE DEPARTMENT OF BIOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT THE

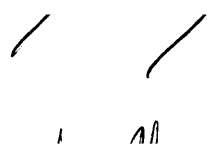
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2016

© 2016 Massachusetts Institute of Technology.
All rights reserved.

Signature redacted

Signature of Author.....



Kevin E. Knockenhauer
Department of Biology
October 2, 2015

Signature redacted

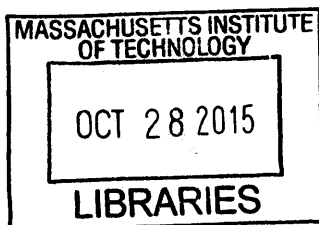
Certified by.....

Thomas U. Schwartz
Professor of Biology
Thesis Supervisor

Signature redacted

Accepted by.....

Amy E. Keating
Professor of Biology and Biological Engineering
Co-Director, Biology Graduate Program



ARCHIVES

Structural and Biophysical Characterization of Membrane-Coating Proteins from the Nuclear Pore and the Primary Cilium

by

Kevin E. Knockenhauer

Submitted to the Department of Biology on October 2, 2015
in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Abstract

A hallmark of eukaryotes is an endomembrane system that spatially separates cellular processes into discrete compartments. Macromolecular transport between these compartments canonically involves the fission and fusion of membrane-bound vesicles. Transport to and from the nucleus is a notable exception to this vesicular pathway. Additionally, transport to the primary cilium is not well characterized and is therefore of interest.

The nuclear envelope comprises a double membrane that fuses at points to produce pores, which link the nucleoplasm with the cytoplasm. Coating these openings are nuclear pore complexes (NPCs), which regulate all nucleocytoplasmic trafficking by anchoring barrier forming FG-repeat proteins. At 60-120 MDa, the NPC is the largest macromolecular complex in the cell. However, it is a modular assembly formed by ~30 different nucleoporins (Nups) arranged into stable subassemblies. One such module is the ~600 kDa Y complex, which forms a Y shape and is the best characterized NPC subcomplex with crystal structures accounting for ~90% of its total mass. The molecular details of how the short arms of the Y contact the long stem, in a region called the 'hub', were not known. We solved the structure of this last major missing piece of the Y complex to 4.1 Å. This hub structure revealed unexpected curvature, allowed us to build the first atomic resolution composite of the mostly complete Y, and led to a novel higher order assembly model of the Y complex in the intact NPC.

The ciliary membrane is topologically contiguous with the plasma membrane yet functionally distinct, due to a unique complement of integral membrane proteins. The Bardet-Biedl syndrome protein complex (BBSome) functions in the establishment or maintenance of this unique composition. The BBSome is ~500 kDa octamer comprising BBS1, 2, 4, 5, 7, 8, 9, and BBIP10. To date, only the N-terminal domain of BBS1 had been structurally determined. In order to more completely characterize this complex, we have solved the structure of the BBS9 N-terminal β -propeller to 1.8 Å and determined its oligomeric state in solution. This structure has allowed us to identify a putative interaction site and characterized a disease relevant mutation on BBS9.

Thesis Supervisor: Thomas U. Schwartz

Title: Professor of Biology

Table of Contents

Acknowledgements.....	5
Chapter 1: Introduction.....	6
Introduction to the nuclear pore complex.....	9
Nup82 complex.....	11
Nsp1 (Nup62) complex.....	12
Nuclear basket.....	13
Ndc1 complex.....	14
Nic96 complex.....	15
Y complex.....	17
Introduction to the BBSome.....	24
Footnotes.....	28
References.....	28
Figures.....	37
Chapter 2: Atomic structure of the Y complex of the nuclear pore.....	40
Introduction.....	41
Results.....	42
Discussion.....	50
Methods.....	51
PDB accession code.....	59
Acknowledgements.....	59
References.....	60
Tables.....	65
Figures.....	68

Chapter 3: Structural characterization of Bardet-Biedl syndrome 9 protein (BBS9).....	79
Introduction.....	80
Results.....	81
Discussion.....	89
Methods.....	90
PDB accession code.....	95
Acknowledgements.....	96
References.....	96
Tables.....	100
Figures.....	101
Chapter 4: Conclusion.....	111
Summary.....	112
Future directions.....	113
References.....	116
Appendix: Allosteric activation of apicomplexan calcium-dependent protein kinases... 	118
Abstract.....	119
Introduction.....	119
Results.....	121
Discussion.....	129
Methods.....	132
PDB accession code.....	144
Acknowledgements.....	145
References.....	145
Tables.....	150
Figures.....	152

Acknowledgements

This body of work would not be possible without, first and foremost, my mentor Thomas U. Schwartz. Thomas has trained me to think like a crystallographer and has taught me how to go from a diffraction pattern to a fully refined structure, arguably the most fun part about what we do. He has not only provided me with sound technical guidance but also much advice and understanding on what comes after graduate school. His door was always open and he was always willing to serve as a sounding board for my many ideas. Also, Thomas had the foresight and wisdom to know when, after a year of negative results, it was appropriate for me to switch thesis projects and I owe my success, at least in part, to this well-timed decision. For all of this I am grateful.

I want to thank Nina Leksa who I first rotated with in the Schwartz lab and who taught me many of the experimental techniques needed to be a successful crystallographer, including molecular cloning. She has provided me ample support and guidance on matters both in and out of the lab, and is a great friend. I would not be here today if it were not for Kasia Sawicka and Sanford Simon, who mentored my undergraduate research at SUNY Stony Brook. Kasia, in particular, provided much guidance in the way of scientific writing and presentation, encouraging me to present at numerous conferences over my three year tenure in the lab.

I want to thank my all of my collaborators, but in particular Kotaro Kelley, Jessica Ingram, Benedikt Marcus, and Sebastian Lourido, without whom much of the work presented here would not be possible. I am grateful to all of the Schwartz lab members, who have made the lab a fun and interesting place to work. I want to thank my thesis committee members, Bob Sauer and Iain Cheeseman, for much help and advice along the way.

I am grateful to my parents for all the inspiration and support they provided and my brother, Glenn, who has kept me grounded these many years and has given me numerous stories to tell. I want to thank my friends and members of the BioMansion who have provided many adventures over the years and have made time outside of the lab an interesting experience. In particular, I thank Greg Newby for being a great roommate, sustainer of sanity, and co-viewer of all things Star Trek. I thank Heather Keys for all the support over the past year, for challenging many ideas and concepts, and for pushing me to be better.

Chapter 1: Introduction

Various membrane coating assemblies are utilized by the eukaryotic cell to carry out basic processes critical for cell survival and the maintenance of homeostasis, including protein transport and lipid exchange between membrane-bound compartments. This includes the clathrin, COPII, and COPI vesicle coats. Despite differences in function, these coats have unifying principles of structure that can be extended to other, more distantly-related membrane coating complexes. These principles have largely been elucidated through X-ray crystallography and cryo-electron microscopy.

Clathrin, perhaps the best studied membrane-coating complex to date, is responsible for receptor-mediated endocytosis from the plasma membrane and coating vesicles that bud off the trans-golgi network, destined for the lysosome (Schekman and Orci, 1996). The clathrin cage is comprised of clathrin heavy and light chain, with adaptin and associated proteins responsible for cargo interaction (Schekman and Orci, 1996). Clathrin heavy chain comprises an N-terminal β -propeller and a C-terminal regular, α -solenoid domain (Haar et al., 1998). Three heavy chain α -solenoids bind end on via their C-terminal regions to form the vertex of the cage and produce the characteristic triskelion shape (Fotin et al., 2004). Clathrin light chain binds, via an α -helical domain, to the heavy chain α -solenoid leg immediately distal to the vertex. Interdigitating triskelions, of the basic assembly described here, interact to form the entire cage which has been solved by cryo-electron microscopy (Fotin et al., 2004).

The COPII lattice, which delivers cargo proteins from the endoplasmic reticulum to the cis-golgi, is formed by the Sec31-Sec13 heterodimer and cargo sorting is performed by an adaptor complex comprising Sec23, Sec24, and Sar1. Unlike clathrin, the COPII vertex element is formed through the homotetramerization of four Sec31 N-terminal β -propeller domains (Fath et al., 2007). The legs of the cage are formed by the C-terminal, irregular α -helical domain of Sec31 and unlike the clathrin α -solenoids, which adopt a high degree of curvature, they are relatively straight. The Sec31 α -helical domain, which is interrupted by a proline-rich region, adopts a U-shaped, fold-back topology that rigidifies the helical stack. Additionally, the Sec13 β -

propeller effectively wedges itself between the Sec31 propeller and helical stack, further reducing possible conformational flexibility. Unlike the interdigitating higher order assembly of clathrin, connections between these vertices are formed by discrete edge elements, comprised of a Sec31-Sec31 homodimer mediated by the central region of the α -helical domain, termed the crown (Fath et al., 2007).

COPI, the vesicle coat that transports cargo in a retrograde direction from the cis-golgi to the ER, was initially proposed to have structural elements in common with both clathrin and COPII beyond the basic domain architecture. The COPI cage was believed to be formed by a trimeric complex comprising α , β' , and ϵ COP (Lee and Goldberg, 2010). The putative cargo adaptor complex comprising β , δ , γ , and ζ COP was defined based on the ability to dissociate it from the cage forming complex by a high salt wash. The domain topology of β' COP is similar to other coat proteins and, as such, contains two β -propellers followed by a C-terminal α -solenoid. α COP comprises an N-terminal β -propeller, a central α -solenoid, which binds the α -solenoid leg of β' COP, and a small C-terminal α -helical domain that interacts with the TPR helical repeats of ϵ COP (Lee and Goldberg, 2010). Based on the crystal structures that revealed these details, Lee et al. proposed a hybrid assembly model in which the cage vertex was formed by β -propeller interactions, as in COPII, contributed by three curved β' COPs in a clathrin-like triskelion (Lee and Goldberg, 2010). However, a recent 13 Å, single particle reconstruction of the assembled COPI coat revealed that it is a unique assembly not formed by triskelions or COPII-like propeller-propeller interactions (Dodonova et al., 2015). The β' COP β -propellers are positioned against the membrane. In fact, there do not appear to be separate coat forming and adaptor complexes, as the β - δ - γ - ζ COP complex bridges individual α - β' - ϵ COP trimers together and is an integral structural component of the coat (Dodonova et al., 2015).

Structural analysis of the clathrin, COPII, and COPI vesicle coats has led to the realization that these coats share common design principles on the atomic level, despite divergent higher order assemblies. Indeed, these coats are formed almost exclusively by β -

propeller domains and extended α -helical stacks. This structural approach has also led to an understanding of how these coats assemble, which in turn dictates how they function to stabilize highly curved membranes and transport cargo.

The goal of my thesis work was to apply a structural approach, primarily through X-ray crystallography, to the nuclear pore complex (NPC) and the Bardet-Biedl syndrome protein complex (BBSome) to understand their assembly. These complexes were chosen for study based on their functional and structural analogies to the canonical vesicle coats. Both the NPC and the BBSome are membrane coating modules, of the nuclear envelope and primary cilium respectively. In addition, β -propeller domains and extended α -helical stacks are heavily represented in their constituent proteins. The challenge associated with these projects stems from the sheer size and complexity of the NPC, given that it regulates all nucleocytoplasmic trafficking in addition to simply coating a membrane, and that so little is structurally known about the BBSome, which transports membrane proteins from the plasma membrane to the primary cilium. In this body of work, an atomic structure of the essential Y-shaped scaffolding complex of the NPC and a structure of BBS9, a component of the BBSome, was sought.

Nuclear pore complex

As the storehouse of the cell's genetic material and the site of transcriptional activity, access to and egress from the nucleus must be tightly regulated. Pores, which perforate the nuclear membrane, are the main gateways linking nucleus and cytoplasm. Transport through these pores is controlled by the NPC, which coats the nuclear envelope with scaffold forming proteins that anchor phenylalanine glycine (FG)-repeat containing proteins. These disordered FG-repeats protrude into the central channel where, through Van der Waals interactions, they form a size-selective hydrogel that serves as a bidirectional molecular sieve (Hülsmann et al., 2012; Frey et al., 2006; Ribbeck and Görlich, 2001). Small molecules and proteins smaller than ~40 kDa freely diffuse through the NPC barrier. However, larger protein cargoes require a

nuclear localization signal for entry that is bound by a nuclear transport receptor (NTR), or karyopherin, that in turn binds FG-repeats. By disrupting FG-FG interactions, the NTR effectively melts its way through the hydrogel and, unexpectedly, large NTR-bound cargoes traverse the pore faster than small, non-FG interacting cargo (Ribbeck and Görlich, 2001). The driving force for nucleocytoplasmic transport is a gradient of the small GTPase Ran. A pool of Ran-GTP exists in the nucleus, due to the presence of the Ran guanine nucleotide exchange factor (GEF), which binds NTRs and causes cargo dissociation. An NTR loaded with Ran-GTP is competent to bind nuclear localized proteins with a nuclear export signal so that it can traverse the NPC. In the cytoplasm, Ran-GTPase activating proteins (GAPs) induce GTP hydrolysis and complex dissociation (reviewed by Cook et al., 2007). mRNA export occurs by a distinct mechanism mediated in part by the cytoplasmic, NPC-associated DEAD-box ATPase Dpb5 (reviewed by Folkmann et al., 2011). The NPC is capable of accommodating large cargoes such as ribosomal subunits and herpes B virus capsids, up to 39 nm in size (Panté and Kann, 2002).

With a mass of roughly 60-120 MDa, hundreds of proteins per complex, and a diameter of roughly 120 nm, the NPC is one of the largest macromolecular assembly in the cell. Given this and its central role in eukaryotic cell biology, the structure of the assembled NPC has long been sought after. Several groups are working to obtain a structure of the intact NPC through cryo-electron tomography (cryo-ET) on native nuclear envelopes from several metazoan species (Bui et al., 2013; Elad et al., 2009; Eibauer et al., 2015). These studies have revealed important ultrastructural details of the NPC, including that three discrete, stacked rings form the complex (Fig. 1A, B). Although the latest cryo-ET reconstruction achieved ~20 Å resolution (Eibauer et al., 2015), without higher resolution crystal structures of individual components their placement, and the details of their contacts, within the assembled NPC will prove difficult to determine. Simplifying this ambitious goal of complete structural determination is the fact that the NPC is a repeating and modular assembly, formed by ~30 different proteins (nucleoporins

or nups) that can be fractionated into immunoprecipitation stable modules, or subcomplexes. The eight-fold rotational symmetry of the yeast and vertebrate NPC observed in early EM micrographs is a hallmark of this repeating structure (Akey and Radermacher, 1993; Rout and Blobel, 1993). These subcomplexes are the: Nup82 complex, Nsp1 complex, Nup145N complex, nuclear basket, Ndc1 complex, Nic96 complex, and Y complex (Fig. 2). As an orthogonal and complementary approach to the study of intact NPCs, our laboratory and others have undertaken a divide-and-conquer strategy with regard to a structural solution, tackling one subcomplex at a time (Schwartz, 2005; Brohawn et al., 2009). The following sections are a broad survey of what has been elucidated, thus far, about the structure and assembly of the NPC subcomplexes.

Nup82 complex

The Nup82 complex is localized to the cytoplasmic ring of the NPC and, through its FG-repeat containing proteins, serves as the first barrier to nuclear import. Additionally, the Nup82 complex functions as a scaffold for the recruitment of factors that mediate mRNA transport. The core complex comprises: scNup159 (hsNup214), scNup82 (hsNup88) and scNsp1 (hsNup62),¹ which bind to one another primarily via coiled-coils in their C-terminal domains to form a 1:1:1 complex based on size exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALLS) (Belgareh et al., 1998; Gaik et al., 2015). scNup159 is formed by an N-terminal β -propeller, a predicted-to-be α -helical C-terminal domain (CTD), and an intervening disordered ~800 amino acid segment (Weirich et al., 2004). The scNup159 β -propeller recruits the DEAD-Box ATPase Dpb5, which terminates mRNA export through dissociation of the messenger RNP complex, and in turn binds the export factor Gle1 (Hodge et al., 1999). The atomic details of this interaction have been determined crystallographically (Montpetit et al., 2011). scNup42, an FG-repeat containing Nup, has not been structurally characterized but is known to bind Gle1 (Strahm et al., 1999). The disordered region of scNup159 comprises FG-repeats and, at least in

S. cerevisiae, several unstructured binding sites for dynein light chain (Dyn2) implicating the Nup82 complex in tethering the NPC to microtubules and proper nucleus positioning within the cell (Stetter et al., 2007). Additionally, scNup159 dimerizes via its CTD (Gaik et al., 2015). scNup82 comprises an N-terminal β -propeller and a C-terminal predicted α -helical domain (Yoshida et al., 2011). Crystal structures of the scNup82 β -propeller bound to the C-terminal domains of hsNup98 (ortholog of scNup145N) and scNup116 (absent from metazoa), both of the Nup145N subcomplex, reveal at least one intersubunit contact that is responsible for integrating the Nup82 complex into the greater NPC (Yoshida et al., 2011; Stuwe et al., 2012). hsNup98 also binds the Y complex and its yeast ortholog Nup145N has been shown to interact with scNup157 of the Nic96 complex (Lutzmann et al., 2005; Hodel et al., 2002). scNsp1 contains an unstructured, N-terminal FG-repeat domain and a C-terminal coiled coil domain, a fragment of which has been solved from the metazoan ortholog rnNup62 (Solmaz et al., 2011). Recent EM characterization of a recombinant scNup159 CTD, scNup82, scNsp1 CTD, and Dyn2 complex resulted in a negative-stain, electron tomography reconstruction where the complex appears “P” shaped (Gaik et al., 2015). Although the authors went on to dock the available crystal structures into this density, this placement is still speculative. Nup358, which is absent from yeast, binds to the Nup82 complex in metazoa and forms 36nm long filaments that project into the cytoplasm and serve as initial docking sites for karyopherin-bound cargo destined for the nucleus (Delphin et al., 1997).

Nsp1 (Nup62) complex

The Nsp1 complex localizes to the central (spoke) ring of the NPC and forms the transport barrier in the central channel. This stable complex of three proteins, isolated from *R. norvegicus* liver nuclei, was the first characterized Nup subcomplex (Finlay et al., 1991). The Nsp1 complex comprises: scNsp1 (hsNup62), scNup49 (hsNup58), and scNup57 (hsNup54) (Finlay et al., 1991; Brohawn et al., 2009). Though scNsp1 is also a component of the Nup82

complex, described earlier, it binds scNup82 and scNup57 in a competitive manner (Bailer et al., 2001). scNsp1, scNup49, and scNup57 have N-terminal disordered FG-repeat domains and C-terminal coiled coils, through which they interact with one another (Guan et al., 1995; Brohawn et al., 2009). Equilibrium analytical ultracentrifugation (AUC) and a crystal structure of the reconstituted metazoan complex comprising the full length coiled coil domains of Nup62, Nup58, and Nup54 demonstrate that they form a 1:1:1 trimeric complex (Ulrich et al., 2014; Chug et al., 2015). The Nup62 complex structure is formed by a parallel heterotrimeric coiled coil followed by a second coiled coil region that folds back onto the first to form an antiparallel, six helix bundle (Chug et al., 2015). Such an elaborate structure likely serves additional purposes in the NPC besides bringing the Nup62 complex together. For one, the entire Nup62 complex is required to interact with the Nic96 coiled coil domain, thereby tethering this subcomplex to the NPC scaffold (Chug et al., 2015).

Nuclear basket

The nuclear basket is not part of the core, three ring structure of the NPC but is instead located under the nucleoplasmic ring, where it extends far into the nucleus. Primarily comprised of Tpr in metazoa and Mlp/Mlp2 in yeast (Brohawn et al., 2009), the basket is formed by a series of flexibly tethered coiled coil domains, similar to beads-on-a-string, based on structure prediction and limited proteolysis experiments (Niepel et al., 2013). These findings extend to the vertebrate nuclear basket as well (Krull et al., 2004). However, no atomic resolution structural information is currently available on Tpr or Mlp1/2. The basket is thought to keep the transport channel free of heterochromatin and it interacts with a host of nuclear proteins, including the desumoylating enzyme Ulp1 in yeast, transiting mRNPs, the proteasome, and telomere silencing factors (Li and Hochstrasser, 2000; Niepel et al., 2013). Interestingly, the nuclear basket is absent from NPCs located in the nucleolus, possibly to better accommodate large ribosomal subunits (Galy et al., 2004).

Ndc1 complex

The Ndc1 complex comprises four proteins with predicted transmembrane helices and as such represents the main tether between the soluble NPC scaffold and the nuclear envelope, into which these proteins embed. The Ndc1 complex is located next to the central (spoke) ring of the NPC. Conserved between yeast and metazoa are two transmembrane Nups: scNdc1 (hsNdc1) and scPom33 (hsTMEM33)². Although Pom33 is conserved, its interaction with the Ndc1 complex has not been shown (Kabachinski and Schwartz, 2015). Yeast and metazoa each have two additional transmembrane Nups that interact with Ndc1 but have no clear homolog in the other clade: scPom121 and scPom34, in yeast, and hsPom152 and hsGp210 in metazoa (Rothballer and Kutay, 2013). It is possible that the poor conservation of Ndc1 complex members between yeast and metazoa is due to the different functional requirements of open vs. closed mitosis on these transmembrane Nups.

No component of the Ndc1 complex has been structurally determined. However, structural prediction and topology mapping experiments, such as proteolysis protection assays, have revealed certain details about their folds. For instance, Ndc1 contains 6 transmembrane helices and a soluble, C-terminal globular domain. Both the N- and C- termini are exposed to the transport channel and are capable of interacting with other NPC subcomplexes (reviewed by Rothballer and Kutay, 2013). Higher resolution structural information on Ndc1 subcomplex members can reveal the atomic details of how the assembled pore is tethered, as it is known that scNup53 and scNup170 of the Nic96 complex bind scNdc1 and scPom121, respectively (Marelli et al., 2001; Onischenko et al., 2009). Additionally, how the NPC is re-assembled after open mitosis in metazoa through the recruitment of Ndc1 containing vesicles by the chromatin binding protein ELYS, a member of the Y complex, is still an open question (Rasala et al., 2008).

Nic96 complex

The scNic96 complex, or the hsNup93 complex in metazoa, is a ~0.5 MDa module located in the central spoke ring of the NPC which serves to bridge the FG-repeat network in the central channel, in the form of the Nsp1 complex, to the nuclear membrane. This subcomplex comprises scNic96 (hsNup93), scNup157/170 (hsNup155), scNup53/59 (hsNup35), and scNup188 (hsNup188) or scNup192 (hsNup205) (reviewed by Brohawn et al., 2009). The compositional heterogeneity of the Nic96 complex is apparent. A whole genome duplication event occurred during yeast evolution, which produced the scNup157/170 and scNup53/59 paralogs that exist as a single gene product in metazoa (Wolfe and Shields, 1997). These paralogs have significantly diverged in sequence, with scNup53/59 only 33% identical and scNup157/70 44% identical, yet they are very similar structurally according to homology modeling. scNup188 and scNup192, which diverged prior to the whole genome duplication event, are structurally related and bind competitively to the Nic96 complex, based on interaction with the same helix of Nic96 located immediately after its N-terminal coiled coil domain (Flemming et al., 2012; Stuwe et al., 2014).

scNic96 comprises an N-terminal coiled coil domain, which has not been structurally characterized but is known to bind to the Nsp1 subcomplex, and a C-terminal α -helical domain that has been solved crystallographically (Bailer et al., 2001; Jeudy and Schwartz, 2007; Schrader et al., 2008). The irregular topology of this α -helical domain is structurally related to Sec31 of the COPII vesicle coat, demonstrating the evolutionary relationship between these membrane coating complexes (Devos et al., 2004; Brohawn et al., 2008). scNup157/170 is a structural homolog of scNup133, a member of the Y complex, and as such contains an N-terminal β -propeller flexibly tethered to an irregular α -helical stack that is highly dissimilar to scNic96 (Flemming et al., 2009; Whittle and Schwartz, 2009; Seo et al., 2013). Studies of Nup188 and the related Nup192 from several organisms, including *C. thermophilum*, *M. thermophila*, and *S. cerevisiae*, by negative stain EM revealed an overall S-shape that is highly

reminiscent of karyopherins (Amlacher et al., 2011; Flemming et al., 2012; Sampathkumar et al., 2013; Andersen et al., 2013). Much like karyopherins, Nup188 and Nup192 are flexible and various conformations, from 'open' to 'closed', have been observed by negative stain EM and confirmed to exist in solution by small angle X-ray scattering (SAXS) (Sampathkumar et al., 2013; Andersen et al., 2013). Crystal structures have revealed the details of this fold (Sampathkumar et al., 2013; Andersen et al., 2013; Stuwe et al., 2014). Nup188 comprises an N-terminal irregular α -helical domain that forms a closed, right-handed superhelical ring, a C-terminal regular α -helical repeat domain, and an intervening α -helical segment that has not been solved crystallographically (Andersen et al., 2013). Interestingly, an SH3-like motif is inserted within the N-terminal domain that is absent from Nup192 (Andersen et al., 2013). It is curious that despite this additional functionality, scNup192 is essential in *S. cerevisiae* but scNup188 is not (Kabachinski and Schwartz, 2015). *Stuwe et al.* captured a partially open state of the scNup192 N-terminal domain, which is open to different extents in the two copies in the asymmetric unit (ASU). Hinge helices that bisect the α -helical ring allow it to open through a rigid body motion of the two halves (Stuwe et al., 2014). scNup53 and scNup59 contain N- and C-terminal disordered regions connected by an intervening homodimerization domain, based on a structure of the ortholog hsNup35 (Handa et al., 2006). A C-terminal amphipathic α -helix is additionally present that binds Ndc1 and the nuclear membrane (Marelli et al., 2001; Onischenko et al., 2009). Nup192 and Nic96 bind to the N-terminal disordered region of Nup53 (Amlacher et al., 2011; Stuwe et al., 2014). Nup170 and Kap121, a karyopherin, bind to the C-terminal disordered region (Amlacher et al., 2011; Lusk et al., 2002).

It is surprising in light of these numerous interactions, revealed by pulldown experiments, that the double deletion of Nup53 and Nup59 in *S. cerevisiae* is viable (Kabachinski and Schwartz, 2015). Clearly, other interactions between Nic96 complex members must exist. Despite individual crystal structures of Nic96 complex members, no binary or higher order assemblies have been solved. Towards the goal of an assembly structure, the Nic96

heterotetramer, comprising Nic96, Nup170, Nup53, and Nup192, was recently reconstituted from the thermophilic fungus *Chaetomium thermophilum* (Amlacher et al., 2011). A single particle or random conical tilt (RCT) EM reconstruction this assembled Nic96 complex will define its currently unknown ultrastructure. Other open questions are: to what extent does the incorporation of Nup188 vs. Nup192 affect the function of the Nic96 complex? Is this incorporation stochastic or are defined mixtures of Nup188 and Nup192 containing Nic96 complexes present in the same NPC? Perhaps different NPCs, with specifically tailored functions such as transport of ribosomal subunits, have Nic96 complexes with only one of these two proteins.

Y complex

The Y, or scNup84, complex localizes to the cytoplasmic and nucleoplasmic rings of the NPC and in *S. cerevisiae*, comprises seven proteins: scNup120 (hsNup160), scNup85 (hsNup85), scNup145C (hsNup96), scNup84 (hsNup107), scNup133 (hsNup133), scSeh1 (hsSeh1), and scSec13 (hsSec13) (Brohawn et al., 2009). In addition to serving as a docking site for transport barrier forming FG-Nups, the Y complex also makes contacts with the nuclear envelope to presumably stabilize the high curvature of the membrane. scNup145N, an FG-repeat containing Nup, is known to bind to the unstructured N-terminal peptide of scNup145C via its C-terminal autocatalytic domain based on a crystal structure of the metazoan orthologs (Hodel et al., 2002). scNup145 is translated as a single polypeptide that is subsequently autocatalytically cleaved into scNup145N and scNup145C. Nup133, in *H. sapiens*, has been shown to have a membrane binding ALPS motif (Berke et al., 2004).

The Nup84 complex forms a Y-shape, as revealed by negative stain TEM, with Nup120, Nup85, and Seh1 comprising the short arms of the Y and Nup145C, Sec13, Nup84, and Nup133 forming the long stem (Lutzmann et al., 2002). The Y complex is the best structurally characterized NPC subcomplex to date, with numerous EM reconstructions of the complete Y

and crystal structures accounting for 90% of the total mass of the 575 kDa complex (Fig. 3A). Prior to this work, the molecular details of how the short arms of the Y contact the long stem were not known (Fig. 3B).

Crystal structures reveal that Nup120 is formed by an N-terminal domain containing a 7-bladed β -propeller, with a helical bundle insertion between blades 6 and 7, that is integrated into an irregular α -helical stack and a C-terminal α -helical domain comprising HEAT repeats with a regular topology (Bilokapic and Schwartz, 2012; Leksa et al., 2009; Seo et al., 2009; Liu et al., 2012). The N-terminal domain, which forms a rigid unit, is juxtaposed to the elongated C-terminal domain (CTD) by 90° (Bilokapic and Schwartz, 2012). It is through this CTD that Nup120 interacts with Nup85 and Nup145C, based on gel filtration binding experiments (Bilokapic and Schwartz, 2012). Seh1 and Sec13 are both open, 6-bladed β -propellers that interact with their binding partners, Nup85 and Nup145C respectively, through the contribution of a 7th blade in *trans* (Brohawn et al., 2008; Brohawn and Schwartz, 2009). Sec13 is also a component of the COPII vesicle coat (Fath et al., 2007). Apart from its N-terminal β -stranded insertion blade, necessary for Seh1 binding, Nup85 is an all α -helical protein with an irregular, U-shaped topology (Brohawn et al., 2008; Debler et al., 2008). The N terminus is in the middle of the helical stack, which traces upward before folding back on itself and tracing downward, past the N terminus to the carboxyl end. This double-layering of helices adds rigidity to the stack and prevents the kind of exaggerated, super helical flexing possible in regular α -solenoids, such as that observed in HEAT repeat proteins (Brohawn et al., 2008; Kappel et al., 2010). This topology and domain architecture of Nup85, termed the ancestral coatomer element 1 (ACE1), is shared by Nic96, Nup85, and Nup145C within the NPC, as well as Sec16 and Sec31 of the COPII vesicle coating system (Brohawn et al., 2008; Whittle and Schwartz, 2010). These ACE1 folds can be further broken down into three subdomains: the trunk (middle domain), the crown (top), and the tail (bottom). These subdomains were initially identified based on their susceptibility to proteolysis, hence the crystallization of numerous ACE1 proteins with their tail

domains deleted (Brohawn et al., 2008; Brohawn and Schwartz, 2009; Debler et al., 2008; Nagy et al., 2009; Hsia et al., 2007). In fact, the structure of the Nup85 and Nup145C tail domains, both of which interact with Nup120, was not known prior to the work presented here. The Nup84 tail domain, from the human ortholog hsNup107, has been solved in complex with a hsNup133 fragment (Boehmer et al., 2008). Additionally, the structure of scNup84 bound to scNup145C, in a crown to crown interface, is solved (Brohawn and Schwartz, 2009; Nagy et al., 2009). Nup133 and its closest homolog Nup157/170, of the Nic96 complex, share a unique fold that is highly dissimilar to Nup120 and the ACE1 proteins of the Y (Whittle and Schwartz, 2009). hsNup133 comprises an N-terminal 7-bladed β -propeller and an irregular, C-terminal α -helical stack (Berke et al., 2004; Whittle and Schwartz, 2009). This α -helical domain can be partitioned into four structural elements: 1. an N-terminal plane, formed by the flat arrangement of a dozen long helices in *H. sapiens* 2. a series of loosely-packed, short helices with a zig-zag topology 3. four helices arranged in a tight helical bundle that serve as the Nup107 binding interface and 4. a compact C-terminal subdomain (Whittle and Schwartz, 2009; Boehmer et al., 2008; Sampathkumar et al., 2011).

Subsequent structural analysis by electron microscopy has revealed that the Y complex is quite flexible (Lutzmann et al., 2002; Kampmann and Blobel, 2009; Bui et al., 2013). This flexibility is most notable in the scNup84-scNup133 stem, which can vary in conformation from bent to nearly straight at this resolution. Even greater stem flexibility has been observed for the *H. sapiens* Y complex (Bui et al., 2013). Elastic network modelling by Whittle and Schwartz demonstrated that the zig-zag portion of the helical domain of hsNup133 can serve as a hinge and allow the N-terminal plane to flex, along the direction of the stack, relative to the hsNup107 binding helices (Whittle and Schwartz, 2009). Additionally, the N-terminal β -propeller is flexibly tethered to the C-terminal α -helical domain, based on SAXS and chemical crosslinking data (Whittle and Schwartz, 2009; Kim et al., 2014b; Shi et al., 2014). Movement at these two

junctions likely explains the flexibility observed by EM. The extent to which this flexibility is required for NPC assembly or function is not known.

The composition of the Y complex varies across eukaryotes. Although the minimal Y has long been assumed to be a heptamer based on sequence conservation, metazoa have additional stably-associated Y complex members including hsNup37, hsNup43, and hsELYS (Neumann et al., 2010). However, the conserved Y complex elements are highly similar in structure, based on crystal structure comparison and modelling, despite low sequence identity. Notable elaborations include hsNup160, which contains an additional 40 kDa predicted-to-be α -helical domain, C-terminal of the putative Nup145C binding site, that is absent from ScNup120. These species-specific additions to the minimal Y, which were identified in a proteomic screen of the mammalian NPC (Cronshaw et al., 2002), are believed to largely decorate the complex without affecting its core structure, based on structure modelling. Nup37, a seven-bladed β -propeller, stably interacts with the Y complex and binds Nup160 at the hinge between the N-terminal domain and the C-terminal domain, likely reducing conformational flexibility (Loiodice et al., 2004; Bilokapic and Schwartz, 2012). hsNup43 is a seven-bladed β -propeller that stably associates with the human Y complex and binds hsNup85, though its exact binding site has not been experimentally determined (Loiodice et al., 2004; Kim et al., 2014a). ELYS is a ~250 kDa protein that contains an N-terminal seven-bladed β -propeller which has been solved crystallographically, a central α -helical domain that interacts with the Y by binding Nup160 and Nup37, and a disordered C-terminal domain (Rasala et al., 2006; Bilokapic and Schwartz, 2012; Bilokapic and Schwartz, 2013). ELYS functions by binding to AT-rich chromatin, through its disordered domain, during open mitosis to recruit Pom121-Ndc1 containing vesicle and initiate NPC reassembly (Rasala et al., 2008). Given this, the absence of ELYS from lower eukaryotes that undergo closed mitosis is unsurprising. Although unicellular eukaryotes such as *S. pombe*, but not *S. cerevisiae*, have a homolog to the ELYS central helical domain they lack the C-terminal disordered domain and N-terminal β -propeller (Bilokapic and Schwartz, 2012).

It is tempting to speculate on what affect, if any, these species-specific additions and subtractions of Y complex Nups have on the overall assembly, stoichiometry, and dimensions of the NPC. Early measurements on isolated NPCs indicated that the *Xenopus* NPC was ~45nm larger in height and ~50nm larger in diameter than the *S. cerevisiae* NPC (Akey and Radermacher, 1993; Yang et al., 1998). A similar height difference of ~40nm has also been observed between the *Dictyostelium discoideum* and *Xenopus* NPC in cryo-electron tomography reconstructions from intact nuclei and native nuclear envelope spreads, respectively (Elad et al., 2009). However, the diameter of the vertebrate and non-vertebrate NPC was roughly the same at ~120nm (Elad et al., 2009), indicating that the difference observed between isolated NPCs was likely a technical artifact. Also, mass estimates for the yeast and metazoan (xenopus) NPC are quite different, at ~45-70 MDa and ~125 MDa, respectively (Rout and Blobel, 1993; Rout et al., 2000; Cronshaw et al., 2002; Reichelt et al., 1990). One possibility is that these species-specific size differences are due to a different number of Y complexes per NPC, where higher order Y assemblies depend on a protein absent from lower eukaryotes.

Complicating this issue, the number of Y complexes per NPC is still debated, although it is believed to be a multiple of eight given the eight-fold rotational symmetry of the yeast and vertebrate NPC observed in EM micrographs (Akey and Radermacher, 1993; Rout and Blobel, 1993). Quantitative immunoblotting in yeast and quantification of SDS-PAGE band intensities from rat liver nuclei yielded estimates of 16 Y complexes per NPC (Rout et al., 2000; Cronshaw et al., 2002). Further, direct counting of GFP-tagged Nups in yeast, by observing discrete photobleaching events with SPEED microscopy, supports 16 Y complexes per NPC (Mi et al., 2015). However, a study in human cell lines combining quantitative mass spectrometry with super resolution microscopy yielded 32 Y complexes per NPC (Ori et al., 2013).

Towards understanding the higher order Y complex assembly in the NPC, Beck and colleagues have generated a 32 Å cryo electron tomography (ET) reconstruction of the entire

human NPC (Fig. 1A, B, Bui et al., 2013). Based on the docking of a 35 Å random conical tilt, negative stain EM reconstruction of the isolated human Y complex vertex, lacking Nup133, into this cryo-ET density, these authors propose a double Y ring model. This model predicts two symmetric double Y rings, one on the cytoplasmic face and the other on the nucleoplasmic face of the NPC, that each contain two parallel, eight-membered Y complex rings arranged in a head-to-tail fashion for a total of 32 Ys per NPC. On each face, the parallel inner and outer Y rings are staggered or shifted in register with respect to each other, and as such contribute a set of non-equivalent contacts to one another (Bui et al., 2013). Although a dimeric Y complex has been recently observed in negative-stain EM, this assembly is inconsistent with the proposed two ring model (Thierbach et al., 2013). Recent fluorescence anisotropy and super-resolution microscopy studies, although supporting a head-to-tail arrangement of neighboring Ys, lacked the resolution to distinguish between a single (8 copy) or double (16 copy) Y complex ring per face of the NPC (Kampmann et al., 2011; Szymborska et al., 2013). Additionally, these methods assume that the Y rings on each face of the NPC are symmetrical, which is not necessarily the case.

A recent 20 Å cryo-ET reconstruction of the *Xenopus* NPC, by Medalia and colleagues, principally supports the Beck double ring model through fitting of the human Y complex reconstruction but only for the cytoplasmic ring (Fig. 1C, Eibauer et al., 2015). The authors note that at this higher resolution the apparent symmetry between the cytoplasmic pore ring and nucleoplasmic pore ring, observed in the human reconstruction, breaks down. Medalia and colleagues also observe that the apparent composition of the inner and outer Y complex of the double ring is different (Eibauer et al., 2015). However, at this resolution it is speculative to comment on which proteins contribute to this difference, if real. The absence of Nup133 and ELYS, the largest addition to the metazoan Y, in this docking approach are caveats to be considered.

Fitting is tentative at best here given that we now know the random conical tilt, Y complex reconstruction used by Bui et al. suffers from the negative-stain flattening problem, based on the data presented in this thesis. Additionally, the use of crystal structures in docking is challenging given the low resolution of the tomograms and the gap in resolution between it and the plethora of available, high resolution structures. Although these crystal structures reveal clear structural differences between the Y complex components, at 20 Å these extended, topologically distinct α -helical stacks are indistinguishable from one another. Exacerbating the fitting problem further, prior to this work we did not yet have a complete atomic structure of an individual Y. Specifically, how the short arms of the Y contact the long stem, in a region referred to as the 'hub', was not known. The work presented in this thesis has led to a structure of the Nup120-Nup85-Nup145C-Sec13 Y complex hub and has allowed us to construct a near complete, atomic resolution composite structure of the entire Y.

One interesting possibility to consider is that different assembly states, and as a consequence different stoichiometries, exist in metazoa compared to yeast. One can speculate that a double head-to-tail Y ring, of 32 copies exists in metazoa, whereas a single ring of 16 copies is present in yeast, given the disparate stoichiometry measures and that Bui et al. observe an XL-MS crosslink between hsNup43 and hsNup96 (scNup145C) that by definition has to be species specific (Nup43 is absent from yeast). The authors conclude that this crosslink is between the inner and outer Y copies. However, this is speculative given that the exact binding site of Nup43 on the Y complex is not known. I find it likely that the basic assembly mode and stoichiometry of Y complexes is conserved across all eukaryotes, given the structural conservation of these Nups on the atomic level.

BBSome

The cilium is a finger-like protrusion from the plasma membrane formed by nine radially arranged microtubule doublets termed the axoneme, which is nucleated from a basal body (Nachury et al., 2010). Cilia can be classified into two functional groups: motile cilia, or flagella, are used for fluid movement or cell propulsion whereas non-motile, or primary, cilia are used for signal transduction (Nachury et al., 2010). This includes the developmentally important hedgehog signaling pathway in vertebrates and phototransduction, earning this organelle the nickname the 'cellular antenna' (Nachury et al., 2010; Ishikawa and Marshall, 2011; Milenkovic et al., 2009). Motile cilia are present in animals and protists but are generally absent from fungi and plants (Nachury et al., 2007). Primary cilia are found in nearly all cell types of vertebrates, except for those derived from a hematopoietic lineage, and are absent from all invertebrate cells except neurons (Pazour and Witman, 2003).

The cilium is a privileged cellular compartment with a unique composition of soluble and membrane proteins (Nachury et al., 2010). Similar to the nuclear membrane, which is a specialized extension of the endoplasmic reticulum, the ciliary membrane is contiguous with the plasma membrane yet functionally distinct. Specificity of the ciliary membrane appears largely due to the exclusion of non-resident proteins and the retention of bona fide ciliary proteins (Hu et al., 2010; Chih et al., 2011). Resident plasma membrane proteins tagged with GFP fail to enter the cilium, whereas short-time scale fluorescent recovery after photobleaching (FRAP) experiments show no recovery of a bleached ciliary reporter protein indicating a lack of exchange with the plasma membrane (Hu et al., 2010). A lateral diffusion barrier at the base of the primary cilium, formed in part by a ring of the Septin2 oligomeric GTPase, has been characterized (Hu et al., 2010; Chih et al., 2011). A precedent for septins in the formation of membrane diffusion barriers has already been established in *S. cerevisiae*, specifically at the mother-bud neck during cell division (Gladfelter et al., 2001). Unexpectedly, the soluble diffusion

barrier into the primary cilium may be formed, in part, by nucleoporins (Kee et al., 2012). Immunofluorescence studies have revealed the localization of several Nups, including the FG-containing hsNup62, to the base of the primary cilium (Kee et al., 2012). Additional evidence for a 'ciliary pore complex' include the participation of importin β 2 and Ran-GTP in the localization of a motor protein to the primary cilium (Dishinger et al., 2010).

The mutation of proteins involved in cilia function leads to a host of human genetic disorders termed ciliopathies, which have symptoms ranging from retinal degeneration and cognitive defects to kidney cysts and polydactyly (Fliegauf et al., 2007). Bardet-Biedl syndrome (BBS) is one such autosomal recessive disorder that is caused by mutation in one of 14 BBS genes identified in patients (Fliegauf et al., 2007; Jin et al., 2010). Seven of these genes, BBS1, 2, 4, 5, 7, 8 and 9, along with BBIP10, form an immunoprecipitation stable complex termed the BBSome (Nachury et al., 2007; Loktev et al., 2008).

The BBSome, through multiple functions, is involved in the establishment or maintenance of the unique protein composition of the ciliary membrane (Jin et al., 2010; Wei et al., 2012). Indeed, the BBSome binds directly to the ciliary targeting signal of SSTR3, a ciliary G-protein coupled receptor, which fails to reach neuronal cilia in BBS4 and BBS2 knockout mice (Jin et al., 2010; Berbari et al., 2008). The BBSome also forms distinct punctae on liposomes dependent on the GTP-bound, small GTPase Arl6, leading to the hypothesis that the BBSome is a membrane coating complex (Jin et al., 2010). Canonical vesicle coats are similarly recruited to membranes via small GTPases that insert into membranes in a GTP-dependent manner (Schekman and Orci, 1996). However unlike the canonical coats, the BBSome localizes to its putative acceptor compartment, the cilium, and not the vesicle donor compartment (Schekman and Orci, 1996; Jin et al., 2010). This observation and the failure of the BBSome to deform liposomes led to a hypothesis in which the complex is predicted to form a planar, polymerized coat that transports membrane proteins from the plasma membrane into the ciliary membrane by assisted lateral diffusion and without membrane fission or fusion (Jin et al., 2010).

The BBSome additionally serves as an adaptor complex by bridging the intraflagellar transport (IFT) complex, which interacts with microtubule motor proteins, to the integral membrane protein cargo. In line with this function, the BBSome assembles IFT complexes at the ciliary base for anterograde transport up the axoneme and recycles IFT particles at the ciliary tip for retrograde movement down the axoneme (Wei et al., 2012). Data consistent with this view are the interaction of BBS1 with WDR19 of the IFT-B subcomplex, trafficking of the BBSome along the axoneme at the same rate as IFT complexes, failure of IFT subcomplexes to bind to one another in BBS7 or 8 mutants, and a BBS1 missense mutation that leads to compromised retrograde transport of IFT-B and accumulation of this complex at the ciliary tip (Wei et al., 2012; Nachury et al., 2007; Ou et al., 2005). The BBSome also functions in the maintenance of the axoneme in G0 cells. The disassembly of cilia is necessary for proper cell cycle progression, as the cilium is nucleated from the mother centriole (Li et al., 2011). Disassembly is regulated by the centrosomal mitotic kinase Aurora A, which phosphorylates and activates histone deacetylase 6 (HDAC6, Pugacheva et al., 2007). HDAC6 in turn deacetylates acetylated alpha-tubulin, which is necessary for ciliary resorption (Loktev et al., 2008; Pugacheva et al., 2007). BBIP10 (BBSome interacting protein of 10 kDa) has been shown to associate with HDAC6 and inhibit tubulin deacetylation (Loktev et al., 2008).

Consistent with its putative role as a membrane coating and protein transport assembly, structure prediction has demonstrated that the domain topology of the BBSome building blocks are similar to COPI, COPII, and clathrin vesicle coats (Jin et al., 2010). BBS2, 7, and 9 are all predicted to be composed of an N-terminal β -propeller, a gamma adaptin ear (GAE) sandwich domain, a mixed platform domain, and a C-terminal globular α -helical domain (Jin et al., 2010). No binding contacts of BBS2 with components external to the BBSome have been described. BBS7 interacts with the CCT/TRiC chaperonin in what is believed to be an early step in BBSome assembly (Seo et al., 2010; Zhang et al., 2012). BBS9 binds the cytoplasmic protein LZTFL1, which regulates ciliary entry by inhibiting it (Seo et al., 2011). BBS1 similarly comprises

an N-terminal β -propeller, which has been solved in complex with the Arl6-GTPase, and a C-terminal GAE domain which binds the Rab8-GEF (Mourão et al., 2014; Jin et al., 2010; Nachury et al., 2007). This interaction with Arl6 is responsible for tethering the BBSome to membranes, but is not the only contact to lipids. BBS5, which comprises two predicted pleckstrin homology-like domains, has been shown to bind phosphatidylinositol phosphate (PIP) *in vitro* (Jin et al., 2010; Nachury et al., 2007). BBS4 and BBS8 are both extended α -helical stacks composed of TPR repeats (Jin et al., 2010). Both proteins bind PCM-1, which is responsible for localizing the BBSome to the base of the primary cilium at centriolar satellites (Kim et al., 2004; Ansley et al., 2003).

No ultrastructure of the BBSome has been observed. Although we do not know what the assembled complex looks like, binding experiments have led to a preliminary understanding of the interaction network within an assembled BBSome. Co-IP experiments, following co-transfection, revealed that BBS1, BBS2, BBS4, BBS5, and BBS8 all interact with BBS9 (Nachury et al., 2007). In contrast, BBS7 binds stoichiometrically to BBS2 but does not interact with BBS9 (Nachury et al., 2007). Based on pulldowns using binding-deficient BBS point mutants, a BBS2-BBS7-BBS9 core complex has been proposed that additionally may serve as an assembly intermediate (Zhang et al., 2012). It is because of this apparent centrality of BBS9 to BBSome complex assembly that I targeted it for structural and biophysical analysis. The work presented in this thesis has led to a high resolution structure of the BBS9 N-terminal domain.

Footnotes

¹Nucleoporins (Nups) are most commonly named based upon their molecular weight, either predicted or determined by MALDI-TOF mass spectrometry, because they were initially identified as unknown proteins with a defined size in pulldown experiments. This nomenclature, although descriptive, is confusing when discussing Nups for different species because even though two Nups are orthologs they often have different names as a result of molecular weight differences. This difference is apparent between yeast and metazoa. Where appropriate, I will name both the *S. cerevisiae* and corresponding *H. sapiens* Nup name. I will default to the *S. cerevisiae* name only for clarity, unless stated otherwise.

²Members of the Ndc1 complex do not follow the standard nucleoporin nomenclature. Ndc1, nuclear division cycle protein 1. Pom, pore membrane protein of x kDa. TMEM33, transmembrane protein of 33 kDa. GP210, glycoprotein of 210 kDa.

