Structural Analyses of the T1L Reovirus Attachment Protein σ1 and the *Phytophthora sojae* Transglutaminase GP42

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Dipl.-Biochem. Kerstin Reiß

aus St. Ingbert

Tübingen

2013

Tag der mündlichen Prüfung:	03.05.2013
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Thilo Stehle
2. Berichterstatter:	Prof. Terence S. Dermody, M.D.

Abstract

This thesis structurally analyzes the attachment protein σ1 of Type 1 reovirus and the GP42 transglutaminase (TGase) from the plant pathogen *Phytophthora sojae*. Reoviruses serve as model systems to study virus-host interactions and are currently tested in clinical trials as oncolytic agents. The reovirus outer capsid protein σ1 mediates initial attachment of the virus to its host cell and is therefore a major determinant of target cell selection and viral tropism. The specific interaction between the trimeric σ1 protein and carbohydrates on the surface of host cells is thought to cause these serotype-dependent differences in tropism and pathogenesis. By glycan array analyses with purified σ1 protein, this study identified the carbohydrate receptor of Type 1 Lang (T1L) reovirus, the GM2 glycan. The interaction between the GM2 tetrasaccharide and T1L σ1 in solution was confirmed by STD NMR spectroscopy. Subsequent determination of the crystal structure of the C-terminal part of T1L σ1 in complex with the GM2 glycan allowed us to understand the specific recognition of the GM2 glycan by T1L σ1 at an atomic level and to define residues required for functional carbohydrate engagement. The GM2 glycan serves as a serotype-specific receptor for T1L reovirus. Thus, the identification of the glycan receptor of Type 1 reovirus and the structural analyses of the σ1-carbohydrate complex help to explain differences in tropism between reovirus serotypes and expand knowledge of virus-host interaction and viral attachment mechanisms.

In addition, the N-terminal tail domain of σ1 was structurally investigated. The σ1 tail carries a heptad repeat pattern characteristic for coiled coil formation. Two different segments of the tail domain were crystallized, and both structures were determined at high resolution. The structures confirmed that a large portion of the T1L σ1 tail folds into an α-helical coiled coil and revealed interactions that stabilize the trimeric coiled coils. The structures furthermore help to establish reovirus σ1 as a platform into which helical elements from other proteins can be inserted and presented for antibody recognition. Such chimeric reoviruses could be used as vaccines to trigger the production of antibodies against conserved viral epitopes, such as the A helix of the influenza virus hemagglutinin.

In the second project, the crystal structure of the TGase GP42 from the parasitic plant pathogen, *Phytophthora sojae*, was solved. *Phytophthora* species are oomycetes that cause huge economic losses by infecting a variety of plants, including potato, cacao, oak, and soybean. Interestingly, all *Phytophthora* species tested so far possess an enzyme homologous to GP42, including a conserved peptide of 13 amino acids. This peptide, Pep-13, is recognized by the plant immune system and elicits an immune response in host and non-host plants. The GP42 crystal structure demonstrates that Pep-13 forms a β-strand on the surface of the TGase, where most residues are well accessible for components of the plant immune system. Furthermore, the GP42 crystal structure suggests that the active site of the enzyme comprises a catalytic triad composed of Cys290, His291, and Asp328. Although there is no sequence homology between GP42 and human or bacterial TGases, GP42 shares a central core fold with papain-like cysteine proteases. Sequence comparisons between GP42 and homologs from other oomycetes showed that conserved residues cluster around the active site of the enzyme. Thus, the crystal structure of GP42 serves as a model for TGases from related oomycetes. Additionally, a search for sequence homologs of GP42 outside the oomycete taxonomic group identified GP42-like proteins in marine *vibrio* bacteria, suggesting an evolutionary relationship between these organisms and *Phytophthora*.

Zusammenfassung

Diese Dissertation umfasst die strukturelle Charakterisierung des Kapsidproteins σ1 von Typ 1 Reovirus und der Transglutaminase (TGase) GP42 des Pflanzenpathogens *Phytophthora sojae*. Reoviren dienen als Modellsystem zur Erforschung der Interaktion zwischen Virus und Wirt. Darüber hinaus werden sie in klinischen Studien zur Krebstherapie eingesetzt, da sie bestimmte Tumorzellen gezielt lysieren können. Das Reoviruskapsidprotein σ1 vermittelt das initiale Andocken des Virus an die Wirtszelle und beeinflusst dadurch, welche Zellen infiziert werden. Ausschlaggebend, ob eine Zelle tatsächlich infiziert wird, sind möglicherweise die spezifischen Wechselwirkungen zwischen σ1 und Zuckerstrukturen auf der Zelloberfläche. Im Rahmen dieser Dissertation wurde die Zuckerverbindung identifiziert, an die Reovirus Typ 1 Lang (T1L) spezifisch andockt, um anschließend die Wirtszelle infizieren zu können. Es handelt sich dabei um das verzweigte Tetrasaccharid GM2. Die Bindung von GM2 an T1L σ1 in Lösung wurde mittels STD NMR-Spektroskopie nachgewiesen. Zudem wurde die Kristallstruktur von T1L σ1 im Komplex mit GM2 gelöst. Anhand der Struktur wurden Aminosäuren identifiziert, die für die Zuckerbindung essentiell sind und es konnte erklärt werden, warum gerade GM2 von T1L σ1 selektiv gebunden wird. Basierend auf den strukturellen Daten kann nun nachvollzogen werden, warum GM2 nur von Serotyp 1 Reoviren erkannt wird. Außerdem liefert diese Arbeit grundlegende Einblicke in den Erkennungsprozess der Wirtszelle durch Viren.