References

- Akey, C.W., and Radermacher, M. (1993). Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. *The Journal of Cell Biology* 122, 1-19.
- Amlacher, S., Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D.P., Arumugam, M., Bork, P., and Hurt, E. (2011). Insight into Structure and Assembly of the Nuclear Pore Complex by Utilizing the Genome of a Eukaryotic Thermophile. *Cell* 146, 277.
- Andersen, K.R., Onischenko, E., Tang, J.H., and Kumar, P. (2013). Scaffold nucleoporins Nup188 and Nup192 share structural and functional properties with nuclear transport receptors. *Elife* 2, e00745.
- Ansley, S.J., Badano, J.L., Blacque, O.E., Hill, J., Hoskins, B.E., Leitch, C.C., Kim, J.C., Ross, A.J., Eichers, E.R., Teslovich, T.M., et al. (2003). Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. *Nature* 425, 628-33.
- Bailer, S.M., Balduf, C., and Hurt, E. (2001). The Nsp1p carboxy-terminal domain is organized into functionally distinct coiled-coil regions required for assembly of nucleoporin subcomplexes and nucleocytoplasmic transport. *Mol. Cell. Biol.* 21, 7944-55-7955.
- Belgareh, N., Snay-Hodge C., Pasteau, F., Dagher, S., Cole, C.N., and Doye, V (1998). Functional Characterization of a Nup159p-containing Nuclear Pore Subcomplex. *Molecular Biology of the Cell* 9, 3475-3492.
- Berberi, N.F., Lewis, J.S., Bishop, G.A., Askwith, C.C., and Mykytyn, K. (2008). Bardet-Biedl syndrome proteins are required for the localization of G protein-coupled receptors to primary cilia. *Proceedings Of The National Academy Of Sciences* 105, 4242-6.
- Berke, I.C., Boehmer, T., Blobel, G., and Schwartz T.U. (2004). Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *The Journal Of Cell Biology* 167, 591.
- Bilokapic, S., and Schwartz, T.U. (2012). Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proceedings Of The National Academy Of Sciences* 109, 15241.

- Bilokapic, S., and Schwartz, T.U. (2013). Structural and Functional Studies of the 252 kDa Nucleoporin ELYS Reveal Distinct Roles for Its Three Tethered Domains. *Structure* 21, 572.
- Boehmer, T., Jeudy, S., Berke, I.C., and Schwartz, T.U. (2008). Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Mol. Cell* 30, 721-31.
- Brohawn, S.G., and Schwartz, T.U. (2009). Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nature Structural and Molecular Biology* 16, 1173-7.
- Brohawn, S.G., Leksa, N.C., Spear, E.D., Rajashankar, K.R., and Schwartz, T.U. (2008). Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* 322, 1369-73-1373.
- Brohawn, S.G., Partridge, J.R., and Schwartz, T.U. (2009). The Nuclear Pore Complex Has Entered the Atomic Age. *Structure* 17, 1156.
- Bui, K.H., Appen, von, A., DiGuilio, A.L., Ori, A., Sparks, L., Mackmull, M., Bock, T., Hagen, W., Andrés-Pons, A., Glavy, J.S., et al. (2013). Integrated Structural Analysis of the Human Nuclear Pore Complex Scaffold. *Cell* 155, 1233.
- Chih, B., Liu, P., Chinn, Y., Chalouni, C., Komuves, L.G., Hass, P.E., Sandoval, W., and Peterson, A.S. (2011). A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain. *Nature Cell Biology* 14, 61-72.
- Chug, H., Trakhanov, S., Hülsmann, B., Pleiner, T., and Gorlich D. (2015). Crystal structure of the metazoan Nup62• Nup58• Nup54 nucleoporin complex. *Science express*.
- Cook, A., Bono, F., Jinek, M., and Conti, E. (2007). Structural Biology of Nucleocytoplasmic Transport. *Annual Review Of Biochemistry* 76, 647.
- Cronshaw, J.M., Krutchinsky, A.N., Zhang, W., Chait, B.T., and Matunis M.J. (2002). Proteomic analysis of the mammalian nuclear pore complex. *The Journal Of Cell Biology* 158, 915-927.
- Debler, E.W., Ma, Y., Seo, H., Hsia, K., Noriega, T.R., Blobel, G., and Hoelz, A. (2008). A fence-like coat for the nuclear pore membrane. *Mol. Cell* 32, 815-26.
- Delphin, C., Guan, T., Melchior, F., and Gerace, L. (1997). RanGTP targets p97 to RanBP2, a filamentous protein localized at the cytoplasmic periphery of the nuclear pore complex. *Molecular Biology Of The Cell* 8, 2379-90.
- Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B.T., Sali, A., and Rout, M.P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *Plos Biol.* 2, e380.
- Dishinger, J.F., Kee, H.L., Jenkins, P.M., Fan, S., Hurd, T.W., Hammond, J.W., Truong, Y.N., Margolis, B., Martens, J.R., and Verhey, K.J. (2010). Ciliary entry of the kinesin-2 motor KIF17 is regulated by importin-beta2 and RanGTP. *Nature Cell Biology* 12, 703-10.
- Dodonova, S.O., Diestelkoetter-Bachert, P., Appen, von, A., Hagen, W.J.H., Beck, R., Beck, M., Wieland, F., and Briggs, J.A.G. (2015). A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science* 349, 195.

- Eibauer, M., Pellanda, M., Turgay, Y., Dubrovsky, A., Wild, A., and Medalia, O. (2015). Structure and gating of the nuclear pore complex. *Nature Communications* 6, 7532.
- Elad, N., Maimon, T., Frenkiel-Krispin, D., and Medalia, O. (2009). Structural analysis of the nuclear pore complex by integrated approaches. *Curr. Opin. Struct. Biol.* 19, 226-32.
- Fath, S., Mancias, J.D., Bi, X., and Goldberg, J. (2007). Structure and organization of coat proteins in the COPII cage. *Cell* 129, 1325-36.
- Finlay, D.R., Meier, E., Bradley, P., Horecka, J., and Forbes, D.J. (1991). A complex of nuclear pore proteins required for pore function. *The Journal Of Cell Biology* 114, 169-83.
- Flemming, D., Devos, D.P., Schwarz, J., Amlacher, S., Lutzmann, M., and Hurt, E. (2012). Analysis of the yeast nucleoporin Nup188 reveals a conserved S-like structure with similarity to karyopherins. *Journal Of Structural Biology* 177, 99.
- Flemming, D., Sarges, P., Stelter, P., Hellwig, A., Böttcher, B., and Hurt, E. (2009). Two structurally distinct domains of the nucleoporin Nup170 cooperate to tether a subset of nucleoporins to nuclear pores. *The Journal Of Cell Biology* 185, 387-95-395.
- Fliegau, M., Benzing, T., and Omran, H. (2007). When cilia go bad: cilia defects and ciliopathies. *Nat. Rev. Mol. Cell Biol.* 8, 880-93.
- Folkmann, A.W., Noble, K.N., Cole, C.N., and Wentz, S.R. (2011). Dbp5, Gle1-IP6 and Nup159: a working model for mRNP export. *Nucleus* 2, 540-8.
- Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S.C., Kirchhausen, T., and Walz, T. (2004). Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* 432, 573-9.
- Frey, S., Richter, R.P., and Görlich, D. (2006). FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* 314, 815-7.
- Gaik, M., Flemming, D., Appen, von, A., Kastiris, P., Mücke, N., Fischer, J., Stelter, P., Ori, A., Bui, K.H., Bassler, J., et al. (2015). Structural basis for assembly and function of the Nup82 complex in the nuclear pore scaffold. *The Journal Of Cell Biology* 208, 283.
- Galy, V., Gadad, O., Fromont-Racine, M., Romano, A., Jacquier, A., and Nehrbass, U. (2004). Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* 116, 63-73.
- Gladfelter, A.S., Pringle, J.R., and Lew, D.J. (2001). The septin cortex at the yeast mother-bud neck. *Curr. Opin. Microbiol.* 4, 681-9.
- Guan, T., Muller, S., Klier, G., Blevitt, J.M., Haner, M., Paschal, B., Aebi, U., and Gerace, L. (1995). Structural analysis of the p62 complex, an assembly of O-linked glycoproteins that localizes near the central gated channel of the nuclear pore complex. *Molecular Biology Of The Cell* 6, 1591-1603.
- Haar, ter, E., Musacchio, A., Harrison, S.C., and Kirchhausen, T. (1998). Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker. *Cell* 95, 563-73.

- Handa, N., Kukimoto-Niino, M., Akasaka, R., Kishishita, S., Murayama, K., Terada, T., Inoue, M., Kigawa, T., Kose, S., Imamoto, N., et al. (2006). The crystal structure of mouse Nup35 reveals atypical RNP motifs and novel homodimerization of the RRM domain. *J. Mol. Biol.* 363, 114-24.
- Hodel, A.E., Hodel, M.R., Griffis, E.R., Hennig, K.A., Ratner, G.A., Xu, S., and Powers, M.A. (2002). The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Mol. Cell* 10, 347-58.
- Hodge, C.A., Colot, H.V., Stafford, P., and Cole C.N. (1999). Rat8p/Dbp5p is a shuttling transport factor that interacts with Rat7p/Nup159p and Gle1p and suppresses the mRNA export defect of xpo1-1 cells. *The EMBO Journal* 18, 5778.
- Hsia, K., Stavropoulos, P., Blobel, G., and Hoelz, A. (2007). Architecture of a coat for the nuclear pore membrane. *Cell* 131, 1313-26-1326.
- Hu, Q., Milenkovic, L., Jin, H., Scott, M.P., Nachury, M.V., Spiliotis, E.T., and Nelson, W.J. (2010). A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. *Science* 329, 436-9.
- Hülsmann, B.B., Labokha, A.A., and Görlich, D. (2012). The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell* 150, 738-51.
- Ishikawa, H., and Marshall, W.F. (2011). Ciliogenesis: building the cell's antenna. *Nat. Rev. Mol. Cell Biol.* 12, 222-34.
- Jeudy, S., and Schwartz, T.U. (2007). Crystal structure of nucleoporin Nic96 reveals a novel, intricate helical domain architecture. *Journal Of Biological Chemistry* 282, 34904-12-34912.
- Jin, H., White, S.R., Shida, T., Schulz, S., Aguiar, M., Gygi, S.P., Bazan, J.F., and Nachury, M.V. (2010). The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. *Cell* 141, 1208-19.
- Kabachinski, G., and Schwartz, T.U. (2015). The nuclear pore complex – structure and function at a glance. *Journal Of Cell Science* 128, 423.
- Kampmann, M., and Blobel, G. (2009). Three-dimensional structure and flexibility of a membrane-coating module of the nuclear pore complex. *Nature Structural and Molecular Biology* 16, 782.
- Kampmann, M., Atkinson, C.E., Mattheyses, A.L., and Simon, S.M. (2011). Mapping the orientation of nuclear pore proteins in living cells with polarized fluorescence microscopy. *Nature Structural and Molecular Biology* 18, 643-9.
- Kappel, C., Zachariae, U., Dölker, N., and Grubmüller, H. (2010). An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. *Biophysical Journal* 99, 1596-603.
- Kee, H.L., Dishinger, J.F., Blasius, T.L., Liu, C., Margolis, B., and Verhey, K.J. (2012). A size-exclusion permeability barrier and nucleoporins characterize a ciliary pore complex that regulates transport into cilia. *Nature Cell Biology* 14, 431-7.

- Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014a). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings Of The National Academy Of Sciences* *111*, E2453-61-E2461.
- Kim, J.C., Badano, J.L., Sibold, S., Esmail, M.A., Hill, J., Hoskins, B.E., Leitch, C.C., Venner, K., Ansley, S.J., Ross, A.J., et al. (2004). The Bardet-Biedl protein BBS4 targets cargo to the pericentriolar region and is required for microtubule anchoring and cell cycle progression. *Nat. Genet.* *36*, 462-70.
- Kim, S.J., Fernandez-Martinez, J., Sampathkumar, P., Martel, A., Matsui, T., Tsuruta, H., Weiss, T.M., Shi, Y., Markina-Inarrairaegui, A., Bonanno, J.B., et al. (2014b). Integrative Structure–Function Mapping of the Nucleoporin Nup133 Suggests a Conserved Mechanism for Membrane Anchoring of the Nuclear Pore Complexes. *Molecular and Cellular Proteomics* *13*, 2911.
- Krull, S., Thyberg, J., and Björkroth, B. (2004). Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. *Molecular Biology of the Cell* *15*, 4261-4277.
- Lee, C., and Goldberg, J. (2010). Structure of coatamer cage proteins and the relationship among COPI, COPII, and clathrin vesicle coats. *Cell* *142*, 123-32.
- Leksa, N.C., Brohawn, S.G., and Schwartz, T.U. (2009). The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture. *Structure* *17*, 1082-1091.
- Li, A., Saito, M., Chuang, J., Tseng, Y., Dedesma, C., Tomizawa, K., Kaitsuka, T., and Sung, C. (2011). Ciliary transition zone activation of phosphorylated Tctex-1 controls ciliary resorption, S-phase entry and fate of neural progenitors. *Nature Cell Biology* *13*, 402-11.
- Li, S.J., and Hochstrasser, M. (2000). The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. *Mol. Cell. Biol.* *20*, 2367-77.
- Liu, X., Mitchell, J.M., Wozniak, R.W., Blobel, G., and Fan, J. (2012). Structural evolution of the membrane-coating module of the nuclear pore complex. *Proceedings Of The National Academy Of Sciences* *109*, 16498-503-16503.
- Loktev, A.V., Zhang, Q., Beck, J.S., Searby, C.C., Scheetz, T.E., Bazan, J.F., Slusarski, D.C., Sheffield, V.C., Jackson, P.K., and Nachury, M.V. (2008). A BBSome subunit links ciliogenesis, microtubule stability, and acetylation. *Dev. Cell* *15*, 854-65.
- Loïodice, I., Alves, A., Rabut, G., Van Overbeek, M., Ellenberg, J., Sibarita, J., and Doye, V. (2004). The entire Nup107-160 complex, including three new members, is targeted as one entity to kinetochores in mitosis. *Molecular Biology Of The Cell* *15*, 3333-44-3344.
- Lusk, C.P., Makhnevych, T., Marelli, M., Aitchison, J.D., and Wozniak, R.W. (2002). Karyopherins in nuclear pore biogenesis: a role for Kap121p in the assembly of Nup53p into nuclear pore complexes. *The Journal Of Cell Biology* *159*, 267-78-278.
- Lutzmann, M., Kunze, R., Buerer, A., Aebi, U., and Hurt, E. (2002). Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *The EMBO Journal* *21*, 387-397.

- Lutzmann, M., Kunze, R., Stangl, K., Stelter, P., Tóth, K.F., Böttcher, B., and Hurt, E. (2005). Reconstitution of Nup157 and Nup145N into the Nup84 complex. *Journal Of Biological Chemistry* 280, 18442-51.
- Marelli, M., Lusk, C.P., Chan, H., Aitchison, J.D., and Wozniak, R.W. (2001). A link between the synthesis of nucleoporins and the biogenesis of the nuclear envelope. *The Journal Of Cell Biology* 153, 709-24-724.
- Mi, L., Goryaynov, A., Lindquist, A., Rexach, M., and Yang, W. (2015). Quantifying Nucleoporin Stoichiometry Inside Single Nuclear Pore Complexes In vivo. *Scientific Reports* 5, 9372.
- Milenkovic, L., Scott, M.P., and Rohatgi, R. (2009). Lateral transport of Smoothed from the plasma membrane to the membrane of the cilium. *The Journal Of Cell Biology* 187, 365-74.
- Montpetit, B., Thomsen, N.D., Helmke, K.J., Seeliger, M.A., Berger, J.M., and Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* 472, 238-42.
- Mourão, A., Nager, A.R., Nachury, M.V., and Lorentzen, E. (2014). Structural basis for membrane targeting of the BBSome by ARL6. *Nature Structural and Molecular Biology* 21, 1035-41.
- Nachury, M.V., Seeley, E.S., and Jin, H. (2010). Trafficking to the ciliary membrane: how to get across the periciliary diffusion barrier? *Annual Review of Cell and Developmental Biology* 26, 59-87.
- Nachury, M.V., Loktev, A.V., Zhang, Q., Westlake, C.J., Peränen, J., Merdes, A., Slusarski, D.C., Scheller, R.H., Bazan, J.F., Sheffield, V.C., et al. (2007). A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* 129, 1201-13.
- Nagy, V., Hsia, K., Debler, E.W., Kampmann, M., Davenport, A.M., Blobel, G., and Hoelz, A. (2009). Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proceedings Of The National Academy Of Sciences* 106, 17693-8-17698.
- Neumann, N., Lundin, D., and Poole, A.M. (2010). Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *Plos ONE* 5, e13241.
- Niepel, M., Molloy, K.R., Williams, R., Farr, J.C., Meinema, A.C., Vecchiotti, N., Cristea, I.M., Chait, B.T., Rout, M.P., and Strambio-De-Castillia, C. (2013). The nuclear basket proteins Mlp1p and Mlp2p are part of a dynamic interactome including Esc1p and the proteasome. *Molecular Biology Of The Cell* 24, 3920.
- Onischenko, E., Stanton, L.H., Madrid, A.S., Kieselbach, T., and Weis, K. (2009). Role of the Ndc1 interaction network in yeast nuclear pore complex assembly and maintenance. *The Journal Of Cell Biology* 185, 475-91-491.
- Ori, A., Banterle, N., Iskar, M., Andres-Pons, A., Escher, C., Bui, H.K., Sparks, L., Solis-Mezarino, V., Rinner, O., Bork, P., et al. (2013). Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. *Molecular Systems Biology* 9, 648.
- Ou, G., Blacque, O.E., Snow, J.J., Leroux, M.R., and Scholey, J.M. (2005). Functional coordination of intraflagellar transport motors. *Nature* 436, 583-7.

Pazour, G.J., and Witman, G.B. (2003). The vertebrate primary cilium is a sensory organelle. *Current Opinion In Cell Biology* 15, 105.

Pugacheva, E.N., Jablonski, S.A., Hartman, T.R., Henske, E.P., and Golemis, E.A. (2007). HEF1-dependent Aurora A activation induces disassembly of the primary cilium. *Cell* 129, 1351-63.

Rasala, B.A., Orjalo, A.V., Shen, Z., Briggs, S., and Forbes, D.J. (2006). ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. *Proceedings Of The National Academy Of Sciences* 103, 17801-6-17806.

Rasala, B.A., Ramos, C., Harel, A., and Forbes, D.J. (2008). Capture of AT-rich chromatin by ELYS recruits POM121 and NDC1 to initiate nuclear pore assembly. *Molecular Biology Of The Cell* 19, 3982-96-3996.

Reichelt, R., Holzenburg, A., Buhle, E.L., Jarnik, M., Engel, A., and Aebi, U. (1990). Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. *The Journal Of Cell Biology* 110, 883-94.

Ribbeck, K., and Görlich, D. (2001). Kinetic analysis of translocation through nuclear pore complexes. *The EMBO Journal* 20, 1320-30.

Rothbaler, A., and Kutay, U. (2013). Poring over pores: nuclear pore complex insertion into the nuclear envelope. *Trends Biochem. Sci.* 38, 292-301.

Rout, M.P., and Blobel, G. (1993). Isolation of the yeast nuclear pore complex. *The Journal Of Cell Biology* 123, 771-83-783.

Rout, M.P., Aitchison, J.D., Suprpto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. *The Journal Of Cell Biology* 148, 635-51-652.

Sampathkumar, P., Gheyi, T., and Miller, S.A. (2011). Structure of the C-terminal domain of *Saccharomyces cerevisiae* Nup133, a component of the nuclear pore complex. *Proteins: Structure* 79, 1672-1677.

Sampathkumar, P., Kim, S.J., Upla, P., Rice, W.J., Phillips, J., Timney, B.L., Pieper, U., Bonanno, J.B., Fernandez-Martinez, J., Hakhverdyan, Z., et al. (2013). Structure, Dynamics, Evolution, and Function of a Major Scaffold Component in the Nuclear Pore Complex. *Structure* 21, 560.

Schekman, R., and Orci, L. (1996). Coat Proteins and Vesicle Budding. *Science* 271, 1526.

Schrader, N., Stelter, P., Flemming, D., Kunze, R., Hurt, E., and Vetter, I.R. (2008). Structural basis of the nic96 subcomplex organization in the nuclear pore channel. *Mol. Cell* 29, 46-55.

Schwartz, T.U. (2005). Modularity within the architecture of the nuclear pore complex. *Curr. Opin. Struct. Biol.* 15, 221-6.

Seo, H., Blus, B.J., Jankovic, N.Z., and Blobel, G. (2013). Structure and nucleic acid binding activity of the nucleoporin Nup157. *Proceedings Of The National Academy Of Sciences* 110, 16450-5-16455.

Seo, H., Ma, Y., Debler, E.W., Wacker, D., Kutik, S., Blobel, G., and Hoelz, A. (2009). Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proceedings Of The National Academy Of Sciences* *106*, 14281-6-14286.

Seo, S., Baye, L.M., Schulz, N.P., Beck, J.S., Zhang, Q., Slusarski, D.C., and Sheffield V.C. (2010). BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. *Proceedings of the National Academy of Sciences* *107*, 1488-1493.

Seo, S., Zhang, Q., Bugge, K., Breslow, D.K., Searby, C.C., Nachury, M.V., and Sheffield, V.C. (2011). A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothed. *Plos Genet.* *7*, e1002358.

Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S.J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M.P., and Chait, B.T. (2014). Structural Characterization by Cross-linking Reveals the Detailed Architecture of a Coatomer-related Heptameric Module from the Nuclear Pore Complex. *Molecular & Cellular Proteomics* *13*, 2927-2943.

Solmaz, S.R., Chauhan, R., Blobel, G., and Melčák, I. (2011). Molecular Architecture of the Transport Channel of the Nuclear Pore Complex. *Cell* *147*, 590.

Stelter, P., Kunze, R., Flemming, D., Höpfner, D., Diepholz, M., Philippsen, P., Böttcher, B., and Hurt, E. (2007). Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nature Cell Biology* *9*, 788-96.

Strahm, Y., Fahrenkrog, B., Zenklusen, D., Rychner, E., Kantor, J., Rosbash, M., and Stutz, F. (1999). The RNA export factor Gle1p is located on the cytoplasmic fibrils of the NPC and physically interacts with the FG-nucleoporin Rip1p, the DEAD-box protein Rat8p/Dbp5p and a new protein Ymr255p. *The EMBO Journal* *18*, 5761.

Stuwe, T., Lin, D.H., Collins, L.N., Hurt, E., and Hoelz, A. (2014). Evidence for an evolutionary relationship between the large adaptor nucleoporin Nup192 and karyopherins. *Proceedings Of The National Academy Of Sciences* *111*, 2530.

Stuwe, T., Borzyskowski, von, L.S., Davenport, A.M., and Hoelz, A. (2012). Molecular basis for the anchoring of proto-oncoprotein Nup98 to the cytoplasmic face of the nuclear pore complex. *J. Mol. Biol.* *419*, 330-46.

Szyzborska, A., de Marco, A., Daigle, N., Cordes, V.C., Briggs, J.A.G., and Ellenberg, J. (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* *341*, 655-8-658.

Thierbach, K., Appen, von, A., Thoms, M., Beck, M., Flemming, D., and Hurt, E. (2013). Protein Interfaces of the Conserved Nup84 Complex from *Chaetomium thermophilum* Shown by Crosslinking Mass Spectrometry and Electron Microscopy. *Structure* *21*, 1672.

Ulrich, A., Partridge, J.R., and Schwartz, T.U. (2014). The stoichiometry of the nucleoporin 62 subcomplex of the nuclear pore in solution. *Molecular Biology Of The Cell* *25*, 1484.

Wei, Q., Zhang, Y., Li, Y., Zhang, Q., Ling, K., and Hu, J. (2012). The BBSome controls IFT assembly and turnaround in cilia. *Nature Cell Biology* *14*, 950-7.

- Weirich, C.S., Erzberger, J.P., Berger, J.M., and Weis, K. (2004). The N-terminal domain of Nup159 forms a beta-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. *Mol. Cell* 16, 749-60.
- Whittle, J.R.R., and Schwartz, T.U. (2009). Architectural Nucleoporins Nup157/170 and Nup133 Are Structurally Related and Descend from a Second Ancestral Element. *Journal Of Biological Chemistry* 284, 28442.
- Whittle, J., and Schwartz, T.U. (2010). Structure of the Sec13–Sec16 edge element, a template for assembly of the COPII vesicle coat. *The Journal Of Cell Biology*.
- Wolfe, K.H., and Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708-13.
- Yang, Q., Rout, M.P., and Akey, C.W. (1998). Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell* 1, 223-34-234.
- Yoshida, K., Seo, H., Debler, E.W., Blobel, G., and Hoelz, A. (2011). Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proceedings Of The National Academy Of Sciences* 108, 16571-6.
- Zhang, Q., Yu, D., Seo, S., and Stone, E.M. (2012). Intrinsic protein-protein interaction-mediated and chaperonin-assisted sequential assembly of stable Bardet-Biedl syndrome protein complex, the BBSome. *Journal of Biological Chemistry* 287, 20625-20635.

Figures

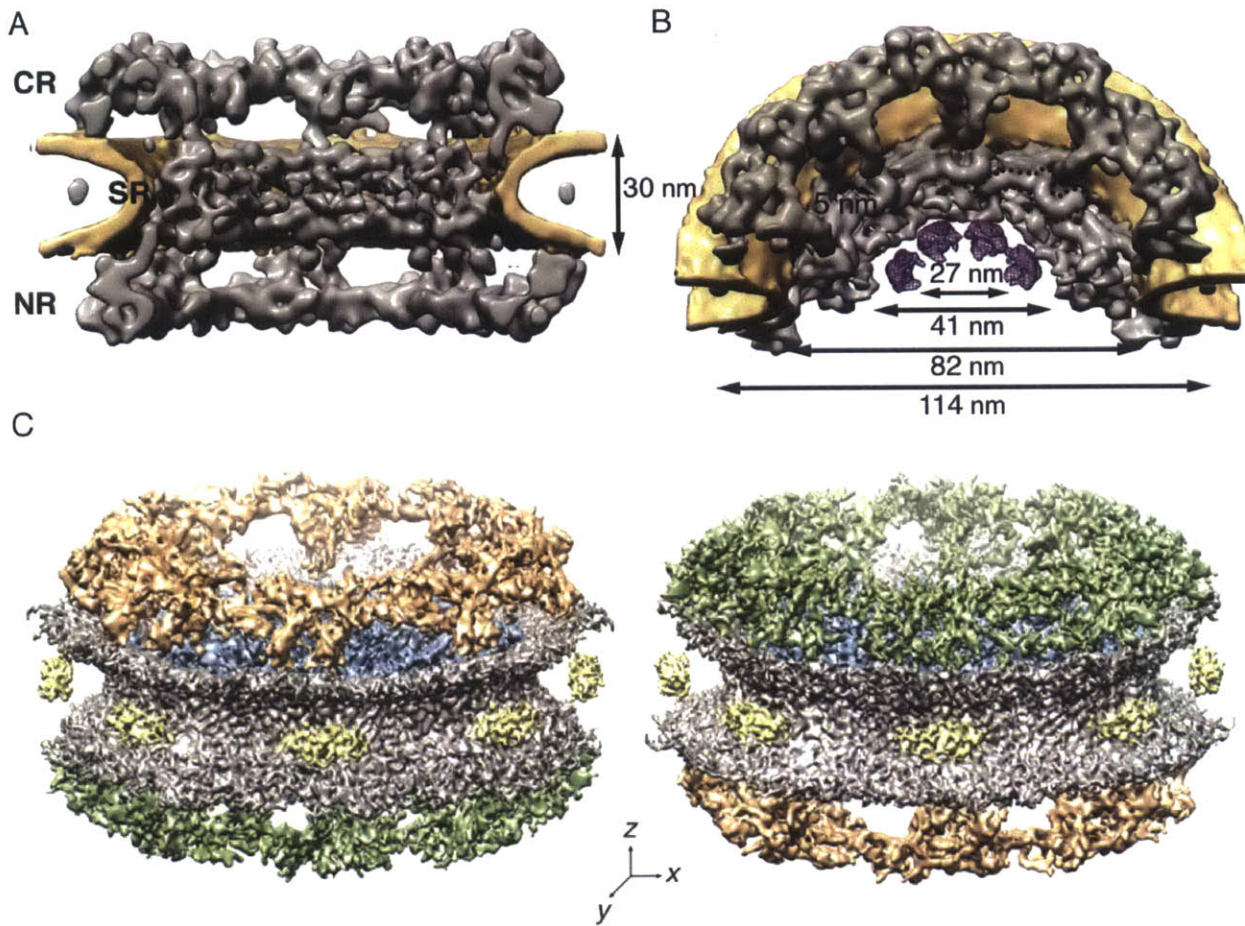


Figure 1. Cryo-ET reconstructions of the metazoan NPC. (A) Side-on cut-away view of the 32 Å cryo-ET reconstruction of the human NPC from Bui et al., 2013, depicting the three structural NPC rings. CR, cytoplasmic ring. SR, spoke ring. NR, nucleoplasmic ring. Protein density (grey) and the membrane (brown) are shown. (B) Top-down view of the human NPC from Bui et al., 2013. Dimensions of the NPC are labeled. (C) 20 Å cryo-ET reconstruction of the *Xenopus* NPC from Eibauer et al., 2015. The two side-on views are rotated 180° w.r.t. each other. The CR (gold), SR (blue), CR (green), membrane (grey), and periplasmic density (yellow) are shown.

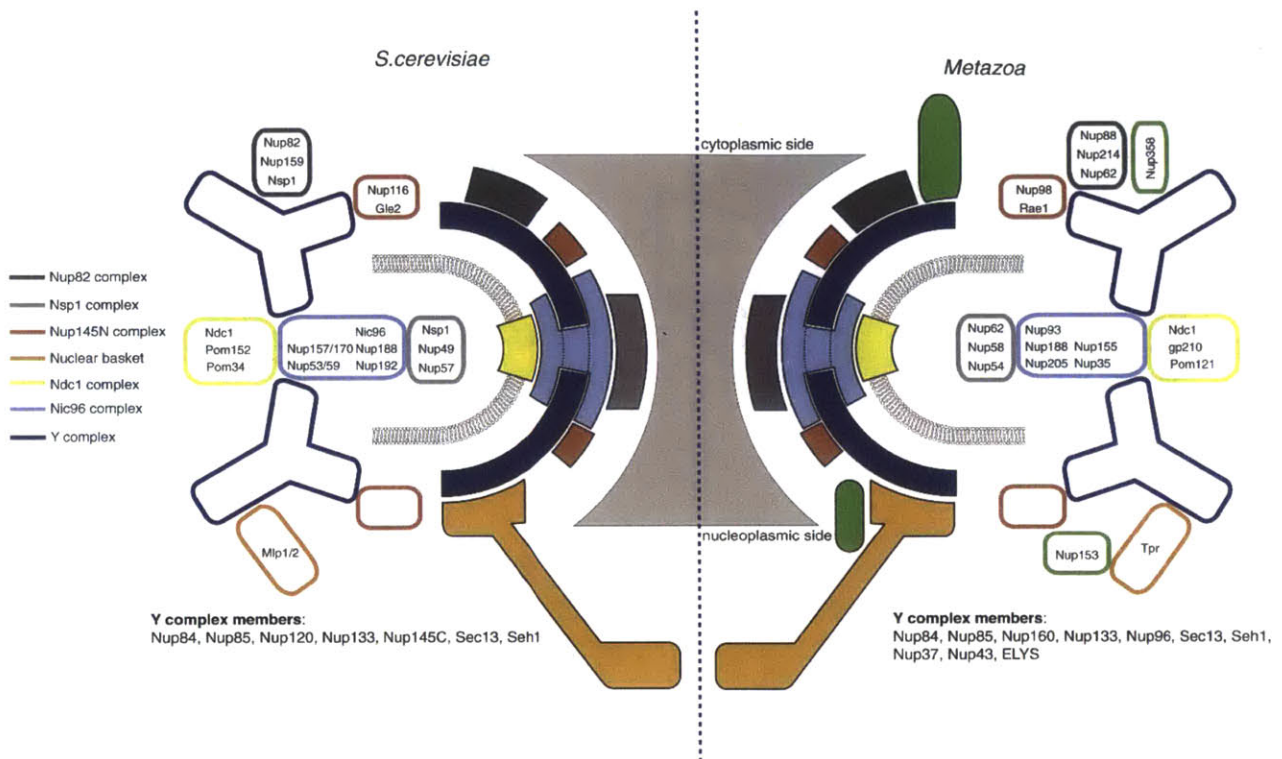


Figure 2. Schematic of NPC subcomplexes. Shown is a side-on, cut away view of the NPC from *S. cerevisiae* (left) and metazoa (right). Complexes are colored according to the legend and are placed based on their relative position in the NPC. Figure is modified from Brohawn et al., 2009.

Chapter 2: Atomic structure of the Y complex of the nuclear pore

The material presented in this chapter was adapted, with permission, from the following publication:

Kelley, K.*, Knockenhauer, K.E.*, Kabachinski, G., and Schwartz, T.U. (2015). Atomic structure of the Y complex of the nuclear pore. *Nature Structural & Molecular Biology* 22, 425-431.

*Equal Contribution

Author contributions: T.U.S., K.K. and K.E.K. designed the study. K.K. and K.E.K. performed the experiments. K.K., K.E.K. and T.U.S. analyzed the data. G.K. performed and analyzed the fitness tests. K.K., K.E.K., G.K. and T.U.S. interpreted the structure and wrote the manuscript.

Introduction

In all eukaryotic cells transcription and translation are physically separated between the nucleus and cytoplasm. This allows for distinct gene expression control mechanisms, for example in cell differentiation and development, unavailable to prokaryotes. Nuclear pore complexes (NPCs), which perforate the nuclear envelope and act as the main transport gate, therefore have a fundamental role in cellular homeostasis (Hetzer and Wente, 2009; Strambio-de-Castillia et al., 2010). The NPC is a modular, donut-shaped assembly of ~30 different proteins (nucleoporins or nups), arranged in multiples of eight around a central axis that is aligned with the main transport channel (Brohawn et al., 2009). Nups can be classified as (i) architectural nups, which form the stable scaffold of the NPC, (ii) peripheral nups with various degrees of mobility, and (iii) nups with the characteristic phenylalanine-glycine (FG)-repeat elements in disordered extensions that form the permeability barrier.

To gain mechanistic insight into NPC function, considerable effort has been undertaken to determine the NPC structure at high resolution. Due to its enormous size of ~40–120 MDa (Reichelt et al., 1990; Rout et al., 2000; Yang et al., 1998), this will ultimately only be possible with a combination of different visualization techniques, notably X-ray crystallography and cryo-electron microscopy.

One of the main architectural elements of the NPC is the Y-shaped complex, which is essential for NPC formation (Harel et al., 2003; Walther et al., 2003). In *S. cerevisiae*, it is a 7-membered 575 kDa complex composed of Nup84, Nup85, Nup120, Nup133, Nup145C, Sec13 and Seh1 (Lutzmann et al., 2002). Homologs for these 7 members are also found in humans, suggesting conservation within all opisthokonts (Belgareh et al., 2001; Neumann et al., 2010; Vasu et al., 2001). It is widely believed that the Y-complex, together with the heteromeric Nic96 complex, forms the principal NPC scaffold (Alber et al., 2007; Bui et al., 2013; Vollmer and

Antonin, 2014). Thus, the atomic structures of the assembled Y- and Nic96 complexes are important milestones toward understanding the NPC at the highest resolution.

All 7 members of the Y-complex have been studied crystallographically, though only individually or as heterodimeric or trimeric complexes, and typically truncated (Berke et al., 2004; Bilokapic and Schwartz, 2012; Boehmer et al., 2008; Brohawn and Schwartz, 2009; Brohawn et al., 2008; Debler et al., 2008; Hsia et al., 2007; Leksa et al., 2009; Nagy et al., 2009; Sampathkumar et al., 2011; Seo et al., 2009; Whittle and Schwartz, 2009). The branch point or 'hub' of the Y-complex, where the two short arms and the long stalk meet, is structurally the least understood even though it is arguably the most important element (Fig. 1a).

Here we now report the structure of the Y-complex hub, which enables us to combine all the additional fragmentary structures into a highly accurate assembled structure of the Y-complex. We can show that the Y-complex structure is widely conserved among all eukaryotes. Species-specific additions to the Y-complex decorate, but do not principally alter the overall structure.

Results

Structure of the Y-complex hub

We generated a series of structure-based expression constructs containing the elements of Nup120, Nup145C, and Nup85 that were known to directly interact with each other at the Y-complex hub (Bilokapic and Schwartz, 2012; Brohawn et al., 2008; Fernandez-Martinez et al., 2012; Leksa et al., 2009). In addition, these constructs were designed to overlap at least partially with the already structurally characterized Y-complex fragments. We succeeded in obtaining crystals of a heterotetrameric construct containing Nup85₂₅₇₋₁₁₈₁, Nup120₉₅₂₋₁₂₄₁, Nup145C₂₃₃₋₇₉₁, and Sec13 from the thermophilic fungus *Myceliophthora thermophila* (*mt*), which diffracted to 4.1 Å resolution (Table 1). The structure was solved with a combination of

molecular replacement and single anomalous dispersion (SAD) using selenomethionine-derivatized protein (see Methods for details). The crystals have a high solvent content (68%) and the structure has a comparatively high degree of positional disorder (Wilson B factor 144 Å²). Despite the high Wilson B factor, we were able to properly assign the sequence to all four proteins within the assembled complex based on model building guided by selenomethionine positions, homology models, and phylogenetic considerations (Table 2).

The overall structure of the heterotetramer is roughly V-shaped, composed of three helical units, Nup85, Nup120, and Nup145C, and a laterally attached β-propeller, Sec13 (Fig. 1b, Fig. 2). Nup85 and Nup145C form the long sides of the 'V', whereas Nup120 is sandwiched between the two sides and acts as the main connector. Sec13 is bound to Nup145C as previously described in the Nup145C-Sec13-Nup84 structure from *Saccharomyces cerevisiae* (*sc*) (Brohawn and Schwartz, 2009; Nagy et al., 2009), namely by the insertion of a seventh blade into its open, 6-bladed β-propeller. Viewed from the side, it is noticeable that the heterotetramer is substantially bent rather than flat (Fig. 1c). The overall dimensions of the complex are ~125 Å x ~140 Å x ~50 Å.

Nup85 and Nup145C belong to the ancestral coatomer element 1 (ACE1)-class of proteins. These proteins are characterized by a ~65kDa, tripartite helical segment composed of a crown, a trunk and a tail element, which adopts a characteristic fold-back structure involving ~30 α-helices (Brohawn et al., 2008; 2009). ACE1 proteins are exclusively found in the NPC scaffold and the COPII vesicle coat (Whittle and Schwartz, 2010). The scNup85 and scNup145C fragments solved previously both lack the tail elements, which are present in this structure. These tail elements are in direct contact with Nup120. Even though the identity between the Mt and Sc sequences is low (12% for Nup85, 14% for Nup145C) (Table 3), the structures superpose well (Fig. 3). Therefore, we assigned secondary structure elements in the Mt proteins in accord with the published Sc fragments (Brohawn and Schwartz, 2009; Brohawn

et al., 2008). Nup145C has 27 helices in total. Helices $\alpha 1$ – $\alpha 3$ and $\alpha 12$ – $\alpha 20$ form the trunk, $\alpha 4$ – $\alpha 11$ the crown, and $\alpha 21$ to $\alpha 27$ the tail. The entire helical stack of Nup145C has a crescent shape. To stabilize the Sec13 interaction, we fused Sec13 N-terminally to 145C, similar to what was previously done for scSec13-Nup145C (Brohawn and Schwartz, 2009). For this reason, our Nup145C construct is lacking 232 N-terminal, predicted-to-be disordered residues.

Nup85 has 33 helices in total. Helices $\alpha 1$ – $\alpha 3$ and $\alpha 12$ – $\alpha 20$ form the trunk, $\alpha 4$ – $\alpha 11$ the crown, and $\alpha 21$ to $\alpha 30$ the tail. Compared to mtNup145C, the mtNup85 tail domain is longer, and contains 4 additional helices. In *S. cerevisiae*, Nup85 binds Seh1 very similarly to Nup145C binding Sec13, i.e. via an insertion blade that closes the open, 6-bladed β -propeller. Although Sordariomycota, including *M. thermophila* and *C. thermophilum*, also contain a recognizable Seh1 homolog, it does not bind to Nup85 (Thierbach et al., 2013). The structure now reveals that mtNup85 lacks the essential Seh1-binding site, i.e. the insertion blade. Instead, it contains an additional N-terminal helix $\alpha 0$ (Fig. 1, Fig. 4), incompatible with Seh1 binding. In light of this observation, we suggest that Seh1 is not a member of the conserved core of the Y-complex. This core is a heterohexamer rather than a heteroheptamer. The presence of an Seh1 homolog in Sordariomycota is likely due to its function in the GATOR(SEA) complex (Algret et al., 2014; Bar-Peled et al., 2013).

The C-terminal fragment of Nup120 contains 10 helices ($\alpha 23$ – $\alpha 29$, $\alpha 29a$, $\alpha 29b$, $\alpha 30$), 8 of which form a regular stack. This stack superimposes very well (RMSD 2.2 Å over 135 C α positions) with the C terminus of Nup120 from *S. pombe* (Bilokapic and Schwartz, 2012). The C-terminal mtNup120 element is wedged between the tail domains of Nup85 and Nup145C, and forms the core of the hub. The interface between Nup120 and Nup85 buries a ~ 940 Å² surface, whereas the Nup120-Nup145C interface measures ~ 1700 Å². The Nup120-Nup85 interface includes helices $\alpha 27$ and $\alpha 29$ from Nup120 and helices $\alpha 28$ and $\alpha 30$ from Nup85. This arrangement generates a tightly packed four-helix bundle between the two proteins, likely

explaining the high affinity of both proteins for each other. In comparison, the Nup120-Nup145C interface is more complicated, with two major elements contributing contacts. One part of the binding interface is generated by the terminal helix $\alpha 27$ of Nup145C that runs along the narrow ridge of the Nup120 C-terminal helical stack. Helices $\alpha 29a$ and $\alpha 29b$ of Nup120 are outside of the C-terminal stack and clamp down Nup145C, as they are positioned on the outward-facing side of the molecule, thereby forming a second interaction site. Helices $\alpha 29a+b$ are not conserved, but appear to be specific for Sordariomycota.

Sec13 is placed right in the neck of the V-shaped molecule, with contacts to all three helical proteins. The contact area with mtNup145C is largest and very similar to the homologous structure of scNup145C (Brohawn and Schwartz, 2009; Nagy et al., 2009). In addition to the close contacts with the insertion blade, the top surface of the Sec13 β -propeller binds the trunk of the Nup145C-ACE1 element. The Sec13-Nup120 contacts are established through interactions between blades 1 and 2 of Sec13, and the very C-terminal helix $\alpha 30$ of Nup120. Finally, contacts between Sec13 and Nup85 are few, but nonetheless likely important. Taken together, it appears as if the Sec13 position helps to rigidify the interaction between the three helical elements.

Fitness analysis

To evaluate the importance of the hub interface *in vivo*, we used a fitness test in *S. cerevisiae*. C-terminal truncations of the last helix of Nup85, Nup120, and Nup145C, were designed to selectively disrupt the mapped interfaces between the three helical proteins. The Nup85 $\Delta\alpha 30$ mutant had the most severe phenotype and showed drastically reduced growth (Fig. 5). Nup145C $\Delta\alpha 27$ and Nup120 $\Delta\alpha 30$ have progressively milder phenotypes. Nup85 $\Delta\alpha 30$ almost phenocopies the lethal Nup85 knockout (Siniosoglou et al., 1996), suggesting that the Nup85-Nup120 interaction is critical for NPC assembly. For Nup120 and Nup145C, it is likely

that the mapped interfaces are not the exclusive elements that integrate these proteins into the NPC, but that additional contacts exist. The N-terminal extension of Nup145C, past the Sec13 insertion blade and not part of our structure, is likely to play a role in this. However, contacts to other scaffold nucleoporins need to be considered as well. Additionally, although we did not formally quantify the protein levels or test the fold retention of the individual truncated proteins, based on previous *in vivo* (Fernandez-Martinez et al., 2012) and *in vitro* (Bilokapic and Schwartz, 2012; Brohawn et al., 2008; Leksa et al., 2009; Stuwe et al., 2015) experiments, we can assume that our specific truncations are folded correctly and expressed at wild type levels.

Composite high-resolution structure of the Y-complex

With the heterotetrameric hub assembly in hand we set out to build a complete high-resolution, composite structure of the entire Y-complex. The structures of full-length Nup120 (Bilokapic and Schwartz, 2012) and Nup84-Nup145C-Sec13 (Brohawn and Schwartz, 2009) contain overlapping elements with the hub structure, 19 kDa and 85 kDa in size, respectively, and were superposed with high confidence (Fig. 6). This generated a heteropentameric complex, in which Nup84 is the preliminary terminal fragment of the long stalk of the Y. Nup84 is an ACE1 protein, but the published structure lacks the tail domain (residues 443–724) that interacts with Nup133. However, the structure of most of the tail domain of Nup107, the human Nup84 homolog, in complex with Nup133 is known (Boehmer et al., 2008). Therefore, we were able to model full-length Nup84 based on the experimentally known N- and C-terminal fragments, and the homology-modeled structure of the intervening 84-residue segment using other ACE1 domains (see Methods for details). After Nup84 was positioned, Nup133 could be docked based on the Nup107-Nup133 complex structure (Whittle and Schwartz, 2009). Finally, the last element on the long stalk of the Y-complex is the N-terminal Nup133 β -propeller, which is loosely tethered to the C-terminal, α -helical stack element (Berke et al., 2004; Kim et al.,

2014). The resulting composite structure constitutes the conserved, heterohexameric core of the Y-complex (Fig. 7a).

Because the composite model is built from structural elements of four different organisms, we examined to what extent this might affect the overall structure. Therefore, we also built models for the heteroheptameric Y-complex in *S. cerevisiae*, including Seh1, as well as the heterodecameric Y-complex in *H. sapiens*, including Seh1 and Nup37 (Fig. 7b, c). The Seh1 position can be deduced from the Seh1-Nup85 complex structure (Brohawn et al., 2008; Debler et al., 2008), whereas the Nup37 position is known from the Nup37-Nup120 complex structure (Bilokapic and Schwartz, 2012; Liu et al., 2012). The positions of Nup43 and ELYS within the human complex are not well understood and were therefore excluded. By comparison, we observed that the conserved, heterohexameric core changes only in local areas between the three models and that the overall shape and dimensions of the Y-complex appear conserved in all opisthokonts. Organism-specific proteins decorate the Y-core, but do not significantly influence its overall structure otherwise.

Our composite structure reveals that the Y-complex when viewed from the front measures about 20 nm wide and 40 nm high (Fig. 7). This is in good agreement with published EM structures and computational models (Bui et al., 2013; Fernandez-Martinez et al., 2012; Kampmann and Blobel, 2009). However, the principal angles between the three extensions from the hub deviate significantly between our X-ray based structure and the EM structures, which is evident by superimposing the different sets of data (Fig. 8). Strikingly, when viewed from the side or the top, the composite crystal structure reveals a distinct, three-dimensional shape (Fig. 7e, Fig. 8b, f), whereas the previously reported structures were essentially flat (Fig. 8d, h). At the hub, the three extensions, namely the two arms and the stalk, exhibit strong curvature and form a dome-shaped structure. As a result, we measure a thickness of ~8 nm for the Y-complex (Fig. 7e), compared to ~4 nm reported in previous EM analyses. Theoretically, this difference could be the result of flexibility within the Y-complex, which is well documented. In our

composite structure we can now specify the main hinge regions and flexible areas (Fig. 9). These flexible regions are detected by general considerations regarding protein structure, domain boundaries established by limited proteolysis, and flexibility observed in previous crystallographic and EM studies (Berke et al., 2004; Bilokapic and Schwartz, 2012; Boehmer et al., 2008; Bui et al., 2013; Kampmann and Blobel, 2009; Whittle and Schwartz, 2009). The hub itself is rather rigid because four proteins engage in a tightly coordinated interface. To flatten the Y-complex, the helical Nup120 C-terminal domain or the Nup85 ACE1 element would have to bend by nearly 90° with respect to each other. Helical stacks have the propensity to bend, as best exemplified by various nuclear transport receptors (Cook et al., 2007; Süel et al., 2006), but the direction of bending is determined by the helical orientation. For the flattening of the Y, the predicted-to-be bendable elements are, however, oriented in an unfavorable way. This means that the necessary distortions in Nup120 or Nup85 would be energetically costly because of the disruption to the hydrophobic core that they would generate, which we consider to be rather unlikely. Therefore, the simplest explanation for the discrepancy between our composite crystal structure and previous random-conical tilt (RCT) negative stain EM structures is that the latter were artificially flattened in the direction normal to the EM grid, a well-known phenomenon.

Implications for NPC assembly models

Next, we tested whether our composite human Y-complex could be positioned into the recently published 3.2 nm cryo-ET density map (Bui et al., 2013) of the human NPC, which predicted a staggered two-ring, head-to-tail orientation of Y-complexes, symmetrically positioned on the nucleoplasmic and the cytoplasmic face of the NPC. We were able to recapitulate the published results of docking the flat, Y-complex structure determined by RCT negative stain EM (EMD code 2443) (Bui et al., 2013) into the cryo-ET map (Bui et al., 2013). We then tried the same procedure using our highly curved Y-complex structure. We searched with the human Y-complex model omitting Nup133 and the tail domain (residues 658–925) of

Nup107, presumably the most flexible Y-complex elements. Nup133 has an N-terminal β -propeller flexibly connected to a C-terminal helical stack domain (Berke et al., 2004). The helical domain of Nup133 is tripartite, with hinges connecting the three helical segments (Whittle and Schwartz, 2009). A flexible hinge in the Nup107 ACE1 element between the trunk-crown, and tail segments can be postulated based on independent structure-mapping studies performed on the human and yeast homologs (Boehmer et al., 2008; Brohawn and Schwartz, 2009; Nagy et al., 2009). Using this stubbed Y-complex, we found three top numerical solutions (see Methods for details). Two solutions roughly coincided with the outer Y-complex ring postulated by Bui et al. (Bui et al., 2013) (Fig. 10), and the third solution coincided with the inner ring of that study (Fig. 11). While our work was in review, Stuwe et al. reported the docking of a similarly stubbed Y-complex from *S. cerevisiae* based on a 7.4 Å crystal structure, which shows a curved topology consistent with our structure, and arrived at a similar solution (Stuwe et al., 2015). However, when we added the Nup107 tail and Nup133 structures back to the docked Y-complex model of our third solution, in any topologically reasonable way, we observed extensive steric clashes with the neighboring Y-complex that seem highly implausible (Fig. 11). Thus, we did not consider this solution further. Regarding the two top solutions, they are rotated around the hub by approximately 20° relative to each other. In each solution the long stalk could be fitted reasonably well to two different regions in the EM density. Both solutions result in a seemingly closed ring when Nup133 is added, albeit the head-to-tail contact would be different in each case. To fit each solution, the long stalk needs to adopt different conformations, largely by adjusting Nup133, which seems realistic based on the expected flexibility around distinct hinge points (Fig. 10c). Obviously, both solutions cannot coexist due to excessive steric clashes. Therefore, the easiest way to explain our docking results is to suggest that the Y-complex ring is a single rather than a double ring, but that it can adopt at least two conformations. We argue that because of subtomogram-averaging, we might see an overlay of the two, equally and most populated, states of the Y-complex ring in the cryo-ET density.

Discussion

As reasonable as our docking attempts may appear, we would like to caution about the interpretation of these results. First, the available cryo-ET map (EMD-2444) (Bui et al., 2013) is calculated based on assumptions that we still do not know to be necessarily correct. For example, a strict eightfold rotational symmetry is applied, which may be appropriate at nm resolution, but possibly not at atomic resolution. If this symmetry is not true on the atomic level, the calculated map could be intrinsically flawed. Due to the similarity of various scaffold nups on a nanometer scale this is particularly troublesome. Second, docking at ~3 nm resolution is at best tentative and only reasonable to attempt because of the distinct and large size of the Y-complex. It is possible that an entire portion of the Y-complex is so flexible in the NPC assembly that it could simply be averaged out in the cryo-ET study. This could in principal be true for the long stalk as well as most of the Nup120 arm, which can bend perpendicular to the long axis of its C-terminal helical stack domain (Bilokapic and Schwartz, 2012) . Many additional docking solutions would need to be considered, if such possibilities were entertained.

One way of independently confirming a specific assembly model initially derived from docking is to map the assumed contacts between neighboring complexes. Bui et al. attempted this by crosslinking experiments (Bui et al., 2013). Surprisingly, only two such crosslinks (Nup107-Nup133 and Nup43-Nup96) were found. However, these interactions could not be confirmed by additional experiments and, therefore, are insufficient to distinguish between the models. A two-ring model in particular would generate many inter-Y contact sites, thus the paucity of detected crosslinks is unexpected. Taking additional studies into account does not resolve the discrepancy between the reticulated two ring model and a flexible one ring model. For example, the radial distribution of fluorescently labeled scaffold nucleoporins was determined by super-resolution microscopy (Szyzborska et al., 2013). However, the raw localization precision of 20-30 nm attained in that study is insufficient to distinguish between the

different ring models. Stoichiometric considerations, which have major model implications, are also not definitive. For example, a recent study concluded there are 32 copies of Y-complexes per NPC in HeLa cells (Ori et al., 2013), a number supporting a two ring model, whereas earlier studies argued for 16 copies (Cronshaw et al., 2002; Rout et al., 2000), which would favor a one ring model. Recent advancements in super-resolution microscopy should allow for the direct counting individual nucleoporins within an NPC, putting this controversy to rest soon (Ries et al., 2012). Finally, mass estimates for the NPC range from 40 MDa to 120 MDa, depending on the species and technique used (Reichelt et al., 1990; Rout et al., 2000; Yang et al., 1998). Again, it is premature to use this information to definitively validate a specific model. In conclusion, to confirm any model, additional experimental data that have an appropriate resolution are needed. Interaction data that would confirm contacts between neighboring subcomplexes would likely be the most helpful tool. Regardless, the Y-complex structure presented here at least provides a benchmark that any reasonable model needs to be consistent with.

Methods

Construct generation

Nup120₉₅₂₋₁₂₄₁, Nup145C₂₃₃₋₇₉₁, Nup85₂₅₇₋₁₁₈₁, and full-length Sec13 were cloned from *Myceliophthora thermophila* (*Mt*). Nup120₉₅₂₋₁₂₄₁ was C-terminally fused with a 10xHis tag and cloned into a kanamycin resistant, T7-promoter-based bacterial expression vector. To increase stability, full-length Sec13 was fused C-to-N-terminally to Nup145C₁₄₇₋₇₀₅ using a flexible four-residue linker, in analogy to a previously described *S. cerevisiae* construct (Brohawn and Schwartz, 2009). The Sec13-Nup145C fusion was N-terminally tethered to a 3C protease cleavable SUMO tag and cloned into an ampicillin resistant, T7-promoter-based bacterial expression vector. Nup85₂₅₇₋₁₁₈₁ was N-terminally fused with a 3C protease cleavable 10xHis-

8xArg-SUMO tag. The tetrameric complex is referred to as Nup120-Sec13-Nup145C-Nup85 for simplicity.

Protein production and purification

Nup120 and Sec13-Nup145C vectors were co-transformed into *Escherichia coli* LOBSTR-RIL(DE3) (Kerafast) (Andersen et al., 2013) cells and protein production was induced with 0.2 mM IPTG at 18 °C for 12-14 h. Nup85 was expressed separately under identical conditions. Cells were collected by centrifugation at 6,000g, resuspended in lysis buffer (50 mM potassium phosphate (pH 8.0), 500 mM NaCl, 30 mM imidazole, 3 mM β -mercaptoethanol (β ME), 1 mM PMSF) and lysed using a cell disruptor (Constant Systems). The lysate was cleared by centrifugation at 12,500g for 15min. The soluble fraction was incubated with Ni Sepharose 6 Fast Flow beads (GE Healthcare) for 30min on ice. After washing the beads with lysis buffer the protein was eluted (250 mM imidazole, pH8.0, 150 mM NaCl, 3 mM β ME). The Nup85 Ni-eluate was diluted 1:1 with 20 mM Hepes·KOH (pH 7.4), 0.1 mM EDTA, 1 mM dithiothreitol (DTT) and subjected to cation exchange chromatography on a HiTrapS column (GE Healthcare) using a linear NaCl gradient. The Nup120-Sec13-Nup145C Ni-eluate was incubated with 3C and dialyzed overnight at 4 °C (20 mM Hepes·KOH (pH 7.4), 150 mM NaCl, 0.1 mM EDTA, 1 mM DTT). Both samples were further purified separately via size exclusion chromatography on a Superdex S200 16/60 column (GE Healthcare) equilibrated in running buffer (10 mM Tris·HCl (pH 8.0), 150 mM NaCl, 1 mM NaCl, 1mM DTT). Purified 10xHis-8xArg-SUMO-Nup85 was incubated with an excess of purified Nup120-Sec13-Nup145C and the assembled quaternary complex was separated via cation-exchange chromatography. Nup120-Sec13-Nup145C-Nup85 complex was incubated with 3C overnight at 4 °C and subjected to a final purification by size exclusion chromatography on Superdex S200 (GE Healthcare). Selenomethionine (SeMet) derivatized Nup85 and Nup120-Sec13-Nup145C-Nup85 was

prepared as described previously (Brohawn et al., 2008) and purified as the native version, except that the reducing agent concentration (β ME, DTT) was 5 mM in all buffers.

Crystallization

Initial crystals of Nup120-Sec13-Nup145C-Nup85 were obtained at 18 °C in 9 days in sitting drops over a reservoir containing 50 mM Tris·HCl (pH 8.5), 1 M NH_4SO_4 , 10 mM EDTA as an additive reduced nucleation and improved the shape of the crystals. Hanging drops of 1 μl protein at 3 mg/ml and 1 μl of precipitant (50 mM Tris·HCl (pH 8.23), 0.7 M $(\text{NH}_4)_2\text{SO}_4$, 20 mM EDTA) were supplemented on the third day of incubation with 0.2 μl seed stock to yield diffraction quality, thin triangular plates. Selenomethionine-derivatized protein crystallized under identical conditions. Crystals were cryo-protected in mother liquor supplemented with 20 % (v/v) ethylene glycol.

Data collection and structure determination

Data collection was performed at the Advanced Photon Source end station 24-IDC. All data processing steps were carried out with programs provided through SBGrid (Morin et al., 2013). Data reduction was performed using HKL2000 (Otwinowski and Minor, 1997). Sec13 from *S. cerevisiae* (3MZK, chain A) (Whittle and Schwartz, 2010) was used as a search model for molecular replacement (MR). One unique solution was found in a Nup85-only SeMet derivative (space group C2, 1 copy per asymmetric unit). Partial MR phases were then used to find 12 out of 20 possible SeMet positions (8 were in disordered regions) with *Phaser* within the PHENIX suite (Adams et al., 2010) using the MR-single anomalous dispersion (MR-SAD) protocol. An interpretable 4.1 Å experimental electron density map was obtained after solvent modification using *Parrot* from the CCP4 suite (Winn et al., 2011). The backbone of most of the model was traced and the sequence for Nup85 was assigned, aided by the selenium positions as markers and the partial Nup85 models available (PDB codes: 3EWE, 3F3F) (Brohawn et al.,

2008; Debler et al., 2008) Next, this initial model was used as a search model for MR with the dataset obtained from fully SeMet-derivatized protein (space group P1, two copies per asymmetric unit). Two unambiguous solutions were readily found using *Phaser*. Again, MR-SAD was used to find 36 selenium sites (out of 74 possible, with 26 in disordered regions), which served as markers to assign the sequence for Nup120 and Nup145C. To build *mtSec13*, a starting model was generated using *Rosetta* (Terwilliger et al., 2012) and Sec13 from *S. cerevisiae* (62% sequence identity). Model building was carried out with *Coot* (Emsley et al., 2010) and *phenix.refine* was used for refinement. NCS and secondary structure restraints were applied throughout the refinement process. For the final refined structure, Ramachandran values were: 91% favored, 8.2% allowed, and 0.8% outlier. The *MolProbity* (Chen et al., 2010) clashscore was 8.33 and the *MolProbity* percentile was 97%.

Generation of Y-complex composite structures

A composite structure of the minimal, hexameric Y-complex was generated in *Coot* using secondary structure match (SSM) superposition of solved, overlapping crystal structure fragments. Nup85, Sec13, and the Nup120 tail coordinates are the *M. thermophila* orthologs and are from this study. Nup120₁₋₉₄₈ from *S. pombe* (sp) (PDB code: 4FHM) (Bilokapic and Schwartz, 2012) was used to generate a full-length hybrid Nup120 structure together with the Mt fragment. Nup145C is a hybrid structure with the trunk and tail domains from Mt (this study) and the crown domain from *S. cerevisiae*, which allowed for the confident modeling of the Nup145C-Nup84 junction previously solved in *S. cerevisiae* (PDB code: 3JRO, 3IKO) (Brohawn and Schwartz, 2009; Nagy et al., 2009). Nup84 is a hybrid structure: with the crown and trunk domains from *S. cerevisiae* (PDB code: 3JRO) (Brohawn and Schwartz, 2009) and the tail domain from the *H. sapiens* Nup107 ortholog (PDB code: 3I4R) (Whittle and Schwartz, 2009). Four helices between the trunk and tail domains on Nup84 were modeled from the corresponding helices solved in Nic96 (PDB code: 2QX5, 2RFO) (Jeudy and Schwartz, 2007;

Schrader et al., 2008) using Phyre2 one-to-one threading (Kelley and Sternberg, 2009). Nic96 was chosen as the template because superposition between its trunk and tail domains and the corresponding regions on Nup84 yielded the lowest RMSD of all solved ACE1 domain proteins. Additionally, the *BackPhyre* (Kelley and Sternberg, 2009) structure prediction server suggests that Nic96 is the closest Nup homolog to Nup84 and vice versa, despite low sequence identity of 8%. The Nup133 stacked α -helical domain and its junction with Nup107 are from *H. sapiens* (PDB code: 3I4R) (Whittle and Schwartz, 2009). The interface between the Nup133 β -propeller (PDB code: 1XKS) (Berke et al., 2004) and the stacked α -helical domain is not known but is expected to be flexible, based on limited proteolysis data (data not shown).

Composite structures of the heptameric *S. cerevisiae* and the octameric *H. sapiens* Y-complexes were generated in *Coot* using SSM superposition onto the minimal composite structure of solved and modeled structure fragments generated by one-to-one threading with the program *Phyre2*. For the *S. cerevisiae* Y-complex composite, scNup120₇₁₄₋₁₀₃₆ was modeled with spNup120 (PDB code: 4FHM) (Bilokapic and Schwartz, 2012) as the template. scNup85₄₉₅₋₇₄₄ was modeled with mtNup85. scNup145C₅₃₄₋₇₁₂ was modeled with mtNup145C. scNup133 was modeled with hsNup133 (PDB code: 1XKS, 3I4R) (Berke et al., 2004; Whittle and Schwartz, 2009). scNup84₄₂₅₋₇₂₆ was modeled with hsNup107 and scNic96 as described above. For the human Y-complex composite, hsNup160 was modeled with spNup120 (PDB code: 4FHM) (Bilokapic and Schwartz, 2012) as the template. hsNup85₁₋₄₁₁ was modeled with scNup85 (PDB codes: 3EWE, 3F3F) (Brohawn et al., 2008; Debler et al., 2008). hsNup85₄₁₅₋₆₅₂ was modeled with mtNup85. hsNup96₁₋₇₃₂ was modeled with scNup145C (PDB code: 3JRO, 3IKO) (Brohawn and Schwartz, 2009; Nagy et al., 2009). hsNup96₇₃₃₋₉₃₇ was modeled with mtNup145C. hsNup107₁₋₆₅₇ was modeled with scNup84 and scNic96 as described above. The composite structures are available upon request.

Fitting composite Y-complex structures into single particle 3-D reconstruction EM maps of Y-complexes

Composite *H. sapiens* (omitting Nup133 and Nup107₆₅₈₋₉₂₅) and *S. cerevisiae* Y-complex structures were fitted into the published EM maps for the respective species (EMD codes: 2443, 5152) (Bui et al., 2013) using the Fitmap tool from *Chimera* (Pettersen et al., 2004). 1000 trials were run with an apparent resolution of 33 Å for human and 35 Å for the *S. cerevisiae* composite structure. For fitting the human composite model into EMD-2443 (Bui et al., 2013), the best solution had a correlation score of 0.8235 with 116/1000 hits (Fig. 8c). The next best solution (not shown) had a correlation score of 0.8219 with 8/1000 hits. For fitting the yeast composite model into EMD-5152 (Bui et al., 2013), the best solution had a correlation score of 0.6992 with 18/1000 hits (Fig. 8g). The next best solution (not shown) had a correlation score of 0.6374 with 19/1000 hits.

Fitting composite *H. sapiens* and *S. cerevisiae* Y-complex structures into 3-D reconstruction EM tomography map of the entire NPC

A procedure similar to the one outlined in Bui et al. (Bui et al., 2013) was followed to fit both human and yeast composite Y-complex structures into the published EM map of the human NPC (EMD code: 2444) (Bui et al., 2013) using *Chimera*. To highlight features that are invariant between the cytoplasmic and nuclear ring, a consensus map was generated by multiplication of a binarized copy rotated by 180 degrees with respect to the original EM map, as in Bui et al. (Bui et al., 2013). Subsequently, a membrane mask was constructed and subtracted, resulting in a membrane free consensus map, which was used for all fitting calculations. To avoid excluding alternate conformations resulting from flexibility of the long stalk of the Y-complex, Nup133 was omitted from the computational fitting procedure and was fitted manually. 30000 trials were run with an apparent resolution of 35 Å for both human and yeast composite Y-complex structures. For the human composite model, the best solution

(conformation 1, Fig. 10b) had a correlation score of 0.7918 with 179/30000 hits. The second best solution (conformation 2, Fig. 10b) had a correlation score of 0.7661 with 252/30000 hits. The third best solution (inner ring, Fig. 11) had a correlation score of 0.7516 with 94/30000 hits. Other lower scoring solutions were not considered further based on steric considerations. Similar results were obtained when using the yeast composite Y-complex structure as a search model.

Multiple sequence alignment generation

Multiple sequence alignments were generated for all components of the hub structure using sequences representative of the eukaryotic tree of life (Ciccarelli et al., 2006; Fritz-Laylin et al., 2010). Sequences, except for Nup85 and Nup120, were obtained via *BLAST* searches and aligned using *Muscle* in *JalView* (Waterhouse et al., 2009). Nup85 and Nup120 multiple sequence alignments were obtained from Neumann et al. (Neumann et al., 2010) pruned, and realigned using *ClustalW*.

Yeast Plasmid Construction

The entire *NUP145*, *NUP85*, and *NUP120* genes with their endogenous promoters and terminators (~1000 nucleotides before the start codon and ~400 after the stop codon, respectively) were separately cloned into the multiple cloning site of the centromeric YCplac33 shuttle vector. Additionally, the entire *NUP145* and *NUP85* genes with their endogenous promoters and terminators were separately cloned into the multiple cloning site of the centromeric prs315 vector. All cloning was performed following the standard Gibson assembly method (Gibson et al., 2009). The constructed wild type vectors used in this study were named as follows: GKYP01 (*NUP120*, *URA3*, *CEN*), GKYP02 (*NUP85*, *URA3*, *CEN*), GKYP03 (*NUP145*, *URA3*, *CEN*), GKYP04 (*NUP85*, *LEU2*, *CEN*), and GKYP05 (*NUP145*, *LEU2*, *CEN*).

Vectors where the last C-terminal alpha helix of Nup120 ($\Delta\alpha30$, $\Delta1016-1037$), Nup85 ($\Delta\alpha30$, $\Delta719-744$), and Nup145C ($\Delta\alpha27$, $\Delta689-712$) has been deleted were also generated using GKYP01, GKYP04, and GKYP05, respectively. The deletion vectors were created by introducing an early stop codon using the Quikchange site-directed mutagenesis method. The constructed deletion vectors used in this study were named as follows: GKYP06 (*NUP120* $\Delta\alpha30$, *URA3*, *CEN*), GKYP07 (*NUP85* $\Delta\alpha30$, *LEU2*, *CEN*), and GKYP08 (*NUP145* $\Delta\alpha27$, *LEU2*, *CEN*).

Yeast Strain Construction

Yeast strains used in this study are listed in Table 4. The Δ *NUP120* haploid strain was obtained from the *S. cerevisiae* deletion consortium (background BY4742, #14906), transformed with either the YCplac33 empty vector (negative control), GKYP01 (positive control), or GKYP06 (truncation), and selected on plates lacking uracil. The diploid Δ *NUP85* and Δ *NUP145* strains were also obtained from the *S. cerevisiae* deletion consortium (background BY4743, #26840 and #24459, respectively), transformed with either GKYP02 or GKYP03, respectively, and selected on uracil negative plates. The transformed diploids were then sporulated and subjected to tetrad dissection. The resulting Δ *NUP85* haploids containing GKYP02 were transformed with either prs315 empty vector (negative control), GKYP04 (positive control), or GKYP07 (truncation), whereas the Δ *NUP145* haploids containing GKYP03 were transformed with either prs315 empty vector (negative control), GKYP05 (positive control), or GKYP08 (truncation). The Δ *NUP85* and Δ *NUP145* haploid transformations were selected on plates lacking leucine.

Cell Growth Analysis

The Δ *NUP120* strains containing YCplac33, GKYP01, or GKYP06 were grown as a liquid culture in YPD overnight shaking at 30 °C. The cultures were then diluted in YPD to an OD₆₀₀ of 0.1 and aliquoted into a 96-well plate (100 μ l of culture / well). The plate was placed into a

Synergy 2 multi-mode microplate reader (BioTek) and the OD₆₀₀ of all the wells containing a culture were monitored every 15 min for 24 h. The plate was shook continuously and kept at 30 °C.

The $\Delta NUP85$ haploid strains containing GKYP02 and prs315, GKY04, or GKY07 and the $\Delta NUP145$ haploid strains containing GKYP03 and prs315, GKY05, or GKY08 were grown as a liquid culture in YPD overnight shaking at 30 °C. The cultures were then diluted in either SC medium (containing all 20 amino acids, uracil and 2% glucose) or SC medium containing 5-fluoroorotic acid (5-FOA) at 1 mg/ml. The plate was placed into a Synergy 2 multi-mode microplate reader and the OD₆₀₀ of all the wells containing a culture were monitored every 15 min for 24 h. The plate was shook continuously and kept at 30 °C.

The growth of three separate clones of each strain was tested in quintuplicates. The data was graphed in Prism (GraphPad Software). All error bars represent the standard deviation of the mean.

PDB accession code

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 4YCZ.

Acknowledgements

The X-ray crystallography work was conducted at the APS NE-CAT beamlines, which are supported by award GM103403 from the National Institute of General Medical Sciences, US National Institutes of Health (NIH). Use of the APS is supported by the US Department of Energy, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. We thank K. Rajashankar (APS NE-CAT) for help in phasing the structure; E. Brignole (MIT) for help with generating the cryo-ET consensus map; L. Berchowitz and A. Amon (MIT) for help with the *in*

in vivo fitness analysis. Research was supported by the US NIH under grant number R01GM77537 (T.U.S.) and T32GM007287 (K.K. and K.E.K.) and the US National Science Foundation Graduate Research Fellowship under Grant No. 1122374 (K.E.K.).

References

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
- Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007). The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701.
- Algret, R., Fernandez-Martinez, J., Shi, Y., Kim, S.J., Pellarin, R., Cimermancic, P., Cochet, E., Sali, A., Chait, B.T., Rout, M.P., et al. (2014). Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway. *Mol. Cell. Proteomics* **13**, 2855–2870.
- Andersen, K.R., Leksa, N.C., and Schwartz, T.U. (2013). Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **81**, 1857–1861.
- Bar-Peled, L., Chantranupong, L., Cherniack, A.D., Chen, W.W., Ottina, K.A., Grabiner, B.C., Spear, E.D., Carter, S.L., Meyerson, M., and Sabatini, D.M. (2013). A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. *Science* **340**, 1100–1106.
- Belgareh, N., Rabut, G., Bai, S.W., van Overbeek, M., Beaudouin, J., Daigle, N., Zatssepina, O.V., Pasteau, F., Labas, V., Fromont-Racine, M., et al. (2001). An evolutionarily conserved NPC subcomplex, which redistributes in part to kinetochores in mammalian cells. *J. Cell Biol.* **154**, 1147–1160.
- Berke, I.C., Boehmer, T., Blobel, G., and Schwartz, T.U. (2004). Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J. Cell Biol.* **167**, 591–597.
- Bilokapic, S., and Schwartz, T.U. (2012). Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proc. Natl. Acad. Sci. USA* **109**, 15241–15246.
- Boehmer, T., Jeudy, S., Berke, I.C., and Schwartz, T.U. (2008). Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Mol. Cell* **30**, 721–731.
- Brohawn, S.G., and Schwartz, T.U. (2009). Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat. Struct. Mol. Biol.* **16**, 1173–1177.

- Brohawn, S.G., Leksa, N.C., Spear, E.D., Rajashankar, K.R., and Schwartz, T.U. (2008). Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* 322, 1369–1373.
- Brohawn, S.G., Partridge, J.R., Whittle, J.R.R., and Schwartz, T.U. (2009). The nuclear pore complex has entered the atomic age. *Structure* 17, 1156–1168.
- Bui, K.H., Appen, von, A., DiGuilio, A.L., Ori, A., Sparks, L., Mackmull, M.-T., Bock, T., Hagen, W., Andrés-Pons, A., Glavy, J.S., et al. (2013). Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* 155, 1233–1243.
- Chen, V.B., Arendall, W.B.I., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21.
- Ciccarelli, F.D., Doerks, T., Mering, von, C., Creevey, C.J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287.
- Cook, A., Bono, F., Jinek, M., and Conti, E. (2007). Structural biology of nucleocytoplasmic transport. *Annu Rev Biochem* 76, 647–671.
- Cronshaw, J.M., Krutchinsky, A.N., Zhang, W., Chait, B.T., and Matunis, M.J. (2002). Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* 158, 915–927.
- Debler, E.W., Ma, Y., Seo, H.-S., Hsia, K.-C., Noriega, T.R., Blobel, G., and Hoelz, A. (2008). A fence-like coat for the nuclear pore membrane. *Mol. Cell* 32, 815–826.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
- Fernandez-Martinez, J., Phillips, J., Sekedat, M.D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J.D., Williams, R., Stokes, D.L., Chait, B.T., Sali, A., et al. (2012). Structure-function mapping of a heptameric module in the nuclear pore complex. *J. Cell Biol.* 196, 419–434.
- Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J.B., Carpenter, M.L., Field, M.C., Kuo, A., Paredez, A., Chapman, J., Pham, J., et al. (2010). The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140, 631–642.
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.
- Harel, A., Orjalo, A.V., Vincent, T., Lachish-Zalait, A., Vasu, S., Shah, S., Zimmerman, E., Elbaum, M., and Forbes, D.J. (2003). Removal of a single pore subcomplex results in vertebrate nuclei devoid of nuclear pores. *Mol. Cell* 11, 853–864.
- Hetzer, M.W., and Wenthe, S.R. (2009). Border control at the nucleus: biogenesis and organization of the nuclear membrane and pore complexes. *Dev. Cell* 17, 606–616.

Hsia, K.-C., Stavropoulos, P., Blobel, G., and Hoelz, A. (2007). Architecture of a coat for the nuclear pore membrane. *Cell* *131*, 1313–1326.

Jeudy, S., and Schwartz, T.U. (2007). Crystal structure of nucleoporin Nic96 reveals a novel, intricate helical domain architecture. *J. Biol. Chem.* *282*, 34904–34912.

Kampmann, M., and Blobel, G. (2009). Three-dimensional structure and flexibility of a membrane-coating module of the nuclear pore complex. *Nat. Struct. Mol. Biol.* *16*, 782–788.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* *4*, 363–371.

Kim, S.J., Fernandez-Martinez, J., Sampathkumar, P., Martel, A., Matsui, T., Tsuruta, H., Weiss, T.M., Shi, Y., Markina-Inarrairaegui, A., Bonanno, J.B., et al. (2014). Integrative structure-function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. *Mol. Cell. Proteomics* *13*, 2911–2926.

Leksa, N.C., Brohawn, S.G., and Schwartz, T.U. (2009). The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture. *Structure* *17*, 1082–1091.

Liu, X., Liu, X., Mitchell, J.M., Wozniak, R.W., Blobel, G., Fan, J., and Fan, J. (2012). Structural evolution of the membrane-coating module of the nuclear pore complex. *Proc. Natl. Acad. Sci. USA* *109*, 16498–16503.

Lutzmann, M., Kunze, R., Buerer, A., Aebi, U., and Hurt, E. (2002). Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *Embo J.* *21*, 387–397.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. *eLife* *2*, e01456.

Nagy, V., Hsia, K.-C., Debler, E.W., Kampmann, M., Davenport, A.M., Blobel, G., and Hoelz, A. (2009). Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proc. Natl. Acad. Sci. USA* *106*, 17693–17698.

Neumann, N., Lundin, D., and Poole, A.M. (2010). Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *PLoS ONE* *5*, e13241.

Ori, A., Banterle, N., Iskar, M., Andrés-Pons, A., Escher, C., Khanh Bui, H., Sparks, L., Solis-Mezarino, V., Rinner, O., Bork, P., et al. (2013). Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* *9*, 648.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* *276*, 307–326.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* *25*, 1605–1612.

- Reichelt, R., Holzenburg, A., Buhle, E.L., Jarnik, M., Engel, A., and Aebi, U. (1990). Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. *J. Cell Biol.* *110*, 883–894.
- Ries, J., Kaplan, C., Platonova, E., Eghlidi, H., and Ewers, H. (2012). A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. *Nat. Methods* *9*, 582–584.
- Rout, M.P., Aitchison, J.D., Suprpto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* *148*, 635–651.
- Sampathkumar, P., Gheyi, T., Miller, S.A., Bain, K.T., Dickey, M., Bonanno, J.B., Kim, S.J., Phillips, J., Pieper, U., Fernandez-Martinez, J., et al. (2011). Structure of the C-terminal domain of *Saccharomyces cerevisiae* Nup133, a component of the nuclear pore complex. *Proteins* *79*, 1672–1677.
- Schrader, N., Stelter, P., Flemming, D., Kunze, R., Hurt, E., and Vetter, I.R. (2008). Structural basis of the nic96 subcomplex organization in the nuclear pore channel. *Mol. Cell* *29*, 46–55.
- Seo, H.-S., Ma, Y., Debler, E.W., Wacker, D., Kutik, S., Blobel, G., and Hoelz, A. (2009). Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc. Natl. Acad. Sci. USA* *106*, 14281–14286.
- Siniosoglou, S., Wimmer, C., Rieger, M., Doye, V., Tekotte, H., Weise, C., Emig, S., Segref, A., and Hurt, E.C. (1996). A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. *Cell* *84*, 265–275.
- Strambio-de-Castillia, C., Niepel, M., and Rout, M.P. (2010). The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat. Rev. Mol. Cell. Biol.* *11*, 490–501.
- Stuwe, T., Correia, A.R., Lin, D.H., Paduch, M., Lu, V.T., Kossiakoff, A.A., and Hoelz, A. (2015). Nuclear pores. Architecture of the nuclear pore complex coat. *Science* *347*, 1148–1152.
- Süel, K.E., Cansizoglu, A.E., and Chook, Y.M. (2006). Atomic resolution structures in nuclear transport. *Methods* *39*, 342–355.
- Szymborska, A., de Marco, A., Daigle, N., Cordes, V.C., Briggs, J.A.G., and Ellenberg, J. (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* *341*, 655–658.
- Terwilliger, T.C., DiMaio, F., Read, R.J., Baker, D., Bunkóczi, G., Adams, P.D., Grosse-Kunstleve, R.W., Afonine, P.V., and Echols, N. (2012). phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J. Struct. Funct. Genomics* *13*, 81–90.
- Thierbach, K., Appen, von, A., Thoms, M., Beck, M., Flemming, D., and Hurt, E. (2013). Protein interfaces of the conserved Nup84 complex from *Chaetomium thermophilum* shown by crosslinking mass spectrometry and electron microscopy. *Structure* *21*, 1672–1682.

- Vasu, S., Shah, S., Orjalo, A., Park, M., Fischer, W.H., and Forbes, D.J. (2001). Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. *J. Cell Biol.* *155*, 339–354.
- Vollmer, B., and Antonin, W. (2014). The diverse roles of the Nup93/Nic96 complex proteins - structural scaffolds of the nuclear pore complex with additional cellular functions. *Biol. Chem.* *395*, 515–528.
- Walther, T.C., Alves, A., Pickersgill, H., Loiodice, I., Hetzer, M., Galy, V., Hülsmann, B.B., Köcher, T., Wilm, M., Allen, T., et al. (2003). The conserved Nup107-160 complex is critical for nuclear pore complex assembly. *Cell* *113*, 195–206.
- Waterhouse, A.M., Waterhouse, A.M., Procter, J.B., Procter, J.B., Martin, D.M.A., Martin, D.M.A., Clamp, M., Clamp, M., Barton, G.J., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* *25*, 1189–1191.
- Whittle, J.R.R., and Schwartz, T.U. (2009). Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element. *J. Biol. Chem.* *284*, 28442–28452.
- Whittle, J.R.R., and Schwartz, T.U. (2010). Structure of the Sec13-Sec16 edge element, a template for assembly of the COPII vesicle coat. *J. Cell Biol.* *190*, 347–361.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* *67*, 235–242.
- Yang, Q., Rout, M.P., and Akey, C.W. (1998). Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell* *1*, 223–234.

Tables

Table 1. Data collection and refinement statistics

	Nup85(SeMet) ₂₅₇₋₁₁₈₁ , Nup120 ₉₅₂₋₁₂₄₁ , Sec13-Nup145C ₂₃₃₋₇₉₁	Nup85(SeMet) ₂₅₇₋₁₁₈₁ , Nup120(SeMet) ₉₅₂₋₁₂₄₁ , Sec13-Nup145C(SeMet) ₂₃₃₋₇₉₁
Data collection		
Space group	C2	P1
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	104.98, 212.02, 170.64	118.96, 107.67, 163.09
<i>a</i> , <i>b</i> , <i>γ</i> (°)	90, 107.2, 90	74.3, 80.4, 63.2
Resolution (Å)	163 - 4.10 (4.25 - 4.10) ^a	157 - 4.00 (4.14 - 4.00)
<i>R</i> _{sym}	0.19 (1.00)	0.21 (0.97)
<i>R</i> _{p.i.m.}	0.09 (0.57)	0.13 (0.63)
<i>I</i> / <i>σI</i>	8.9 (0.9)	5.3 (0.8)
<i>CC</i> _{1/2}	1.00 (0.68)	1.00 (0.58)
Completeness (%)	98.2 (93.5)	90.3 (81.3)
Redundancy	6.1 (4.7)	3.6 (2.9)
Refinement		
Resolution (Å)	163 - 4.10	
No. reflections	53648	
No. free reflections	3885	
<i>R</i> _{work} / <i>R</i> _{free}	31.9/35.8	
No. atoms	10070	
Protein	10070	
<i>B</i> factors		
Protein	161.7	
r.m.s. deviations		
Bond lengths (Å)	0.002	
Bond angles (°)	0.64	
Ramachandran		
Favored (%)	91.0	
Allowed (%)	8.2	
Outlier (%)	0.8	
Clashscore	8.33	
Molprobrity percentile	97 th	

^aValues in parentheses are for highest-resolution shell. One crystal was used for each dataset.

Table 2. Proportion of residues and side chains resolved in the crystal structure

Protein	<i>M. thermophila</i> to <i>S. cerevisiae</i>		<i>M. thermophila</i> to <i>H. sapiens</i>		<i>S. cerevisiae</i> to <i>H.</i> <i>sapiens</i>	
	Identity	Similarity	Identity	Similarity	Identity	Similarity
Nup85	12%	25%	14%	25%	13%	52%
Nup120	11%	46%	12%	29%	10%	31%
Nup145C	14%	43%	20%	40%	17%	40%
Nup84	17%	54%	17%	51%	19%	56%
Nup133	14%	41%	16%	44%	13%	46%
Sec13	60%	75%	50%	68%	50%	66%
Seh1	25%	54%	22%	52%	34%	76%

Table 3. Sequence Identity and similarity between *M. thermophila*, *S. cerevisiae*, and *H. sapiens* Y-complex nucleoporins

Protein	Fully Resolved Residues ¹	Backbone Resolved Residues	Total Resolved Residues	Total Residues
Sec13	239 (79%)	31 (10%)	270 (89%)	304
Nup145C ₂₃₃₋₇₉₁	257 (46%)	188 (34%)	445 (80%)	559
Nup85 ₂₅₇₋₁₁₈₁	424 (46%)	166 (18%)	590 (64%)	925
Nup120 ₉₅₂₋₁₂₄₁	134 (46%)	63 (22%)	197 (68%)	290

¹Percentage refers to the percent of total residues in the crystallization construct that are fully resolved (with side chains), backbone resolved (no side chains), or the sum of the two (total resolved).

Table 4. Yeast strains used in the study

Plasmid	Description	Parent Plasmid
YCplac22	CEN plasmid with Ura- selection	-----
prs315	CEN plasmid with His- selection	-----
GKYp01	Full length Nup120 + promoter and terminator	YCplac33
GKYp02	Full length Nup85 + promoter and terminator	YCplac33
GKYp03	Full length Nup145 + promoter and terminator	YCplac33
GKYp04	Full length Nup85 + promoter and terminator	prs315
GKYp05	Full length Nup145 + promoter and terminator	prs315
GKYp06	Nup120 (1-1015; Δ α 30) + promoter and terminator	GKYp01
GKYp07	Nup85 (1-718; Δ α 30) + promoter and terminator	GKYp04
GKYp08	Nup145 (1-1293; Δ α 27) + promoter and terminator	GKYp05

Strain	Description	Parent Strain
GKY112	nup120 Δ ::KanMX MAT α his3 Δ leu2 Δ ura3 Δ	-----
GKY113	NUP85/nup85 Δ ::KanMX MAT α /MATA his3 Δ /his3 Δ leu2 Δ /leu2 Δ ura3 Δ /ura3 Δ	-----
GKY114	NUP145/nup145 Δ ::KanMX MAT α /MATA his3 Δ /his3 Δ leu2 Δ /leu2 Δ ura3 Δ /ura3 Δ	-----
GKY115	nup120 Δ ::KanMX with YCplac22	GKY112
GKY116	nup120 Δ ::KanMX with GKYP01	GKY112
GKY117	nup120 Δ ::KanMX with GKYP06	GKY112
GKY121	NUP85/nup85 Δ ::KanMX with GKYP02	GKY113
GKY122	NUP145/nup145 Δ ::KanMX with GKYP03	GKY114
GKY123	nup85 Δ ::KanMX MAT α with GKYP02	GKY113
GKY124	nup145 Δ ::KanMX MAT α with GKYP04	GKY114
GKY125	nup85 Δ ::KanMX MAT α with GKYP02 and prs315	GKY123
GKY125	nup85 Δ ::KanMX MAT α with GKYP02 and GKYP04	GKY123
GKY125	nup85 Δ ::KanMX MAT α with GKYP02 and GKYP07	GKY123
GKY126	nup145 Δ ::KanMX MAT α with GKYP03 and prs315	GKY124
GKY127	nup145 Δ ::KanMX MAT α with GKYP03 and GKYP05	GKY124
GKY128	nup145 Δ ::KanMX MAT α with GKYP03 and GKYP08	GKY124

Figures

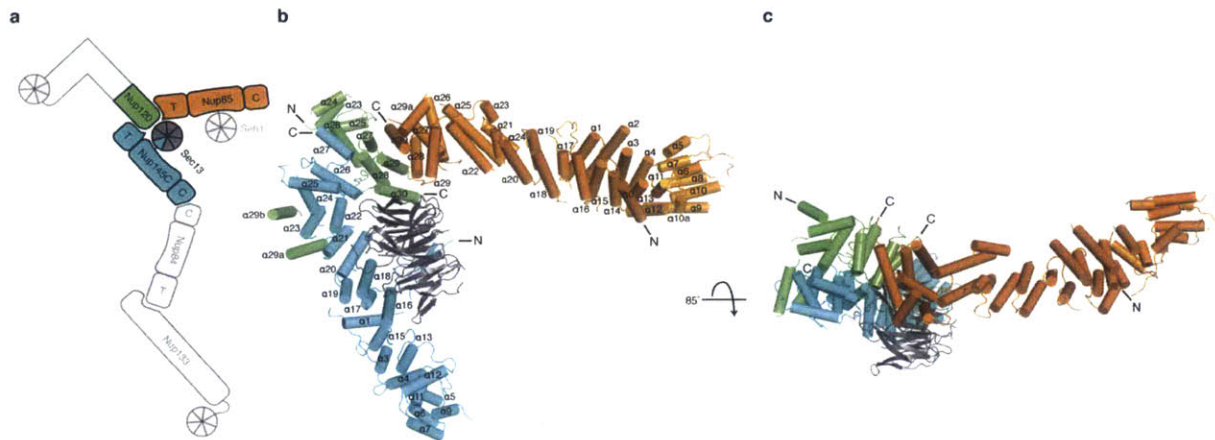


Figure 1. Structure of the *Myceliophthora thermophila* Y-complex hub at 4.1 Å resolution. (a) Schematic of the Y-complex. Regions included in the crystallization construct are colored, other Y-complex regions in grey. Elements of ACE1 fold proteins are indicated: T – tail and C – crown flank the central trunk element. (b) The hub structure is colored as follows: Nup85 (orange), Nup120 (green), Nup145C (cyan), and the Sec13 β -propeller (grey). N and C termini are indicated for the helical proteins. Helices are numbered according to previously solved *S. cerevisiae* fragments (Bilokapic and Schwartz, 2012; Brohawn and Schwartz, 2009; Brohawn et al., 2008). Numbers that include a letter modifier indicate helical elements not present in *S. cerevisiae*. (c) Top-down view of the hub. The N terminus of Nup145C is not indicated because it is obscured by the 85° rotation.

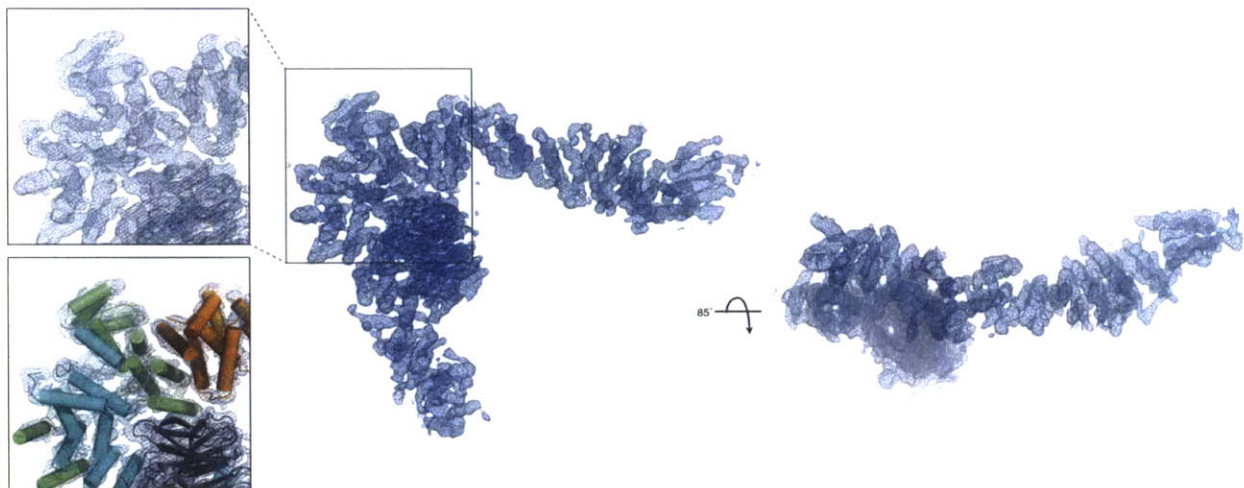


Figure 2. Electron density map for hub structure. Final 2Fo-Fc electron density map for the hub contoured at 1.5σ , shown in the same view as Fig. 1 (Inset, top) zoom-in on the interaction region of the hub. (Inset, bottom) the hub structure is placed into the electron density.

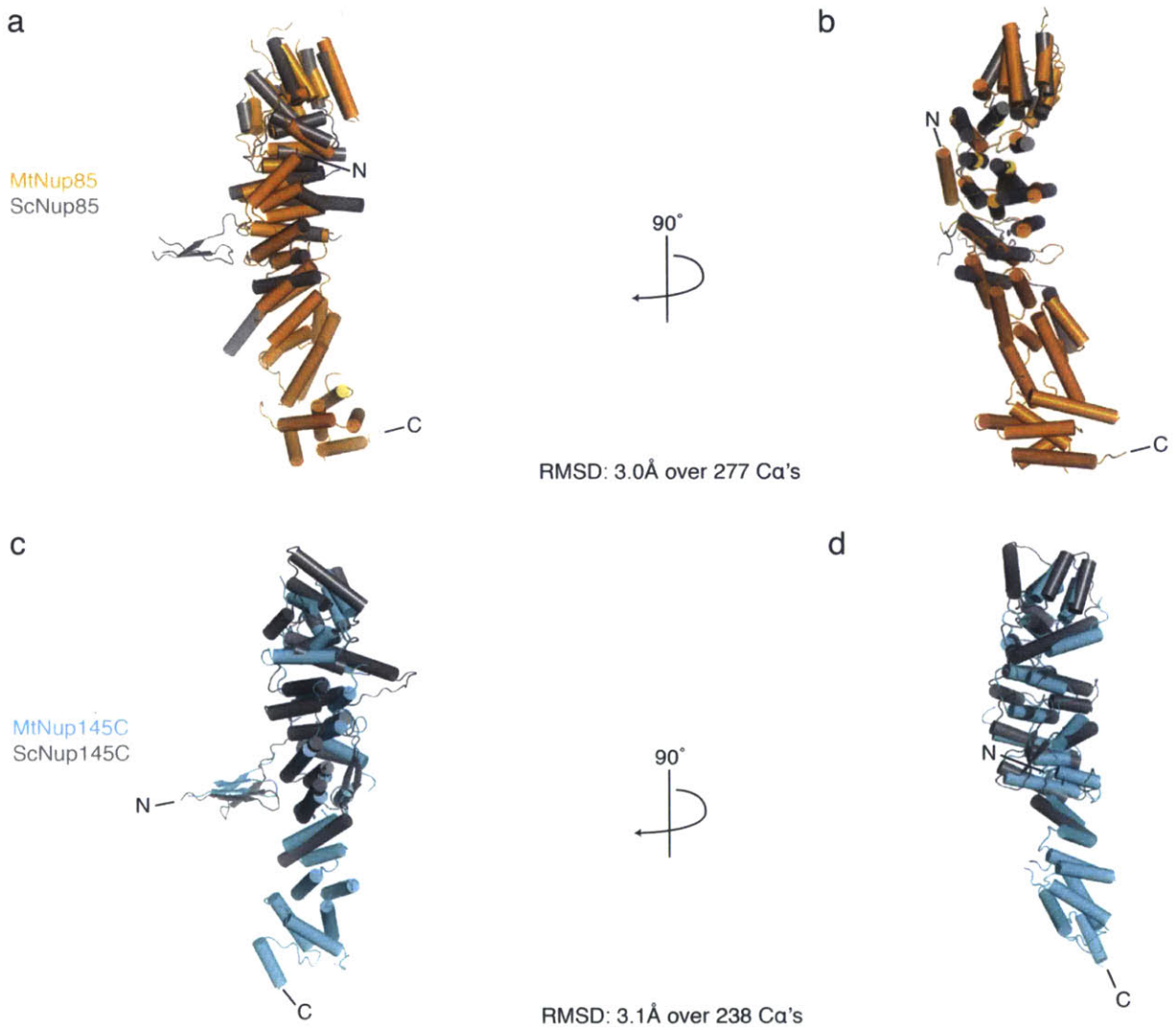


Figure 3. Superposition of *M. thermophila* and *S. cerevisiae* ACE1 proteins of the hub. (a) Overlay of mtNup85 (orange) and scNup85 (grey) reveals structural conservation, despite low (14%) sequence identity. N and C termini of the mtNups are labeled. **(b)** 90° rotation of the superposed structures. **(c)** Overlay of mtNup145C (cyan) and scNup145C (grey) (20% identity). **(d)** 90° rotation of the superposed structures.

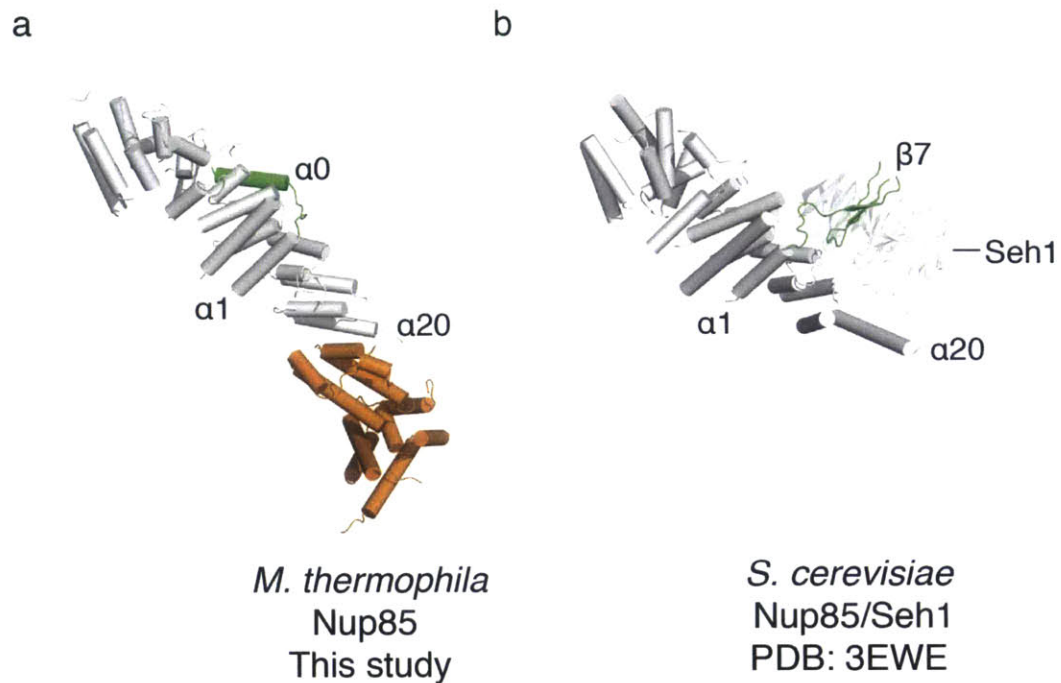


Figure 4. mtNup85 lacks an Seh1 binding motif. (a) mtNup85, where the structural element N-terminal of $\alpha 1$ is an additional helix, $\alpha 0$ (green). The crown and trunk elements, which align to the solved scNup85 fragment, are shown in grey. The tail element, not solved in *S. cerevisiae*, is shown in orange. (b) scNup85, aligned to mtNup85, where the Seh1 insertion blade, $\beta 7$ (green), is N-terminal to $\alpha 1$. Seh1 is labeled and scNup85 is shown in grey.