Neben Untersuchungen des C-terminalen Bereichs des σ1-Proteins und dessen Wechselwirkung mit Zuckerliganden wurde auch die N-terminale Region des Proteins strukturell analysiert. Strukturvorhersagen zufolge bildet dieser Bereich von σ1 ein trimeres α-helikales *coiled coil*, da die Aminosäuresequenz ein dafür charakteristisches Muster aufweist. Im Rahmen dieser Studie wurden die Kristallstrukturen zweier verschiedener Abschnitte dieser Domäne bestimmt. Beide Kristallstrukturen bestätigen das Vorliegen von *coiled coils* in diesen Bereichen. Des Weiteren wurden die Wechselwirkungen analysiert, die diese Strukturelemente stabilisieren. Basierend auf den Kristallstrukturen der N-terminalen Bereiche von T1L σ1 können nun chimäre Proteine generiert werden, in denen helikale Segmente anderer Proteine im Kontext von Reovirus σ1

eingefügt werden. Ein Beispiel für diese Anwendung ist die Insertion der A-Helix des Influenzavirusproteins Hämagglutinin, welche ein konserviertes Strukturelement in Influenzaviren darstellt. Reoviren, die die A-Helix in ihrem Kapsid präsentieren, könnten daher als Impfstoffe eingesetzt werden, um gezielt die Produktion von Antikörpern gegen dieses Strukturelement auszulösen und so dem Patienten einen breiten Schutz gegen Influenzaviren zu gewährleisten.

Das zweite Projekt dieser Dissertation beinhaltete die Strukturaufklärung des Proteins GP42, einer TGase aus dem Pflanzenpathogen *Phytophthora sojae.* Die den Oomyceten angehörenden *Phytophthora* befallen eine Vielzahl an Pflanzen und richten dadurch einen hohen ökonomischen Schaden an. In allen bisher getesteten *Phytophthora*-Arten gibt es ein zu GP42 homologes Enzym. Ein 13 Aminosäuren langes Peptid des GP42, Pep-13, wird von Pflanzen als „fremd“ erkannt und löst daraufhin eine weitreichende Immunantwort aus. Die Kristallstruktur von GP42 zeigt, dass Pep-13 einen β-Strang an der Oberfläche des Proteins bildet, wo es für Komponenten des Immunsystems der Pflanze teilweise zugänglich ist. Gemäß der Kristallstruktur wird das aktive Zentrum des Enzyms vermutlich durch die Aminosäuren Cys290, His291 und Asp328 gebildet. GP42 besitzt keinerlei Sequenzähnlichkeit zu bereits charakterisierten TGasen von Säugetieren oder Bakterien. Allerdings weist die Struktur im Kernbereich des Enzyms Ähnlichkeit zu Cysteinproteasen der Papainfamilie auf. Ein Vergleich der Sequenz von GP42 mit homologen Proteinen aus Oomyceten zeigt, dass sich die konservierten Reste um das aktive Zentrum der TGase häufen. Die Kristallstruktur von GP42 dient daher als Modellstruktur für verwandte Enzyme aus anderen Oomyceten. Eine Suche nach homologen Sequenzen außerhalb dieser taxonomischen Gruppe offenbarte eine unerwartete Sequenzhomologie zwischen GP42 und bestimmten Proteinen aus marinen Vibriobakterien. Eine evolutionäre Verwandtschaft dieser Bakterien und *Phytophthora* ist daher wahrscheinlich.

Table of Contents

Abbreviations

A. thaliana	*Arabidopsis thaliana*
aa	Amino acid(s)
Avr gene	Avirulence gene
BLAST	Basic Local Alignment Search Tool
Bsa	*Bacillus stearothermophilus*
CC, cc	Coiled coil
CD	Circular dichroism
Da	Dalton
DNA	Deoxyribonucleic acid
dsRNA	Double-stranded RNA
E. coli	*Escherichia coli*
ETI	Effector-triggered immunity
Gal	Galactose
GalNAc	*N*-acetylgalactosamine
$(GCN4)_3$	Trimeric version of the dimeric GCN4 leucine zipper
Glc	Glucose
GP, gp	Glycoprotein
HA	Hemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
i.e.	Id est
ISVP	Infectious subvirion particle
JAM-A	Junctional Adhesion Molecule-A
K_D	Dissociation constant
LRR	Leucine-rich repeat
MAMP	Microbe-associated molecular pattern
Neu5Ac	5'-*N*-acetyl neuraminic acid
Neu5Gc	5'-*N*-glycolyl neuraminic acid
NMR	Nuclear magnetic resonance
P.	*Phytophthora*
PAMP	Pathogen-associated molecular pattern
PDB	Protein Data Bank
PRR	Pattern recognition receptor
PTI	PAMP-triggered immunity

R gene	Resistance gene
R protein	Resistance protein
RIN4	RPM1‐interacting protein 4
RNA	Ribonucleic acid
SIGLEC	Sialic acid-recognizing immunoglobulin-like lectin
SIRAS	Single Isomorphous Replacement with Anomalous Scattering
STD	Saturation Transfer Difference
T1L	Type 1 Lang
T2J	Type 2 Jones
T3D	Type 3 Dearing
TGase, TG	Transglutaminase
Tris	Tris(hydroxymethyl)aminomethane

Amino acid	**Three-letter code**	**Single-letter code**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 The T1L reovirus attachment protein σ1

1.1 Introduction

1.1.1 Reovirus

The *Reoviridae* (*R*espiratory *E*nteric *O*rphan viruses) form a group of non-enveloped double-stranded RNA (dsRNA) viruses. When they were initially isolated from humans they were not associated with any disease. Among other members, the family includes the rotaviruses, which cause severe diarrhea in infants, and the genus orthoreovirus. Orthoreoviruses infect mammalians, avians, and reptiles. Orthoreovirus infections in animals are often associated with neurologic disease, pneumonia, and death (Schiff et al, 2007). In contrast, orthoreoviruses cause mainly asymptomatic infections in humans and only rarely lead to mild respiratory or gastrointestinal disease in infants (Tai et al, 2005; Dermody et al, 2012). However, reoviruses serve as a useful model system to study virus-host interactions and pathogenesis. In addition to their relevance in basic research in virology, reoviruses are currently tested in clinical trials as cancer therapeutics, because they preferentially lyse tumor cells and are not pathogenic, which is a benefit (section 1.1.1.3).

1.1.1.1 *Reovirus structure and components*

Mammalian orthoreoviruses, referred to as reoviruses from here on, possess an icosahedral capsid. The viral capsid is composed of two concentric shells: the outer capsid and the inner core (Figure 1). The core encloses ten gene segments of dsRNA. The RNA segments are called S (small), M (medium), and L (large) according to their size, and are numbered sequentially. There are four S genes (S1-S4), three M genes (M1-M3), and three L genes (L1-L3) (Shatkin et al, 1968). The proteins they encode are identified with the Greek letters σ, μ, and λ to denote products of the S, M, and L genes, respectively. The outer capsid, which is composed of the proteins σ1, λ2, μ1, and σ3, mediates viral attachment to host cells and entry of the virus into the cell interior. The capsid itself is formed by μ1, σ3, and λ2. The σ1 protein protrudes from pentameric λ2 turrets and serves as the viral attachment protein (section 1.1.2).

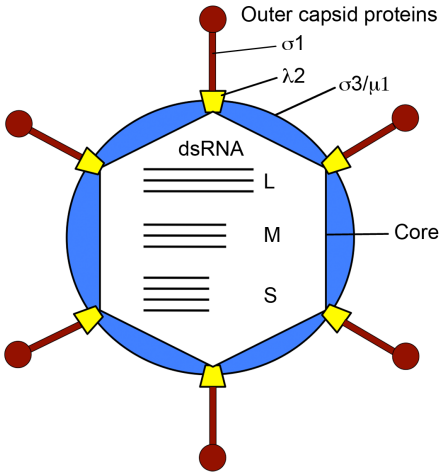

Figure 1: Schematic drawing of the reovirus virion. The outer capsid is composed of the proteins σ1, σ3, λ2, and μ1. The pentameric λ2 protein spans both the outer capsid and the inner core. The viral attachment protein σ1 protrudes from turrets of λ2. The inner core contains the ten dsRNA segments S1-S4, M1-M3, and L1-L3.

The inner core is composed of the proteins λ1, σ2, λ3, and μ2. The functions of the core are transcription of dsRNA into (+) strand RNA, capping of each RNA transcript, and export of mature mRNA into the cytoplasm of the infected cell. The inner core is built by the structural proteins λ1 and σ2 (Reinisch et al, 2000). The protein λ3 is the RNA-dependent RNA-polymerase, which synthesizes the viral mRNA and replicates the reovirus RNA genome (Morozov, 1989). The λ2 protein spans both the inner core and outer capsid. It serves as a guanyltransferase and is therefore thought to catalyze capping of the viral RNA (Cleveland et al, 1986; Seliger et al, 1987; Fausnaugh & Shatkin, 1990; Mao & Joklik, 1991). Additionally, λ2 forms a channel for exit of the viral mRNA into the cytoplasm, where the mRNA is translated into viral proteins (Bartlett et al, 1974).