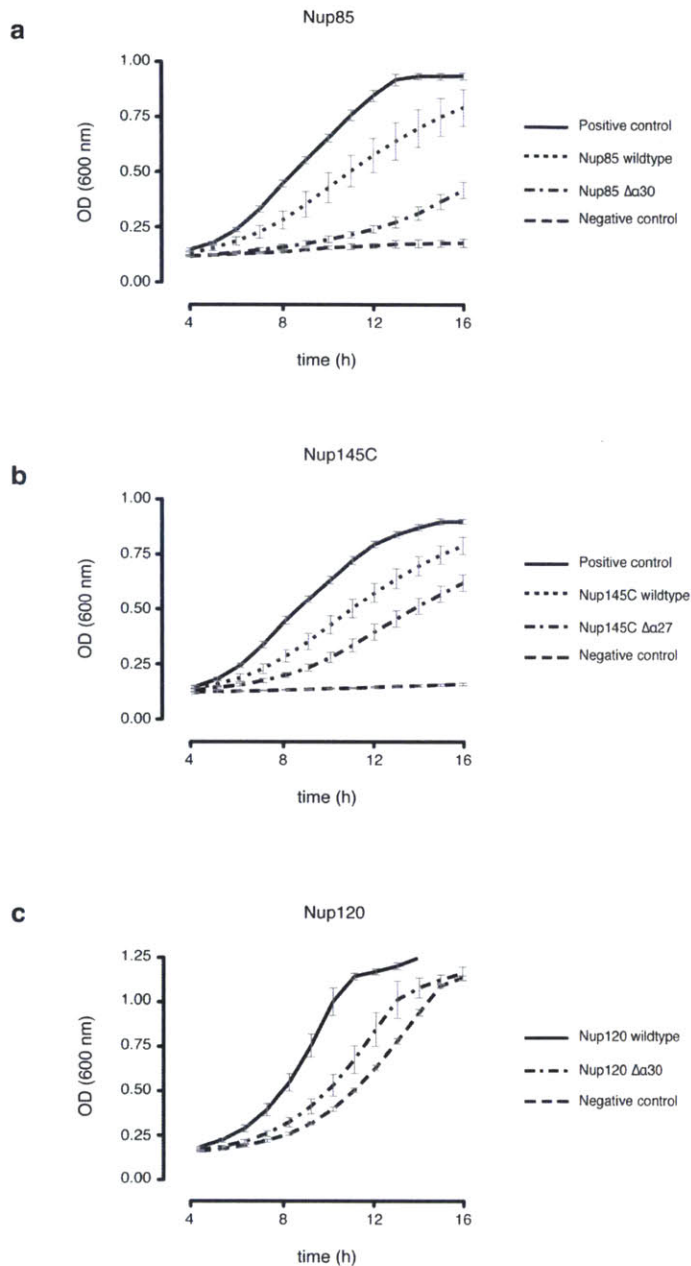


Figure 5. Fitness analysis of hub interactions. (a) Growth curves of *NUP85 Δ* strains carrying *NUP85:URA3* and either empty pRS315 (negative control), Nup85 wildtype, or Nup85 $\Delta\alpha 30$ grown in the presence of 5-FOA. The positive control is the *NUP85 Δ* strain carrying *NUP85:URA3* and empty pRS315 grown in the absence of 5-FOA. (b) Growth curves of *NUP145 Δ* strains carrying *NUP145:URA3* and either empty pRS315 (negative control), Nup145C wildtype, or Nup145C $\Delta\alpha 27$ grown in the presence of 5-FOA. The positive control is the *NUP145 Δ* strain carrying *NUP145:URA3* and empty pRS315 grown in the absence of 5-FOA. (c) The growth curves of *NUP120 Δ* strains carrying YClac33 empty vector (negative control), Nup120 wildtype, or Nup120 $\Delta\alpha 30$ grown in YPD. Four technical replicates ($n=4$ OD measurements) for each of three biological replicates ($n=3$), from separate colonies, were performed at 30 °C for all experiments. All error bars are standard deviation of the mean (s.e.m.).

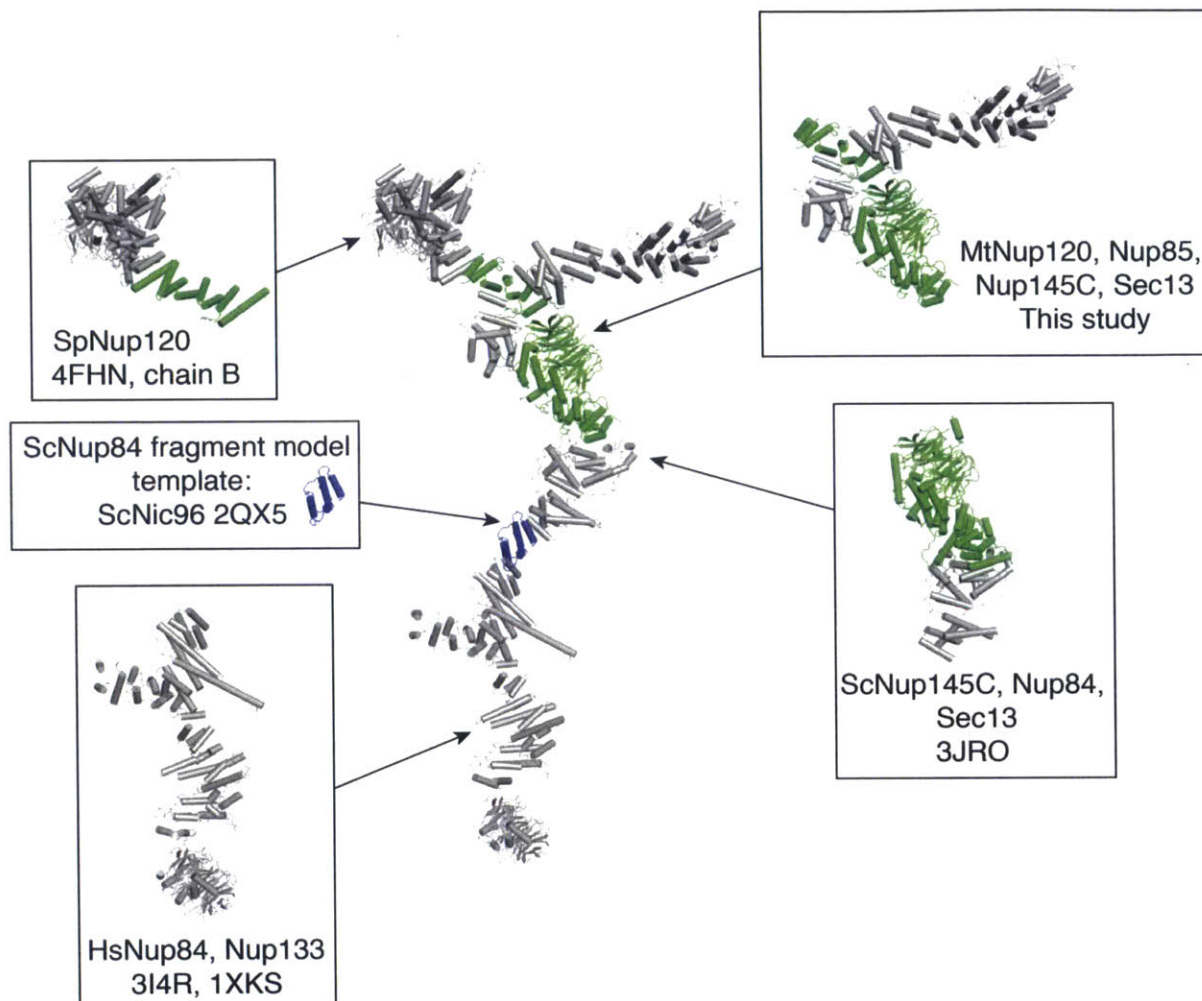


Figure 6. Composite Y-complex structure generated through overlapping crystal structure fragments. Overlapping elements between the hub and previously solved structures are shown in green. Grey elements are non-redundant crystal structures used to generate the composite. The only portion of the composite that is modeled is a four-helix bundle in Nup84 (blue).

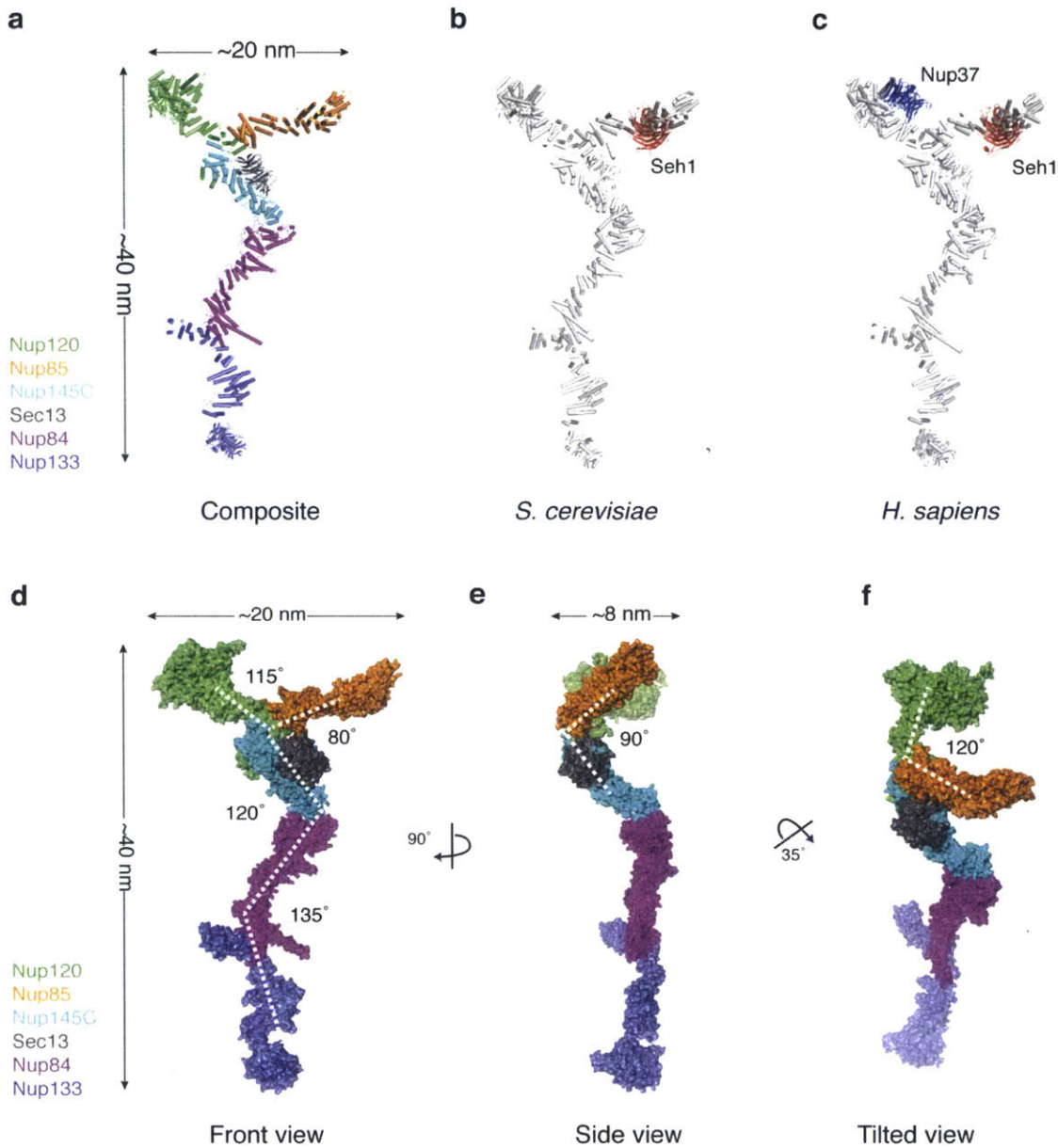


Figure 7. Composite high-resolution structure of the Y-complex. (a) Composite, hexameric Y-complex core constructed from the hub structure (Fig. 1) combined with previously published X-ray crystal structure fragments. Within Nup84, 4 helices were modeled computationally. (b) Composite *S. cerevisiae* Y-complex based on (a), with *S. cerevisiae* sequences threaded onto existing homologous structures. Compared to the universally conserved hexameric Y-complex core shown in (a), Seh1 (red) is an additional component found in many organisms, including yeast. (c) Composite *H. sapiens* Y-complex with *H. sapiens* sequences threaded onto existing homologous structures. Nup37 (blue) is another Y-complex component only found in a subset of eukaryotes, including humans (Neumann et al., 2010). (d) Space filling surface view of the composite, hexameric Y-complex viewed from the front. (e) Side view. (f) Tilted view.

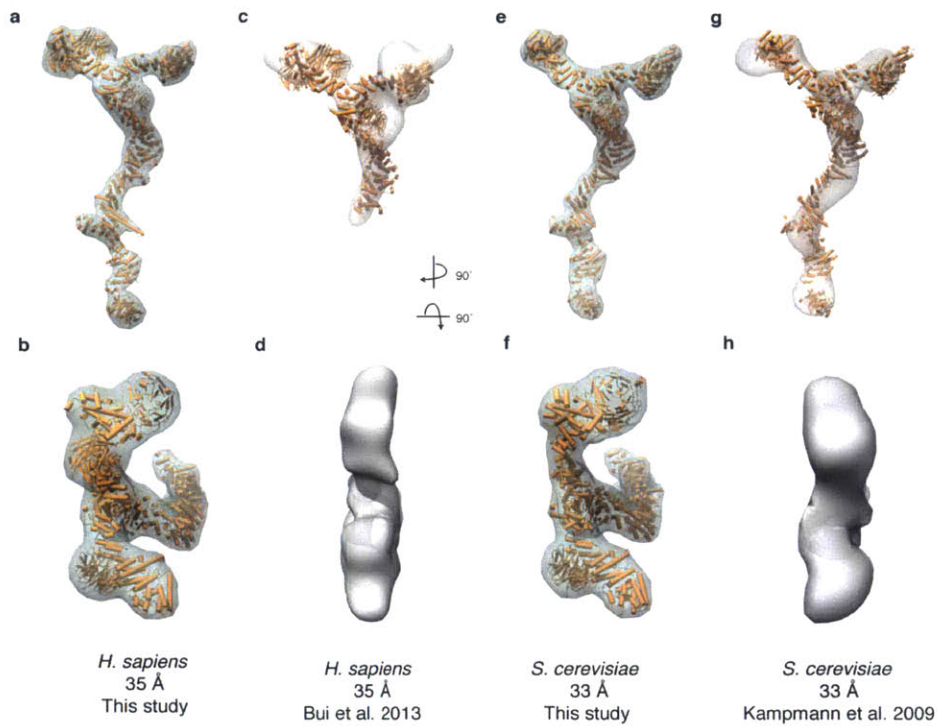


Figure 8. Comparison of the X-ray based, composite Y-complex structure with published 3-D EM reconstruction structures. (a,b) Electron density envelope around the composite *H. sapiens* Y-complex calculated for 35 Å resolution from front (a) or top (b) view. (c,d) 3-D EM reconstruction of the *H. sapiens* Y-complex with an overlay of the composite model, fitted computationally from front (c) or top (d) view. (e,f) Electron density envelope around the composite *S. cerevisiae* Y-complex calculated at 33 Å resolution from front (e) or top (f) view. (g,h) 3-D EM reconstruction of the *S. cerevisiae* Y-complex with an overlay of the composite model, fitted computationally from front (g) or top (h) view.



Figure 9. Flexibility of the Y-complex. Experimentally observed hinge regions of the Y-complex are denoted by dashed lines. (a) (Bilokapic and Schwartz, 2012). (b–c) (Brohawn et al., 2008). (d–e), (Bui et al., 2013) (Berke et al., 2004) (Boehmer et al., 2008) (Whittle and Schwartz, 2009) (Kampmann and Blobel, 2009) (f) (Berke et al., 2004).

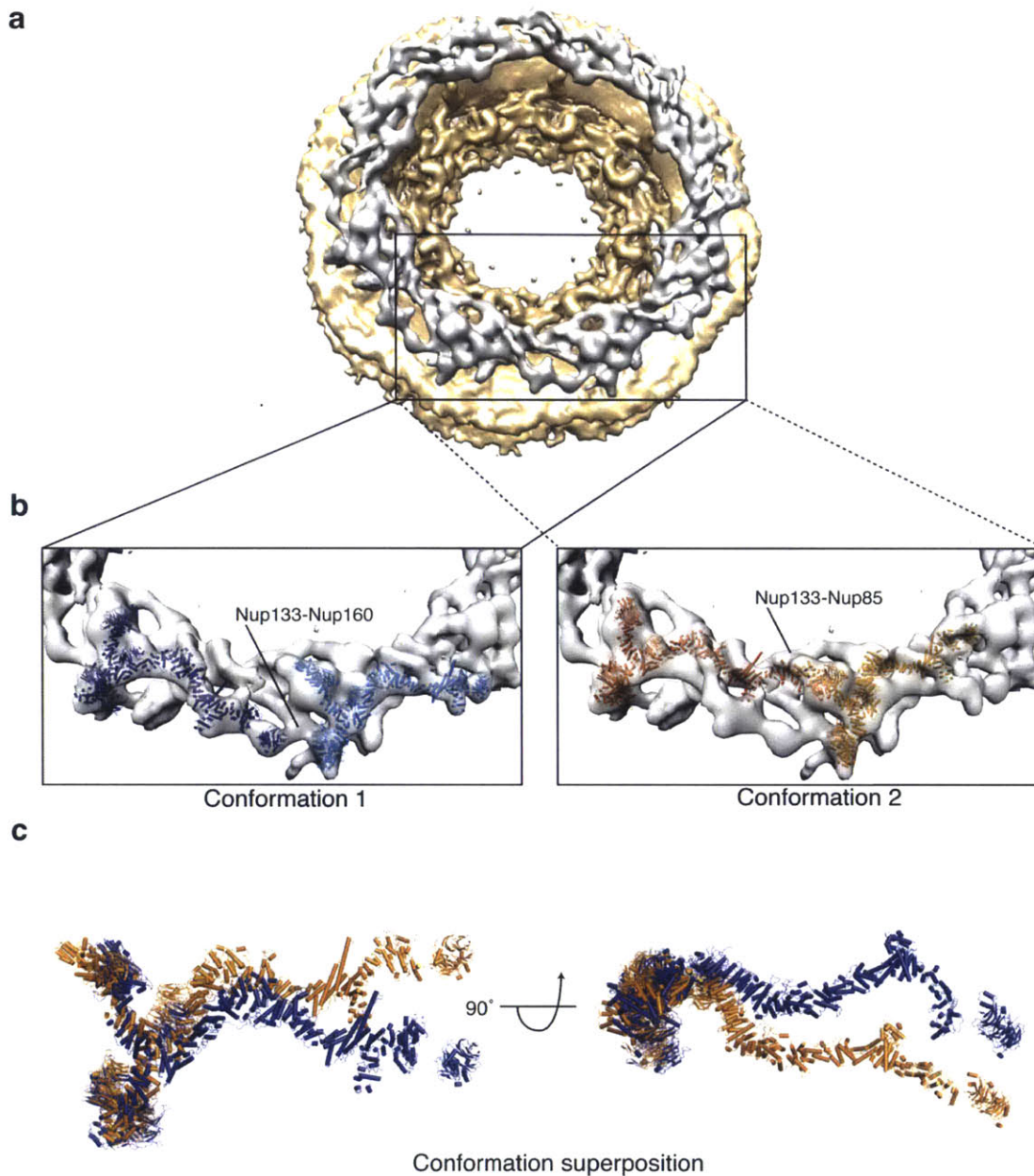
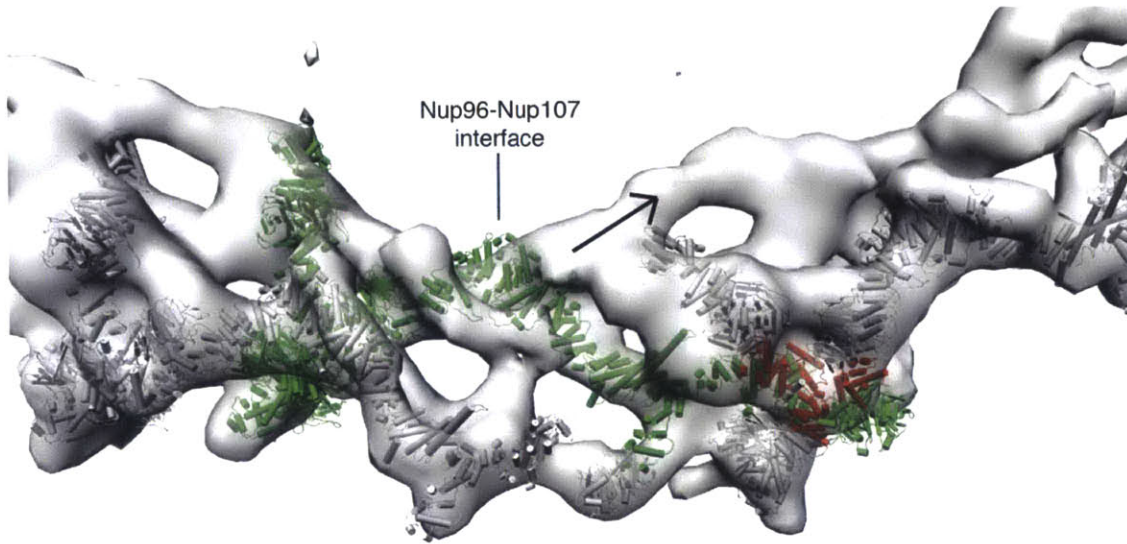
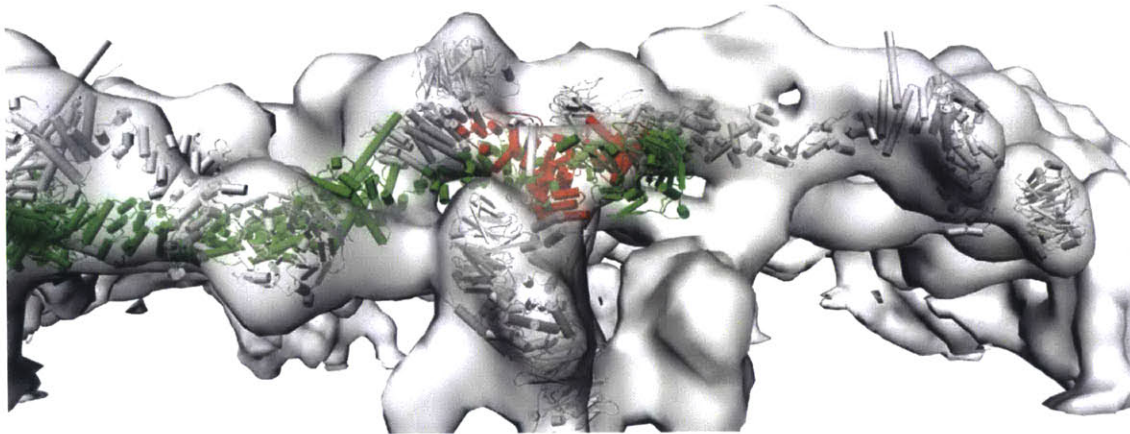


Figure 10. Fitting of the composite *H. sapiens* Y-complex into the cryo-ET map of the entire NPC. (a) Consensus map calculated from the cryo-ET map of the human NPC (EMD code: 2444) (Bui et al., 2013). The cytoplasmic ring density is highlighted in grey for clarity. (b) The top scoring fit of the composite *H. sapiens* Y-complex (conformation 1) and the second top scoring fit (conformation 2) are depicted. (c) A superposition of the two fits from (b) shows that they are related by a $\sim 20^\circ$ rotation about the Y-complex hub, and substantial bending of the long stack.



Bui et al. fit - Top view



Bui et al. fit - Side view

Figure 11. Fitting of the composite *H. sapiens* Y-complex into an inner ring position in the 3-D EM tomography map suffers from significant steric clashes with the outer ring fits. Inner ring fit (3rd top scoring solution) is colored green. One of the outer ring fits (1st solution) is colored grey and regions of steric clash with Nup133 of the inner ring fit are colored red. The arrow (top view) shows where the Y complex stem needs to move in order to match the Bui et al. inner ring placement and avoid steric clashes with the adjacent outer ring. Such stem placement would involve rotation at or around the Nup96-Nup107 interface, which is unlikely due to the energetic cost of disrupting the complex's hydrophobic core.

Chapter 3: Structural characterization of Bardet-Biedl syndrome 9 protein (BBS9).

The material presented in this chapter was adapted, with permission, from the following publication:

Knockenbauer K.E. and Schwartz T.U. (2015). Structural Characterization of Bardet-Biedl Syndrome 9 Protein (BBS9). *J Biol Chem* 290, 19569-19583.

Author contributions: TUS conceived the study. KEK and TUS designed the experiments, analyzed the data, and wrote the manuscript. KEK performed the experiments.

Introduction

The primary cilium is a singular, microtubule-based protrusion from the plasma membrane found in nearly all vertebrate cell types. The microtubule bundle, termed the axoneme, forms the structural scaffold of this finger-like projection and is surrounded by the lipid bilayer except at its base, where it originates from the mother centriole in the cytoplasm (Nachury et al., 2010). Unlike motile cilia, the role of the primary cilium is not locomotion but rather signal transduction and, as a consequence, is essential to Hedgehog and non-canonical Wnt signaling pathways (Singla and Reiter, 2006). Despite being continuous with the plasma membrane, the membrane of the primary cilium has a unique composition of proteins, namely an enrichment of specific G-protein coupled receptors (GPCRs) involved in signaling (Corbit et al., 2005; Nachury et al., 2010). How this organelle maintains the integrity and specificity of its compartment remains an active area of investigation, though a septin diffusion barrier has been shown to play a critical role (Chih et al., 2011; Hu et al., 2010).

One such component to the establishment or maintenance of the unique protein composition of the ciliary membrane is the BBSome, an IP-stable, eight-membered protein complex comprising BBS1, BBS2, BBS4, BBS5, BBS7, BBS8, BBS9, and BBIP10 (Loktev et al., 2008; Nachury et al., 2007). The genes coding for seven of these proteins were originally identified based on their mutation in a rare, autosomal recessive genetic disorder termed Bardet-Biedl syndrome (BBS, ref. (Tobin and Beales, 2009)). BBS is a pleiotropic disease characterized by mental retardation, polycystic kidneys, obesity, vision problems, and hypogonadism (Nachury et al., 2010; Tobin and Beales, 2009). Evidence for BBSome function was seen in BBS2 and BBS4 knockout mice, where a GPCR resident of the ciliary membrane, somatostatin receptor type 3, fails to localize to the primary cilium in neurons (Berbari et al., 2008). The BBSome has been shown to localize to the ciliary membrane and to coat liposomes in an Arl6-GTP dependent manner through thin section electron microscopy, leading to the

hypothesis that the complex forms a membrane protein transport coat similar to COPI or COPII (Jin et al., 2010). Consistent with this hypothesis, proteins of the BBSome and the coat forming elements of COPI, COPII, and clathrin share a similar domain composition and architecture (Jin et al., 2010). Namely, N-terminal β -propeller and C-terminal appendage domains, as is the case for BBS1, BBS2, BBS7, and BBS9, and alpha-solenoid domains, as in BBS4 and BBS8 (Jin et al., 2010). Recently, an additional role of the BBSome in the assembly and recycling of intraflagellar transport particles, which move cargo along the axoneme via microtubule motors, has been identified through a loss of function mutation of BBS1 in *C. elegans* (Wei et al., 2012).

Given the structural homology to vesicle coat proteins on a domain level, we set out to assess if and to what degree this homology held true on the atomic level. To this end, another group has already reported the first high-resolution crystal structure of a BBS protein, the BBS1 N-terminal β -propeller from *C. reinhardtii* in complex with the Arl6-GTPase (Mourão et al., 2014). Here we present the high-resolution crystal structure of the BBS9 N-terminal β -propeller from *H. sapiens* and explain the structural basis of the G141R point mutation that causes BBS in humans (Nishimura et al., 2005). BBS9 knockdown in zebrafish leads to defects in ciliogenesis and neuronal development that the G141R mutant fails to rescue (Veleri et al., 2012).

Results

Structure of the HsBBS9₁₋₄₀₇ N-terminal domain

To structurally characterize *Homo sapiens* BBS9 we recombinantly expressed the protein in *E. coli*. Full length HsBBS9₁₋₈₈₇ suffered from low yield. In order to identify a construct suitable for crystallization, we performed a limited tryptic digest and identified a protease-resistant HsBBS9 fragment corresponding to the N-terminal domain, residues 1-374, via LC-MS/MS (Fig. 1A, data not shown). Subsequent expression, in bacteria, of HsBBS9₁₋₄₀₇ yielded

sufficient material for structural analysis. We obtained crystals of the HsBBS9 N-terminal domain, which diffracted to 2.4 Å (Table 1). Despite knowing that HsBBS9₁₋₄₀₇ likely adopts a β -propeller fold, we found no molecular replacement solutions using various β -propeller templates. We also employed the wide-search molecular replacement (WSMR) protocol, available through the *SBgrid* software consortium (Morin et al., 2013), without success. The structure was ultimately solved using experimental phases that were obtained through single anomalous dispersion (SAD) on selenomethionine-derivatized (SeMet) HsBBS9₁₋₄₀₇ crystals, which diffracted to 1.8 Å (Table 1). The model was refined against the SeMet dataset to a final R_{work} of 16.9% and an R_{free} of 20.0%. Overall the structure is well ordered, with an average protein B-factor of 35.4 Å². Exceptions to this include the protein termini, as residues 1-2 and 370-407 are disordered, as well as several loop regions, where residues 113-117, 219-232, and 254-255 could not be built with confidence.

HsBBS9₁₋₄₀₇ comprises seven anti-parallel, four-stranded β -sheets that are radially arranged around a central, solvent accessible pore (Fig. 1B, C). HsBBS9₁₋₄₀₇ is approximately 40 Å in diameter across its face and 20 Å in height when viewed side-on. HsBBS9₁₋₄₀₇ is a closed, seven-bladed β -propeller of the WD-40 repeat family, which canonically is characterized by motifs 40 amino acids in length that end in tryptophan (W)-aspartate (D) motifs (Fong et al., 1986). Blades one through six (β 1- β 6), numbered from the N to the C terminus, each begin with the inner-most strand (“A”) and trace to the outer-most strand (“D”), which then connects to the A strand of the next blade. β 7 is an exception to this topology because the N-terminal β -strand of the protein serves as the outer-most D strand of this C-terminal, three-stranded β -sheet (A-C). This irregular connectivity, termed a “Velcro” closure, is common among β -propellers (Wall et al., 1995; Xu and Min, 2011). By physically tethering the N- and C-terminal blades together, this Velcro closure likely stabilizes the HsBBS9₁₋₄₀₇ fold by preventing opening of the propeller.

HsBBS9₁₋₄₀₇ contains additional elaborations between the C and D strands of various blades, such as the helical insertion of β 4 and the extended loops of β 1 and β 2.

Structural analysis of the HsBBS9₁₋₄₀₇ β -propeller

With the HsBBS9₁₋₄₀₇ structure in hand, we sought to gain insight into the function of the β -propeller through analogy to structurally similar proteins. We performed structure-based homology searches using the DALI server (Holm and Rosenstrom, 2010). In total, over 100 significant homologs were found (data not shown), as determined by Z scores above a predetermined threshold ($Z > 10$), and greater than half of the hits had an RMSD below 3.0 Å. The large number of hits obtained likely owes to the fact that HsBBS9₁₋₄₀₇ contains fairly few structural elaborations beyond the conserved, core scaffold that defines a propeller fold. Such elaborations would undoubtedly discriminate in the DALI search between various sub-classes of 7-bladed propellers, reducing the number of significant solutions obtained. Despite this, three structural homologs, through their Z scores, stood out above the rest. The top scoring homolog was the recently solved *Chlamydomonas reinhardtii* (Cr) BBS1 N-terminal β -propeller (PDB: 4v0m_B, RMSD 2.8 Å over 296aa, SeqID 12%, $Z=31.5$), which functions by binding the Arl6-GTPase that tethers the BBSome complex to membranes (Mourão et al., 2014). The CrBBS1 N-terminal domain is also a fairly minimal propeller with few structured insertions. As the top solution, the CrBBS1 N-terminal domain and HsBBS9₁₋₄₀₇ are more similar to each other than they are to other propellers. This result likely extends to the other predicted, 7-bladed propellers within the BBSome, namely the BBS2 and BBS7 N-terminal domains. The other two high scoring homologs were the WD40 domain of a *S. cerevisiae* serine/threonine-protein kinase (Vps15, PDB: 3GRE_A, RMSD 2.4 Å over 291aa, SeqID 11%, $Z=28.6$), which binds the protein Gpa1 (Heenan et al., 2009), and a WD40 protein of the Ski complex (Ski8, PDB: 1SQ9_A, RMSD 2.8 Å over 292aa, SeqID 10%, $Z=27.4$), which binds the proteins Ski2 and Ski3 (Brown

et al., 2000). These hits illustrate that β -propellers are among the most common folds in biology, found in a wide range of proteins with diverse functions. Many β -propellers function as platforms for protein-protein interactions (Xu and Min, 2011).

Given the propensity of β -propellers to form protein-protein interactions, we mapped the electrostatic potential onto the surface of HsBBS9₁₋₄₀₇ to see if a charged binding site is apparent. This analysis reveals a large negative patch around the central pore on one face of the propeller (Fig. 2A). To determine if this feature is conserved we created a multiple sequence alignment (MSA) sampling a diverse set of 17 BBS9 homologs and mapped sequence conservation onto the surface of HsBBS9₁₋₄₀₇ (Fig. 2B). This analysis revealed that the conservation of the charged residues responsible for the negative potential and those lining the pore opening, in general, is poor. Before dismissing this patch as an oddity of *Homo sapiens*, we decided to create structural models of the 17 BBS9 β -propellers in the MSA, based on the *H. sapiens* structure, and map their surface electrostatic potential. This negative patch is notably present in BBS9 β -propellers across the MSA (Fig. 3). Although individual sequence positions are not conserved, these homologs have negatively charged residues at different positions and a relative lack of positive charge in the vicinity of the pore. Solvent exposed carbonyl oxygens additionally contribute to the negative potential of this surface.

Structural analysis of HsBBS9 G141R, which causes Bardet-Biedl syndrome (BBS)

Homozygosity mapping in a small, consanguineous family has revealed a single nucleotide polymorphism (SNP) within the BBS9 N-terminal domain that causes BBS in humans (Nishimura et al., 2005). This SNP confers a missense mutation on residue 141, mutating the glycine at that position to an arginine. Gly141 is strictly conserved across the MSA with the exception of *C. elegans* BBS9, which has an Ile at this position and generally poor conservation in the region surrounding this residue (Fig. 4A). With the HsBBS9₁₋₄₀₇ crystal structure at hand,

we can model the G141R mutation into the structure to assess why it is pathological. Gly141 is positioned within the tightly packed, C terminus of the A strand of blade $\beta 3$ (Fig. 4B), on the opposite face of the negatively charged patch described in Fig. 2. Immediately downstream of Gly141 is a highly curved loop that traces perpendicular to the A strand, packing Phe143 into the hydrophobic core of blade $\beta 3$, before changing direction by 180° (Fig. 5A). Analyzing the Ramachandran plot reveals that Gly141 occupies a special position with dihedral angles that, although accessible to glycine, are forbidden to all other amino acids due to steric restraints imposed by the invariant presence of a β carbon atom (Fig. 5B). Additionally, modeling all possible rotamers of arginine in position 141 of HsBBS9₁₋₄₀₇ reveals a steric clash between the β carbon of the arginine side chain and the C δ 2 carbon of the Tyr186 aromatic ring (Fig. 5C). These carbon atoms are only 2.4 Å apart, which results in an overlap of 1 Å when hydrogen atoms are included. Based on these backbone and side chain restraints, we posit that there is an absolute structural requirement for a glycine at position 141.

To test the effect of the G141R mutation in solution, we expressed the 10xHis-8xArg-ScSUMO-3C tagged HsBBS9₁₋₄₀₇ G141R point mutant in *E. coli* and compared the amount of purified protein to the level of wild type, 10xHis-8xArg-ScSUMO-3C tagged HsBBS9₁₋₄₀₇ using a nickel affinity pulldown assay. Compared to wild type, the yield of HsBBS9₁₋₄₀₇ G141R after Ni-affinity purification is marginal (Fig. 6A). This reduction in yield is due to a concomitant decrease in overall expression and solubility. This result is clearer after removal of the 10xHis-8xArg-ScSUMO-tag with 3C protease. Tagged HsBBS9₁₋₄₀₇ runs just under the 66 kDa marker in the same position as the *E. coli* GroEL chaperone, as determined by LC-MS/MS (data not shown), whereas untagged HsBBS9₁₋₄₀₇ runs unobstructed beneath the 3C protease band near the 43 kDa marker. This difference in protein level is reproducibly seen and is not due to clonal variation, as the same result was obtained using five different clones of HsBBS9₁₋₄₀₇ G141R producing *E. coli* (data not shown). To confirm that the observed difference is not an artifact of

bacterial expression, we transiently transfected FLAG tagged HsBBS9₁₋₄₀₇ and HsBBS9₁₋₄₀₇ G141R into the HEK 293T human cell line. Anti-FLAG western blotting of the soluble fraction reveals that HsBBS9₁₋₄₀₇ G141R is barely visible, whereas the wild type protein is detected easily (Fig. 6B).

A prediction of our hypothesis, that a glycine is essential at position 141, is that any mutation regardless of identity will be equally deleterious. To test this prediction, we created a series of point mutants and tested their protein levels in *E. coli* again by analyzing the amount of protein isolated in a nickel pulldown on SDS-PAGE, after 3C cleavage of the affinity tag (Fig. 6C). Consistent with this prediction, mutating Gly141 to alanine results in the same virtual absence of the protein as observed with the G141R mutant. Interestingly, mutating the neighboring residue, Ser142, to a glycine fails to rescue protein levels in the context of both the arginine and alanine mutant at position 141. Additionally, an attempt to remove the predicted side chain steric clash of Arg141 with Tyr186 by mutating the latter to alanine (Y186A) also failed to rescue purifiable protein amounts. In summary, the mutational analysis and the expression tests in two different hosts strongly suggest that the G141R mutation promotes a folding defect, effectively rendering HsBBS9 non-functional.

The HsBBS9₁₋₄₀₇ β -propeller forms a homodimer in the crystal lattice

We analyzed packing contacts within our crystal to see if potentially biologically relevant interactions are recapitulated within the lattice. The metric for relevance here is the predicted size or strength of a particular interaction. One such large contact occurs between two non-crystallographic symmetry (NCS) mates where the β -propellers bind in a side-on fashion (Fig. 7A). The interface is symmetrical and primarily characterized by two domain swaps. In the first, each NCS-symmetry mate contributes a fifth β -strand to blade β 7 of the other HsBBS9 copy, forming a parallel β -sheet. This contributed strand emanates from the loop region between the

C and D strands of blade $\beta 1$ (Fig. 7A; inset). In the second domain swap, the C-terminal loop inserts into a surface pocket, between blades $\beta 6$ and $\beta 7$, on each NCS-symmetry mate. In addition, the N-terminal loop packs against the loop connecting strands A and B of blade $\beta 7$. In total, this dimeric interface buries a surface area of $1,731 \text{ \AA}^2$ and comprises 24 hydrogen bonds and two salt bridges for a predicted free energy change of -23.8 kcal/mol upon binding (Krissinel and Henrick, 2007).

To validate this interaction, we examined sequence conservation on the HsBBS9₁₋₄₀₇ structure. This analysis reveals that the interface of the NCS dimer is moderately conserved, which makes it difficult to rule out this interaction as a crystallization artifact (Fig. 7B). By examining the MSA in detail, it becomes clear that two interface features are poorly conserved (Fig. 8). The first is the Asp8-His339 salt bridge, which can only be formed in a subset of species. Second, the β -strand domain swap is not conserved. The loop that contributes the fifth β -strand to blade $\beta 7$ (residues 53-68 in *Homo sapiens*) is of variable length in the MSA. Specifically, 11 out of 18 species in the alignment have a 4 or 5 amino acid deletion in this loop, likely making it too short to contribute this strand to the neighboring BBS9 β -propeller. Removing this interaction, however, still leaves the majority of Van der Waals contacts and 14 hydrogen bonds within the homodimer interface intact.

Biophysical characterization of HsBBS9 in solution

To determine if HsBBS9₁₋₄₀₇ dimerizes in solution, we performed size exclusion chromatography (SEC) with purified HsBBS9₁₋₄₀₇, yielding a single peak at the approximate size of a monomer. Increasing the concentration of HsBBS9₁₋₄₀₇ had no effect on the elution volume, inconsistent with dimer formation (Fig. 9A). Due to the limited resolution of SEC, we set out to validate this result with equilibrium analytical ultracentrifugation (AUC). We tested a single protein concentration at three different speeds. The data for all three speeds fit globally to an

ideal, non-interacting, 1-component system with a molecular weight of 44,910 Da, indicative of a monomeric species (calculated MW - 45862 Da, (Wilkins et al., 1999), Fig. 9B). These results demonstrate that the HsBBS9₁₋₄₀₇ homodimer does not form in solution.

To determine if other regions of HsBBS9 oligomerize, we performed SEC on the C-terminal half of the protein, HsBBS9₄₀₅₋₈₈₇, which contains three folded domains according to structure prediction and can be recombinantly expressed and purified from *E. coli*. Unlike HsBBS9₁₋₄₀₇, the elution volume of HsBBS9₄₀₅₋₈₈₇ decreases upon increasing protein concentration (Fig. 9C). The observed decrease in elution volume and broadening of the peak is consistent with dimer formation. To verify this result, we performed equilibrium AUC on HsBBS9₄₀₅₋₈₈₇ at a single concentration, in duplicate, spun at three different speeds. The data fit well to a monomer-dimer equilibrium model, with a fitted molecular weight of 55,810 Da (monomer, calculated MW - 54,419 Da, ref. (Wilkins et al., 1999)) and a dissociation constant of 33 μ M (Fig. 9D). This data demonstrates that HsBBS9₄₀₅₋₈₈₇ dimerizes in solution.

We went on to assess whether the N-terminal domain and C-terminal half of HsBBS9 bind to each other. HsBBS9₁₋₄₀₇ and HsBBS9₄₀₅₋₈₈₇ were mixed in a 1:1 molar ratio prior to SEC in conditions of increasing ionic strength (100 mM NaCl, 10% v/v glycerol; 150 mM NaCl; 500 mM NaCl). Under no condition was binding observed (Fig. 10A, B). This indicates that the HsBBS9 N-terminal β -propeller is flexibly tethered to the C-terminal half of the protein. Based on this series of experiments, we can conclude that HsBBS9₁₋₄₀₇ dimerization is a crystallization artifact.

Discussion

Here we analyzed the structural and biophysical properties of BBS9, a component of the BBSome. We show that the N-terminal domain folds into a regular β -propeller. How does this information help in deducing the function of BBS9? The protein acts in context of the octameric BBSome, which is a putative membrane coating complex in the primary cilium. The cell has several such membrane coating systems, like COPI, COPII, clathrin and, more distantly related, the nuclear pore complex (Brohawn et al., 2009). The architecture of these assemblies is principally related, presumably because of a common evolutionary past (Devos et al., 2004), (Mans et al., 2004). Many of the architectural components within these coats are β -propellers, α -helical stacks, or a combination of both. However, the ways in which these elements come together to form the various coats seem to be quite diverse (Fath et al., 2007; Fotin et al., 2004; Lee and Goldberg, 2010). Regarding BBS9, our data indicate that it is possibly involved in the higher-order assembly of a BBSome coat. We observed that the protein has a dimerization interface in its C-terminal half. In the context of building a BBSome coat weak interactions, such as this, might work in an additive manner to build the polymerized assembly. Weak interactions could have evolved such that only small perturbations are required to shift the equilibrium between net polymerization and depolymerization.

In this context, the conserved negative patch on the BBS9 β -propeller domain is particularly intriguing. It is suggestive to propose that it forms a platform for a charge interaction with a peptide or protein of positive character. Other seven-bladed β -propellers are known to bind to positively charged peptides via their negatively charged pores. Examples include β' COP, a COPI vesicle coat component, that binds ER retrieval motifs on cargo proteins (Ma and Goldberg, 2013) and the nucleosome-remodeling factor Nurf55, which binds to the positively charged N-terminus of histone H3 via a negatively charged pore (Schmitges et al., 2011). Such

an electrostatic interaction could be blocked easily by phosphorylation of the binding partner or selective interaction with phospholipids.

Our data indicate that the pathogenic G141R mutation in HsBBS9 likely causes misfolding of the β -propeller domain, which presumably abrogates any potential protein-protein interaction. Whether this interaction is within the BBSome octamer or with an external factor still needs to be understood.

Methods

Cloning

The *Homo sapiens* BBS9 open reading frame was PCR amplified with BamHI (5') and NotI (3') restriction site overhangs from cDNA (open biosystems) and was subcloned into a modified pETDuet-1 vector, containing an N-terminal 10xHis-8xArg-ScSUMO-3C cleavage site tag, via restriction digest. Subsequent inverse PCRs were performed on this vector to remove the coding region for either the N-terminal domain (HsBBS9₁₋₄₀₇) or the C-terminal half of the protein (HsBBS9₄₀₅₋₈₈₇). The resulting linearized vectors were recircularized by enzymatic treatment with T4 phosphonucleotide kinase (NEB) and T4 ligase (NEB). HsBBS9₁₋₄₀₇ G141R and all other point mutations were generated via site-directed mutagenesis.

Protein expression and purification

LOBSTR(DE3)-RIL *E.coli* (ref. (Andersen et al., 2013)) cells transformed with the 10xHis-8xArg-ScSUMO-3C-HsBBS9₁₋₄₀₇ pETDuet-1 vector were grown in LB medium at 37 °C to 0.65 OD. Cultures were shifted to 18 °C for 30 minutes prior to induction with 0.2 mM IPTG. Expression was carried out for 16 hours, after which the cultures were pelleted by centrifugation at 6000 x g for 6 minutes. Cell pellets were resuspended in lysis buffer (50 mM potassium phosphate pH 8.0, 500 mM NaCl, 30 mM imidazole pH 8.0, and 3 mM β -mercaptoethanol) at 4

°C and were lysed by passage through a cell disruptor (constant systems) at 25 kpsi. After lysis, 1 mM PMSF and 250 units of turbonuclease (Eton biosciences) were added to the lysate, which was subsequently centrifuged at 9500 x g for 25 minutes to separate soluble and insoluble fractions. Nickel sepharose 6 fast flow beads (GE healthcare) were washed twice with lysis buffer prior to addition to the soluble fraction. Nickel binding was carried out at 4 °C for 1 hour. After binding, the beads were pelleted, at 2000 x g, and washed three times with lysis buffer. Six column volumes of elution buffer (250 mM imidazole pH 8.0, 150 mM NaCl, and 3 mM β -mercaptoethanol) were used to elute the 10xHis-8xArg-ScSUMO-3C-HsBBS9₁₋₄₀₇. Following elution, the protein was diluted in half with 10 mM potassium phosphate pH 8.0, 1 mM DTT, and 0.1 mM EDTA (buffer A) and loaded onto a SP Sepharose Fast Flow cation exchange column (1 mL or 5 mL depending on prep size, GE healthcare). Following a two column volume wash with buffer A, bound protein was eluted with a 15 column volume gradient of 10 mM potassium phosphate pH 8.0, 1 mM DTT, 0.1 mM EDTA, and 1 M NaCl (buffer B). The peak corresponding to bound 10xHis-8xArg-ScSUMO-3C-HsBBS9₁₋₄₀₇ was pooled, following verification by SDS-PAGE. The N-terminal tag was cut with human rhinovirus 3C (3C) protease, added in a 1:100 ratio, and dialyzed overnight into 10 mM potassium phosphate pH 8.0, 150 mM NaCl, 1 mM DTT, and 0.1 mM EDTA. The dialyzed sample was run over a second SP Sepharose Fast Flow column and the flow through fraction was collected. After the flow through was concentrated, the sample was run over an Superdex S200 gel filtration column equilibrated in 10 mM Tris/HCl pH 7.4, 150 mM NaCl, 1 mM DTT, and 0.1 mM EDTA (GF buffer 1). The peak corresponding to HsBBS9₁₋₄₀₇ was pooled and concentrated after SDS-PAGE analysis. Full length HsBBS9₁₋₈₈₇ was identically expressed and purified. HsBBS9₄₀₅₋₈₈₇ was identically expressed and purified, with the exception that gel filtration was performed in 10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP, and 0.1 mM EDTA (GF buffer 2). Selenomethionine-derivatized (SeMet) HsBBS9₁₋₄₀₇ was expressed as previously described (Brohawn et al., 2008) and purified identically to native HsBBS9₁₋₄₀₇, with the exception that 10 mM β -mercaptoethanol was used in the nickel

purification and 5 mM DTT in subsequent steps. The HsBBS9₁₋₄₀₇ point mutant purifications were performed identically to the wild type HsBBS9₁₋₄₀₇ nickel purification.