1.1.1.2 *Viral uptake and disassembly*

After viral attachment to host cells, reoviruses are internalized. This process is probably triggered by the interaction between integrin-binding sequences in the pentameric λ2 protein and β1 integrins (Maginnis et al, 2006; Maginnis et al, 2008). Reovirus uptake takes place by clathrin-dependent endocytosis (Ehrlich et al, 2004; Maginnis et al, 2006). Following uptake, the virus-loaded vesicles are transported along microtubules and finally accumulate in late endosomes, where proteolytic uncoating of the virus occurs (Borsa et al, 1979; Borsa et al, 1981; Sturzenbecker et al, 1987; Rubin et al, 1992; Ebert et al, 2002; Mainou & Dermody, 2012). The process of viral uncoating is schematically depicted in

Figure 2. The first intermediate, the infectious subvirion particle (ISVP), is characterized by removal of σ3 and conformational changes in the outer capsid proteins σ1 and μ1 (Figure 2A, B). In addition, the μ1 protein is autocatalytically cleaved into the N- and C-terminal fragments μ1N and μ1C. μ1C is further degraded by host cell proteases into particle-associated fragments δ and ϕ (Nibert & Fields, 1992). ISVPs are subsequently processed into a second uncoating intermediate, the ISVP* (Figure 2C). In the ISVP*, the σ1 protein and the μ1N fragment are released, and conformational changes in μ1C take place that expose hydrophobic residues. These hydrophobic sequences can penetrate the endosomal membrane to finally release the transcriptionally active core particles (Figure 2D) into the cytoplasm, where the viral proteins are translated and new virus particles assemble (Chandran & Nibert, 1998; Chandran et al, 2002; Odegard et al, 2004).

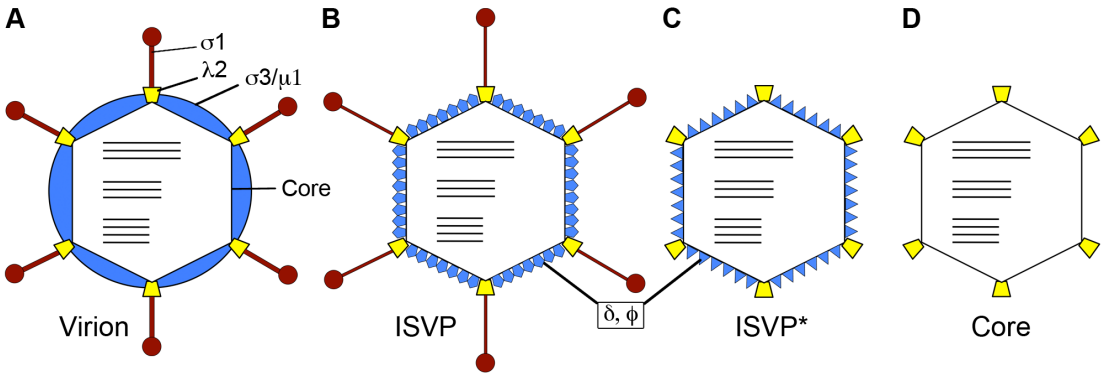

Figure 2: Reovirus disassembly. Schematic representation of a virion (**A**). After attachment and uptake, virions are processed in endosomes at low pH by host cell proteases. Uncoating intermediates are the ISVP (**B**) and the ISVP* (**C**). Exposure of hydrophobic residues facilitates penetration of the endosomal membrane and release of the transcriptionally active core (**D**) into the cytoplasm. Colors were chosen as in figure 1.

1.1.1.3 *Reoviruses as oncolytic agents*

In addition to their role as a model system to study virus-host interactions, reoviruses are being evaluated in clinical trials as potential oncolytic agents due to their cytotoxic effect on certain transformed cell lines (Hashiro et al, 1977). Reovirus infections of adults usually lead only to mild symptoms, which is an advantage over other viral cancer treatment strategies. In fact, reoviruses (Reolysin®, Oncolytics Biotech, Inc.) are non-toxic in human subjects, regardless

if they are administered intratumorally (Forsyth et al, 2008; Morris et al, 2012) or systemically (White et al, 2008; Yap et al, 2008), and even when applied at high doses.

Phase II and III clinical trials of Reolysin® are currently ongoing, both with Reolysin® alone and in conjunction with conventional chemotherapeutics or radiotherapy (Kyula et al, 2012; Maitra et al, 2012). In its first phase III clinical trial, Reolysin® was applied in combination with carboplatin/paclitaxel in patients with platin-refractory, relapsed or metastatic head and neck cancer (Kyula et al, 2012). The results are being analyzed now. A synergistic effect of reovirus and standard chemotherapy has been demonstrated in cell culture and animal models (Yap et al, 2008). The results of phase II and III clinical trials of Reolysin® will be critical for further development of oncolytic therapy with reovirus.

Reoviruses show a high potential in oncolytic therapy. Nevertheless, the underlying mechanism of the cytotoxic effect of reovirus on tumor cells is still poorly understood. It is known that reoviruses target tumor cells with an upregulated Ras pathway (Coffey et al, 1998; Strong et al, 1998). However, it is likely that other factors, for example the abundance of cellular receptors for reovirus on tumor cells, also contribute to this process.

1.1.2 **The reovirus attachment protein σ1**

The three reovirus serotypes are represented by prototype strains Type 1 Lang (T1L), Type 2 Jones (T2J), and Type 3 Dearing (T3D), which were named after the respective patients from which they were isolated. Reovirus infections are usually asymptomatic in humans, but they generate distinct disease patterns in newborn mice. Mice therefore serve as animal models to study reovirus pathogenesis and spread. In the nervous system of newborn mice, the reovirus prototype strains differ in their routes of spread and cell tropism. While T1L spreads by hematogenous routes and infects ependymal cells leading to non-lethal hydrocephalus in mice (Weiner et al, 1977; Weiner et al, 1980), T3D spreads via hematogenous and neural routes. This serotype infects neurons,

causing lethal encephalitis (Weiner et al, 1977; Weiner et al, 1980; Tyler et al, 1986; Morrison et al, 1991; Antar et al, 2009; Boehme et al, 2011a). These serotype-dependent differences are causally linked to sequence variations within outer capsid protein σ1 (Weiner et al, 1980; Tyler et al, 1986).

1.1.2.1 *Domain organization and structure of σ1*

The reovirus outer capsid protein σ1 is a trimeric fiber of about 480 Å in length (Furlong et al, 1988; Fraser et al, 1990) that mediates attachment of the virus to its target cell (Weiner et al, 1980; Lee et al, 1981). Approximately 20 amino acids at the N-terminus of σ1 anchor the protein into turrets of λ2, while the remaining portion of the protein protrudes from the viral surface (Figure 1). σ1 can be divided into three structurally and functionally distinct domains: the tail, the body, and the head (Figure 3) (Bassel-Duby et al, 1987; Duncan et al, 1990; Nibert et al, 1990; Lupas et al, 1991). The sequence of the tail region carries a heptad repeat pattern and was therefore predicted to form a trimeric α-helical coiled coil (sections. 1.1.3, 1.3.3) (Bassel-Duby et al, 1985; Nibert et al, 1990; Lupas et al, 1991). The N-terminal tail is followed by the body region, which predominantly consists of β-spiral repeats (Chappell et al, 2000). Apart from reovirus σ1, this motif has been observed thus far only in the structures of the adenovirus attachment protein fiber (van Raaij et al, 1999), the bacteriophage PRD1 P5 (Merckel et al, 2005), and the avian reovirus attachment protein σC (Zhang et al, 2005). In case of T3D σ1, the body includes a short α-helical segment (Reiter et al, 2011). The C-terminus of σ1 folds into the globular head domain, which has a β-barrel-like structure. The σ1 protein incorporates two regions of flexibility. The first one connects the tail and the body domain, and the second one is located beneath the σ1 head (Furlong et al, 1988; Fraser et al, 1990; Chappell et al, 2002). Optimal length and flexibility of both regions are important for σ1 functionality (Bokiej et al, 2012).

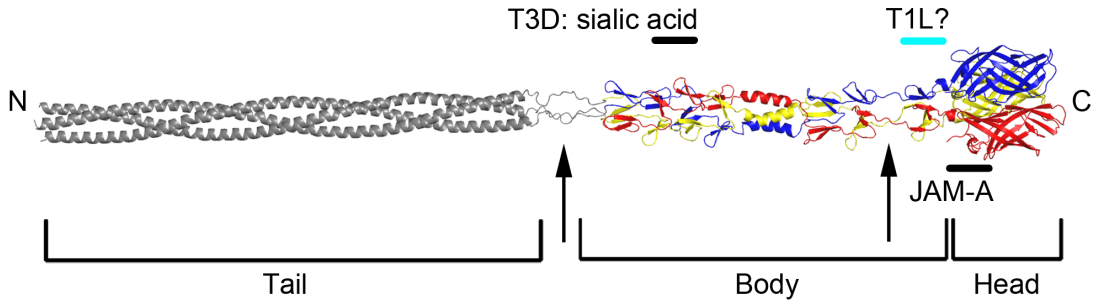

Figure 3: Domain organization of σ1. A model of the N-terminal tail that is predicted to form an α-helical coiled coil is shown in gray. The crystal structure of the T3D σ1 body and head (PDB code 3S6X) is shown as ribbon drawing in red, blue, and yellow. N- and C-termini are labeled. The JAM-A binding site of all reovirus serotypes and the sialic acid binding site of T3D σ1 are indicated. The predicted carbohydrate binding site of T1L σ1 is highlighted in cyan. Regions of flexibility are indicated with black arrows.

Several crystal structures exist for different parts of T3D σ1. The first σ1 crystal structure, which was determined at 2.6 Å resolution, comprises the T3D σ1 head domain and the three most C-terminal β-spiral repeats, i.e., amino acids (aa) 246-455 (Chappell et al, 2002). The head domain is constructed from two Greek-key motifs that face each other and that consist of four β-strands each (β-strands A-D and E-H, respectively). The two Greek-key motifs are linked by a short α-helix. The crystal structure revealed an unexpected evolutionary relationship between σ1 and the adenovirus attachment protein, fiber (Chappell et al, 2002). The crystal structure of the T3D σ1 head domain and the most C-terminal β-spiral (aa 293-455) at 1.75 Å resolution is the highest resolution structure of σ1 available so far (Schelling et al, 2007). This structure showed in detail that the σ1 head contains an unusual arrangement of aspartic acids that stabilize the trimer, but that may induce a conformational change upon exposure to acidic conditions. The more recent crystal structure of the head and the complete body domain (aa 170-455) at 2.2 Å resolution provides the most complete model of T3D σ1 determined so far (Reiter et al, 2011). This structure revealed that the T3D σ1 body domain is not only composed of β-spiral repeats but also includes an α-helical segment.