Limited proteolysis

22 μ L of 0.3 mg/mL HsBBS9₁₋₈₈₇ was digested with 2 μ L of trypsin, at one of six different concentrations, in GF buffer 1. Trypsin was used at 0.1 mg/mL or diluted 1:3, 1:9, 1:27, 1:81, or 1:243 with GF buffer 1 from the 0.1 mg/mL stock. After 30-minute incubation at room temperature, the reactions were quenched with 6 μ L of SDS-PAGE loading dye and boiled for 3 minutes. SDS-PAGE was performed on these samples and a band of ~40 kDa was produced upon treatment with 1:243 diluted trypsin that was stable through treatment with 1:9 diluted trypsin. This band was excised, placed in 50% methanol, and subjected to LC-MS/MS.

Crystallization

Initial crystals of native HsBBS9₁₋₄₀₇, set up at 7.2 mg/mL, were obtained via vapor diffusion with 10% w/v PEG 4000, 0.1 M MES pH 6.5, and 0.2 M MgCl₂ as part of the PEGII suite (Qiagen) in a 96-well, sitting drop tray. Single, albeit multi-layered, three-dimensional crystals were observed after 2 days at 18 °C. Crystals were transferred into a cryoprotectant solution containing the crystallization condition with 20% v/v PEG 200 and cryocooled in liquid nitrogen. SeMet-derivatized HsBBS9₁₋₄₀₇ crystals, set up at 6.8 mg/mL, were obtained in 10% w/v PEG 4000, 0.1 M tri-sodium citrate pH 6.1, and 0.15 M ammonium sulfate in 24 well, hanging drop plates at 18 °C. Compared to the native crystals, the SeMet crystals were optically superior (i.e. they appeared singular and unlayered) and grew larger to a final size of 0.1 mm. The SeMet crystals were transferred into a cryoprotectant solution containing the crystallization condition with 20% v/v glycerol and cryocooled in liquid nitrogen.

Structure determination

Datasets were collected at the Advanced Photon Source user end station 24. A complete native dataset was collected to 2.4 Å and a complete SeMet dataset, at the selenium peak wavelength, was collected to 1.8 Å. Data reduction was performed in *HKL2000* (Otwinowski and Minor, 1997). Molecular replacement attempts were performed in *phaser-MR*, part of the PHENIX suite (Adams et al., 2010). The phase problem was solved using single anomalous dispersion (SAD) and selenium positions were determined in *HYSS*, run as part of the *phenix AutoSol* program, for the SeMet dataset. The anomalous data extended to 2.5 Å, as assessed by *phenix xtriage*. The initial model was built using *phenix AutoBuild* with the structure factors from the SeMet dataset because it was superior to the native data. Subsequent manual building was performed in *Coot* (Emsley et al., 2010) and *phenix.refine* was used for refinement.

Structure analysis

The HsBBS9₁₋₄₀₇ non-crystallographic symmetry (NCS) homodimer interface was analyzed using PDBePISA (Krissinel and Henrick, 2007). Sequence conservation was mapped onto the surface of HsBBS9₁₋₄₀₇ using ConSurf (Ashkenazy et al., 2010) and a multiple sequence alignment (MSA) generated in Jalview (Waterhouse et al., 2009) with the Muscle alignment algorithm. Sequences of HsBBS9₁₋₄₀₇ homologs were obtained via NCBI BLAST searches (Altschul et al., 1990). The structure of HsBBS9 homologs was modeled using Phyre 2 one-to-one threading (Kelley and Sternberg, 2009). Electrostatic potential was mapped onto the surface of HsBBS9₁₋₄₀₇ and the modeled homologs using the APBS plugin (Baker et al., 2001) in PyMol (DeLano, 2010). The G141R mutation was modeled into the wild type HsBBS9₁₋₄₀₇ structure using the mutagenesis tool in *Coot*.

Expression test of HsBBS9₁₋₄₀₇ constructs in human cell culture

2x10⁶ HEK 293T cells were grown in DMEM+10% inactivated fetal bovine serum (IFS) in 10 cm dishes. 24 hours later, cells were transfected with 1 µg DNA of either wild type HsBBS9₁₋₄₀₇ or HsBBS9₁₋₄₀₇ G141R cloned into a pRK5-FLAG expression plasmid, using polyethylenimine (PEI). 48 hours post-transfection cells were harvested in triton lysis buffer (40 mM Hepes pH 7.4, 1% Triton X-100, 10 mM sodium pyrophosphate, 10 mM β-glycerophosphate, 2.5 mM MgCl₂, and 1 complete protease inhibitor tablet (Roche). After a 20-minute incubation the soluble and insoluble fractions were separated by a 10-minute centrifugation step at 21,000 x g. The soluble fractions were loaded onto an SDS-PAGE gel followed by western blotting with anti-FLAG (Cell Signaling Technologies) and anti-Raptor (for loading control) antibodies.

Analytic 3C proteolysis of 10xHis-8xArg-ScSUMO-3C-HsBBS9₁₋₄₀₇ point mutants

2 µL of 1 mg/mL 3C protease was added to 48 µL of nickel eluate and the reaction was allowed to proceed for 30 minutes at 37 °C. The reaction was quenched by the addition of 5x SDS-loading dye followed by boiling for 5 minutes. 10 µL of boiled sample was loaded onto an SDS-PAGE gel for analysis.

Size exclusion chromatography (SEC) experiments

The HsBBS9₁₋₄₀₇ concentration-dependent shift experiment was performed in GF buffer 2. 100 µL of HsBBS9₁₋₄₀₇ was loaded onto a Superdex S75 10/300 size exclusion column at 10 mg/mL, 5 mg/mL, 2.5 mg/mL, or 1 mg/mL. The runs were performed in direct succession and dilutions were performed immediately prior to the size exclusion run. HsBBS9₄₀₅₋₈₈₇ concentration dependent shift experiment was identically performed, except a Superdex S200 10/300 column was used. The initial HsBBS9₁₋₄₀₇, HsBBS9₄₀₅₋₈₈₇ binding test was performed in GF buffer 1. HsBBS9₁₋₄₀₇ and HsBBS9₄₀₅₋₈₈₇ were mixed in an equimolar ratio and, after incubation at 4 °C, were loaded onto a Superdex S200 10/300 column equilibrated in GF buffer

1. For the low salt binding test, HsBBS9₁₋₄₀₇ and HsBBS9₄₀₅₋₈₈₇ were mixed in an equimolar ratio in 10 mM Tris/HCl pH 7.4, 100 mM NaCl, 10% glycerol, 1 mM DTT, and 0.1 mM EDTA prior to loading onto an S200 10/300 column equilibrated in the same buffer. For the high salt binding test, equimolar amounts of HsBBS9₁₋₄₀₇ and HsBBS9₄₀₅₋₈₈₇ were mixed in 10 mM Tris/HCl pH 7.4, 500 mM NaCl, 1 mM DTT, and 0.1 mM EDTA prior to loading onto an appropriately equilibrated Superdex S200 10/300 column. The data was plotted in Prism (GraphPad Software).

Equilibrium analytical ultracentrifugation

Equilibrium analytical ultracentrifugation (AUC) on HsBBS9₁₋₄₀₇ and HsBBS9₄₀₅₋₈₈₇ was performed, in GF buffer 2, in an Optima XL-1 analytical ultracentrifuge (Beckman Coulter) using an AN-50 titanium rotor (Beckman Coulter) with a 6-channel, epon rectangular centerpiece (1.2 cm, Beckman Coulter). $A_{280\text{nm}}$ measurements were taken once the samples reached equilibrium and buffer contribution to the absorbance was removed using a reference cell, which contained GF buffer 2. HsBBS9₁₋₄₀₇ at 0.37 mg/mL, was spun to 9,500 rpm, 12,000 rpm, and 18,000 rpm. Five absorbance scans were performed at each speed, with 3 replicates per scan. HsBBS9₄₀₅₋₈₈₇ at 1.2 mg/mL, in duplicate, was spun to 9,500 rpm, 14,000 rpm, and 17,000 rpm. Three absorbance scans were performed at each speed, with 3 replicates per scan. The equilibrium absorbance data was analyzed using Ultrascan II (Demeler, 2005) and fit using nonlinear least squares fitting.

PDB accession code

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 4YD8.

Acknowledgements

We thank R. Saxton for performing the HEK 293T transient transfection and subsequent western blot, N. Leksa for experimental advice, S. Bilokapic for help with crystal fishing and E. Spooner for performing the LC-MS/MS. The X-ray crystallography work was conducted at the APS NE-CAT beamlines, which are supported by award GM103403 from the National Institute of General Medical Sciences, NIH. Use of the APS is supported by the US Department of Energy, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. The Biophysical Instrumentation Facility for the Study of Complex Macromolecular Systems (NSF-007031) is gratefully acknowledged.

References

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 213–221.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* *215*, 403–410.
- Andersen, K.R., Leksa, N.C., and Schwartz, T.U. (2013). Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* *81*, 1857–1861.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research* *38*, W529–W533.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* *98*, 10037–10041.
- Berbari, N.F., Lewis, J.S., Bishop, G.A., Askwith, C.C., and Mykytyn, K. (2008). Bardet-Biedl syndrome proteins are required for the localization of G protein-coupled receptors to primary cilia. *Proc. Natl. Acad. Sci. U.S.A.* *105*, 4242–4246.
- Brohawn, S.G., Leksa, N.C., Spear, E.D., Rajashankar, K.R., and Schwartz, T.U. (2008). Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* *322*, 1369–1373.

Brohawn, S.G., Partridge, J.R., Whittle, J.R.R., and Schwartz, T.U. (2009). The Nuclear Pore Complex Has Entered the Atomic Age. *Structure* 17, 1156–1168.

Brown, J.T., Bai, X., and Johnson, A.W. (2000). The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. *Rna* 6, 449–457.

Chih B., Liu, P., Chinn, Y., Chalouni, C., Komuves, L.G., Hass, P.E., Sandoval, W., and Peterson, A.S. (2011). A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain. *Nature Cell Biology* 14, 61–72.

Ciccarelli, F.D., Doerks, T., Mering, von, C., Creevey, C.J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287.

Corbit, K.C., Aanstad, P., Singla, V., Norman, A.R., Stainier, D.Y.R., and Reiter, J.F. (2005). Vertebrate Smoothed functions at the primary cilium. *Nature Cell Biology* 437, 1018–1021.

DeLano, W.L. (2010). The PyMOL Molecular Graphics System, Version 1.3r1. Schrodinger, LLC, New York.

Demeler, B. (2005). UltraScan: a comprehensive data analysis software package for analytical ultracentrifugation experiments. *Modern Analytical Ultracentrifugation: Techniques and Methods* 210–229.

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B.T., Sali, A., and Rout, M.P. (2004). Components of Coated Vesicles and Nuclear Pore Complexes Share a Common Molecular Architecture. *Plos Biol* 2, e380.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.

Fath, S., Mancias, J.D., Bi, X., and Goldberg, J. (2007). Structure and Organization of Coat Proteins in the COPII Cage. *Cell* 129, 1325–1336.

Fong, H.K., Hurley, J.B., Hopkins, R.S., Miake-Lye, R., Johnson, M.S., Doolittle, R.F., and Simon, M.I. (1986). Repetitive segmental structure of the transducin beta subunit: homology with the CDC4 gene and identification of related mRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 83, 2162–2166.

Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S.C., Kirchhausen, T., and Walz, T. (2004). Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* 432, 573–579.

Fritz-Laylin, L.K., Prochnik, S.E., Ginger, M.L., Dacks, J.B., Carpenter, M.L., Field, M.C., Kuo, A., Paredes, A., Chapman, J., Pham, J., et al. (2010). The Genome of *Naegleria gruberi* Illuminates Early Eukaryotic Versatility. *Cell* 140, 631–642.

Heenan, E.J., Vanhooke, J.L., Temple, B.R., Betts, L., Sondek, J.E., and Dohlman, H.G. (2009). Structure and Function of Vps15 in the Endosomal G Protein Signaling Pathway. *Biochemistry* 48, 6390–6401.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research* 38, W545–W549.

Hu, Q., Milenkovic, L., Jin, H., Scott, M.P., Nachury, M.V., Spiliotis, E.T., and Nelson, W.J. (2010). A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. *Science* 329, 436–439.

Jin, H., White, S.R., Shida, T., Schulz, S., Aguiar, M., Gygi, S.P., Bazan, J.F., and Nachury, M.V. (2010). The Conserved Bardet-Biedl Syndrome Proteins Assemble a Coat that Traffics Membrane Proteins to Cilia. *Cell* 141, 1208–1219.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4, 363–371.

Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* 372, 774–797.

Lee, C., and Goldberg, J. (2010). Structure of Coatomer Cage Proteins and the Relationship among COPI, COPII, and Clathrin Vesicle Coats. *Cell* 142, 123–132.

Loktev, A.V., Zhang, Q., Beck, J.S., Searby, C.C., Scheetz, T.E., Bazan, J.F., Slusarski, D.C., Sheffield, V.C., Jackson, P.K., and Nachury, M.V. (2008). A BBSome Subunit Links Ciliogenesis, Microtubule Stability, and Acetylation. *Developmental Cell* 15, 854–865.

Ma, W., and Goldberg, J. (2013). Rules for the recognition of dilysine retrieval motifs by coatomer. *The EMBO Journal* 32, 926–937.

Mans, B.J., Anantharaman, V., Aravind, L., and Koonin, E.V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3, 1612–1637.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. *eLife* 2, e01456.

Mourão, A., Nager, A.R., Nachury, M.V., and Lorentzen, E. (2014). Structural basis for membrane targeting of the BBSome by ARL6. *Nat. Struct. Mol. Biol.* 21, 1035–1041.

Nachury, M.V., Loktev, A.V., Zhang, Q., Westlake, C.J., Peränen, J., Merdes, A., Slusarski, D.C., Scheller, R.H., Bazan, J.F., Sheffield, V.C., et al. (2007). A Core Complex of BBS Proteins Cooperates with the GTPase Rab8 to Promote Ciliary Membrane Biogenesis. *Cell* 129, 1201–1213.

Nachury, M.V., Seeley, E.S., and Jin, H. (2010). Trafficking to the Ciliary Membrane: How to Get Across the Periciliary Diffusion Barrier? *Annu. Rev. Cell Dev. Biol.* 26, 59–87.

Nishimura, D.Y., Swiderski, R.E., Searby, C.C., Berg, E.M., Ferguson, A.L., Hennekam, R., Merin, S., Weleber, R.G., Biesecker, L.G., Stone, E.M., et al. (2005). Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. *Am. J. Hum. Genet.* 77, 1021–1033.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology*, C.J. Carter, and R.M. Sweet, eds. (New York: Elsevier), pp. 307–326.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell* 42, 330–341.

Singla, V., and Reiter, J.F. (2006). The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science* 313, 629–633.

Tobin, J.L., and Beales, P.L. (2009). The nonmotile ciliopathies. *Genet Med* 11, 386–402.

Veleri, S., Bishop, K., Dalle Nogare, D.E., English, M.A., Foskett, T.J., Chitnis, A., Sood, R., Liu, P., and Swaroop, A. (2012). Knockdown of Bardet-Biedl Syndrome Gene BBS9/PTHB1 Leads to Cilia Defects. *PLoS ONE* 7, e34389.

Wall, M.A., Coleman, D.E., Lee, E., Iñiguez-Lluhi, J.A., Posner, B.A., Gilman, A.G., and Sprang, S.R. (1995). The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* 83, 1047–1058.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.

Wei, Q., Zhang, Y., Li, Y., Zhang, Q., Ling, K., and Hu, J. (2012). The BBSome controls IFT assembly and turnaround in cilia. *Nature Cell Biology* 14, 950–957.

Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552.

Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. *Protein Cell* 2, 202–214.

Tables

Table 1. Data collection and refinement statistics

	BBS9 ₁₋₄₀₇ SeMet derivative	BBS9 ₁₋₄₀₇ native
Data collection		
Space group	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	136.42, 81.88, 85.65	136.14, 81.62, 85.49
<i>a</i> , <i>b</i> , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	44.70 - 1.80 (1.83 - 1.80) ^a	50.00 - 2.40 (2.44 - 2.40)
<i>R</i> _{sym}	0.09 (0.99)	0.15 (0.57)
<i>R</i> _{<i>p.i.m.</i>}	0.05 (0.60)	0.09 (0.40)
<i>I</i> / σ <i>I</i>	23.3 (1.1)	8.8 (1.2)
CC _{1/2}	1.00 (0.58)	0.99 (0.67)
Completeness (%)	99.9 (99.9)	95.2 (78.0)
Redundancy	3.7 (3.7)	3.3 (2.7)
Refinement		
Resolution (Å)	36.94 - 1.80	
No. reflections	171157	
No. free reflections	3941	
<i>R</i> _{work} / <i>R</i> _{free}	16.9/20.0	
No. atoms	6141	
Protein	5424	
Water	717	
<i>B</i> factors		
Protein	35.4	
Water	48.4	
r.m.s. deviations		
Bond lengths (Å)	0.008	
Bond angles (°)	1.119	
Ramachandran		
Favored (%)	97.3	
Allowed (%)	2.7	
Outlier (%)	0	
Clashscore	5.6	
Molprobrity percentile	93 rd	

^aValues in parentheses are for highest-resolution shell. One crystal was used for each dataset.

Figures

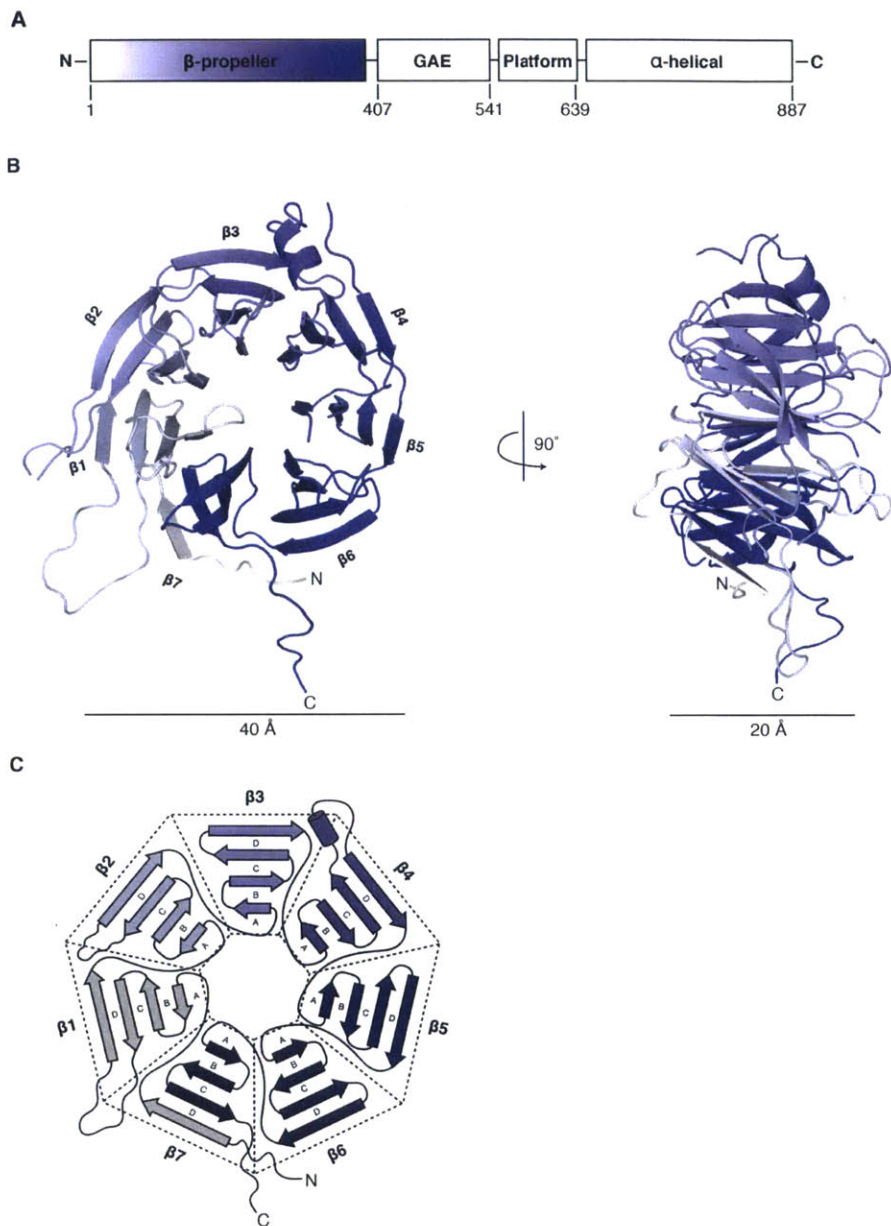


Figure 1. The HsBBS9₁₋₄₀₇ N-terminal domain is a WD40 β -propeller. (A) Schematic diagram of the domain architecture of HsBBS9 with the relevant domain boundaries numbered. The N-terminal β -propeller domain, solved here, is colored white to blue from N to C terminus. Following the β -propeller is a γ -adaptin ear (GAE) β -sandwich domain, a mixed α/β platform domain, and a C-terminal α -helical domain (Jin et al., 2010). (B) Cartoon representation of the HsBBS9₁₋₄₀₇ β -propeller viewed from the face and, as a result of a 90° rotation, the side. The structure is colored white to blue from N to C terminus as in (A). β -strands are depicted as arrows, helices as ribbons, and loops as lines. The individual blades are labeled. (C) Detailed schematic of HsBBS9 tertiary structure, colored as in (A).

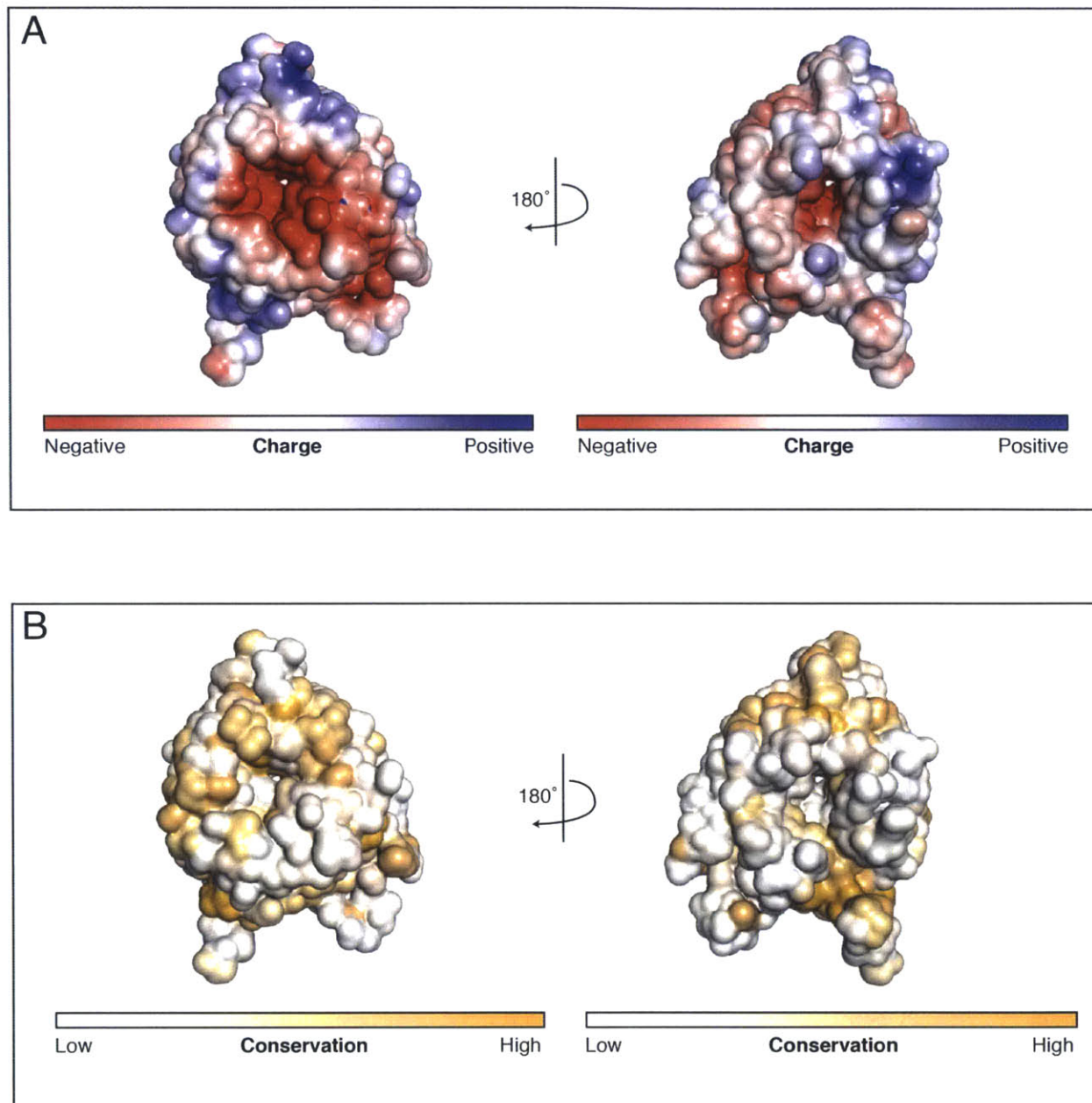


Figure 2. A negative patch exists around the pore on one face of the HsBBS9₁₋₄₀₇ β -propeller. (A) Electrostatic potential mapped onto the solvent-accessible surface of HsBBS9₁₋₄₀₇. Potential is colored from -5 kT/e (negative, red) to +5 kT/e (positive, blue). Both faces, rotated by 180° with respect to each other, are shown. (B) Sequence conservation mapped onto the solvent-accessible surface of HsBBS9₁₋₄₀₇. Conservation is colored from white (not conserved) to orange (conserved) and views are as in (A).

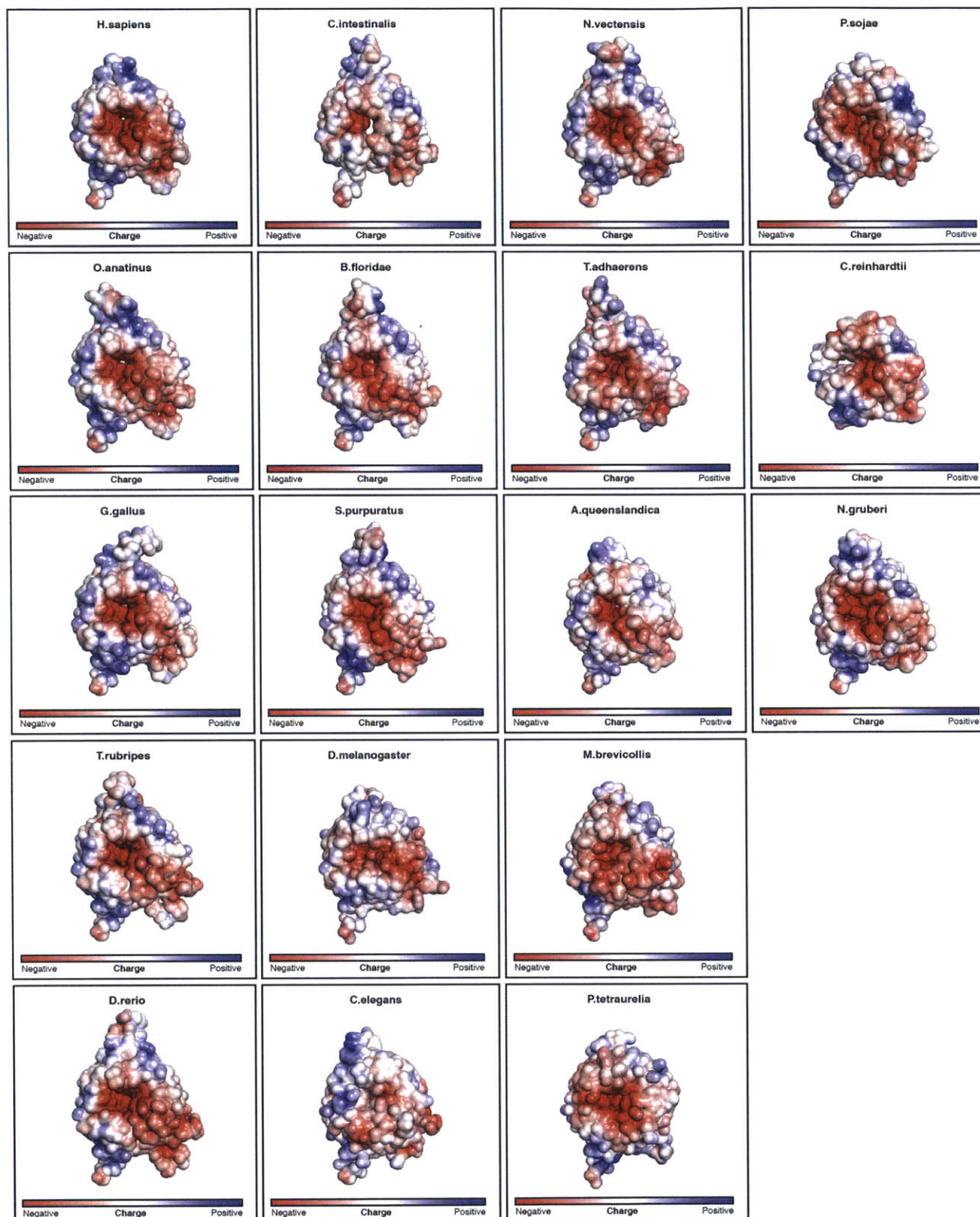


Figure 3. Electrostatic potential mapped onto the solvent-accessible surface of HsBBS9₁₋₄₀₇ and the modeled structures of the 17 homologs from the MSA, depicted in Fig. 8. Views and coloring are identical to Fig. 2A (left), with the electrostatic potential calculated from -5 kT/e to +5 kT/e.

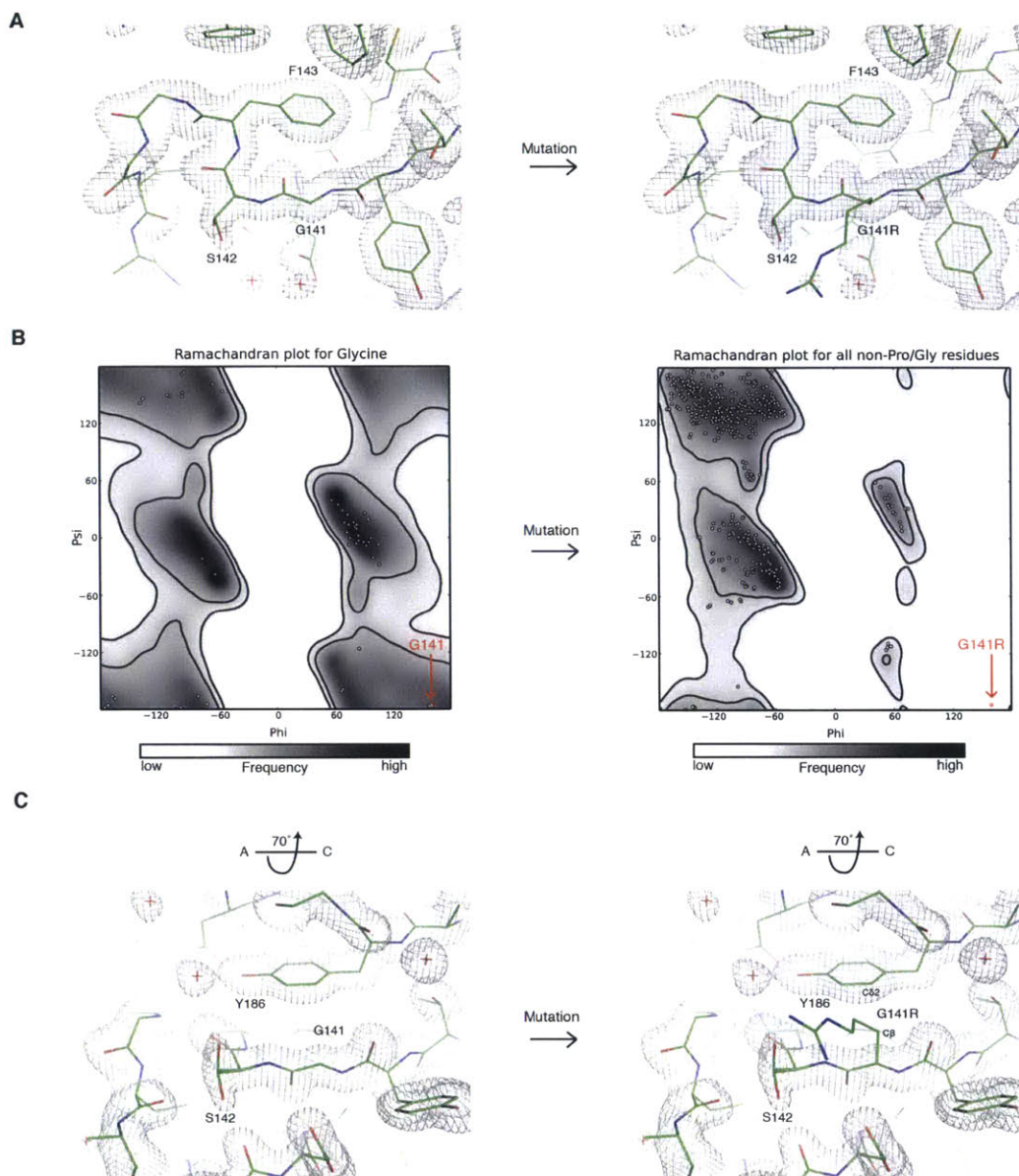


Figure 5. Detailed view of the position of glycine 141, which reveals steric restraints on the identity of this residue. (A) View of main chain topology up- and downstream of glycine 141. The protein chain (green sticks) is shown for glycine 141 (left) or the G141R mutation (right) modeled as the most sterically permissive rotamer (fewest clashes). The final 2Fo-Fc electron density map is represented as a grey mesh. (B) Ramachandran plots for glycine (left) or non-proline/glycine residues such as arginine (right). White dots show the distribution of phi and psi angles for residues in the HsBBS9₁₋₄₀₇ structure. The position of glycine 141 and the G141R point mutant is shown (red dot). Favored regions (inner black line) and allowed regions (outer black line) are shown and the plot is colored from white to black according to frequency of occurrence in the Top500 database of high-resolution protein structures. These plots were generated in PHENIX (Adams et al., 2010). (C) View of the G141R-Y186 steric clash, rotated by 70° around the x axis relative to (A). The arginine side chain is pointing out of the page. The atoms in close contact are labeled.

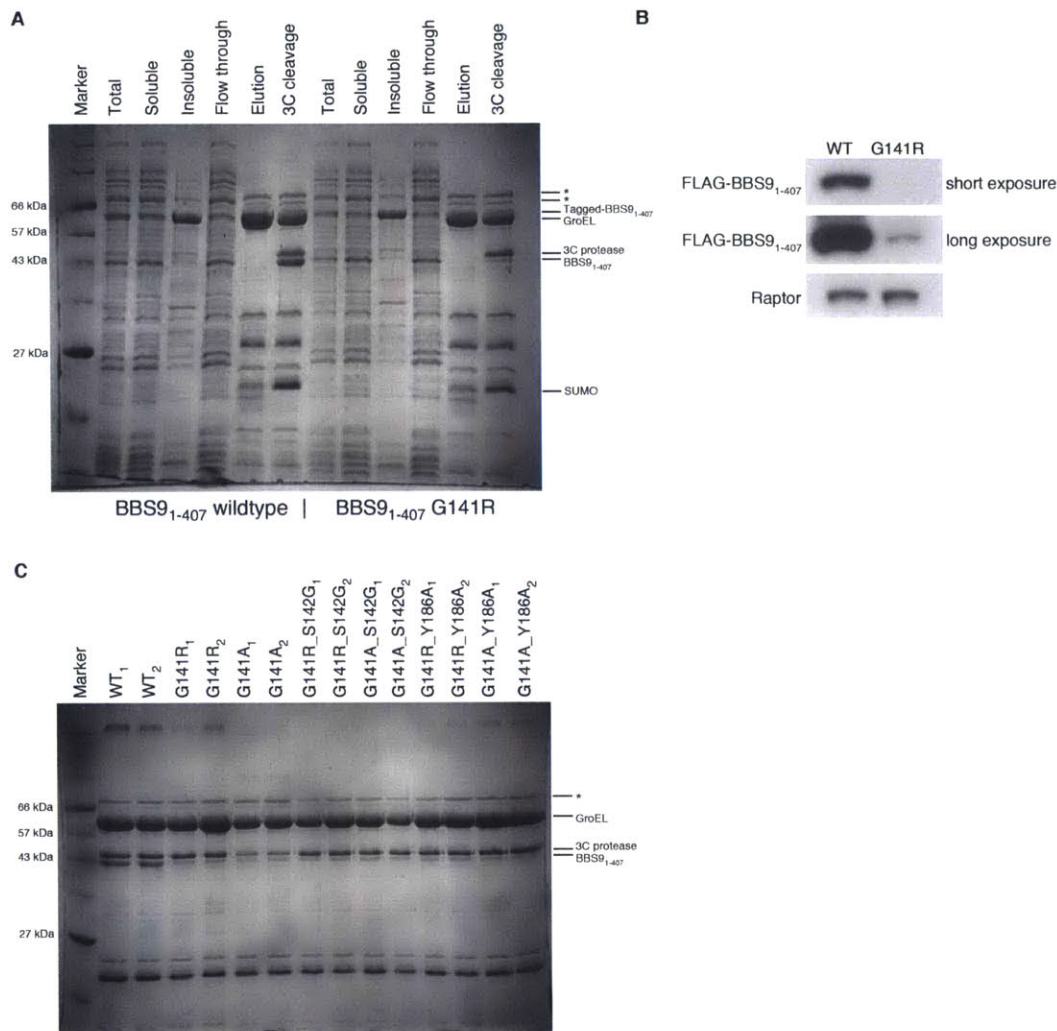
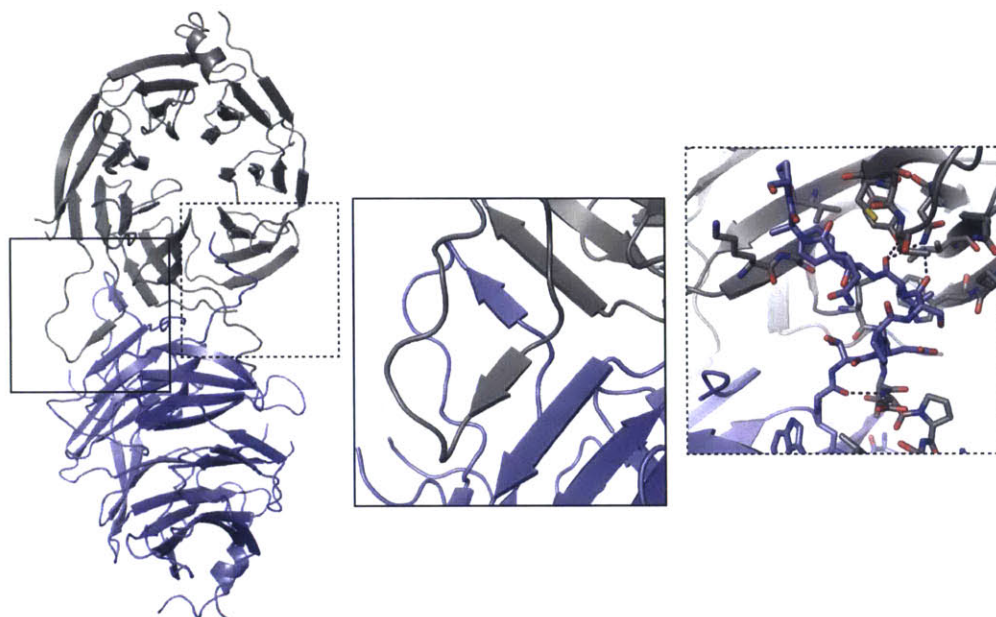


Figure 6. Soluble protein levels of HsBBS9₁₋₄₀₇ G141R are substantially reduced compared to wild type. (A) Coomassie-stained SDS-PAGE gel of nickel affinity pulldown experiment, comparing purified protein levels of wild type (WT) HsBBS9₁₋₄₀₇ (left) and HsBBS9₁₋₄₀₇ G141R (right) expressed in *E. coli*. Lane labels refer to samples taken throughout the nickel affinity purification: Total – crude lysate, Soluble – supernatant after lysate clarification, Insoluble – pellet after lysate clarification, Flow through – soluble sample that did not bind to nickel, Elution – soluble sample that bound nickel and was eluted with imidazole, 3C cleavage – eluate treated with 3C protease for 30 minutes at 37 °C for tag removal. In the elution fractions, the largest two bands (labeled with *) are *E. coli* contaminants. Other bands are labeled accordingly. (B) Anti-FLAG western blot of soluble fraction from expression test of FLAG-tagged, wild type HsBBS9₁₋₄₀₇ (left) and FLAG-tagged HsBBS9₁₋₄₀₇ G141R (right) in HEK 293T cells. Two exposure times are shown, for clarity. Anti-Raptor antibody is used for the Raptor loading control. (C) Coomassie-stained SDS-PAGE gel from nickel affinity pulldown of HsBBS9₁₋₄₀₇ point mutants in *E. coli*. Shown are the 3C-treated, nickel elution fractions for two different clones of wild type (WT), disease mutant (G141R), and designed point mutants (labeled accordingly, above gel).

A



B

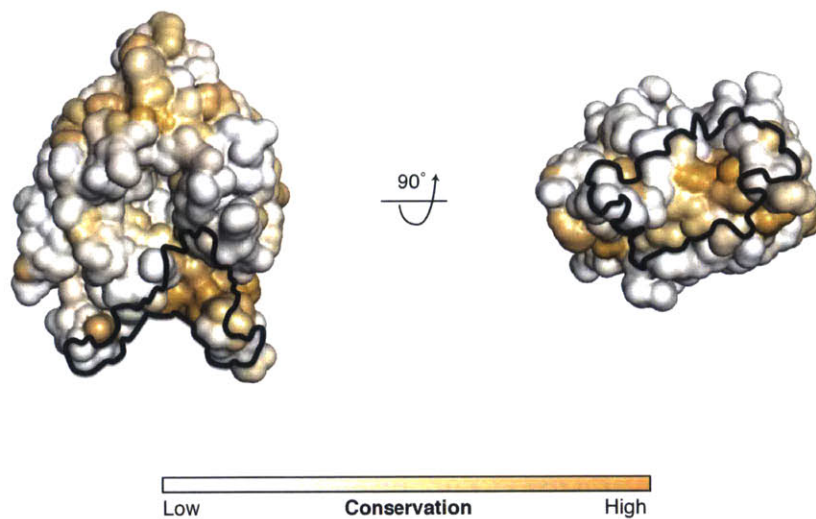


Figure 7. The HsBBS9₁₋₄₀₇ β -propeller forms a homodimer in the crystal lattice. (A) Overview of the HsBBS9₁₋₄₀₇ homodimer. Each NCS copy (one shown in blue, the other in grey) is depicted as a cartoon representation. *Inset left (solid line)*, close-up view of the β -strand domain swap. *Inset right (dotted line)*, close-up view of the C-terminal loop domain swap, with interacting residues shown as sticks. Black dotted lines depict hydrogen bonds. The inset views are rotated relative to the overview for clarity. **(B)** Conservation mapped, from low (white) to high (orange), onto a solvent-accessible surface representation of HsBBS9₁₋₄₀₇ in the same orientation as the grey copy in (A). The binding site of the symmetry mate is outlined (black). A 90° rotation of this starting orientation is shown, for clarity.



Figure 8. Multiple sequence alignment (MSA) of the BBS9 N-terminal domain from 13 BBS9-containing eukaryotes. Residues are colored according to sequence identity, from white to blue. Numbered above the MSA, in red, are residues that participate in sequence-specific interactions within the HsBBS9 homodimer interface (hydrogen bonds and salt bridges). Residue ranges marked by black lines above the MSA denote amino acids buried by the interface (Van der Waals interactions). Species were chosen for inclusion in the MSA based on Bork et al. and Dawson et al. (Cicarelli et al., 2006; Fritz-Laylin et al., 2010), to obtain a diverse sampling of eukaryotes and are ordered according to increasing distance from *Homo sapiens*. BBS9 is notably absent from land plants and fungi. The alignment comprises 13 metazoa (7 chordata, 1 arthropoda, 1 nematoda, 1 echinodermata, 1 placozoa, 1 cnidaria, and 1 porifera), 1 excavata, 1 protist (choanoflagellate), 1 plantae (green algae/chlorophyta), and 2 protozoa (chromalveolata).

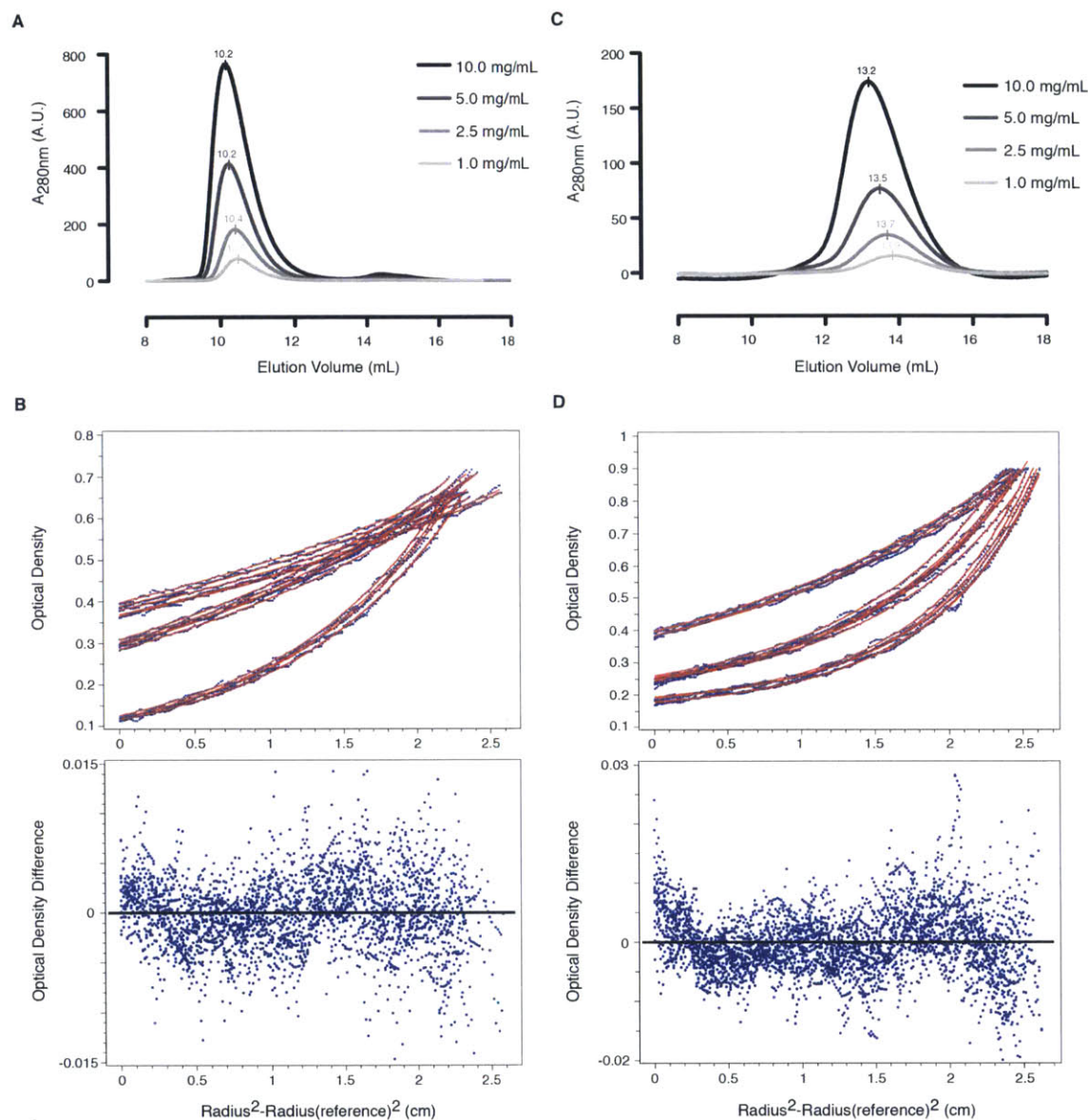
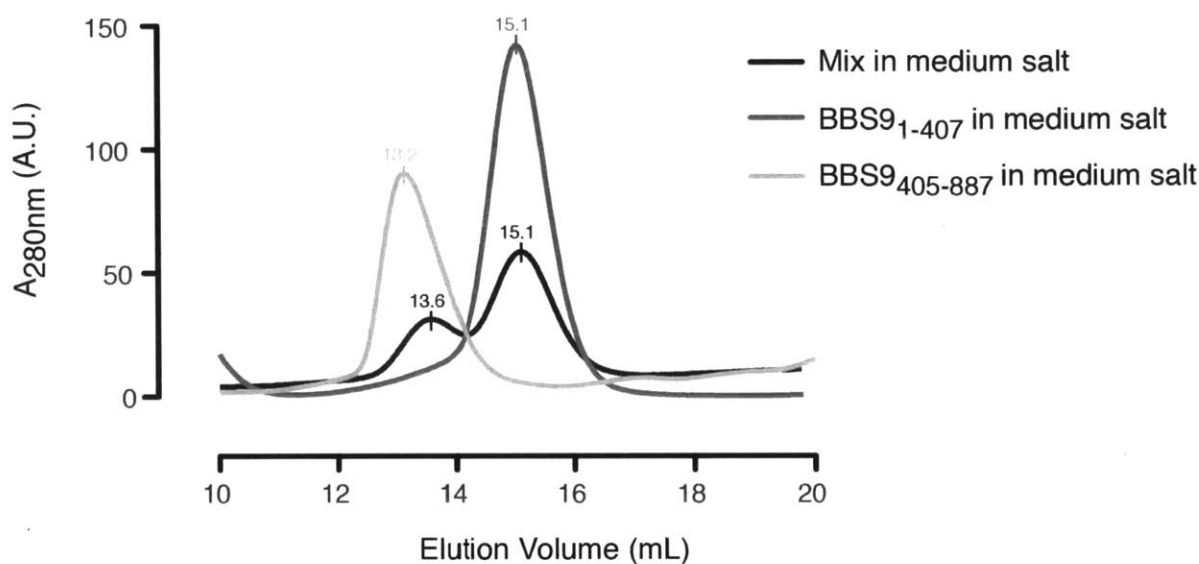


Figure 9. The HsBBS9₁₋₄₀₇ β -propeller does not form a homodimer in solution, unlike the HsBBS9₄₀₅₋₈₈₇ C-terminal half. (A) Overlaid absorbance traces from SEC experiment on HsBBS9₁₋₄₀₇ at 1 mg/mL (light grey), 2.5 mg/mL (grey), 5 mg/mL (dark grey), and 10 mg/mL (black). Elution volume is indicated above each peak. (B, top) Absorbance data, at 280 nm, for HsBBS9₁₋₄₀₇ from equilibrium AUC experiment plotted against cell radius position in centimeters (see methods for details). Individual absorbance measures are depicted as blue dots and non-linear least squares fits are shown (red lines). Data and fits are plotted for HsBBS9₁₋₄₀₇ at 9,500 rpm (top), 12,000 rpm (middle), and 18,000 rpm (bottom). (B, bottom) Residuals for model fit to data. The zero position is indicated by a black line. (C) Overlaid absorbance traces from SEC experiment on HsBBS9₄₀₅₋₈₈₇ at 1 mg/mL (light grey), 2.5 mg/mL (grey), 5 mg/mL (dark grey), and 10 mg/mL (black). (D) Absorbance data, at 280 nm, for HsBBS9₄₀₅₋₈₈₇ from equilibrium AUC experiment plotted as in (B). Data and fits are plotted for HsBBS9₄₀₅₋₈₈₇ at 9,500 rpm (top), 14,000 rpm (middle), and 17,000 rpm (bottom).