1.1.2.2 *Interaction with JAM-A*

The outer capsid protein σ1 mediates the attachment of the virus to host cells. By specifically interacting with receptor molecules, σ1 selects the target cells of the virus and is thereby a major determinant of viral tropism and disease outcome. σ1 binds with high affinity to Junctional Adhesion Molecule-A (JAM-A) (Barton et al, 2001b). JAM-A serves as a receptor for all reovirus strains tested so far (Campbell et al, 2005). JAM-A is a homodimeric molecule and a member of the immunoglobulin superfamily (Martin-Padura et al, 1998; Williams et al, 1999; Prota et al, 2003). The protein is a component of tight junctions (Liu et al, 2000) and is also involved in immune responses and inflammation processes. JAM-A is composed of a short N-terminal signal peptide, two immunoglobulin-like extracellular domains, D1 and D2, a membrane-spanning region, and a short C-terminal cytoplasmic domain. The membrane-distal, N-terminal D1 domain contains the dimer interface. σ1 binds to the JAM-A D1 domain with a dissociation constant (K_D) in the low nM range. The crystal structure of a complex between T3D σ1 and JAM-A D1 was determined at 3.4 Å resolution (Kirchner et al, 2008). In this complex, σ1 disrupts the JAM-A homodimer by binding to the dimer interface. Thus, each σ1 monomer engages one JAM-A monomer (Figure 4). The affinity between two JAM-A D1 domains in the JAM-A dimer is about 1000-fold lower than the affinity between JAM-A and σ1 (Vedula et al, 2008). The interactions between JAM-A and σ1 mainly involve polar interactions, with numerous hydrogen bonds and two salt bridges. The JAM-A binding site is formed by the loop and the short 3_{10} helix connecting β-strands D and E. In addition, JAM-A binds to an α-helix at the lower end of the head domain and a small region of the C-terminal β-spiral (Kirchner et al, 2008).

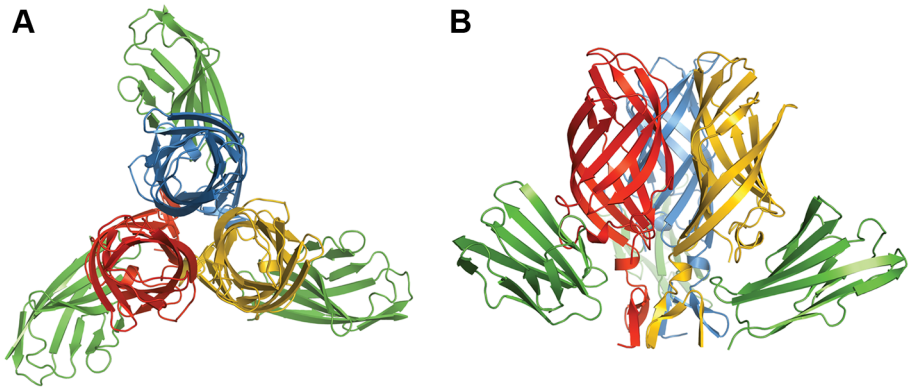

Figure 4: Crystal structure of T3D σ1 in complex with JAM-A D1. Ribbon representation of the complex viewed from the top (**A**) and the side (**B**). The σ1 trimer is depicted in red, blue, and yellow. The JAM-A D1 monomers are shown in green. Modified from Kirchner et al, 2008.

Sequence comparisons between T1L, T2J, and T3D σ1 suggest that the JAM-A binding site is conserved among serotypes. This conclusion was confirmed when the crystal structure of T1L σ1 in complex with JAM-A D1 was determined at 3.2 Å resolution (Kirchner, 2009). The architecture of the T1L σ1-JAM-A complex is highly similar to the T3D σ1-JAM-A complex, and interactions are largely conserved in both structures. It is therefore unlikely that the engagement of JAM-A is responsible for the known differences in cell tropism among serotypes. Furthermore, differences in tropism are maintained when JAM-A knockout mice are infected intracranially with T1L and T3D reovirus (Antar et al, 2009), suggesting that binding to receptors others than JAM-A determines viral tropism. As discussed in section 1.1.2.5, interactions between σ1 and carbohydrates present on the surface of host cells may dictate reovirus tropism.

1.1.2.3 *Carbohydrates*

Carbohydrates, and especially sialic acid-containing glycans, serve as receptors for a variety of viruses and are often major determinants of viral tropism. Eukaryotic cells are decorated with different forms of carbohydrates. Carbohydrates can occur in context of glycoproteins, in which the glycan is attached to a protein via N- or O-linkage, or they can be connected to lipids, which anchor the sugar chain into the cell membrane. These structures are referred to as glycosphingolipids (Fuster & Esko, 2005).