A



B

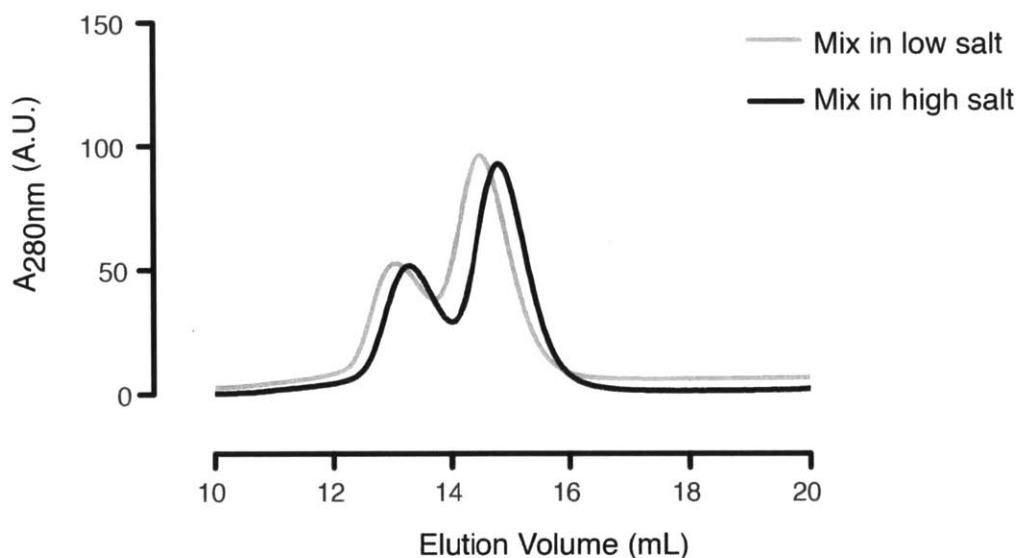


Figure 10. HsBBS9₁₋₄₀₇ and HsBBS9₄₀₅₋₈₈₇ do not bind to each other on SEC. (A)

Absorbance trace overlay from SEC on HsBBS9₁₋₄₀₇ (dark grey), HsBBS9₄₀₅₋₈₈₇ (light grey), and the two mixed together in a 1:1 ratio (black) in medium salt buffer (150 mM NaCl, see methods section for complete buffer composition). Peak elution volumes are noted above the curves and are approximately the same after mixing, indicating a lack of complex formation. **(B)**

Absorbance trace overlay from SEC on HsBBS₁₋₄₀₇ and HsBBS₄₀₅₋₈₈₇ mixed in low salt buffer (100 mM NaCl and 10% v/v glycerol, light grey) and high salt buffer (500 mM NaCl, black). A net shift of both peaks occurs as a result of buffer differences. A third peak of smaller elution volume does not form, indicating that the proteins do not bind to each other.

Chapter 4: Conclusion

Summary

The work presented in this thesis has led to a 4.1 Å structure of the Nup120-Nup85-Nup145C-Sec13 Y complex hub from the thermophilic fungus *Myceliophthora thermophila*. This structure revealed that the minimal Y complex is in fact a hexamer and demonstrated a dramatic curvature of the Y complex arms not seen in previous EM reconstructions of the Y, which were notably flat. We attribute this flatness to a commonly observed negative stain artifact. Using this hub structure, we were able to assemble a nearly complete, atomic resolution composite of the complete Y complex. Modeling based on this composite suggested that the core, hexameric Y complex structure is conserved across eukaryotes and the additional Y components in metazoa largely decorate this core. Docking of this composite structure into the 32 Å cryo-ET reconstruction of the entire human NPC, from Bui et al., allowed us to propose a new higher-order assembly model for the Y complex (Bui et al., 2013). Specifically, we proposed a single head-to-tail Y ring on each face of the NPC, that is capable of adopting two different, mutually exclusive stem conformations in the assembled pore that are equally represented in the averaged cryo-ET density. At the very least, this new model illustrates that many reasonable fitting solutions are possible from docking approaches and additional experimental evidence is required to rule out certain models.

This body of work has also led to a 1.8 Å crystal structure of the BBS9 N-terminal β-propeller. Analysis of this structure revealed clear homology to the BBS1 β-propeller and a conserved, negatively charged patch on one face of the propeller that could be a binding site for an interacting protein or ligand. The G141R point mutation in BBS9 that causes Bardet-Biedl syndrome is located within this N-terminal β-propeller. We demonstrated that this mutant β-propeller fails to express, in *E. coli* and human cells, and mutational analysis reveals an absolute requirement for a glycine at this position. This requirement is structural, based on the unique main chain dihedral angles that only glycine can adopt and side chain steric restraints that cause even an alanine to clash with neighboring residues. Although the BBS9 β-propeller

forms a convincing dimer in the crystal lattice, we show that it is a monomer in solution by analytical ultracentrifugation (AUC). By the same technique, we demonstrated that the C-terminal half of BBS9, comprising the GAE, α/β platform, and α -helical domains, weakly dimerizes in solution. The BBS9 β -propeller does not interact with the C-terminal half of the protein and rather is flexibly tethered, based on gel filtration binding experiments.

Future directions

Y complex

With a structure of the Y complex hub in hand, the only remaining portion of the Y that has not been structurally characterized, from any organism, is an 8.5 kDa segment on scNup84 (hsNup107). Solving this fragment, which is predicted to fold into four α -helices and is located between the trunk and tail subdomains, will allow the construction of a truly complete Y complex composite structure, without any modeled portions. To this end, I currently have crystals of the scNup84_{FL}-scNup133 _{Δ propeller} complex that diffract to 8 Å and contain this missing piece. Optimization of these crystals, to increase the resolution of diffraction, is required to obtain a structure solution.

The human Y complex is an attractive next target for structural characterization by x-ray crystallography, as we want to understand how the additional Nups present in metazoa, but absent from *S. cerevisiae*, incorporate into the Y. We can assess to what extent, if any, these additional components distort the core Y complex structure and provide experimental validation for the modeling that suggests these components merely decorate the core scaffold. Although the structure of Nup37 bound to Nup120 from *S. pombe* has been solved (Bilokapic and Schwartz, 2012), structures are not available for Nup43 bound to Nup85 or the ELYS central α -helical domain, by itself or bound to Nup120-Nup37. Additionally, the exact binding sites of these components are not known. How the hsNup160 (scNup120) C-terminal α -helical domain, absent in *S. cerevisiae*, incorporates into the hub region is also an open question. These

structures will additionally lead to the goal of an atomic resolution, composite structure of the complete human Y complex. Such a composite will allow for more confident fitting into cryo-ET density maps from intact metazoan NPCs. Docking, in order to determine the Y complex arrangement in the assembled NPC, will become more accurate as the resolution gap between the tomograms and the crystal structures shrinks, primarily through higher resolution tomograms. Recent advances in microscope technology will make this possible.

Toward the goal of understanding species-specific differences in NPC structure, a cryo-ET reconstruction of the intact *S. cerevisiae* NPC would be particularly useful. The dimensions of the yeast NPC could be directly compared to the metazoan NPC and docking could be performed with the composite Y structure to determine if the same set of fitting solutions are obtained for the yeast and metazoan NPC. After all, it is formally possible that the higher order assembly modes are different in these organisms, i.e. a single Y ring vs. double Y ring per face of the NPC.

To understand the higher order assembly of the Y complex in the context of an intact NPC, we need to define the inter Y complex interaction network. Currently, there are very few inter Y complex contacts described in the literature. This includes the scNup133-scNup120 interaction observed by Hoelz and colleagues and a hsNup43-hsNup96 mass spec crosslink observed by Bui et al. that cannot exist in the yeast NPC due to the absence of Nup43 (Seo et al., 2009; Bui et al., 2013). One path forward is to use nanobodies, derived from heavy-chain only antibodies raised in Alpacas, specific for *S. cerevisiae* Y complex Nups to probe NPC assembly by disrupting interactions *in vivo* (De Meyer et al., 2014). Additionally, nanobodies that do not disrupt NPC assembly, but incorporate into the pore, can be used as site specific labels for individual Nups in cryo-ET reconstructions of the NPC, allowing for definitive positional assignment. Functionalization of these nanobodies by fusing them with the non-specific biotin ligase, BirA*, or a method such as crosslinking mass spectrometry on nanobody-bound NPCs

can yield proximity information on which Nups are near the known nanobody binding site (Roux et al., 2012).

One possibility as to why direct inter Y and inter subcomplex contacts have been difficult to identify, besides the problem of potentially low affinity interactions, is that few direct interactions are formed between the Y and other members. Rather, it is possible that interactions are mediated through the cohesive FG network. After all, it has been shown that FG-repeats, specifically Nup100, Nup116, and Nsp1, bind to Nup188 and Nup192 of the Nic96 complex (Andersen et al., 2013). We have preliminary data that demonstrates FG-binding to the Y complex as well. Nup145N, an FG-Nup whose metazoan ortholog spontaneously forms hydrogels in solution, is localized throughout the NPC (Hülsmann et al., 2012; Schmidt and Görlich, 2015). Interactions between Nup145N or its metazoan ortholog with the Y complex, Nic96 complex, and Nup82 complex have all been observed (Lutzmann et al., 2005; Hodel et al., 2002; Stuwe et al., 2012). Perhaps these interactions are necessary for higher order NPC assembly and are not only required for transport barrier formation. Although we find this possibility unlikely, identifying FG-binding sites on scaffold Nups and determining the function of this binding could prove useful for understanding NPC assembly and function more generally.

With a near-complete Y complex structure in hand, the next major goal for NPC biologists is to obtain an assembly structure of the Nic96 complex. Although atomic resolution structures are available for the individual components, it is currently unknown how these proteins come together and what this assembled complex looks like. This makes even tentative placement of the complex into the current cryo-ET density maps impossible. It is likely that inherent flexibility in the complex has confounded structure solution thus far by both crystallography and EM.

BBSome

Solving the structure of the BBS9 N-terminal β -propeller is a first step in the structural determination of the assembled BBSome. Much additional work is required to assess, both on

the atomic and ultrastructural level, how similar this coat complex is to the clathrin, COPI, and COPII vesicle coats. We also do not yet know how the BBSome interacts with cargo, intraflagellar transport complexes, or the interactions that govern protomer and higher order coat assembly. In order to solve crystal structures of other BBS components, either in complex or isolation, to answer these questions a source other than recombinant expression in *E.coli* is required. Despite extensive effort, I was unable to obtain sufficient soluble material of the other BBSome complex members for crystallization. It would be worthwhile to attempt expression in insect or human cells. Additionally, *in vitro* reconstitution of the entire complex is desired so that it can be studied by negative stain or cryo-EM in order to obtain a single particle reconstruction of the intact BBSome. Recombinant coexpression of all eight components should also be attempted. If recombinant efforts fail, isolation of a TAP-tagged BBSome from *Tetrahymena thermophila* could yield sufficient material for structural analysis. I suggest *Tetrahymena* on the basis that it is a genetically tractable organism, it has hundreds of motile cilia per cell, and it has the full core complement of BBS proteins with the exception of BBIP10 (Nachury et al., 2007; Kataoka et al., 2010). Additionally, these *Tetrahymena* BBS proteins are smaller than their human orthologs, which could make them more amenable to crystallographic analysis.

References

- Andersen, K.R., Onischenko, E., Tang, J.H., and Kumar, P. (2013). Scaffold nucleoporins Nup188 and Nup192 share structural and functional properties with nuclear transport receptors. *Elife* 2, e00745.
- Bilokapic, S., and Schwartz, T.U. (2012). Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proceedings of the National Academy Of Sciences* 109, 15241.
- Bui, K.H., Appen, von, A., DiGuilio, A.L., Ori, A., Sparks, L., Mackmull, M., Bock, T., Hagen, W., Andrés-Pons, A., Glavy, J.S., et al. (2013). Integrated Structural Analysis of the Human Nuclear Pore Complex Scaffold. *Cell* 155, 1233.
- De Meyer, T., Muyldermans, S., and Depicker, A. (2014). Nanobody-based products as research and diagnostic tools. *Trends Biotechnol.* 32, 263-70.

- Hodel, A.E., Hodel, M.R., Griffis, E.R., Hennig, K.A., Ratner, G.A., Xu, S., and Powers, M.A. (2002). The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Mol. Cell* *10*, 347-58.
- Hülsmann, B.B., Labokha, A.A., and Görlich, D. (2012). The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell* *150*, 738-51.
- Kataoka, K., Schoeberl, U.E., and Mochizuki, K. (2010). Modules for C-terminal epitope tagging of Tetrahymena genes. *Journal of Microbiological Methods* *82*, 342.
- Lutzmann, M., Kunze, R., Stangl, K., Stelter, P., Tóth, K.F., Böttcher, B., and Hurt, E. (2005). Reconstitution of Nup157 and Nup145N into the Nup84 complex. *Journal of Biological Chemistry* *280*, 18442-51.
- Nachury, M.V., Loktev, A.V., Zhang, Q., Westlake, C.J., Peränen, J., Merdes, A., Slusarski, D.C., Scheller, R.H., Bazan, J.F., Sheffield, V.C., et al. (2007). A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* *129*, 1201-13.
- Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology* *196*, 801-10.
- Schmidt, H.B., and Görlich, D. (2015). Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *Elife* *4*, e04251.
- Seo, H., Ma, Y., Debler, E.W., Wacker, D., Kutik, S., Blobel, G., and Hoelz, A. (2009). Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proceedings of the National Academy of Sciences* *106*, 14281-14286.
- Stuwe, T., Borzyskowski, von, L.S., Davenport, A.M., and Hoelz, A. (2012). Molecular basis for the anchoring of proto-oncoprotein Nup98 to the cytoplasmic face of the nuclear pore complex. *J. Mol. Biol.* *419*, 330-46.

Appendix: Allosteric activation of apicomplexan calcium-dependent protein kinases

The material presented in this chapter was adapted, with permission, from the following publication:

Ingram J.R.*, Knockenhauer K.E.*, Markus B.M.*, Mandelbaum J., Ramek A., Shan Y., Shaw D.E., Schwartz T.U., Ploegh H.L., and Lourido S. (2015). Allosteric activation of apicomplexan calcium-dependent protein kinases. *Proc Natl Acad Sci USA* 112, E4975-4984.

*Equal contribution

Author contributions: J.R.I., K.E.K., B.M.M., A.R., Y.S., and S.L. designed research; J.R.I., K.E.K., B.M.M., J.M., A.R., Y.S., and S.L. performed research; D.E.S. participated in overseeing computational research; J.R.I., K.E.K., B.M.M., A.R., Y.S., T.U.S., H.L.P., and S.L. analyzed data; and J.R.I., K.E.K., B.M.M., T.U.S., H.L.P., and S.L. wrote the paper.

Abstract

Calcium-dependent protein kinases (CDPKs) comprise the major group of Ca^{2+} -regulated kinases in plants and protists. It has long been assumed that CDPKs are activated, like other Ca^{2+} -regulated kinases, by derepression of the kinase domain (KD). However, we found that removal of the autoinhibitory domain from *Toxoplasma gondii* CDPK1 is not sufficient for kinase activation. From a library of heavy chain-only antibody fragments (VHHs) we isolated an antibody (1B7) that binds *Tg*CDPK1 in a conformation-dependent manner and potently inhibits it. We uncovered the molecular basis for this inhibition by solving the crystal structure of the complex and simulating, through molecular dynamics, the effects of 1B7-kinase interactions. In contrast to other Ca^{2+} -regulated kinases, the regulatory domain of *Tg*CDPK1 plays a dual role, inhibiting or activating the kinase in response to changes in Ca^{2+} concentrations. We propose that the regulatory domain of *Tg*CDPK1 acts as a molecular splint to stabilize the otherwise inactive KD. This dependence on allosteric stabilization reveals a novel susceptibility in this important class of parasite enzymes.

Introduction

Calcium-dependent protein kinases (CDPKs) are major transducers of Ca^{2+} signaling in plants and protists. CDPKs regulate many essential processes in medically important parasites, including the apicomplexans responsible for malaria (*Plasmodium* spp.), toxoplasmosis (*Toxoplasma gondii*), and cryptosporidiosis (*Cryptosporidium parvum*). In these parasites, CDPKs control Ca^{2+} -dependent secretion, parasite motility, gametogenesis, as well as invasion and egress from infected cells (reviewed in ref. Lourido and Moreno, 2015). Their absence from mammalian genomes makes CDPKs attractive drug targets. Inhibition of *Tg*CDPK1 phenocopies a conditional knock-down of the kinase by blocking Ca^{2+} -regulated secretion (Lourido et al., 2010). The closest paralog of *Tg*CDPK1, *Tg*CDPK3, regulates the initiation of *T.*

gondii egress from host cells (Garrison et al., 2012; Lourido et al., 2012; McCoy et al., 2012), and its homolog in *P. falciparum*, PfCDPK1, is essential (Azevedo et al., 2013).

CDPKs comprise an N-terminal kinase domain (KD), linked by a junction region to four EF hands that form a calmodulin-like domain (CaM-LD) and bind Ca^{2+} . Studies of plant CDPKs suggest that the key mechanism of regulation is inhibition of the KD by the junction region preceding the EF hands, predicted to act as a pseudo substrate (Harmon et al., 1994; Harper et al., 1994). The junction region interacts with the CaM-LD, and Ca^{2+} binding alters this interaction, presumably to relieve inhibition of the KD (Chandran et al., 2006; Christodoulou et al., 2004). Structures of *Tg*CDPK1 corroborated these predictions (Ojo et al., 2010; Wernimont et al., 2010), and showed the extreme Ca^{2+} -dependent rearrangement of the CaM-LD, which warranted its distinction as a new regulatory domain, dubbed the CDPK activation domain (CAD) (Wernimont et al., 2010). In the absence of Ca^{2+} , the junction domain and the first α -helix of EF hand I form a single long α -helix that extends from the base of the KD, occupies the substrate-binding site and occludes the catalytic pocket (Wernimont et al., 2010). This structure is reminiscent of the autoinhibited state of mammalian Ca^{2+} /calmodulin-dependent protein kinases (CaMKs) (Rellos et al., 2010; Rosenberg et al., 2005), reinforcing the notion that autoinhibition is the major mechanism of CDPK regulation.

Here we show that removal of the regulatory domain of *Tg*CDPK1 is not sufficient to activate the kinase. To identify new tools to probe the mechanism of kinase activation, we turned to single domain antibody fragments, or VHHS, derived from alpaca (*Vicugna pacos*) heavy chain-only antibodies. In screening phage display libraries of VHHS from alpacas naturally infected with *T. gondii*, we isolated a single domain antibody (VHH 1B7) against *Tg*CDPK1. This antibody recognizes the kinase in a conformation-dependent manner and selectively inhibits the Ca^{2+} -bound form. Co-crystallization of the antibody with *Tg*CDPK1 has

yielded new insights into the activation of these enzymes, and revealed a novel mechanism for CDPK inhibition.

Results

Removal of the regulatory domain of *Tg*CDPK1 is not sufficient for kinase activation

By analogy to other Ca^{2+} -regulated kinases, it has been assumed that apicomplexan CDPKs are activated by de-repression of the KD upon Ca^{2+} -binding to the regulatory domain (CAD) (Ojo et al., 2010; Wernimont et al., 2010; Zhao et al., 1994). We reasoned that expression of *Tg*CDPK1 without the predicted repressive domains would yield a constitutively active kinase. We expressed two constructs that either removed the CAD completely (residues 1–314) or retained the autoinhibitory α -helix (CH1; residues 1–336). Like the full-length enzyme, both constructs expressed well as soluble proteins, but neither showed measurable catalytic activity (Fig. 1A).

To examine the effect of severing the regulatory domain from the active kinase, we introduced the human rhinovirus 3C protease cleavage site at the junction of the KD and the CAD of *Tg*CDPK1, to generate the kinase referred to as “3C”. Treatment of the mutant enzyme with the protease separated the two domains of *Tg*CDPK1-3C, but did not cleave the wild-type enzyme (Fig. 2A). Although introduction of the protease site decreased activity somewhat, cleaving the connection between the CAD and the KD abolished kinase activity (Fig. 2B). Separating the KD from the CAD by size exclusion chromatography (SEC), or inclusion of a 5-fold molar excess of the CAD failed to yield appreciable kinase activity (Fig. 1B). We conclude that tethering the CAD to the KD is necessary for *Tg*CDPK1 activity.

Identification of single domain antibodies to probe *Tg*CDPK1 structure

Single domain antibodies derived from the heavy chain-only antibodies of alpacas (VHHs) are useful crystallization chaperones that often stabilize transient conformations (De Meyer et al., 2014). We identified VHHs that bind *Tg*CDPK1 from alpacas in our cohort that had antibodies against *T. gondii* proteins (Fig. 3A). VHH phage display libraries from the seropositive animals were pooled and panned against recombinant *Tg*CDPK1. VHH clones that showed reactivity against *Tg*CDPK1 by ELISA were enriched for the sequence represented by clone 1B7 (Fig. 3B). 1B7 was sub-cloned into a bacterial cytoplasmic expression vector, produced in *E. coli* and purified to homogeneity (Fig. 3C, D). By immunoblot, 1B7 recognized both endogenous and recombinant versions of *Tg*CDPK1 and its closest paralog *Tg*CDPK3 (Fig. 4A), with a limit of detection below 15 ng (Fig. 4A, Fig. 5). Measured against a standard curve of recombinant proteins, these kinases are expressed at ~4–7 femtograms per cell, equivalent to 40,000–80,000 copies.

1B7 dynamically senses *Tg*CDPK1 conformation

To determine whether 1B7 could sense Ca^{2+} -dependent conformational changes (Ojo et al., 2010; Wernimont et al., 2010), we performed immunoprecipitations (IPs) in the presence of either CaCl_2 or the cation chelator EGTA. 1B7 required Ca^{2+} to deplete lysates of its cognate antigens (Fig. 4B). Chelation of Ca^{2+} with EGTA readily eluted the bound CDPKs from the 1B7 beads, which could be reversed by further Ca^{2+} addition, demonstrating the reversibility of the interaction (Fig. 6A). To identify parasite proteins retrieved by 1B7, the eluted samples were analyzed by mass spectrometry. The two paralogs, *Tg*CDPK1 and *Tg*CDPK3, were prominently represented with 26 unique peptides each (52% and 41% sequence coverage, respectively), and were the only CDPKs found (Table 1). 1B7 thus specifically recognizes the Ca^{2+} -bound form of *Tg*CDPK1 and *Tg*CDPK3.

We next determined whether 1B7 could form a stable complex with *Tg*CDPK1. Using 1B7 labeled with a carboxytetramethylrhodamine (TAMRA) fluorophore at its C terminus, we tracked 1B7 binding to recombinant *Tg*CDPK1 by SEC. When pre-incubated with *Tg*CDPK1 in the presence of Ca^{2+} , 1B7-TAMRA co-eluted with the kinase, well resolved from 1B7-TAMRA alone (Fig. 4C). Thus, 1B7 binds a conformation of *Tg*CDPK1 that appears in the presence of Ca^{2+} , and releases the kinase upon Ca^{2+} chelation.

1B7 recognizes an intermediate conformation of *Tg*CDPK1

To explore the conformational changes imposed by Ca^{2+} on *Tg*CDPK1 and their influence on 1B7 binding, we mutated the Ca^{2+} -binding EF hands of the kinase. The four EF hands of canonical CDPKs are segregated into two lobes, proposed to have differing Ca^{2+} affinities (Christodoulou et al., 2004). To isolate the contribution of each lobe, we mutated the first invariant Asp in each EF-hand loop in either lobe (N lobe Asp368Ala/Asp415Ala, and C lobe Asp451Ala/Asp485Ala). Both the C-lobe mutant and wild-type kinase bound immobilized 1B7 in a Ca^{2+} -dependent manner with indistinguishable affinities, although binding was abolished by mutation of the N lobe (Fig. 4D). The N-lobe mutations also prevented 1B7 recognition by immunoblot (Fig. 6B). The N-lobe EF hands are thus crucial for 1B7 binding.

We next compared the precise $[\text{Ca}^{2+}]$ necessary for 1B7 binding and kinase activation. Using buffers with defined free $[\text{Ca}^{2+}]$, we measured the amount of kinase bound by 1B7 and separately, the phosphorylation of a heterologous substrate. Binding of *Tg*CDPK1 by 1B7 occurred at a low free $[\text{Ca}^{2+}]$, with an EC_{50} of 300–350 nM (95% C.I.), 5-fold lower than that required for activation (Fig. 4E). The discrepancy between the $[\text{Ca}^{2+}]$ required for binding and that required for activation suggests that 1B7 binds a *Tg*CDPK1 conformation that is distinct from the active state of the kinase.

1B7 is a specific inhibitor of *Tg*CDPK1 and related kinases

We wondered whether 1B7-bound CDPKs would retain activity in the presence of Ca^{2+} . We measured kinase activity in increasing concentrations of 1B7, or an irrelevant VHH. 1B7 inhibited both kinases—*Tg*CDPK1 and *Tg*CDPK3—with an IC_{50} of ~40 nM, and achieved complete inhibition at approximately equimolar concentrations of antibody and enzyme (Fig. 7A).

Though $\text{ATP}_{\gamma}\text{S}$ is a substrate of many parasite kinases, 6-Fu- $\text{ATP}_{\gamma}\text{S}$ is used exclusively by *Tg*CDPK1 in parasite lysates (Lourido et al., 2013). Specificity is conferred by the unusually large ATP-binding pocket of *Tg*CDPK1 that harbors a Gly at a key position typically occupied by a bulky residue, which prevents other kinases—like *Tg*CDPK3—from using 6-Fu- $\text{ATP}_{\gamma}\text{S}$ as a substrate. Thiophosphorylation can be detected following alkylation using the monoclonal antibody 51-8 (Lourido et al., 2013). We measured thiophosphorylation in total cell extracts treated with increasing concentrations of 1B7. The presence of 1B7 robustly inhibited *Tg*CDPK1-mediated thiophosphorylation but did not affect overall kinase activity (Fig. 7B).

We also probed lysates from erythrocytic stages of two different *P. falciparum* strains. A single band of the correct molecular weight for the homologs of *Tg*CDPK1 and *Tg*CDPK3, was detected in each lysate (Fig. 7C). Accordingly, 1B7 also inhibits recombinant *P. falciparum* CDPK1 (Fig. 7A). 1B7 is thus a potent and specific inhibitor of *Tg*CDPK1 and related apicomplexan kinases.

1B7 interacts directly with the regulatory domain of *Tg*CDPK1

To understand the Ca^{2+} -dependency of 1B7 binding and kinase inhibition, we obtained crystals of the *Tg*CDPK1:1B7 complex that diffracted to 2.94 Å (Table 2). We solved the structure by molecular replacement and single anomalous dispersion phasing using selenomethionine-derivatized *Tg*CDPK1. The structure was refined to a final R_{work} of 22.6% and

an R_{free} of 26.2%. We used the structures of *Tg*CDPK1 previously solved in the Ca^{2+} -bound (PDB: 3HX4) and Ca^{2+} -depleted (PDB: 3KU2, 3V51) states (Ojo et al., 2010; Wernimont et al., 2010) for comparison.

The antibody engages the N lobe of the CAD (Fig. 8A, Fig. 9A), consistent with loss of antibody binding to the N-lobe mutant (Fig. 4D, Fig. 6B). All four EF hands are occupied by Ca^{2+} in the Ca^{2+} /1B7-bound structure (Fig. 9B). The Ca^{2+} dependency of the interaction is explained by 1B7's engagement of residues that participate in Ca^{2+} coordination and form the Ca^{2+} -binding regions of EF hands I and II. The Ca^{2+} -binding loop of EF hand II is ordered in the presence of 1B7, in contrast to the structures derived in the absence of Ca^{2+} (Ojo et al., 2010; Wernimont et al., 2010). Asn53, Asn55, and Ser58 of 1B7 CDR2 form a hydrogen-bonding network with the carbonyl oxygens of Gly371 and Asp372 in the loop segment of EF hand I. CDR3 of 1B7 contacts $\alpha 2$ of the N lobe via Thr100-Lys350 and Arg101-Ser349 hydrogen bonds, and $\alpha 3$ (the first helix of EF hand I) via an Arg101-Thr361 hydrogen bond, a Glu102-His365 salt bridge, and a Phe103-Phe364 aromatic ring stack. CDR3 also contacts the loop segment and second helix of EF hand II ($\alpha 6$) through Tyr107, which forms a bidentate hydrogen bond to Asp417 and Glu426 on the CAD, and a hydrogen bond between the amide of Ala106 and the Glu423 side chain (Fig. 8B; see Fig. 9C-E for density maps). There are additional Van der Waals contacts that result in a total buried surface area of 790 \AA^2 between 1B7 and *Tg*CDPK1.

Having defined the interactions between *Tg*CDPK1 and 1B7, we explored the basis of 1B7's specificity. We generated a sequence alignment of known 1B7 binders (*Tg*CDPK1, *Tg*CDPK3, and *Pf*CDPK1) together with CDPKs from *T. gondii*, *P. falciparum* and *C. parvum*. Of the six residues involved in sequence-specific interactions, Lys350 and Thr361 are conserved in the 1B7-binding CDPKs, and variable in the others (Fig. 10).

Allosteric inhibition of *Tg*CDPK1 by 1B7

We aligned the three solved *Tg*CDPK1 structures by their KDs to determine the mechanism through which 1B7 inhibits the kinase. The CADs from the Ca^{2+} -depleted and Ca^{2+} /1B7-bound conformations occupy the opposite face of the KD from where the CAD lies in the Ca^{2+} -bound state (Fig. 11A). The first helix of the CAD (CH1) sits in the same cleft of the KD in the Ca^{2+} /1B7-bound and Ca^{2+} -depleted structures, although the helix terminates earlier in the Ca^{2+} /1B7-bound structure. Superposition of the CAD domains shows that the Ca^{2+} /1B7-bound CAD closely resembles the Ca^{2+} -bound conformer (Fig. 11B).

The Ca^{2+} -depleted conformation of *Tg*CDPK1 has been proposed to be inhibited by an autoinhibitory triad between Lys338 in CH1 and the catalytic Glu135 and Asp138, and steric occlusion of the substrate binding site (Ojo et al., 2010; Wernimont et al., 2010). In contrast, Lys338 in our Ca^{2+} /1B7-bound structure participates in interactions within the CAD. We observed the formation of a disulfide bond linking Cys247 in the KD to Cys505 in the CAD in our Ca^{2+} /1B7-bound structure (Fig. 12A). Addition of a reducing agent did not affect 1B7 inhibition of kinase activity (Fig. 12B), although crystals were not obtained in the presence of 1 mM DTT. It is therefore likely that the Ca^{2+} /1B7-bound CAD is sampling many conformations in solution, including the state captured by the oxidized crystal structure, making steric occlusion of the substrate-binding pocket an unlikely mechanism for 1B7-mediated inhibition of *Tg*CDPK1. The Ca^{2+} /1B7-bound kinase may be inactive because 1B7 and the KD bind to the same face of the CAD in the active conformation (Fig. 11C). This would imply an unprecedented role for the interaction of the CAD with the KD during activation.

Molecular dynamics show that the KD of *Tg*CDPK1 is intrinsically unstable

Long-timescale molecular dynamics (MD) simulations have been used to assess the intrinsic stability of various tyrosine kinases and inform their mechanisms of activation (Shan et al., 2012). To model the interaction of the CAD and the KD we performed MD simulations using

the Ca²⁺-bound conformation of *Tg*CDPK1 as the starting point. Simulations performed in the absence of the CAD showed that the KD diverged from that of the Ca²⁺-bound, active structure (Fig. 13A). The major changes were in the rearrangement of the activation loop and in an approximate 30° turn in the α C helix of the KD to a so-called “ α C-out” conformation that dissolves the catalytically important salt bridge between Lys80 and Glu99, and marks the inactive state of many protein kinases. We noted a clear preference for the inactive state when the CAD was removed from the KD (Fig. 13B, C).

The “regulatory” spine (R-spine) is a series of non-contiguous hydrophobic residues (Leu103, Leu114, His172, and Phe196 in *Tg*CDPK1) that are aligned in the active conformations of all protein kinases (Taylor et al., 2012). Misalignment of the R-spine residues is a key feature of the inactive states found in the MD simulation in the absence of the CAD, and the Ca²⁺/1B7-bound structure (Fig. 13D). The R-spine was similarly disrupted in the Ca²⁺-depleted structure in contrast to its perfect alignment in the Ca²⁺-bound conformation (Fig. 13D). The Ca²⁺-bound CAD appears to impart an active conformation on the KD by stabilizing the position of the α C helix required for alignment of the R-spine and catalysis.

The CAD acts as a molecular splint to stabilize the KD of *Tg*CDPK1

During activation, the CAD imposes a 15° rotation in the relative orientations of the N and C lobes of the KD, suggesting that points of attachment to both kinase lobes are needed to generate torque upon Ca²⁺ binding. Having established that integrity of the tether between the KD and the CAD is necessary (Fig. 2B), we investigated whether the flexibility or length of the tether would affect kinase activity. We replaced 9 residues from the tether with a flexible linker of equal length (9AA) or one 11 residues longer (20AA). We observed a loss of kinase activity proportional to the length and presumed flexibility of the tether (Fig. 13E). The greater loop

entropy associated with the longer linkers presumably disfavors proper positioning of CAD and KD (Zhou, 2004).

Interaction between the CAD and an N-terminal extension of the KD, also visible in the Ca^{2+} -bound structure (Wernimont et al., 2010), could provide the proposed tether to the N-lobe of the KD. Superposition of the Ca^{2+} -bound and $\text{Ca}^{2+}/1\text{B7}$ -bound structures showed that the same groove in the Ca^{2+} -bound CAD was engaged by both the N-terminal extension and CDR3 of 1B7 (Fig. 14). We therefore asked whether the N-terminal extension of *Tg*CDPK1 participates in kinase activation. Of the 50 residues in the N-terminal extension, only the final 20 are observed in the Ca^{2+} -bound structure, and only the final 6 in both inactive structures. Whereas truncation of the first 35 residues minimally affected kinase activity, elimination of 44 residues from the N-terminal extension led to a complete loss of function (Fig. 13F).

We identified two key residues important for this interaction: Phe39 and Val40. We mutated Phe39 and Val40 individually and in combination to Ala, and measured their effects. Val40Ala shifted the EC_{50} of the kinase by ~10-fold, whereas Phe39Ala strongly inhibited kinase activity, as did the double mutation (Fig. 13G). The N-terminal extension is therefore crucial for the activation of *Tg*CDPK1.

Allosteric activation is required for *Tg*CDPK1 function *in vivo*

To examine the role of the N-terminal extension *in vivo*, we used a conditional knockout (cKO) of *Tg*CDPK1 (Lourido et al., 2010). This strain uses a tetracycline trans-activator to express an HA9-tagged allele of *Tg*CDPK1, in the absence of the endogenous gene (Fig. 15A). Growth of the cKO strain in the presence of anhydrotetracycline (ATc) blocks expression of *Tg*CDPK1, which was used to demonstrate the essentiality of the kinase in the *T. gondii* lytic cycle (Lourido et al., 2010). The cKO can be complemented with an allele of the kinase constitutively expressed under the endogenous promoter of *Tg*CDPK1 (Fig. 15A). We generated strains complemented either with wild type, or with the Phe39Ala mutant incapable of

allosteric activation. We confirmed regulation of the HA9-tagged allele, and constitutive expression of the complementing, Myc-tagged alleles by immunoblot (Fig. 15B). To determine whether the expressed enzymes were active, we performed thiophosphorylation reactions, using 6-Fu-ATP γ S as a substrate, with lysates from parasites grown for 48 hours in the presence of ATc. We observed robust Ca²⁺-dependent thiophosphorylation in the wild-type complemented strain. However, consistent with our analysis of the recombinant enzymes, no *Tg*CDPK1 activity was observed in the Phe39Ala mutant strain (Fig. 15C).

We next measured plaque formation to assess functional complementation of the cKO. All strains formed plaques of similar size when grown in the absence of ATc (Fig. 15D). However, plaque formation was impaired when the cKO was grown in the presence of ATc (Lourido et al., 2010). This defect was complemented by constitutive expression of the wild-type allele, but not the Phe39Ala mutant (Figure 7D now Fig. 15D). Taken together, these data demonstrate that the Phe39 residue of *Tg*CDPK1, which mediates allosteric activation, is necessary for the function of the kinase *in vivo*.

Discussion

Unlike many other Ca²⁺-regulated kinases, the KD of *T. gondii* CDPK1 is intrinsically inactive. We identified an alpaca-derived single domain antibody (1B7) that binds to and inhibits *Tg*CDPK1 in a reversible, Ca²⁺-dependent manner. Solving the structure of *Tg*CDPK1 in complex with 1B7 showed a common binding site on the Ca²⁺-bound CAD for both the antibody and the *Tg*CDPK1 N-terminal extension. 1B7 binding prevents the stabilization of the KD by blocking the interaction of the N-terminal extension and the CAD. We refer to this mode of stabilization as a molecular splint because anchoring of the KD to the CAD at both the N and C termini is required for activation.

Ca²⁺-regulated kinases are activated by removal of an inhibitory element from the KD (Harmon et al., 1994; Harper et al., 1994; Rellos et al., 2010; Rosenberg et al., 2005). Animal Ca²⁺/calmodulin-dependent protein kinases (CaMKs) possess intrinsically active KDs, kept inactive by an autoinhibitory region (Rellos et al., 2010; Rosenberg et al., 2005). A similar autoinhibitory region is part of the CAD of CDPKs (Wernimont et al., 2010). Peptide mimetics of this region inhibit both plant (Harmon et al., 1994; Harper et al., 1994) and parasite kinases (Azevedo et al., 2013; Ranjan et al., 2009). Based on observations of plant CDPK truncations that appeared constitutively active and Ca²⁺-independent, CDPK KDs—like those of CaMKs—were thought to be intrinsically active (Harper et al., 1994; Ranjan et al., 2009). These truncations consisted of N-terminal fusions with Glutathione S-transferase (GST), which can dimerize and stabilize the conformation of its fusion partners. For phosphatidylinositol 3'-kinase and the insulin receptor KD, fusion to GST is sufficient for activation (Baer et al., 2001; Yu et al., 1998). In contrast, the KD of *Tg*CDPK1 is inactive on its own. Removal of the CAD, either genetically or proteolytically, fails to activate *Tg*CDPK1. MD simulations showed that the KD adopts an inactive conformation in the absence of the CAD. The KDs of different CDPKs might well differ in stability, but our work suggests that instability is a conserved feature of the CDPKs most closely related to *Tg*CDPK1.

Ca²⁺ binding to the CAD imposes rigidity to its fold. When connected to the N and C termini of the KD, the CAD acts as a molecular splint that supports and stabilizes the KD in its active conformation (Fig. 15E). The CAD therefore exerts a far greater degree of regulation over the KD than assumed previously. Perturbation of either point of contact between the CAD and KD impairs kinase activity. Disruption of the N-terminal extension (Fig. 13F, G) strongly inhibited kinase activity, congruent with studies of *Pf*CDPK1 (Ahmed et al., 2012). We identified Phe39 and Val40 in the N-terminal extension as crucial for kinase activation. The presence of Phe39 is required for *Tg*CDPK1 function *in vivo*, because the Phe39Ala mutant, when expressed at wild-type levels, fails to complement a conditional knockout of the kinase (Fig. 15). The similarity of

these anchor residues amongst diverse CDPKs suggests that the corresponding regions may function like the N-terminal extension of *Tg*CDPK1.

KDs have evolved diverse modes of positive and negative regulation, often through acquisition of additional protein domains or non-covalent interacting partners. The diversity of these regulatory systems often converges on similar themes, including regulation of the α C helix position via a hydrophobic patch in the N lobe of the KD (Jura et al., 2011). Regulatory elements can interact directly with this hydrophobic patch, as is the case with the non-receptor tyrosine kinases Fes and Abl (Filippakopoulos et al., 2008). In the case of Abl, the same regulatory domain can also act to lock the kinase in an inactive conformation (Nagar et al., 2003). Only a few domains are known to exert dual functions in the regulation of kinase activity, by serving both as repressors and activators. In a remarkable convergence, the CAD has adopted such a role in the regulation of *Tg*CDPK1.

The interaction of 1B7 with its target is characteristic of VHHs, which often present large convex paratopes that can access buried epitopes less accessible to conventional antibodies (De Genst et al., 2006). The 1B7 CDR3 protrudes into a groove that forms upon Ca^{2+} -binding to the CAD, displacing the N-terminal extension in the active conformation of the kinase (Wernimont et al., 2010). Features associated with autoinhibition are notably absent from the Ca^{2+} /1B7-bound structure. 1B7 may bind an intermediate state in the activation of *Tg*CDPK1 as inferred from the lower $[\text{Ca}^{2+}]$ required for 1B7 binding, versus kinase activation (Fig. 4E). Preventing the stabilization of the KD is the primary mechanism for the inhibition of *Tg*CDPK1 by 1B7, a trait generalizable at least to the clade of CDPKs to which *Tg*CDPK1 belongs.

CDPKs are obvious therapeutic targets. Development of small molecule inhibitors against them has focused exclusively on ATP-analogs, which often suffer from off-target effects against other host and parasite kinases. Targeting the atypical ATP-binding pocket of *Tg*CDPK1 and related kinases has afforded some measure of specificity. However, a single mutation suffices to confer resistance to these compounds (Lourido et al., 2012). We identify a novel

vulnerability in this family of parasite kinases, exposed by their need for stabilization. Through the use of 1B7, we provide proof of principle that this feature can be exploited to inhibit *Tg*CDPK1 and related kinases, including *Pf*CDPK1, a kinase essential for the erythrocytic stages of the malaria parasite (Azevedo et al., 2013). Conservation of the residues that mediate allosteric activation suggests that this mechanism may extend to *Cryptosporidium* CDPK1 and *Plasmodium* CDPK4. Inhibition of these kinases has already shown promising activity in murine infection models, in the case of *Cp*CDPK1 (Castellanos-Gonzalez et al., 2013), and in blocking the transmission of malaria, in the case of *Pb*CDPK4 (Ojo et al., 2012). We expect that the binding of 1B7 to *Tg*CDPK1 can be used to guide the design of small molecules with the therapeutic potential to allosterically inhibit these essential parasite kinases.

Methods

Protein Purification

Vectors encoding *Tg*CDPK1-6xHis and *Tg*CDPK3-6xHis, were generated as previously described (Lourido et al., 2012). Mutations in the EF hands or the N-terminal extension were introduced using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies), with specific primers designed according to manufacturer's instructions. N-terminally truncated versions of *Tg*CDPK1 were generated using primers P1 and P2 (Δ 35), or P1 and P3 (Δ 44). Mutant kinases with insertions in the junction between kinase domain and CAD were ordered as gBlocks from IDT (B1 for 9AA, B2 for 20AA and B3 for 3C), and cloned into pET22b+(Novagen). Kinase constructs, were transformed into BL21(DE3)V2RpAcYc-LIC+LamP *E. coli*, which express the LamP phosphatase as previously described (Wernimont et al., 2010). Cultures were grown to 0.6 OD₆₀₀ at 37°C, and induced with 1 mM IPTG for 2.5 h. Pelleted cells were lysed in CellLytic B buffer (Sigma) containing 0.1 mg ml⁻¹ lysozyme (Sigma), 650 units benzonase (Sigma) and HALT Protease Inhibitor Cocktail (Thermo-Scientific). Lysates were clarified by

centrifugation at $24 \times g$ for 30 min at 4°C. The soluble fraction was added to prewashed Ni-NTA agarose (Qiagen), and allowed to bind 1 h at 4°C, rotating. The resin was washed with 50 mM sodium phosphate pH 8.0, 500 mM NaCl, 15 mM imidazole and HALT Protease Inhibitor Cocktail, and bound proteins were eluted by increasing imidazole to 500 mM. Eluates were concentrated using Amicon Centrifugal Filters (EMD Millipore) and loaded onto a Superdex 200 26/60 size-exclusion chromatography (SEC) column equilibrated with 50 mM Tris, pH 7.5, and 150 mM NaCl. Peak elution fractions were pooled and concentrated. Protein purity was assessed by SDS-PAGE, and concentrations were determined using the DC Protein Assay (Bio-Rad). Recombinant CDPKs were stored at -80°C after addition of 25% glycerol and 1 mM DTT. For crystallization, the following changes were made. Proteins were expressed in BL21(DE3)-LOBSTR-RIL *E. coli*, which has been optimized for low background in Ni-affinity purifications (Andersen et al., 2013). Cultures were shifted to 18°C prior to induction with 0.2 mM IPTG for 16 h. Cells were lysed using a cell disruptor (Constant Systems) in 50 mM potassium phosphate, pH 8.0, 500 mM NaCl, 30 mM imidazole, 3 mM β -mercaptoethanol, 1 mM PMSF and 250 units of Turbonuclease (Eton Bioscience). The soluble fraction was added to prewashed Ni-Sepharose fast-flow beads (GE Healthcare) and agitated 20 min at 4°C. Protein was eluted with 250 mM imidazole, pH 8.0, 150 mM NaCl, and 3 mM β -mercaptoethanol. SEC was performed in 10 mM Tris, pH 7.4, 150 mM NaCl, 1 mM DTT, and 1 mM CaCl_2 . SeMet-derivatized *Tg*CDPK1 was purified identically, with the exception that 10 mM β -mercaptoethanol and 5 mM DTT were used in purification steps. Upon VHH selection by phage display and ELISA, the 1B7 coding sequence was sub-cloned for expression. First, VHH coding sequences were amplified with primers P4 and P5, which include a sortase A recognition motif. The resulting product was subcloned into the pET30b+ (Novagen) expression vector using NdeI and XhoI sites for inclusion of a C-terminal 6xHis tag. BL21(DE3) *E. coli* containing the plasmid were grown to mid-log phase at 37°C in 2YT plus kanamycin, and induced with 1 mM IPTG overnight

at 30°C. Total soluble protein was collected by French press and cleared at 39 × kg for 30 min, before loading onto Ni-NTA (Qiagen) in 50 mM Tris, pH 7.5, 150 mM NaCl and 10 mM imidazole. Protein was eluted in 50 mM Tris, pH 7.5, 150 mM NaCl, 500 mM imidazole and 10% glycerol, then loaded onto a Superdex 75 10/300 column in 50 mM Tris, pH 7.5, 150 mM NaCl, 10% glycerol. Peak fractions were pooled and concentrated. Purity was assessed by SDS-PAGE. The *Tg*CDPK1 variant carrying a human rhinovirus 3C protease cleavage site was digested with 3C protease for 36 h at 4°C to achieve near-complete digestion. Wild-type kinase underwent the same treatment, to be used as a control. KD and CAD were separated from digested samples by SEC through a Superdex 75 16/60 column equilibrated with 50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM CaCl₂ and 1 mM DTT. Fractions for each peak were pooled and concentrated as described above.