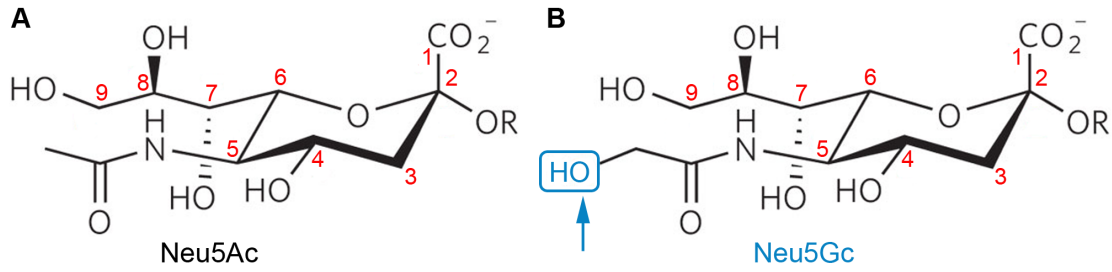

Figure 5: Structure of Neu5Ac (A) and Neu5Gc (B). Neuraminic acid is built from a backbone of nine carbon atoms (numbered in red). The functional groups of the sugar ring are a carboxyl group at C2 and a glycerol chain at C6. In case of Neu5Ac, an *N*-acetyl chain is attached to C5, in Neu5Gc it is an *N*-glycolyl chain. The hydroxyl group that is absent in Neu5Ac and present in Neu5Gc is highlighted in blue. Modified from Varki, 2007.

Sialic acid comprises the family of neuraminic acid derivatives. Neuraminic acids have a common backbone of nine carbon atoms. The most abundant family member is 5'-*N*-acetyl neuraminic acid (Neu5Ac, Figure 5A), which is thought to serve as a precursor for the biosynthesis of all other sialic acids (Varki, 1992). The functional groups of Neu5Ac are a carboxylate at C2, an *N*-acetyl chain at C5, and a glycerol chain at C6. Addition of a hydroxyl group at the *N*-acetyl functionality of Neu5Ac leads to the formation of 5'-*N*-glycolyl neuraminic acid (Neu5Gc, Figure 5B) (Schauer, 1991; Reuter & Gabius, 1996). Neu5Gc is not produced by humans due to a mutation in the gene encoding the enzyme cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase, which catalyzes this synthetic step (Shaw & Schauer, 1989; Takematsu et al, 1994; Chou et al, 1998; Irie et al, 1998). Accordingly, Neu5Gc is absent on human erythrocytes (Muchmore et al, 1998).

The functional group at C5 plays an important role in human evolution, immunology, and susceptibility to pathogens. The absence of Neu5Gc in humans represents a major difference between humans and Great Apes and was one determinant of human evolution (Hayakawa et al, 2001; Chou et al, 2002). Humans even elicit an antibody response against Neu5Gc, which can produce "serum sickness" (Higashi et al, 1977; Merrick et al, 1978). The lack of Neu5Gc and the presence of Neu5Gc-specific antibodies have presumably evolved to protect humans from pathogens that bind Neu5Gc and from enveloped viruses that carry Neu5Gc (Martin et al, 2005; Varki, 2010). However, Neu5Gc can be absorbed from dietary sources and is present in certain cancers and on fetal tissues (Malykh et al, 2001).

Sugar chains of vertebrates and some invertebrates often terminate in sialic acid (Varki, 1993; Traving & Schauer, 1998). Sialic acid is important for embryonic development (Schwarzkopf et al, 2002) and involved in cell-cell recognition. Moreover, sialic acid serves as a signal for self recognition for the immune system. In addition to its function as an intrinsic receptor, sialic acid also serves as an extrinsic receptor for pathogens. Human-specific pathogens often target Neu5Ac on host cells (Varki & Gagneux, 2009). Additionally, some pathogens express or absorb and present Neu5Ac to mask themselves from the adaptive immune system (Varki, 2010). They also use sialic acid-recognizing immunoglobulin-like lectins (Siglecs) to suppress the innate immune response. For example, CD33r Siglecs preferentially bind Neu5Ac (and not Neu5Gc) and thereby recognize "self". Siglecs with occupied Neu5Ac binding sites send dampening signals to the innate immune system (Lajaunias et al, 2005). Pathogens such as *Pseudomonas aeruginosa* or group B *Streptococcus* successfully exploit this regulating mechanism by presenting Neu5Ac on the bacterial cell surface (Carlin et al, 2009; Khatua et al, 2010).