Growth of host cells and parasite strains

T. gondii tachyzoites were maintained by growth in monolayers of human foreskin fibroblasts (HFFs) cultured in DMEM containing 10% tetracycline-free fetal bovine serum (HyClone), 2 mM glutamine, 10 mM HEPES pH 7.5 and 20 µg ml⁻¹ gentamicin, as described (Lourido et al., 2010). The *Tg*CDPK1 conditional knockout (cKO) was generated as previously described (Lourido et al., 2010). The wild-type complementation plasmid was previously generated, and contains a Myc-tagged version of *Tg*CDPK1 downstream of the endogenous promoter sequence of the kinase (Lourido et al., 2010). Phe39 was mutated to Ala in this construct using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies), according to manufacturer's instructions. Mutagenesis and integrity of the entire open reading frame was confirmed by sequencing. The complementing plasmids were co-transfected with pDHFR-TS (Donald and Roos, 1993), and stable integration was selected for by growth in 3 µM pyrimethamine (Sigma). Clones were isolated by limiting dilution and expression of the

transgenes was confirmed by immunoblot. $1.5 \mu\text{g ml}^{-1}$ ATc (Clontech) and was added as indicated to suppress expression of the regulatable, HA9-tagged wild-type allele.

Plaque assays

Assays were performed as previously described (Lourido et al., 2010). Confluent monolayers of HFFs in six-well plates were infected with 400-1000 parasites per well in media with or without $1.5 \mu\text{g ml}^{-1}$ ATc. At 24 h after infection, additional medium was added to decrease the concentration of ATc to $1 \mu\text{g ml}^{-1}$. Monolayers were fixed 7 days after infection and stained with crystal violet.

Immunoblots

Parasites from fully lysed monolayers were lysed directly in SDS-PAGE loading buffer, or in a 50 mM Tris, pH 7.5, 150 mM NaCl, 1% Triton X-100 and HALT Protease Inhibitor Cocktail, prior to measuring protein concentration. Serum was collected from each alpaca prior to immunization and stored at -80°C (Maass et al., 2007). *T. gondii* lysates were run on SDS-PAGE and transferred to nitrocellulose. Membranes were blocked overnight at 4°C in PBS, 5% (w/v) nonfat dry milk. Serum was diluted 1:5,000 in blocking solution and incubated 1 h with the respective section of the blot. Horseradish peroxidase-conjugated goat anti-llama IgG (Bethyl) was used as a secondary antibody at a 1:20,000 dilution in blocking solution. To compare *T. gondii* lysates with recombinant proteins, lysates corresponding to different numbers of parasites were loaded besides dilutions of the recombinant proteins. 1B7 was labeled using the LI-COR IRDye 800CW labeling kit according to manufacturer's instructions, and used at a concentration of $1 \mu\text{g ml}^{-1}$ for immunoblots. Expression of the regulatable and complementing alleles of *TgCDPK1* was similarly assessed by immunoblot following growth of the parasites in the presence or absence of ATc. Immunoblots were probed, as indicated, with Rabbit anti-

TgACT1 (Dobrowolski et al., 1997) diluted 1:10,000, mouse anti-penta HIS (Qiagen) diluted 1:5,000, rabbit-anti-HA9 (Invitrogen) diluted 1:10,000, mouse-anti-Myc (9E10; Sigma) diluted 1:5,000 and mouse anti- α Tubulin (12G10; DSHB) diluted 1:10,000.

VHH Library generation

Alpacas (*Vicugna pacos*) were purchased locally, maintained in pasture, and immunized following a protocol authorized by the Tufts University Cummings Veterinary School IACUC. For each alpaca, total RNA was isolated from $\sim 10^6$ fresh PBLs using the RNeasy Plus Mini Kit (Qiagen), following the manufacturer's instructions. First strand cDNA synthesis was performed using SuperScript III reverse transcriptase (Life Technologies) and a combination of oligo dT, random hexamer or immunoglobulin-specific primers, AICH2 and AICH2.2, as previously described (Maass et al., 2007). Subsequent PCR amplification of VHH sequences and phage library generation followed an established procedure (Maass et al., 2007). This included the use of alpaca-specific primers for VHH gene amplification and the use of a phagemid vector adapted for VHH expression as a pIII fusion. Following transformation into TGI cells (Agilent), the total number of independent clones for each library was estimated to range from 1×10^6 - 10^7 . The resulting phagemid libraries were stored at -80°C .

Generation of M13 phage displaying anti-*T. gondii* VHH library

Several independent libraries were generated from animals immunized with different mixtures of antigens. None of the animals were deliberately immunized with *T. gondii*. Libraries were pooled and used to inoculate 100 ml SOC with $50 \mu\text{g ml}^{-1}$ ampicillin. The culture was grown to mid-log phase, and infected with $100 \mu\text{l } 10^{14}$ PFU ml^{-1} VCSM13 helper phage. Following a 2 h incubation at 37°C , the cells were harvested by centrifugation, and re-suspended in 100 ml 2YT, 0.1% glucose, $50 \mu\text{g ml}^{-1}$ kanamycin and $50 \mu\text{g ml}^{-1}$ ampicillin.

Cultures were incubated overnight at 30°C, then centrifuged for 20 min at 7,700 × g, followed by phage precipitation from the resulting supernatant with 1% PEG-6000, 500 mM NaCl at 4°C, and re-suspension in PBS.

Selection of VHHs by phage display

VHHs were selected by panning against recombinant *Tg*CDPK1. 100 µg recombinant *Tg*CDPK1 was biotinylated by coupling Chromalink NHS-biotin reagent (Solulink) to primary amines for 90 min in 100 mM sodium phosphate, pH 7.4, 150 mM NaCl. Unreacted biotin was removed using a 10 kDa MWCO concentrator (Millipore). Biotin incorporation was monitored, following the manufacturer's guidelines. 100 µl MyOne Streptavidin-T1 Dynabeads (Life Technologies) were blocked in PBS, 2% (w/v) bovine serum albumin (Sigma) for 2 h at 37°C. 20 µg biotinylated *Tg*CDPK1 in PBS was added to the blocked beads, and incubated for 30 min at 25°C, with agitation. The beads were then washed 3 times in PBS, and 200 µl of 10¹⁴ PFU ml⁻¹ M13 phage displaying the VHH library were added in PBS, 2% BSA for 1 h at room temperature. The beads were then washed 15 times with PBS, 0.1% Tween-20. Phage was eluted by the addition of *E. coli* ER2738 (NEB) for 15 min at 37°C, followed by elution with 200 mM glycine, pH 2.2, for 10 min at 25°C. The eluate was neutralized, pooled with the *E. coli* culture, and plated onto 2YT agar plates supplemented with 2% glucose, 5 µg ml⁻¹ tetracycline and 10 µg ml⁻¹ ampicillin, and grown overnight at 37°C. A second round of panning was performed with the following modifications: 2 µg of biotinylated *Tg*CDPK1 was used as bait, and incubated with 2 µl 10¹⁴ PFU/ml M13 phage displaying the first-round VHH library for 15 min at 37°C, followed by extended washes in PBS, 0.1% Tween-20.

ELISA

Following two rounds of phage panning, 96 colonies were isolated in 96-well round-bottom plates and grown to mid-log phase at 37°C in 200 µl 2YT, 10 µg ml⁻¹ ampicillin, 5 µg ml⁻¹ tetracycline, induced with 3 mM IPTG and grown overnight at 30°C. Plates were centrifuged at 12 × kg for 10 min, and 100 µl of supernatant was mixed with an equal volume of PBS, 5% (w/v) nonfat dry milk. This mixture was added to an ELISA plate coated with 1 µg ml⁻¹ TgCDPK1. Following four washes in PBS, 1% Tween-20, anti-llama-HRP antibody (Bethyl) was added at a 1:10,000 dilution in PBS, 5% (w/v) nonfat dry milk for 1 h at 25°C. The plate was developed with fast kinetic TMB (Sigma) and quenched with 1 M HCl. Absorbance at 450 nm was determined in a plate reader (Spectramax; Molecular Devices).

Immunoprecipitation of TgCDPK1 and TgCDPK3

4 mg of 1B7 was coupled to 500 mg of cyanogen bromide-activated Sepharose (Sigma) according to the manufacturer's instructions. Coupling efficiency was monitored by SDS-PAGE. 1B7-Sepharose was stored in 50 mM Tris, pH 7.5, 150 mM NaCl at 4°C. Approximately 10⁹ parasites were harvested, resuspended in lysis buffer (50 mM HEPES, pH 7.4, 100 mM KCl, 1 mM MgCl₂, 10% glycerol and 0.05% Triton X-100), and disrupted with a Dounce homogenizer. 0.25 g 1B7-Sepharose was prepared for each sample by washing with lysis buffer supplemented with either 1 mM EGTA or CaCl₂. Half of the parasite lysate was diluted to 10 ml containing 1 mM EGTA or CaCl₂, and incubated with the respective beads for 1 h at 4°C, with agitation. The resin was then washed with 200 ml of the appropriate lysis buffer supplemented with EGTA or CaCl₂, and eluted with three consecutive 500 µl volumes of lysis buffer containing 1 mM EGTA. The eluted material was analyzed by SDS-PAGE followed by silver staining. Bands were excised, reduced, alkylated and digested with trypsin at 37°C overnight. The resulting peptides were extracted, concentrated and injected by standard reverse-phase

chromatography into a Thermo Orbitrap Elite mass spectrometer operated in a data dependent manner. The resulting fragmentation spectra were correlated against the known database using SEQUEST. Scaffold Q+S (Proteome Software) was used to provide consensus reports for the identified proteins. Immunoprecipitation of recombinant, wild-type and double EF-hand mutant kinases was performed similarly. 5 μ g of kinase were diluted in 1 ml of the respective binding buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X-100, supplemented with either 1 mM EGTA or 1 mM CaCl_2 , respectively) for 1 h. For each binding condition, 1B7-Sepharose beads were washed twice with the respective binding buffer, followed by incubation of \sim 100 μ g 1B7-Sepharose with 1 ml of diluted kinase for 30 min, agitating. Following incubation, the beads were loaded onto 200 μ l filter tips and washed with 3 column volumes of the respective buffer, before eluting with 40 μ l 200 mM glycine, pH 2.2, followed by neutralization with 10 μ l 1 M Tris, pH 9.0. The amount of immunoprecipitated kinase was determined by SDS-PAGE, SYPRO Ruby protein staining (Life Technologies), and fluorescent gel imaging using a Typhoon FLA 9500 laser scanner (GE Healthcare Life Sciences). Densitometry was conducted using ImageQuant TL software (GE Healthcare Life Sciences). Immunoprecipitation of wild-type *TgCDPK1* in the presence of specific $[\text{Ca}^{2+}]_{\text{free}}$ concentrations was performed correspondingly, but with distinct buffer compositions. These binding buffers were made using the Calcium Calibration Buffer Kit (Life Technologies), adjusting the supplied buffer composition to 30 mM MOPS, pH 7.2, 100 mM KCl, 4 mM CaEGTA/ K_2 EGTA, 1mM MgCl_2 and 0.05% Tween-20 in order to more closely mimic physiological conditions and to assure consistency between the conditions used to measure kinase activity and 1B7 binding. The $[\text{Ca}^{2+}]_{\text{free}}$ were determined using WEBMAX STANDARD (Bers et al., 2010). Prior to immunoprecipitation, wild-type *TgCDPK1* was dialyzed against chelex-100 (Bio Rad) treated buffer, to minimize contaminating Ca^{2+} present in the buffers.

Size exclusion chromatography (SEC)

Sortase was used to label 1B7 by incubating 1 mg of purified 1B7 with 1 mg *S. aureus* sortase A and 1 mM GGGK-TAMRA in 50 mM Tris, pH 7.5, 150 mM NaCl and 10 mM CaCl₂. The reaction proceeded overnight at 25°C, and unreacted VHH and sortase A were removed by adsorption onto Ni-NTA agarose (Qiagen). Labeled 1B7 was isolated by SEC, concentrated, and its purity determined as described above. *Tg*CDPK1 and purified TAMRA-labeled 1B7 were incubated at a 1:1 molar ratio in 50 mM Tris, pH 7.5, 150 mM NaCl, supplemented with 1 mM CaCl₂ or 1 mM EGTA for 2 h at 25°C. Samples were then loaded onto a Superdex 75 10/30 column in 50 mM Tris, pH 7.5, 150 mM NaCl, supplemented with 1 mM CaCl₂ or 1 mM EGTA, and absorbance was monitored at both 280 nm and 550 nm.

Kinase assays

Kinase assays were conducted using a peptide-based ELISA, as previously described (Lourido et al., 2012). 96-well plates were coated with Syntide-2 peptide (Calbiochem) diluted to 100 µg ml⁻¹ in carbonate coating buffer (pH 9.6), at 4°C overnight. Following incubation, plates were rinsed with wash buffer (50 mM Tris, pH 7.5, 0.2% Tween-20), and blocked with 3% (w/v) BSA in wash buffer, 2 h at 25°C. Subsequent washes were performed with wash buffer. Reactions were conducted at 30°C for 30 min in kinase buffer (20 mM HEPES, pH 7.0, 10 mM MgCl₂, 1 mM CaCl₂, 0.005% Tween-20, 1 mM ATP). Phosphorylation was detected with mAb MS-6E6 (MBL), followed by HRP-conjugated goat-anti-mouse IgG, developed with the substrate o-phenylenediamine dihydrochloride (Thermo Scientific), and detected by absorbance at 492 nm. Where appropriate, each kinase preparation was individually tested to determine its half maximal effective concentration (EC₅₀). For VHH inhibition assays and Ca²⁺ titration experiments, *Tg*CDPK1 was used at its EC₅₀. For the inhibition assays, 1B7 or a control VHH (2A7) was titrated into the reaction, prior to the addition of 1 mM ATP. Ca²⁺ titration experiments

used reaction buffers with specific $[Ca^{2+}]_{free}$, prepared as in the immunoprecipitation experiments, with 1 mM ATP was added. EC_{50} and IC_{50} values were calculated from a dose-response curve plotted in Prism (GraphPad). Duplicate or triplicate experiments were run for all assays.

Thiophosphorylation assay

Thiophosphorylation was measured as previously described (Lourido et al., 2013). Briefly, *T. gondii* parasites were lysed on ice in the reaction buffer (20 mM HEPES, pH 7.4, 137 mM KCl, 10 mM $MgCl_2$) supplemented with 1% Triton X-100. Individual aliquots of the lysate were combined with an equal volume of the reaction buffer supplemented with 8 mM CaEGTA alone or in combination with different amounts of 1B7. To start the reactions, an equal volume of reaction buffer supplemented with 1 mM GTP, 0.1 mM ATP, and 50 μ M of either ATP γ S or 6-Fu-ATP γ S (BioLog) was added to each sample. Reactions were allowed to proceed for 30 min at 30°C, before alkylating with 2 mM p-nitrobenzyl mesylate (Epitomics) for 1 h. Samples were analyzed by Western blot using a rabbit-anti-thiophosphate ester (51-8; Epitomics) at a dilution of 1:10,000, and an appropriate loading control. Goat-anti-rabbit conjugated to IRdye 680 (Li-Cor) was used to visualize the signal from the thiophosphate ester antibody. To compare *Tg*CDPK1-dependent thiophosphorylation in the different *T. gondii* strains, lysates were prepared from strains grown in 1.5μ g ml⁻¹, 48 h. Reactions were initiated as described above, except CaEGTA was substituted for K₂EGTA in the samples without Ca²⁺.

Crystallization

Recombinant *Tg*CDPK1 and 1B7 were combined in a 1:1.5 molar ratio at room temperature, rotating for 1 h. The mixture was loaded on a Superdex 200 16/60 SEC column to separate complex from excess 1B7. The complex peak was pooled, concentrated, and dialyzed

into 50 mM Tris, pH 7.4, 150 mM NaCl for 16-20 h prior to crystal-tray setup. Initial crystals of *Tg*CDPK1-1B7 complex grew via vapor diffusion at 4 mg ml⁻¹ in 25% (w/v) PEG 3350, 0.2 M (NH₄)₂SO₄, and 0.1 M Bis-Tris, pH 6.5 at 18°C, after one day incubation in a 96-well, sitting drop tray (IndexHT, Hampton Research). Protein and mother liquor were mixed in a 1:1 ratio. The use of 4% (v/v) 1-propanol, as an additive, produced optically superior crystals. Scaling-up to 24-well, hanging drop format and immediate microseeding of the crystallization drop produced diffraction-quality crystals that grew as thin, mechanically separable plate clusters. SeMet-derivatized protein crystals were obtained following this protocol, albeit they grew singular, were smaller, and took several days longer to grow. Individual crystals were serially cryoprotected in mother liquor supplemented with 20% glycerol or 35% PEG 3350 prior to cooling in liquid nitrogen.

Data processing and structure determination

Complete native and SeMet datasets were collected at the Advanced Photon Source user end station 24-IDC. Data reduction was performed in *HKL2000* (Otwinowski and Minor, 1997). Molecular replacement (MR) and single anomalous dispersion (SAD) phasing was performed using *phaser* in the PHENIX suite (Adams et al., 2010), available through *SBGrid* (Morin et al., 2013). An initial MR solution was obtained in the SeMet dataset using the Ca²⁺-depleted kinase domain of *Tg*CDPK1 (PDB: 3KU2, residues 44-313) and a VHH with the complementarity determining regions (CDRs) removed (PDB code: 1BZQ) as search models in *phaser-MR*. No solution was obtained with the active kinase domain (PDB: 3HX4) or the CAD from either Ca²⁺-bound or Ca²⁺-depleted *Tg*CDPK1. Even with weak anomalous data, an MR-SAD search strategy using the MR solution in *phaser-EP* yielded seven selenium sites of the possible 30 (2 copies/asu in P2₁). Only one of these sites was in the placed kinase domains. We took advantage of the fourteen known methionine positions in the two placed kinase domains by inserting selenium atoms at the side chain sulfur positions. Inputting these sites into the

subsequent *phaser-EP* run resulted in the deletion of only a single site, the XYZ, B-factor, and occupancy refinement of all other input sites, and the identification of five new sites in the CAD. Application of the known non-crystallographic symmetry (NCS) operator to the selenium sites resulted in a total of 23 sites. After solvent flattening and NCS map averaging in *parrot*, available through the *CCP4* suite (Winn et al., 2011), clear helical density could be seen corresponding to the missing CAD. Iterative rounds of model building in *Coot* (Emsley et al., 2010), *phaser-EP* map calculation, and *parrot* density modification were performed until the map figure of merit was greater than 0.5. *Phenix.refine* (Adams et al., 2010) was used for model refinement in the SeMet dataset until it was clear that experimental phase restraints were hindering further model improvement. The partially built and refined model was placed into the native P1 unit cell (4 copies/asu) using *phaser-MR*. Sequence assignment was aided by the use of Se sites as positional markers and structural homology to the high resolution, active CAD structure (RMSD: 3.2 Å, 137 Ca's aligned). Iterative building in *Coot* and refinement in *Phenix.refine* were performed in the native dataset to obtain the final, fully refined structure.

Structure analysis

The CDPK1:1B7 interface was analyzed using PDBePISA (Krissinel and Henrick, 2007). A multiple sequence alignment of *Tg*CDPK1 homologs was created using ClustalW in JalView (Waterhouse et al., 2009).

Molecular dynamics simulations

Simulated systems were modeled from the x-ray crystal structure of active, Ca²⁺-bound *Tg*CDPK1 (PDB: 3HX4) (Wernimont et al., 2010). *Tg*CDPK1 with the Ca²⁺-bound CAD present (+CAD, residues 30-503) or removed (-CAD, residues 30-314) were solvated with 150 mM NaCl in cubic simulation boxes of 93 Å or 75 Å per side, respectively. There were approximately 81,000 atoms in each +CAD simulation and 43,000 atoms in each -CAD simulation. The

systems were parameterized using the Amber ff99SB-ILDN (Cornell et al., 1995; Hornak et al., 2006; Lindorff-Larsen et al., 2010) force field, combined with the ff99SB* backbone correction (Best and Hummer, 2009) for proteins, and a TIP3P water model (Jorgensen et al., 1983). After parameterization, the systems were energy-minimized and then equilibrated in a 10 ns simulation with a NPT ensemble at 310°K and 1 bar. In the equilibration, 1 fs time step was used and harmonic positional restraints of a force constant of 1 kcal/mol/Å² were initially applied to the protein backbone atoms. Production simulations were performed starting from the last frame of equilibration simulations with an NPT ensemble at 310°K and 1 bar using the Nosé-Hoover thermostat (Hoover, 1985) with a time constant of 0.05 ps. All hydrogen-containing bonds were constrained using an implementation (Lippert et al., 2007) of the M-SHAKE algorithm (Kräutler et al., 2001). A reversible reference system propagator algorithm (r-RESPA) integrator (Tuckerman et al., 1992) was used. Bonded, Van der Waals, and short-range electrostatic interactions were computed every 2 fs, whereas long-range electrostatic interactions were computed every 6 fs. The short-range electrostatic interactions were calculated at a cutoff of 9.0 Å, and the long-range electrostatic interactions were computed using Gaussian-split Ewald (Shan et al., 2005). All simulations were carried out on either Anton or Anton 2, special purpose supercomputers designed for MD (Shaw et al., 2014).

PDB accession code

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 4YGA.

Acknowledgements

We thank Majida El Bakkouri and Raymond Hui at the Structural Genomics Consortium for providing us with the *Pf*CDPK1 protein and helpful advice, Manoj T. Duraisingh for providing malaria lysates, Charles Shoemaker and Jacqueline Tremblay at Tufts Cummings School of Veterinary Medicine for their assistance with alpaca immunizations and library construction, and Eric Spooner at the Whitehead Institute Proteomics Facility for technical support. The X-ray diffraction data was collected at the APS NE-CAT beamlines supported by the National Institute of General Medical Sciences, NIH (award GM103403). Use of the APS is supported by the US Department of Energy, Office of Basic Energy Sciences (contract DE-AC02-06CH11357). K.E.K received fellowship support from the NIH (grant T32GM007287) and the NSF Graduate Research Fellowship (grant 1122374). B.M.M. received stipend support from the DAAD (PROMOS) under the sponsorship of Christoph Borner at the Institute for Molecular Medicine and Cell Research, and with support from Damaris Braun both at the University of Freiburg. This work was supported in part by the NIH Director's Early Independence Award (grant 1DP5OD017892) to S.L.

References

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 213–221.
- Ahmed, A., Gaadhe, K., Sharma, G.P., Kumar, N., Neculai, M., Hui, R., Mohanty, D., and Sharma, P. (2012). Novel insights into the regulation of malarial calcium-dependent protein kinase 1. *FASEB J* *26*, 3212–3221.
- Andersen, K.R., Leksa, N.C., and Schwartz, T.U. (2013). Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* *81*, 1857–1861.
- Azevedo, M.F., Sanders, P.R., Krejany, E., Nie, C.Q., Fu, P., Bach, L.A., Wunderlich, G., Crabb, B.S., and Gilson, P.R. (2013). Inhibition of *Plasmodium falciparum* CDPK1 by conditional expression of its J-domain demonstrates a key role in schizont development. *Biochem J* *452*,

433–441.

Baer, K., Al-Hasani, H., Parvaresch, S., Corona, T., Rufer, A., Nölle, V., Bergschneider, E., and Klein, H.W. (2001). Dimerization-induced activation of soluble insulin/IGF-1 receptor kinases: an alternative mechanism of activation. *Biochemistry* *40*, 14268–14278.

Bers, D.M., Patton, C.W., and Nuccitelli, R. (2010). A practical guide to the preparation of Ca(2+) buffers. *Methods Cell Biol.* *99*, 1–26.

Best, R.B., and Hummer, G. (2009). Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* *113*, 9004–9015.

Castellanos-Gonzalez, A., White, A.C., Ojo, K.K., Vidadala, R.S.R., Zhang, Z., Reid, M.C., Fox, A.M.W., Keyloun, K.R., Rivas, K., Irani, A., et al. (2013). A novel calcium-dependent protein kinase inhibitor as a lead compound for treating cryptosporidiosis. *J Infect Dis* *208*, 1342–1348.

Chandran, V., Stollar, E.J., Lindorff-Larsen, K., Harper, J.F., Chazin, W.J., Dobson, C.M., Luisi, B.F., and Christodoulou, J. (2006). Structure of the regulatory apparatus of a calcium-dependent protein kinase (CDPK): a novel mode of calmodulin-target recognition. *J Mol Biol* *357*, 400–410.

Christodoulou, J., Malmendal, A., Harper, J.F., and Chazin, W.J. (2004). Evidence for differing roles for each lobe of the calmodulin-like domain in a calcium-dependent protein kinase. *J Biol Chem* *279*, 29092–29100.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* *117*, 5179–5197.

De Genst, E., Silence, K., Decanniere, K., Conrath, K., Loris, R., Kinne, J., Muyldermans, S., and Wyns, L. (2006). Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc Natl Acad Sci USA* *103*, 4586–4591.

De Meyer, T., Muyldermans, S., and Depicker, A. (2014). Nanobody-based products as research and diagnostic tools. *Trends in Biotechnology* *32*, 263–270.

Dobrowolski, J.M., Carruthers, V.B., and Sibley, L.D. (1997). Participation of myosin in gliding motility and host cell invasion by *Toxoplasma gondii*. *Mol Microbiol* *26*, 163–173.

Donald, R.G., and Roos, D.S. (1993). Stable molecular transformation of *Toxoplasma gondii*: a selectable dihydrofolate reductase-thymidylate synthase marker based on drug-resistance mutations in malaria. *Proc Natl Acad Sci USA* *90*, 11703–11707.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 486–501.

Filippakopoulos, P., Kofler, M., Hantschel, O., Gish, G.D., Grebien, F., Salah, E., Neudecker, P., Kay, L.E., Turk, B.E., Superti-Furga, G., et al. (2008). Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. *Cell* *134*, 793–803.

Garrison, E., Treeck, M., Ehret, E., Butz, H., Garbuz, T., Oswald, B.P., Settles, M., Boothroyd, J., and Arrizabalaga, G. (2012). A forward genetic screen reveals that calcium-dependent protein kinase 3 regulates egress in *Toxoplasma*. *PLoS Pathog* 8, e1003049.

Harmon, A.C., Yoo, B.C., and McCaffery, C. (1994). Pseudosubstrate inhibition of CDPK, a protein kinase with a calmodulin-like domain. *Biochemistry* 33, 7278–7287.

Harper, J.F., Huang, J.F., and Lloyd, S.J. (1994). Genetic identification of an autoinhibitor in CDPK, a protein kinase with a calmodulin-like domain. *Biochemistry* 33, 7267–7277.

Hoover, W. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79, 926–935.

Jura, N., Zhang, X., Endres, N.F., Seeliger, M.A., Schindler, T., and Kuriyan, J. (2011). Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Mol Cell* 42, 9–22.

Kräutler, V., van Gunsteren, W.F., and Hünenberger, P.H. (2001). A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* 22, 501–508.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774–797.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–1958.

Lippert, R.A., Bowers, K.J., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossvary, I., and Shaw, D.E. (2007). A common, avoidable source of error in molecular dynamics integrators. *The Journal of Chemical Physics* 126, 046101.

Lourido, S., and Moreno, S.N.J. (2015). The calcium signaling toolkit of the Apicomplexan parasites *Toxoplasma gondii* and *Plasmodium* spp. *Cell Calcium* 57, 186–193.

Lourido, S., Jeschke, G.R., Turk, B.E., and Sibley, L.D. (2013). Exploiting the unique ATP-binding pocket of *Toxoplasma* calcium-dependent protein kinase 1 to identify its substrates. *ACS Chem Biol* 8, 1155–1162.

Lourido, S., Shuman, J., Zhang, C., Shokat, K.M., Hui, R., and Sibley, L.D. (2010). Calcium-dependent protein kinase 1 is an essential regulator of exocytosis in *Toxoplasma*. *Nature* 465,

359–362.

Lourido, S., Tang, K., and Sibley, L.D. (2012). Distinct signalling pathways control *Toxoplasma* egress and host-cell invasion. *Embo J* 31, 4524–4534.

Maass, D.R., Sepulveda, J., Pernthaler, A., and Shoemaker, C.B. (2007). Alpaca (*Lama pacos*) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). *J. Immunol. Methods* 324, 13–25.

McCoy, J.M., Whitehead, L., van Dooren, G.G., and Tonkin, C.J. (2012). TgCDPK3 Regulates Calcium-Dependent Egress of *Toxoplasma gondii* from Host Cells. *PLoS Pathog* 8, e1003066.

Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. *Elife* 2, e01456.

Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., and Kuriyan, J. (2003). Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* 112, 859–871.

Ojo, K.K., Larson, E.T., Keyloun, K.R., Castaneda, L.J., Derocher, A.E., Inampudi, K.K., Kim, J.E., Arakaki, T.L., Murphy, R.C., Zhang, L., et al. (2010). *Toxoplasma gondii* calcium-dependent protein kinase 1 is a target for selective kinase inhibitors. *Nat Struct Mol Biol* 17, 602–607.

Ojo, K.K., Pfander, C., Mueller, N.R., Burstroem, C., Larson, E.T., Bryan, C.M., Fox, A.M.W., Reid, M.C., Johnson, S.M., Murphy, R.C., et al. (2012). Transmission of malaria to mosquitoes blocked by bumped kinase inhibitors. *J. Clin. Invest.* 122, 2301–2305.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology*, C.J. Carter, and R.M. Sweet, eds. (New York: Elsevier), pp. 307–326.

Ranjan, R., Ahmed, A., Gourinath, S., and Sharma, P. (2009). Dissection of mechanisms involved in the regulation of *Plasmodium falciparum* calcium-dependent protein kinase 4. *J Biol Chem* 284, 15267–15276.

Rellos, P., Pike, A.C.W., Niesen, F.H., Salah, E., Lee, W.H., Delft, von, F., and Knapp, S. (2010). Structure of the CaMKII δ /calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biol* 8, e1000426.

Rosenberg, O.S., Deindl, S., Sung, R.-J., Nairn, A.C., and Kuriyan, J. (2005). Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell* 123, 849–860.

Shan, Y., Eastwood, M.P., Zhang, X., Kim, E.T., Arkhipov, A., Dror, R.O., Jumper, J., Kuriyan, J., and Shaw, D.E. (2012). Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* 149, 860–870.

Shan, Y., Klepeis, J.L., Eastwood, M.P., Dror, R.O., and Shaw, D.E. (2005). Gaussian split

Ewald: A fast Ewald mesh method for molecular simulation. *The Journal of Chemical Physics* 122, 054101.

Shaw, D.E., Grossman, J.P., Bank, J.A., Batson, B., Butts, J.A., Chao, J.C., Deneroff, M.M., Dror, R.O., Even, A., Fenton, C.H., et al. (2014). Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer (IEEE Press).

Taylor, S.S., Keshwani, M.M., Steichen, J.M., and Kornev, A.P. (2012). Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 367, 2517–2528.

Tuckerman, M., Berne, B.J., and Martyna, G.J. (1992). Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* 97, 1990–2001.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.

Wernimont, A.K., Artz, J.D., Finerty, P., Lin, Y.-H., Amani, M., Allali-Hassani, A., Senisterra, G., Vedadi, M., Tempel, W., Mackenzie, F., et al. (2010). Structures of apicomplexan calcium-dependent protein kinases reveal mechanism of activation by calcium. *Nat Struct Mol Biol* 17, 596–601.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242.

Yu, J., Zhang, Y., McIlroy, J., Rordorf-Nikolic, T., Orr, G.A., and Backer, J.M. (1998). Regulation of the p85/p110 phosphatidylinositol 3'-kinase: stabilization and inhibition of the p110alpha catalytic subunit by the p85 regulatory subunit. *Mol Cell Biol* 18, 1379–87.

Zhao, Y., Pokutta, S., Maurer, P., Lindt, M., Franklin, R.M., and Kappes, B. (1994). Calcium-binding properties of a calcium-dependent protein kinase from *Plasmodium falciparum* and the significance of individual calcium-binding sites for kinase activation. *Biochemistry* 33, 3714–3721.

Zhou, H.-X. (2004). Loops, Linkages, Rings, Catenanes, Cages, and Crowders: Entropy-Based Strategies for Stabilizing Proteins. *Acc. Chem. Res.* 37, 123–130.

Tables

Table 1. Mass spectrometric analysis of 1B7 IP

Gene ID	Product Description	MW (kDa)	Coverage	Unique Peptides	Total Spectra
TGGT1_301440	TgCDPK1	57	52%	26	70
TGGT1_305860	TgCDPK3	60	41%	26	80
TGGT1_247550	HSP60	61	36%	15	20
TGGT1_225050	Adenosyl-homocysteinase	52	31%	14	30
TGGT1_211680	Protein disulfide isomerase	53	25%	9	14
TGGT1_266960	Beta-tubulin	50	20%	8	10
TGGT1_270510	Asparaginyl-tRNA synthetase	75	15%	8	12
TGGT1_209030	Actin	42	28%	8	9
TGGT1_311240	DnaJ family chaperone	56	16%	7	7
TGGT1_411430	ROP5	61	13%	5	5
TGGT1_253430	Asparagine synthetase	66	9.20%	5	8
TGGT1_305770	ABC transporter	40	5.80%	3	3
TGGT1_233110	IMPDH	101	4.50%	3	3
TGGT1_290670	Leucyl-aminopeptidase	83	4.70%	3	3
TGGT1_219590	RuvB family 1 protein	53	7.70%	3	3
TGGT1_215775	ROP8	62	4.60%	2	2
TGGT1_228490	Hypothetical protein	43	8.90%	2	2
TGGT1_291960	ROP40	58	5.40%	2	2
TGGT1_290200	NAD/NADP dehydrogenase	48	6.00%	2	4
TGGT1_230450	bifunctional GMP synthase	62	4.00%	2	2
TGGT1_294800B	eIF-1 α	49	5.40%	2	3
TGGT1_268890	Citrate synthase I	60	4.10%	2	2
TGVEG_050040	ROP4/7	64	5.20%	2	2
TGGT1_220950	Hypothetical protein	48	5.60%	2	2
TGGT1_215590	Subunit of succinate dehydrogenase	73	3.30%	2	2

Table 2. Data collection and refinement statistics

	TgCDPK1-1B7 native	TgCDPK1-1B7 SeMet derivative
Data collection		
Space group	P1	P2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	76.92, 91.07, 106.64	77.61, 91.92, 107.90
<i>a</i> , <i>b</i> , γ (°)	88.7, 108.3, 90.3	90.0, 108.5, 90.0
Resolution (Å)	73.04 – 2.94 (3.05 – 2.94) ^a	102.3 – 4.09 (4.24 – 4.09)
<i>R</i> _{sym}	0.11 (0.94)	0.30 (0.74)
<i>R</i> _{p.i.m.}	0.06 (0.54)	0.13 (0.38)
<i>I</i> / σ <i>I</i>	10.3 (1.4)	6.0 (1.8)
CC _{1/2}	0.99 (0.70)	0.99 (0.75)
Completeness (%)	93.5 (94.8)	98.2 (95.6)
Redundancy	3.9 (3.8)	5.8 (4.3)
Refinement		
Resolution (Å)	70.67 – 2.94	
No. reflections	53984	
No. free reflections	1715	
<i>R</i> _{work} / <i>R</i> _{free}	22.6/26.2	
No. atoms	17348	
Protein	17332	
Calcium	16	
<i>B</i> factors		
Protein	113.3	
Calcium	84.4	
r.m.s. deviations		
Bond lengths (Å)	0.003	
Bond angles (°)	0.780	
Ramachandran		
Allowed (%)	94.9	
Favored (%)	4.3	
Outlier (%)	0.8	
Clashscore	9.05	
Molprobability percentile	97 th	

^aValues in parentheses are for highest-resolution shell. One crystal was used for each dataset.

Figures

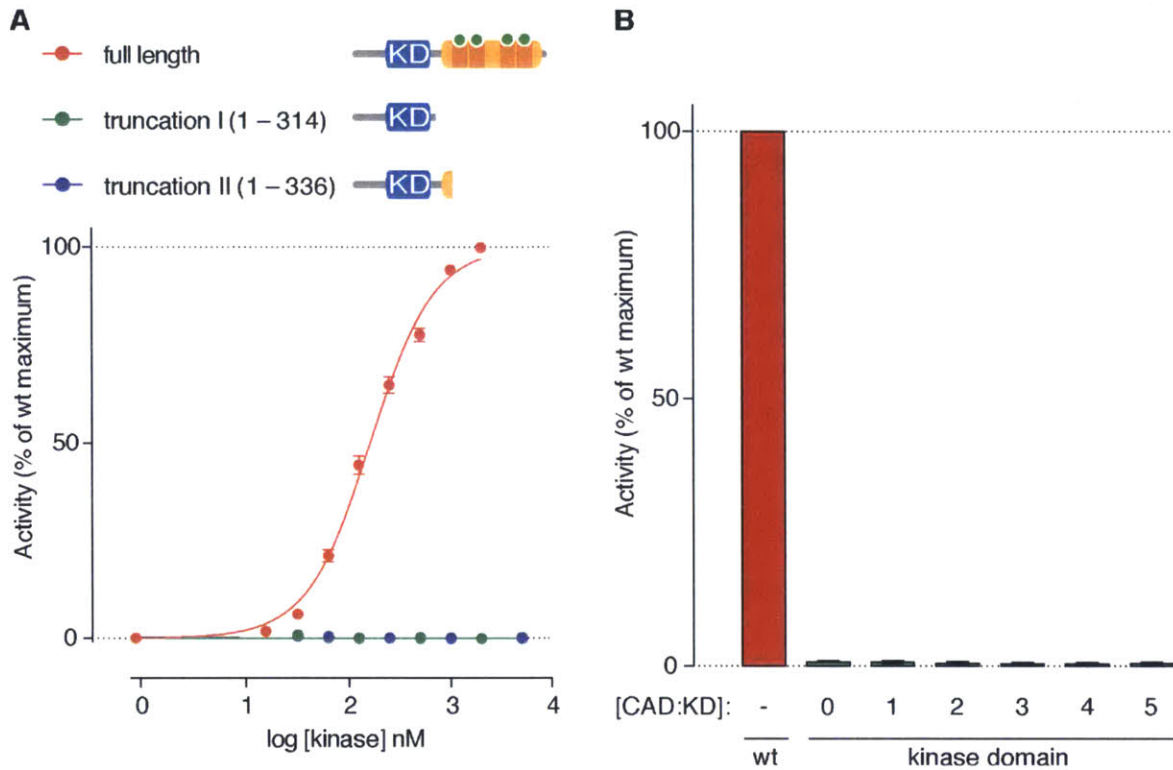


Figure 1. *TgCDPK1* truncations lacking the regulatory domain are catalytically inactive.

(A) Two constructs were expressed and assayed for kinase activity: truncation I lacked the CAD completely, whereas truncation II retained the autoinhibitory α -helix. The constructs' sequence coverage of wt-enzyme amino acids is indicated in parentheses. Values normalized to maximal activity. Mean \pm s.d., representative experiment. (B) The isolated *TgCDPK1* kinase domain cannot be activated by the CAD in trans. The purified kinase domain (KD) yielded no measurable activity even at the 5 μ M tested. Addition of purified CAD to the KD did not lead to any measurable kinase activity. Values normalized to maximum wt activity. Mean \pm s.e.m., $n=3$ independent experiments.

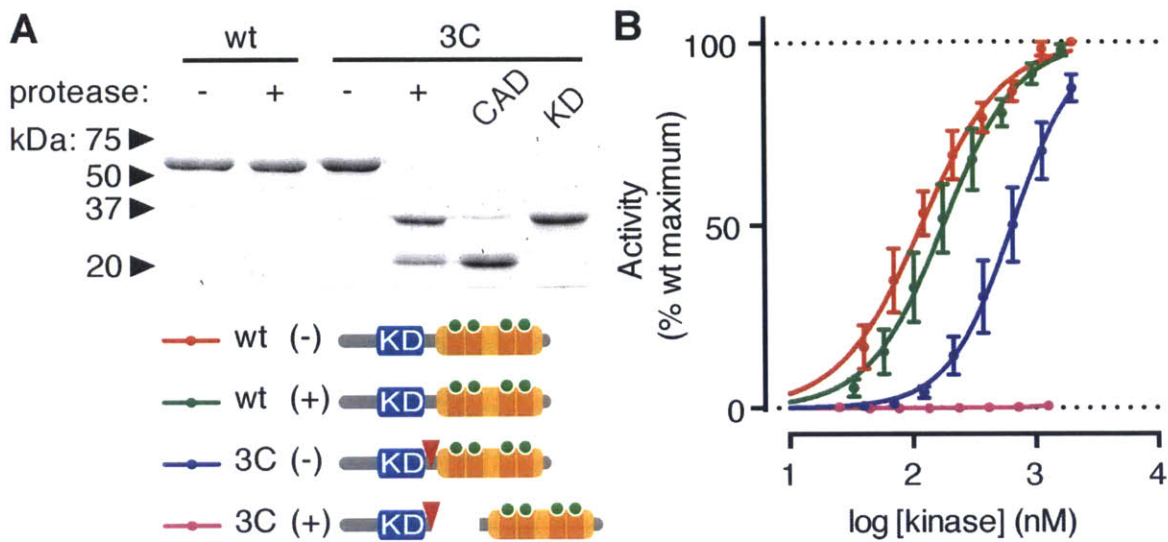


Figure 2. *Tg*CDPK1 is not activated by removal of its regulatory domain. (A) Introduction of a 3C protease cleavage site between the CAD and KD of *Tg*CDPK1, in the kinase referred to as “3C”, enables separation of the two domains upon protease treatment. The wild-type (wt) enzyme was unaffected by the protease treatment. The CAD and KD could be further isolated from the digested 3C. (B) Kinase activity of wt and 3C, before (-) and after (+) protease treatment. Mean \pm s.e.m., $n = 3$ independent experiments.

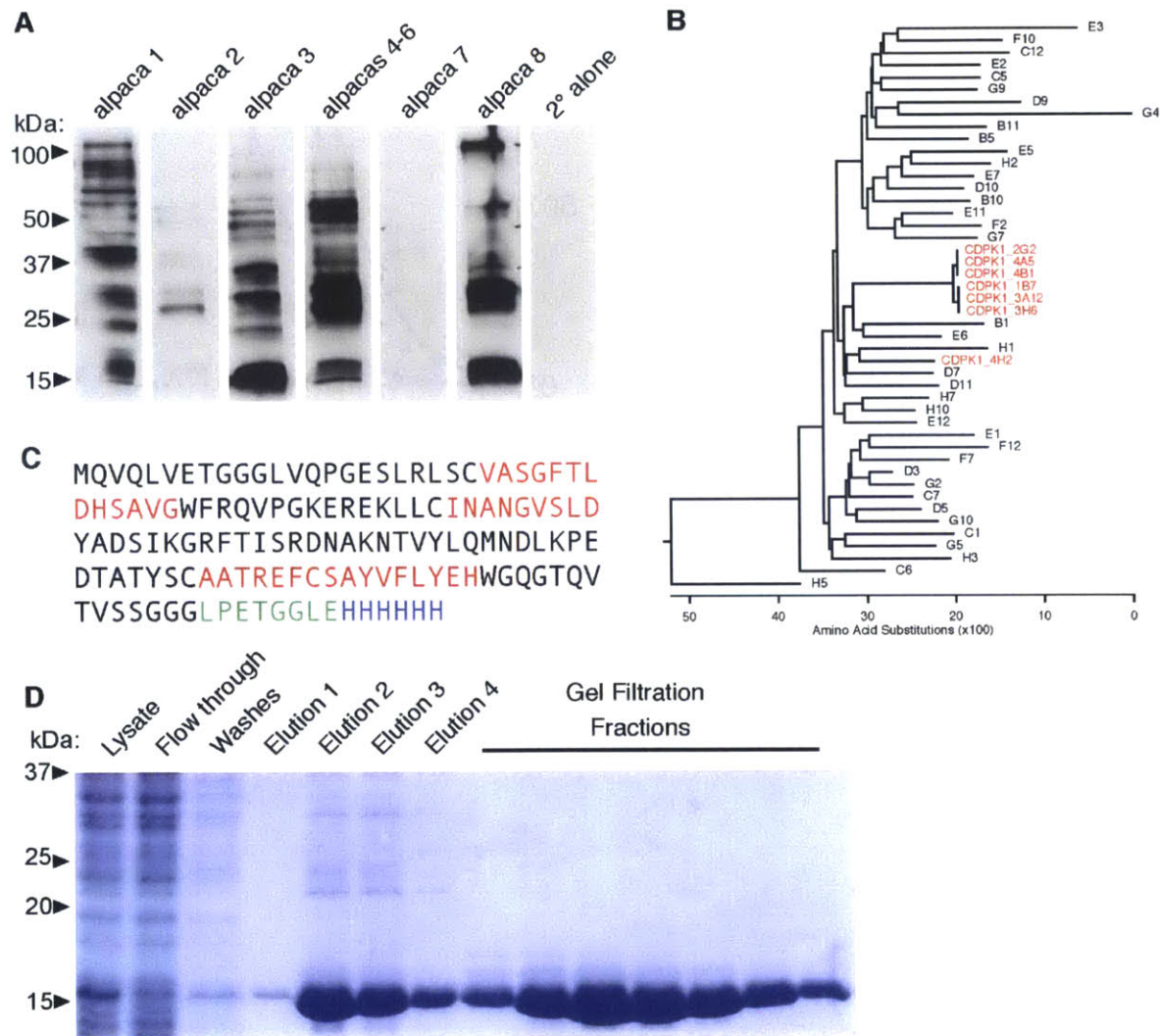


Figure 3. Identification of heavy chain-only antibodies to study the structure of *TgCDPK1*. (A) Individual or pooled serum from alpacas was tested for reactivity against *T. gondii* lysate by immunoblot. The secondary antibody showed no reactivity on its own. (B) Panning of a combined VHH phage display library from alpacas 1, 2, 3, 4 and 8 against immobilized *TgCDPK1* enriched for a group of nearly identical sequences from which 1B7 was selected. Sequences from 96 clones prior to enrichment are shown for comparison. (C) Sequence of 1B7 highlighting the variable loops (red), the sortase substrate motif (green) and the 6xHis-tag (blue). (D) Expression and purification of 1B7. Following the induction of 1B7 expression by IPTG, a lysate was prepared. The lysate was collected on a Ni-NTA column, and a sample of the flow through was collected. The column was washed and 1B7 was eluted with buffer containing 500 mM imidazole. The eluted material was further enriched by SEC, and peak fractions were collected and concentrated. Samples collected during the purification were resolved by SDS-PAGE and stained with Coomassie for total protein.

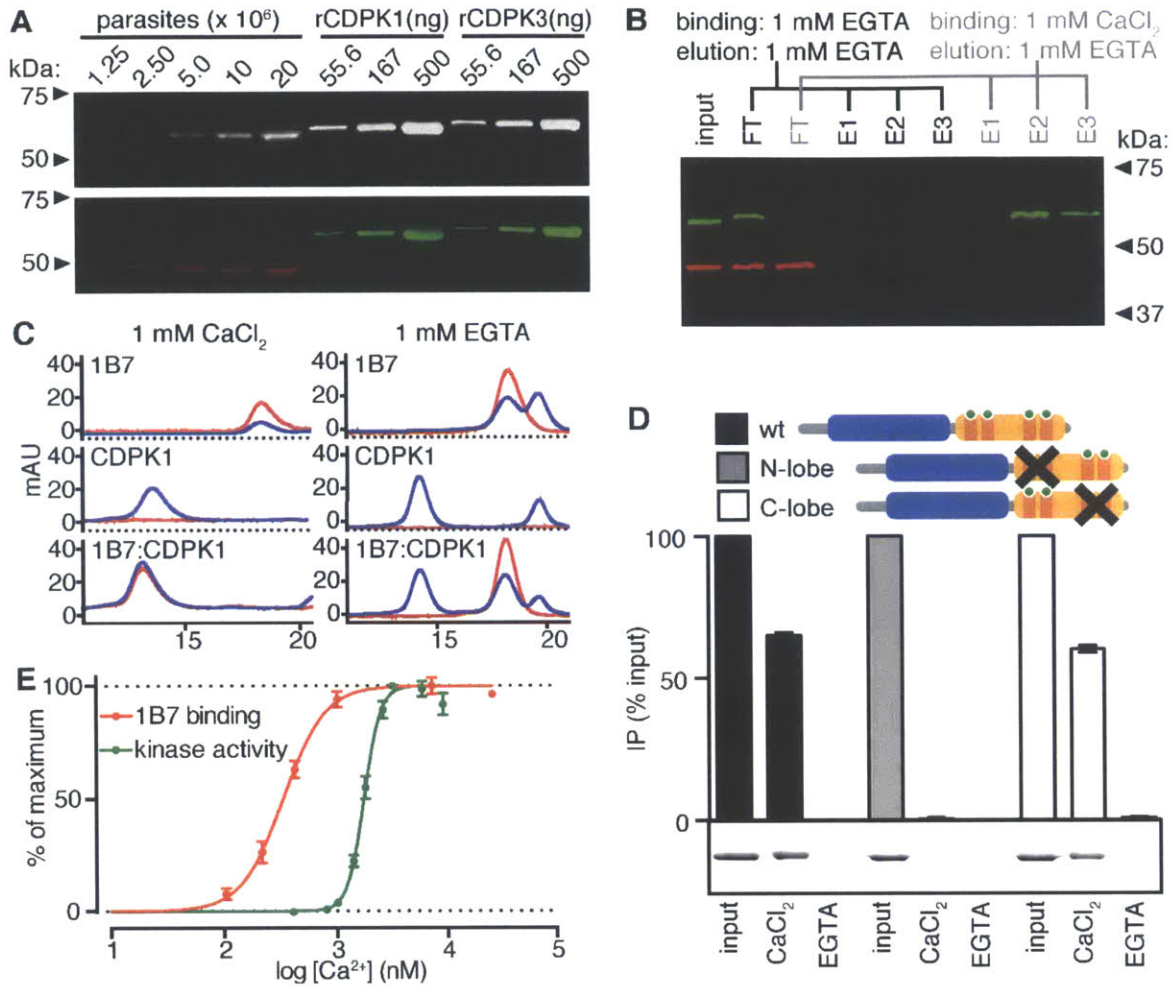


Figure 4. 1B7 recognizes a calcium-dependent conformation of *Tg*CDPK1. (A) IRDye800-labeled 1B7 was used to probe decreasing concentrations of total *T. gondii* lysate or recombinant kinase. Samples were separately probed for parasite actin (ACT1; red) or the 6xHis-tag on recombinant proteins (HIS; green). (B) Incubation of parasite lysates with immobilized 1B7 depletes *Tg*CDPK1/*Tg*CDPK3 in the presence of CaCl₂ but not in EGTA. The immunoprecipitated material could be eluted from the beads by subsequent incubation with EGTA. ACT1 is included as a loading control. (C) Size exclusion chromatography of 1B7-TAMRA and *Tg*CDPK1, alone or incubated at equimolar concentrations, in buffer containing 1 mM CaCl₂ or EGTA. Absorbance recorded at 280 nm for total protein (blue) and 550 nm for 1B7-TAMRA (red). (D) IP of recombinant kinases by 1B7 in the presence of 1 mM CaCl₂ or EGTA. The domain structure of each kinase, indicating the KD and each of the EF hands (I–IV), is depicted for wild type (wt), D368A/D415A (N-lobe mutant), and D451A/D485A (C-lobe mutant). Means ± s.e.m., *n* = 3 independent experiments. Insert shows the results from a representative experiment. (E) Calcium dependency of *Tg*CDPK1 activation or 1B7 binding. Values normalized to maximal activity or binding. Means ± s.e.m., *n* = 3 independent experiments.

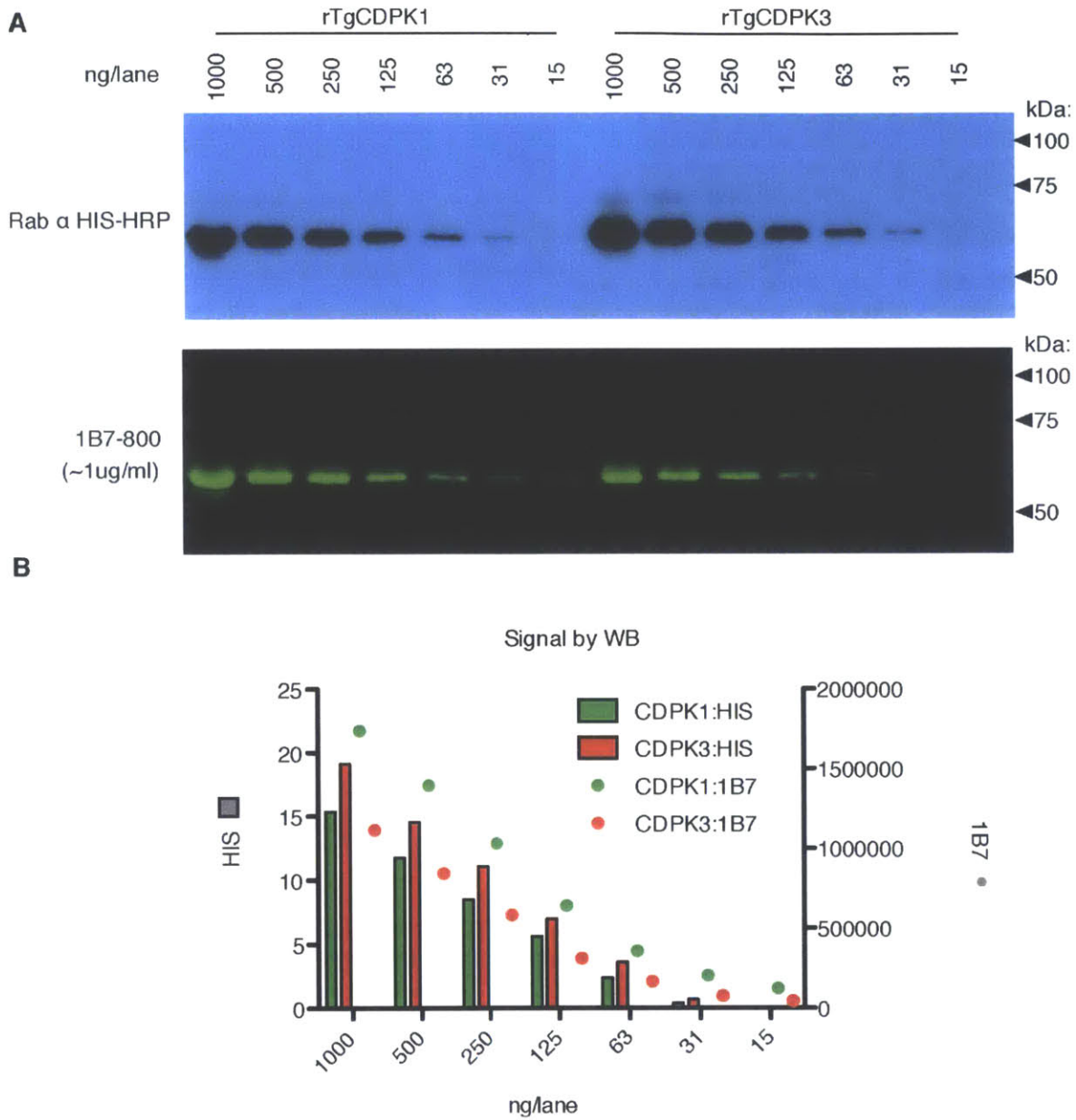


Figure 5. Extended dilution series of recombinant TgCDPKs. (A) Western blots using either IRdye800-labeled 1B7 or an antibody recognizing the 6xHis-tag of the recombinant proteins. (B) Densitometric analysis of western blots showing similar recognition of both kinases by 1B7.

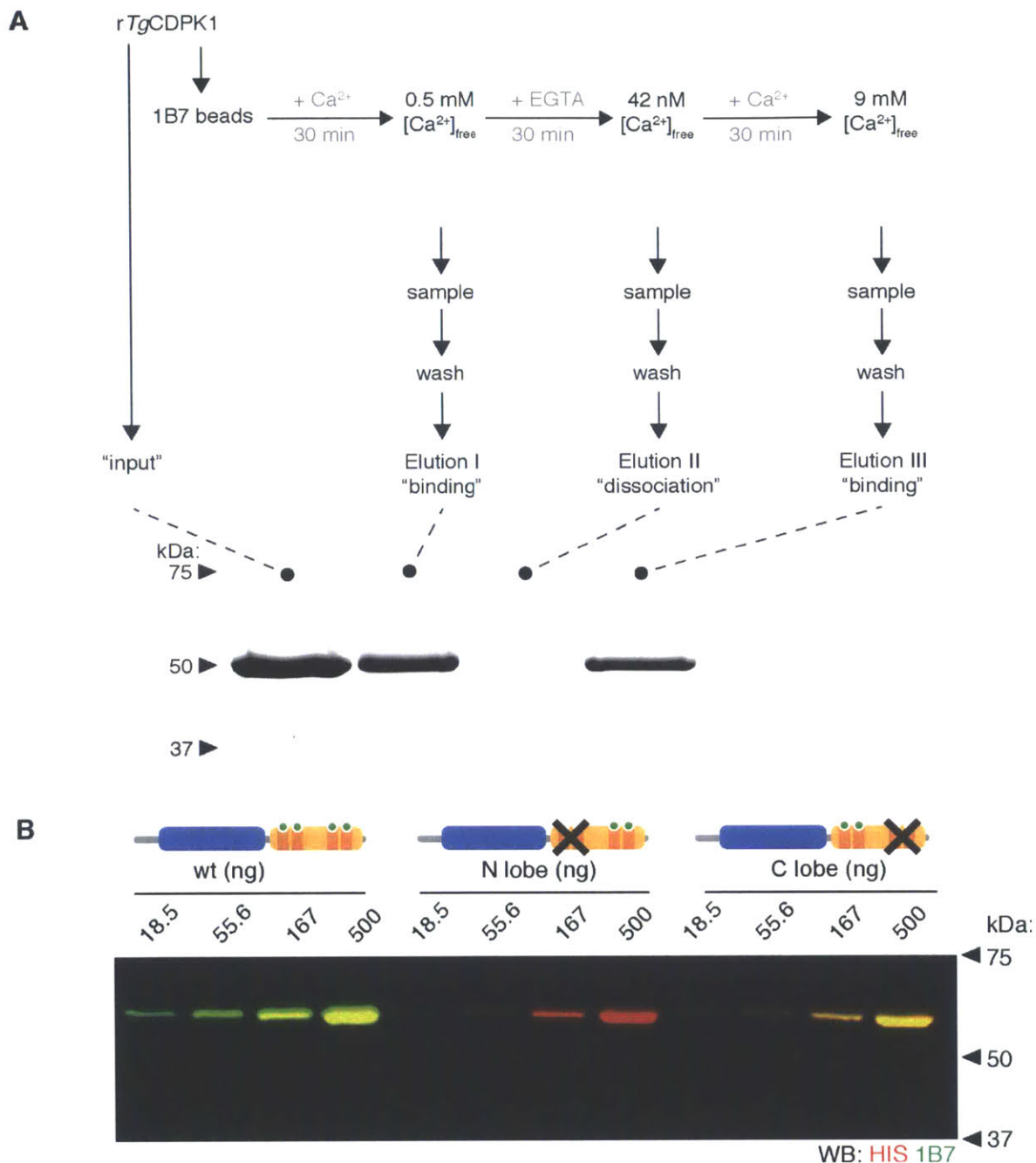


Figure 6. Ca^{2+} -dependency of 1B7 binding. (A) 1B7 binds *TgCDPK1* reversibly. Precipitation of recombinant *TgCDPK1* by covalently-immobilized 1B7. The experimental set-up is diagrammed on top. Kinase was incubated with 1B7-beads and the free $[\text{Ca}^{2+}]$ was modulated by adding EGTA or CaCl_2 , as indicated. Samples were taken after each free $[\text{Ca}^{2+}]$ change to determine whether complex formation had occurred, as measured by the presence of *TgCDPK1* in the eluate. (B) Immunoblot comparing the detection of different kinases by IRDye800-labeled 1B7 (green). The domain structure of each kinase, indicating the KD and each of the EF hands (I–IV), is depicted for wild type (wt), D368A/D415A (N-lobe mutant), and D451A/D485A (C-lobe mutant). Detection of recombinant kinases by their His-tags (red) is included as a loading control.

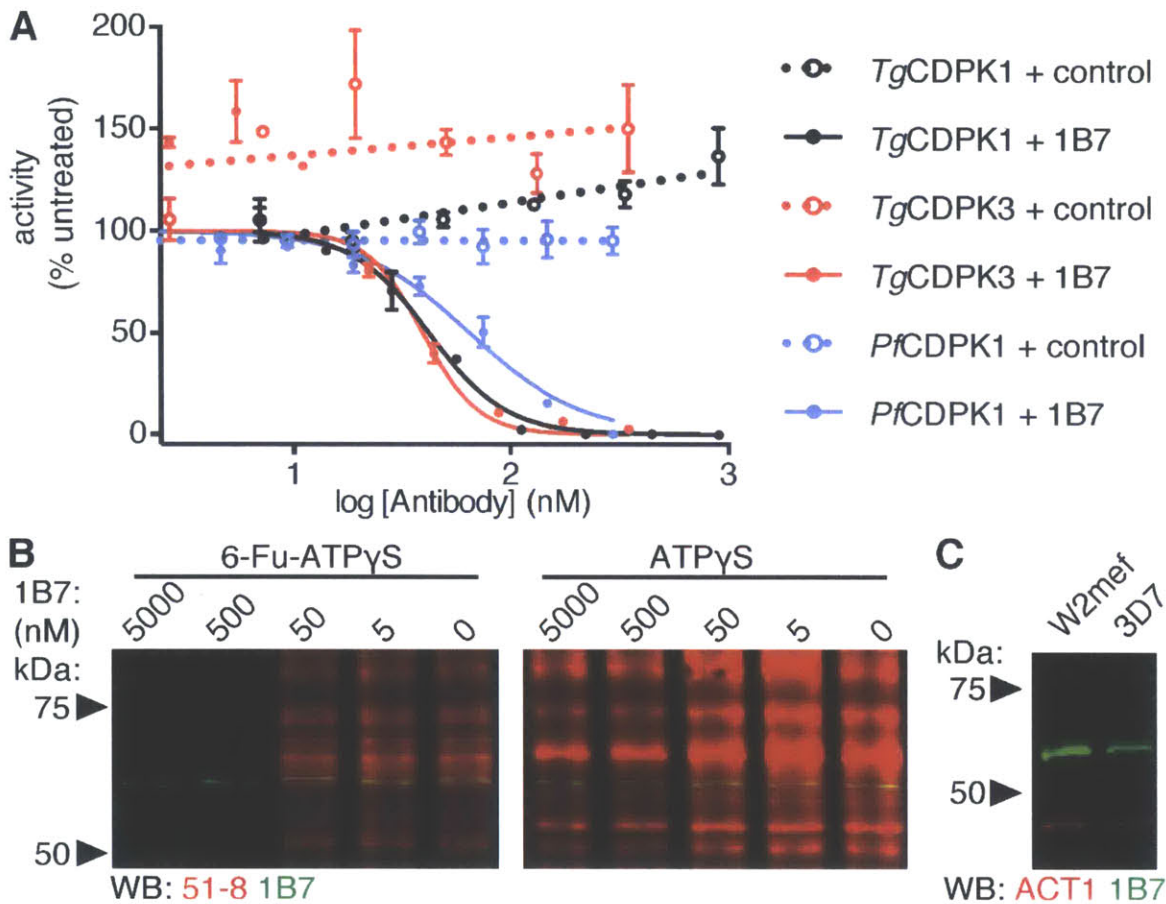


Figure 7. 1B7 inhibits *TgCDPK1* and related kinases. (A) *In vitro* kinase assay with increasing concentrations of 1B7. Mean \pm s.e.m., $n = 3$ independent experiments. (B) *TgCDPK1*-dependent (6-Fu-ATP γ S) and total (ATP γ S) thiophosphorylation in parasite lysates incubated with varying 1B7 concentrations. Thiophosphorylation is visualized by western blot with rabbit mAb 51-8 (red); 1B7 staining is shown as a loading control (green). (C) Immunoblot of *P. falciparum* lysate from strain 3D7 or W2mef probed with 1B7-IRDye800 (green) or an antibody recognizing parasite actin (ACT1; red).

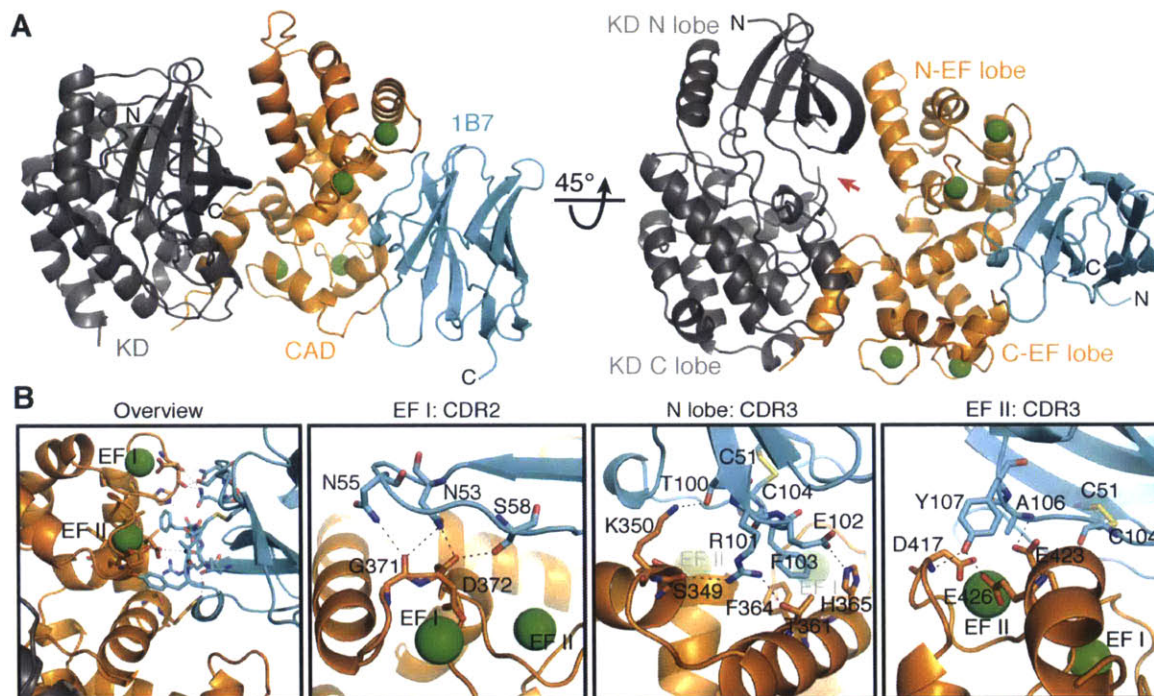


Figure 8. 1B7 stabilizes a novel conformation of *Tg*CDPK1. (A) Cartoon depicting the structure of the complex with the catalytic KD (KD; gray), the calmodulin-like, CDPK activation domain (CAD; orange), bound calcium ions (green spheres) and 1B7 (blue). The position of the active site pocket is indicated (red arrow). (B) Specific interactions within the CAD:1B7 binding interface. Hydrogen bonds and salt bridges are depicted as black dotted lines. Interacting residues are shown as sticks for the CAD and 1B7. Cys51, in the beta strand immediately preceding CDR2, and Cys104, in CDR3, form a stabilizing, intra-chain disulfide in 1B7. CDR1 on 1B7 makes no contacts with *Tg*CDPK1.

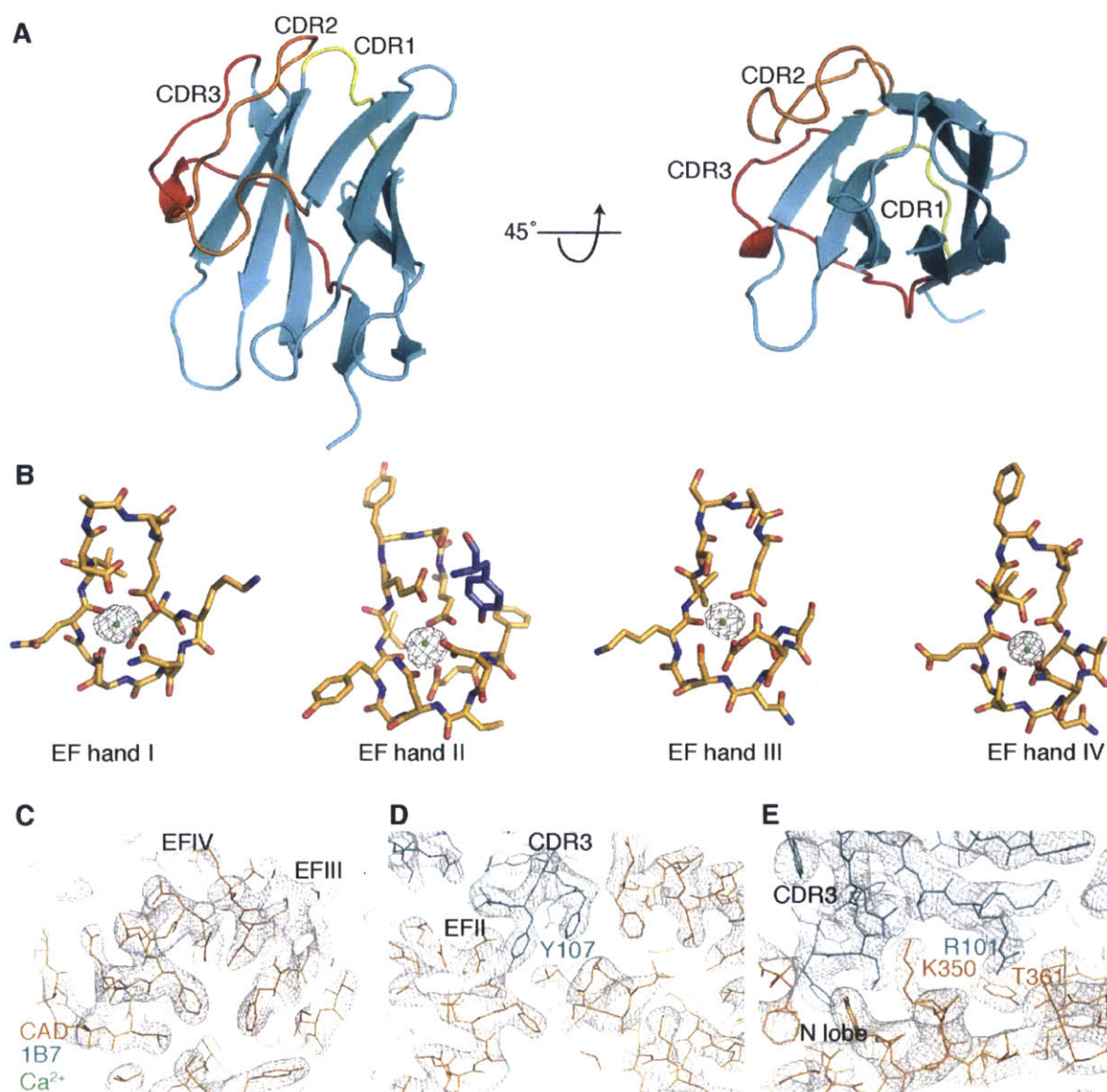


Figure 9. Details of the Ca²⁺/1B7-bound TgCDPK1 structure. (A) Complementarity-determining regions (CDRs) of 1B7. CDR1 (yellow), CDR2 (orange), and CDR3 (red) of 1B7 are colored. The orientation of 1B7 shown is the same as in Figure 4A. (B) Ca²⁺ occupancy of the TgCDPK1 EF hands in the presence of 1B7. Stick representation of the four EF hands that comprise the CAD. The built Ca²⁺ ions are depicted as green spheres. The electron density shown is from a Sigma-A weighted difference (mFo-DFc) omit map for Ca²⁺ contoured at 6 σ . (C–E) Representative images of the final 2mFo-DFc electron density map contoured at 1 σ . 1B7 (teal) and the CAD (orange) are shown as sticks and calcium ions (green) are depicted as dots. (C) View of the hydrophobic core of the CAD C lobe. (D) Interaction site between CDR3 of 1B7 and EF-hand II of the CAD. (E) Interaction site between CDR3 and the N-lobe of the CAD.

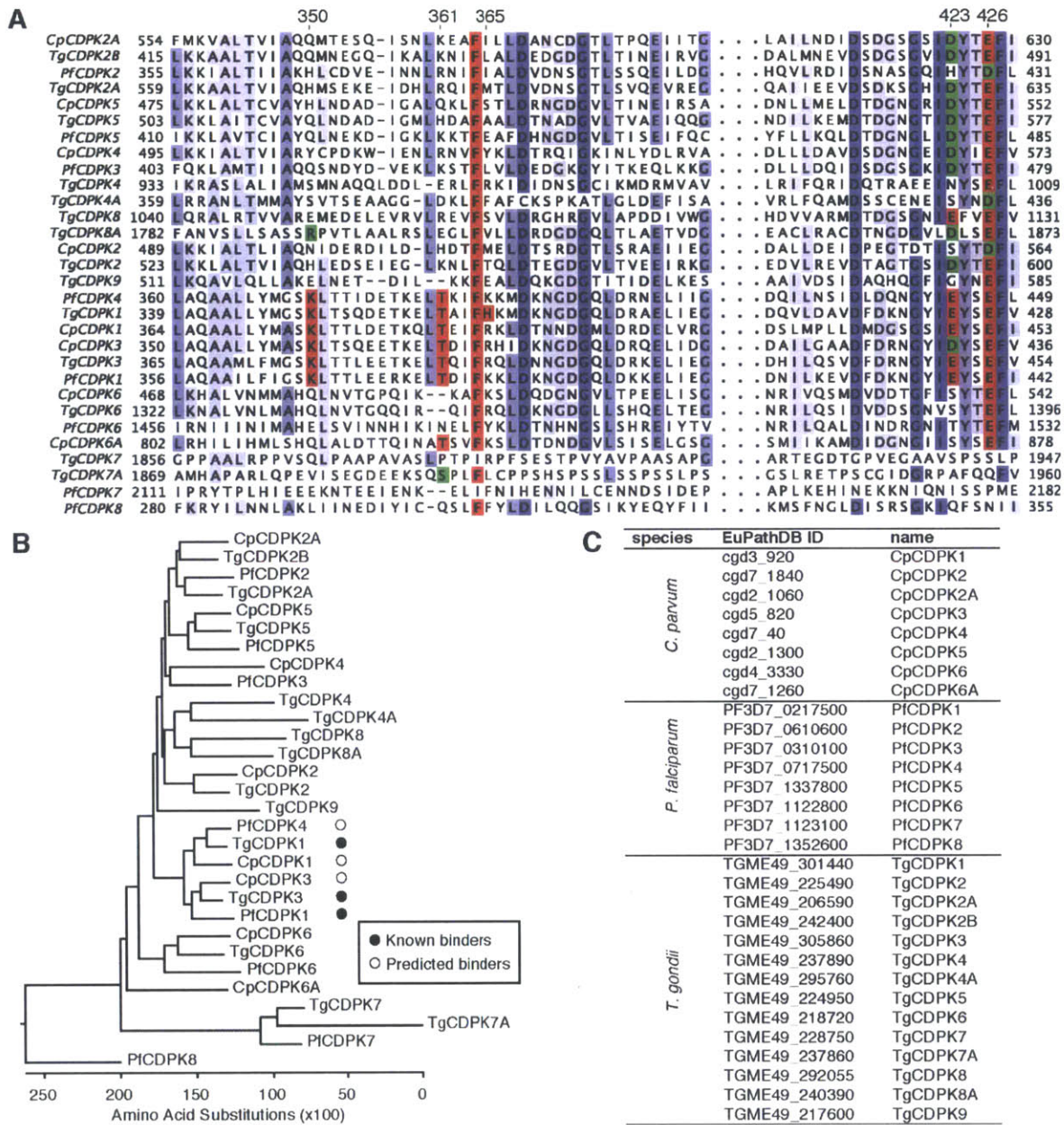


Figure 10. Sequence comparison of Apicomplexan CDPKs. (A) Multiple sequence alignment (MSA) of the 1B7-binding region of the CAD of 30 CDPKs from *T. gondii*, *Plasmodium falciparum* and *Cryptosporidium parvum*. Residues known to interact with 1B7 at the six sequence-specific CAD positions are highlighted in red, and numbered according to their position in *TgCDPK1*. Conservative mutations expected to still interact with 1B7, are highlighted in green. The remaining residues are highlighted white to blue according to sequence identity. Non-conserved insertions omitted for clarity. Phe364 and Glu426 are conserved in nearly all CDPKs. His365 is only present in *TgCDPK1*. Glu423 or the conservative aspartate mutant are present in most CDPKs. Positions 350 and 361 are considered the important sequence determinants. Although non-1B7 binding CDPKs may have a conservative mutation (*i.e.* K350R or T361S) at an individual position, none of these homologs have the correct identity at both of the sites. (B) Phylogenetic tree of *TgCDPK1* homologs. (C) Table of sequences used.

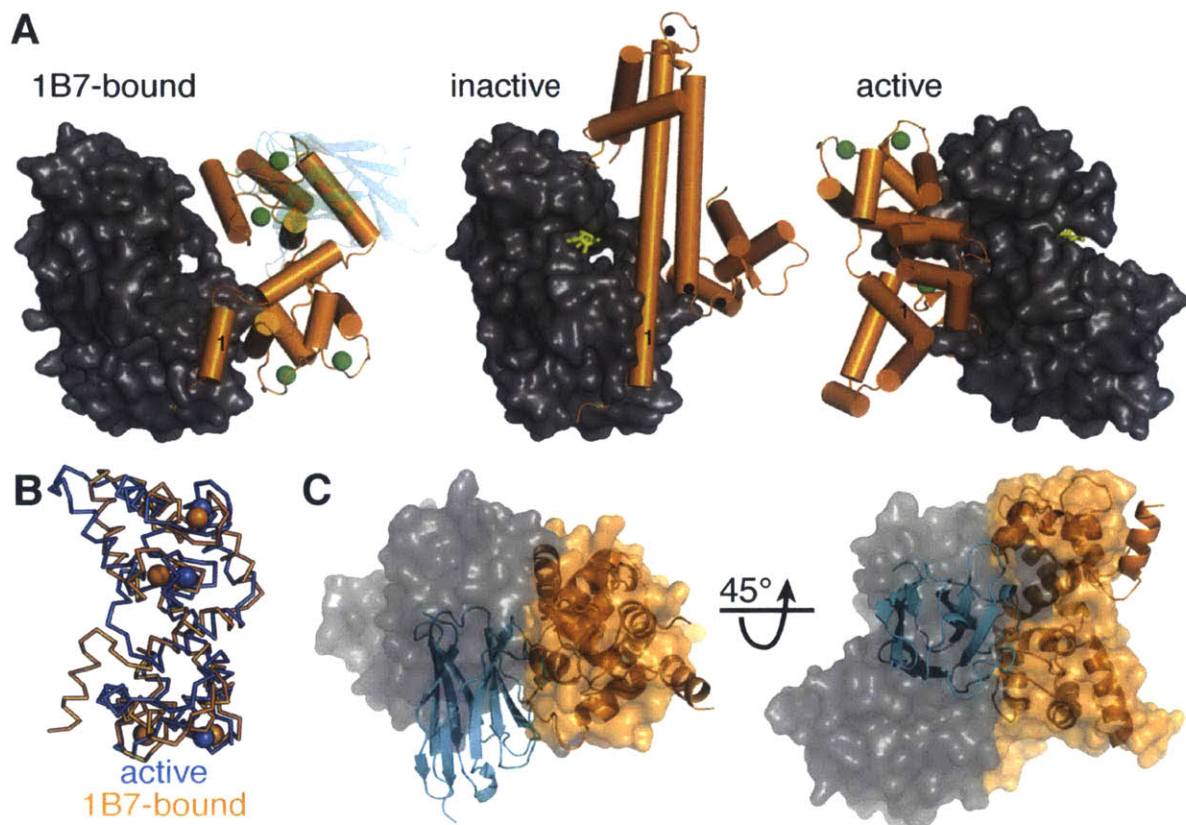


Figure 11. 1B7 prevents reorganization of the CAD to its activation-associated conformation. (A) Comparison of 1B7-bound *TgCDPK1* to the inactive (3KU2) and active (3HX4) conformations. The three *TgCDPK1* conformations are aligned by their KDs (gray), depicting the CAD (orange), 1B7 (blue), bound calcium (green) and unknown cations (black). ANP (*phosphoaminophosphonic acid-adenylate ester*) is built as yellow sticks to indicate the active site. The first helix of the CAD (CH1) is labeled with “1”. (B) Aligned carbon trace of the CADs from the active (blue) and 1B7-bound structures (orange). (C) Superposition of 1B7 and the bound CAD as ribbon diagrams over a space-filling model of the active structure.

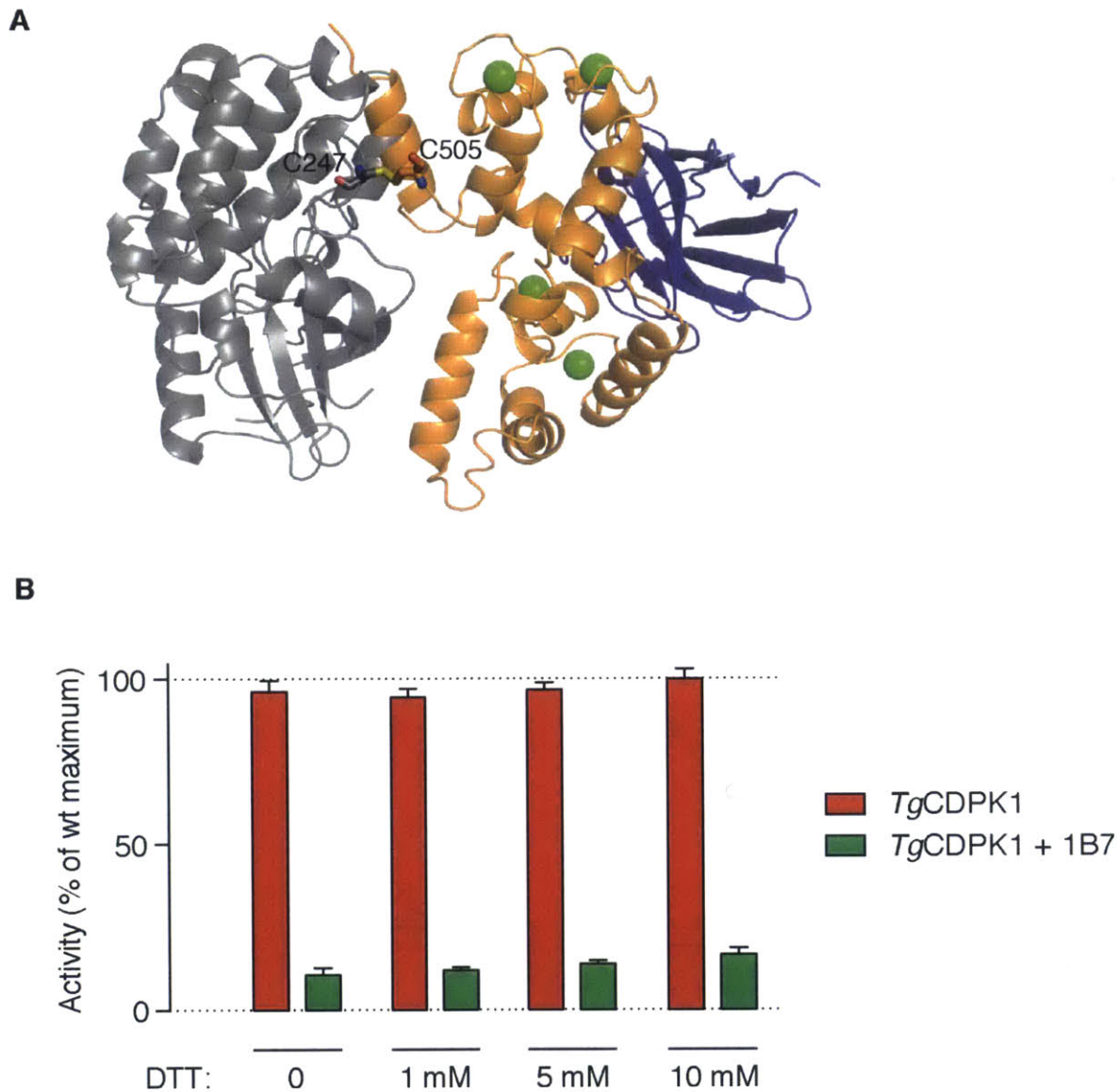


Figure 12. A disulfide bond stabilizes the structure of Ca^{2+} /1B7-bound *TgCDPK1*. (A) Disulfide bond (yellow) between Cys247 of the kinase domain (gray) and Cys505 of the CAD (orange) rendered as sticks in the Ca^{2+} /1B7-bound *TgCDPK1* structure. (B) 1B7 inhibition of *TgCDPK1* is not affected by reducing conditions. Addition of up to 10 mM DTT did not affect kinase activity or inhibition by 1B7. Kinase and 1B7 were assayed at 100 nM and 150 nM, respectively. Values normalized to maximal wt activity. Mean \pm s.d., representative experiment.

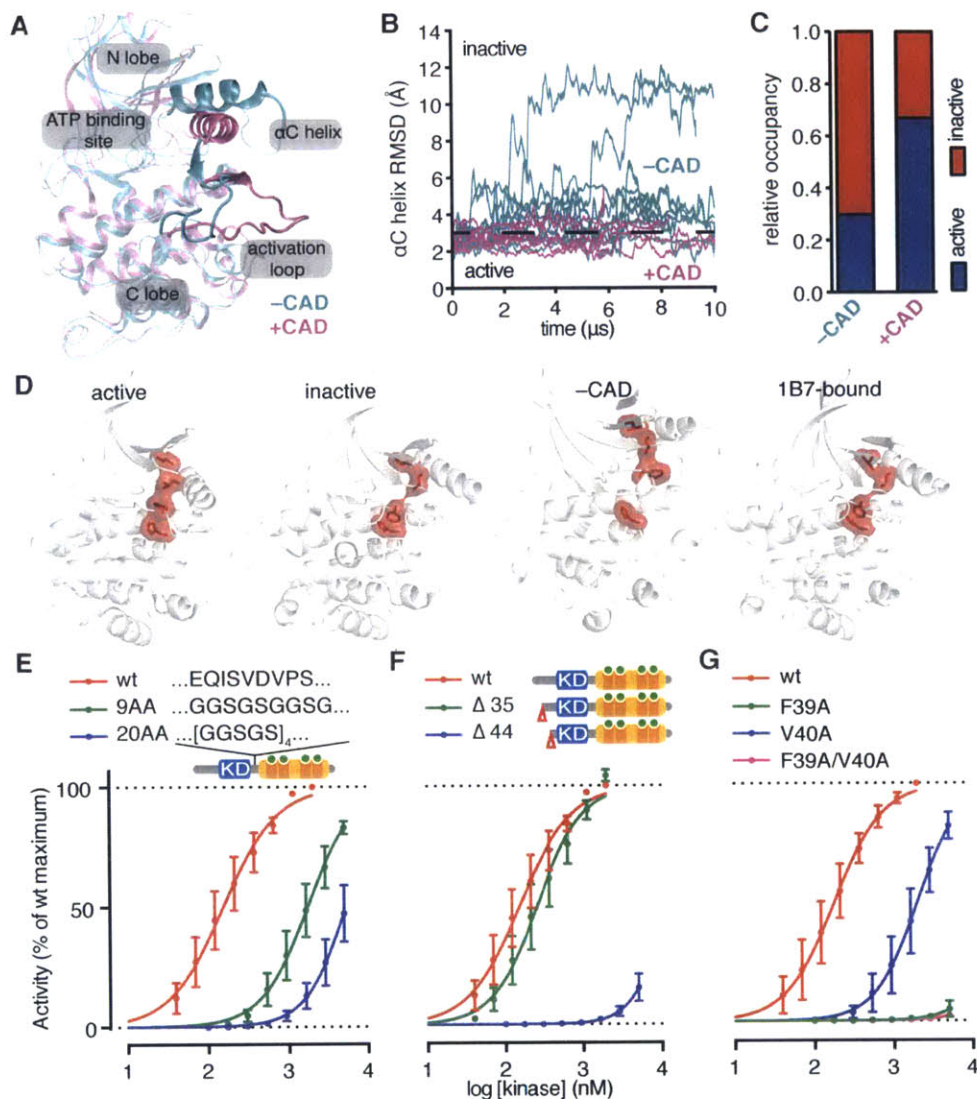


Figure 13. The KD of *TgCDPK1* is intrinsically inactive and stabilized by the Ca^{2+} -bound CAD. (A) Superposition of the active conformation of the *TgCDPK1* KD with an inactive conformation obtained from a simulation without the CAD, in which the KD readily departed from the (starting) active conformation. The superposition is obtained using the C-lobe backbone atoms. (B) RMSD trace of the α C helix relative to the (starting) active conformation in simulations of the *TgCDPK1* KD with or without the CAD. Conformations with RMSD less than 3Å are considered active. (C) The relative occupancy of the catalytically active conformation in simulations of the *TgCDPK1* KD with or without the CAD. As shown, the presence of the CAD stabilizes the active conformation. (D) Structure of the KD highlighting the integrity of the regulatory spine (red) in the Ca^{2+} -bound (active) conformation, and its disruption in the Ca^{2+} -depleted (inactive), the simulation without the CAD (-CAD), and the Ca^{2+} /1B7-bound (1B7-bound) conformations. (E) Kinase activity of *TgCDPK1* mutants carrying different flexible linkers between the KD and CAD domains. (F) Kinase activity of different N-terminal truncations removing the first 35 ($\Delta 35$) or 44 ($\Delta 44$) amino acids preceding the KD. (G) Kinase activity of point mutants in the N-terminal extension, alone or in combination. All kinase experiments plotted as means \pm s.e.m., $n = 3$ independent experiments.

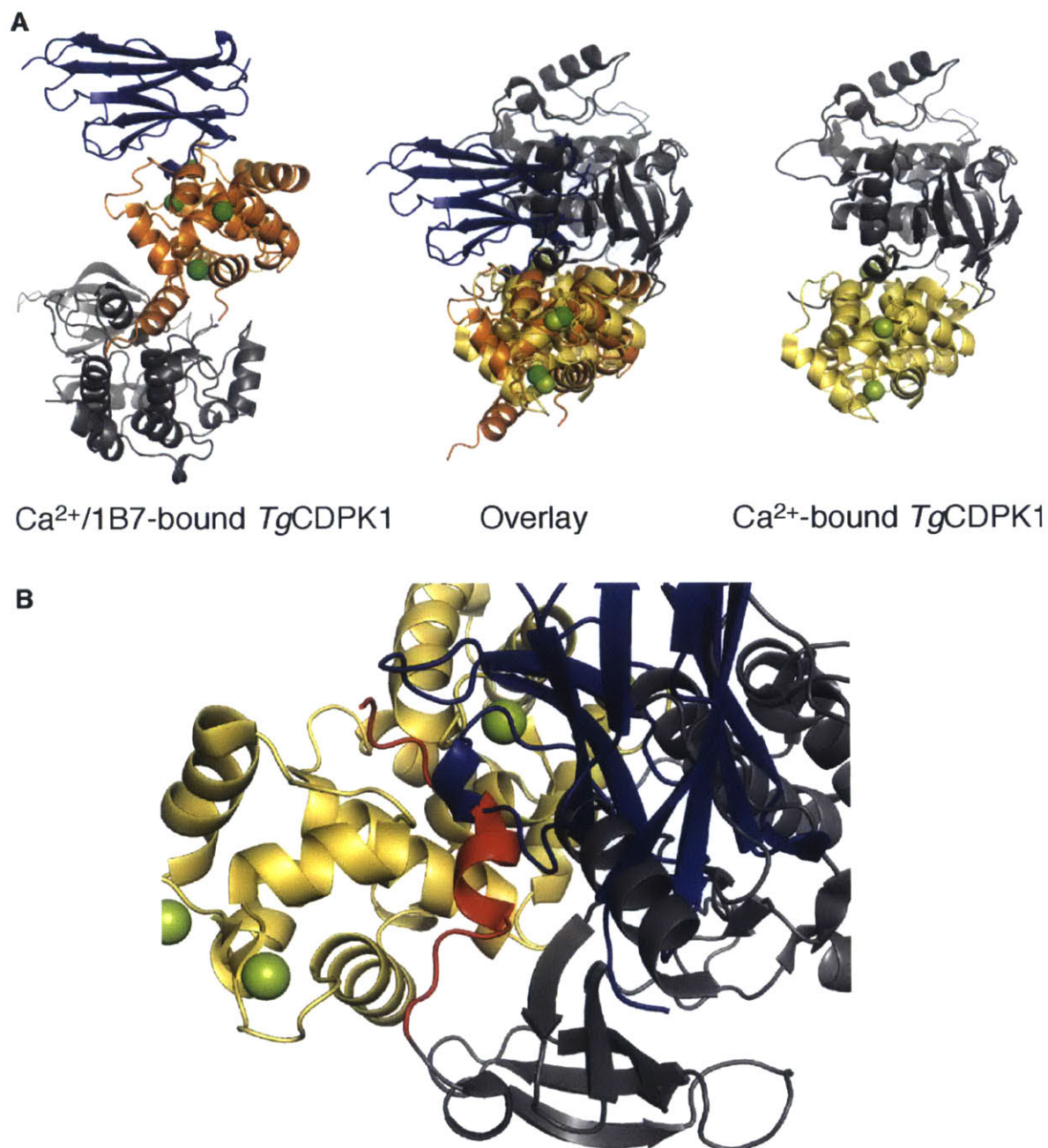


Figure 14. Superposition of the Ca²⁺/1B7-bound and Ca²⁺-bound structures. (A) The two structures are shown individually and superimposed. For clarity, the kinase domain of the Ca²⁺/1B7-bound structure is not shown in the superimposed image. (B) Detailed view of the predicted clash between the N-terminal extension (red) in the active, Ca²⁺-bound structure and CDR3 of the VHH (blue). The kinase domain (gray) and CAD (yellow) from the active structure are shown.

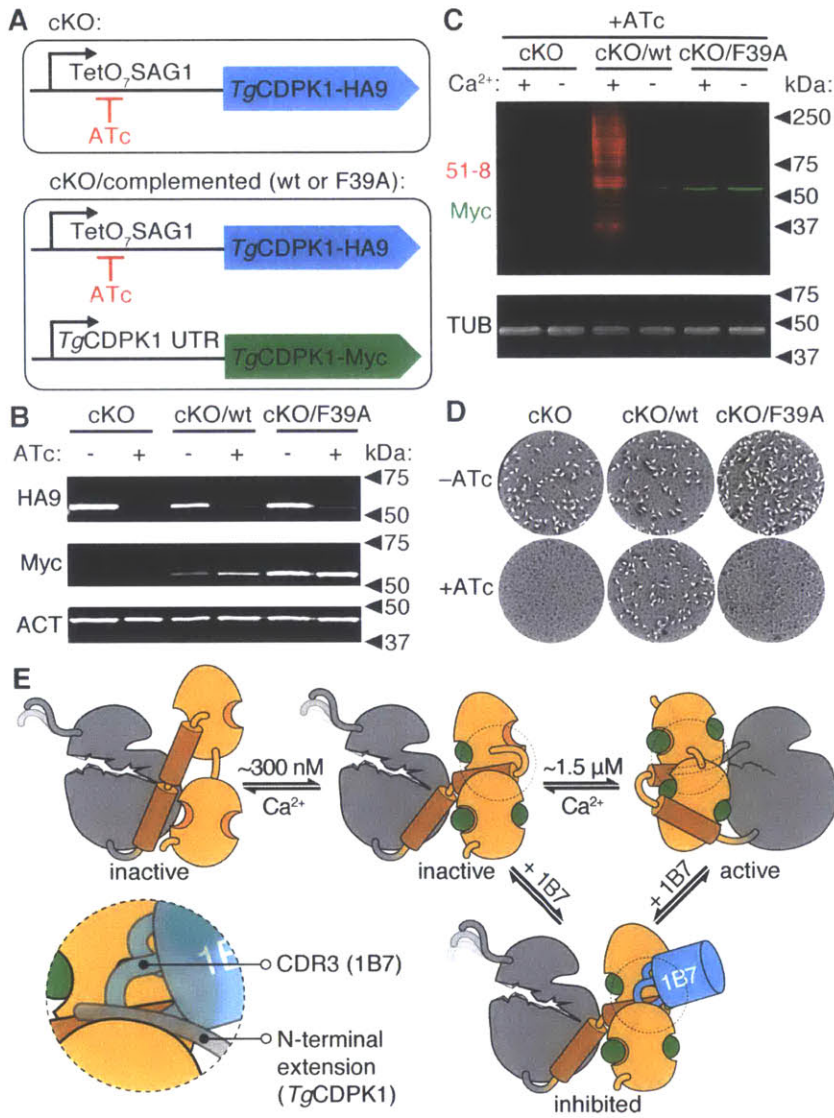


Figure 15. Allosteric activation is required for *TgCDPK1* activity *in vivo*. (A) Strategy depicting the *TgCDPK1* conditional knockout (cKO) carrying an HA9-tagged allele that can be shut down by addition of ATc. The complemented strains additionally carry a constitutive Myc-tagged allele (wt or F39A), expressed under the endogenous *TgCDPK1* promoter. (B) Immunoblot of regulatable (HA9) or constitutive (Myc) *TgCDPK1* alleles in the different strains grown in the presence or absence of ATc for 48 h. Parasite actin (ACT) is included as a loading control. (C) *TgCDPK1*-dependent thiophosphorylation (51-8; red) in lysates from the various parasite strains following growth in the presence of ATc. Reactions were performed in the presence or absence of Ca²⁺. Tubulin is included as a loading control. (D) Plaque formation by the various strains in the presence or absence of ATc. (E) Model for the activation of *TgCDPK1* and its inhibition by 1B7. The KD (gray) is intrinsically inactive and its catalytic site is occluded by the CAD (orange). At ~300 nM Ca²⁺, the CAD is partially occupied by Ca²⁺ and can bind 1B7 (blue). Higher Ca²⁺ concentrations are required for full Ca²⁺ occupancy of the CAD and stabilization of the KD, which leads to activation. The region of the CAD competitively bound by 1B7 and the N-terminal extension is circled and highlighted in the inset.