1.1.2.4 *Sialic acid-based glycans as viral receptors*

Many viruses, such as species D adenoviruses (Burmeister et al, 2004), reoviruses (Reiter et al, 2011), several polyomaviruses (Stehle et al, 1994; Neu et al, 2008; Neu et al, 2010; Neu et al, 2012), and influenza virus A (Weis et al, 1988) engage carbohydrates terminating in Neu5Ac for cell attachment. As glycan composition, accessibility, and density differ between organisms, cell types, and tissues, the interaction between virus and carbohydrate can dictate viral tropism. The well-known case of influenza-carbohydrate interactions demonstrates that in addition to the composition of the carbohydrate receptor, the linkage between the individual sugar moieties critically contributes to receptor specificity (Rogers & Paulson, 1983). The influenza virus protein hemagglutinin (HA) mediates binding of the virus to Neu5Ac. Strains that infect humans preferentially bind to Neu5Ac connected via $\alpha(2,6)$-linkage to Galactose (Gal), which dominates on the surface of epithelial cells in the airways. In contrast, avian influenza viruses use Neu5Ac $\alpha(2,3)$-linked to Gal for cell attachment. A crossover of bird influenza to humans takes place when mutations

in HA enable engagement of α(2,6)-linked Neu5Ac. Thus, the specific recognition of α(2,3)- and α(2,6)-linked Neu5Ac defines both host range and cell tropism of the virus (Rogers et al, 1983; Daniels et al, 1984; Baum & Paulson, 1990; Chandrasekaran et al, 2008).

The affinity between viral attachment proteins and carbohydrates are usually low, with a K_D value often in the mM range (Sauter et al, 1989; Stehle & Harrison, 1996; Burmeister et al, 2004; Neu et al, 2008). However, on a virus particle many carbohydrate binding sites are present, which allow the binding between virus and cell to be amplified through multivalency (Barton et al, 2001a; Neu et al, 2011). Thus, the interaction between virus and ligand is not determined by the affinity of a single viral protein chain for its carbohydrate ligand but by the avidity of the complete virus particle for many sugar molecules on the host cell surface.

1.1.2.5 *Interaction between σ1 and carbohydrates*

Interactions between viruses and sialic acid can be assessed by determination of the hemagglutination capacity of the virus (Hirst, 1941; Hirst, 1942; Burnet et al, 1946; Bovarnick & de Burgh, 1947), a property depending on the binding of a virus to carbohydrates on the surface of erythrocytes (Green & Woolley, 1947; Kathan et al, 1961). Reoviruses can agglutinate red blood cells of several mammalian species (Lerner et al, 1963). However, the reovirus serotypes differ in their hemagglutination properties. While T1L agglutinates human erythrocytes well, T3D agglutinates human erythrocytes less efficiently than T1L but is able to agglutinate bovine red blood cells (Gomatos & Tamm, 1962). This finding suggests that both T1L and T3D bind to carbohydrates but that the particular glycan they recognize differs. The specific interaction between the σ1 protein and carbohydrates on the host cell surface may be responsible for the differences in tropism among the reovirus serotypes.

T3D binds to glycophorin A, an erythrocyte protein with glycosylation terminating in α(2,3)-linked Neu5Ac (Gentsch & Pacitti, 1987; Pacitti & Gentsch, 1987). T3D binds to sialylated glycophorin with an avidity of 10^{-9} M (Barton et al, 2001a). In contrast to T3D, T1L reovirus does not bind to glycophorin A. From

these data it was concluded that T1L does not bind Neu5Ac (Pacitti & Gentsch, 1987; Barton et al, 2001a). On the other hand, infection of M cells can be diminished by pretreatment with neuraminidase, an enzyme that specifically removes Neu5Ac (Helander et al, 2003), suggesting that T1L recognizes Neu5Ac in at least some contexts.

To define the respective carbohydrate binding sites of T1L and T3D σ1, chimeric proteins were synthesized and capacity to agglutinate human and bovine erythrocytes quantified (Chappell et al, 2000). These experiments suggested that the carbohydrate binding site of T1L σ1 is located within the neck region of the protein, i.e., at the C-terminal part of the body, close to the head domain (Figure 3). For T3D σ1, the carbohydrate binding site was predicted to lie in the N-terminal part of the body, next to the midpoint of the σ1 fiber. Crystal structures of T3D σ1 in complex with sialylated compounds confirmed that the carbohydrate binding site of T3D σ1 was mapped correctly (Reiter et al, 2011). T3D σ1 was crystallized in complex with α(2,3)-sialyllactose, α(2,6)-sialyllactose, and α(2,8)-disialyllactose. Thus, T3D σ1 recognizes Neu5Ac in different linkages. The carbohydrate binding site is located on a loop connecting β-spirals 2 and 3 (Figure 6A, B). In each of the three structures, Neu5Ac contributes most of the interactions, while the glucose (Glc) and Gal moieties only form minimal contacts with the protein. A central contact between Neu5Ac and T3D σ1 is a bidentate salt bridge between R202 of σ1 and the carboxyl group of Neu5Ac (Figure 6C).

Based on the crystal structures, effects of carbohydrate engagement on the neuropathology of T3D reovirus were assessed. Analyses of wild type T3D reovirus and a σ1 mutant that does not bind Neu5Ac showed that carbohydrate engagement increases the neurovirulence of the virus (Frierson et al, 2012). Thus, the capacity of the virus to bind to its carbohydrate receptor directly influences its pathogenesis.

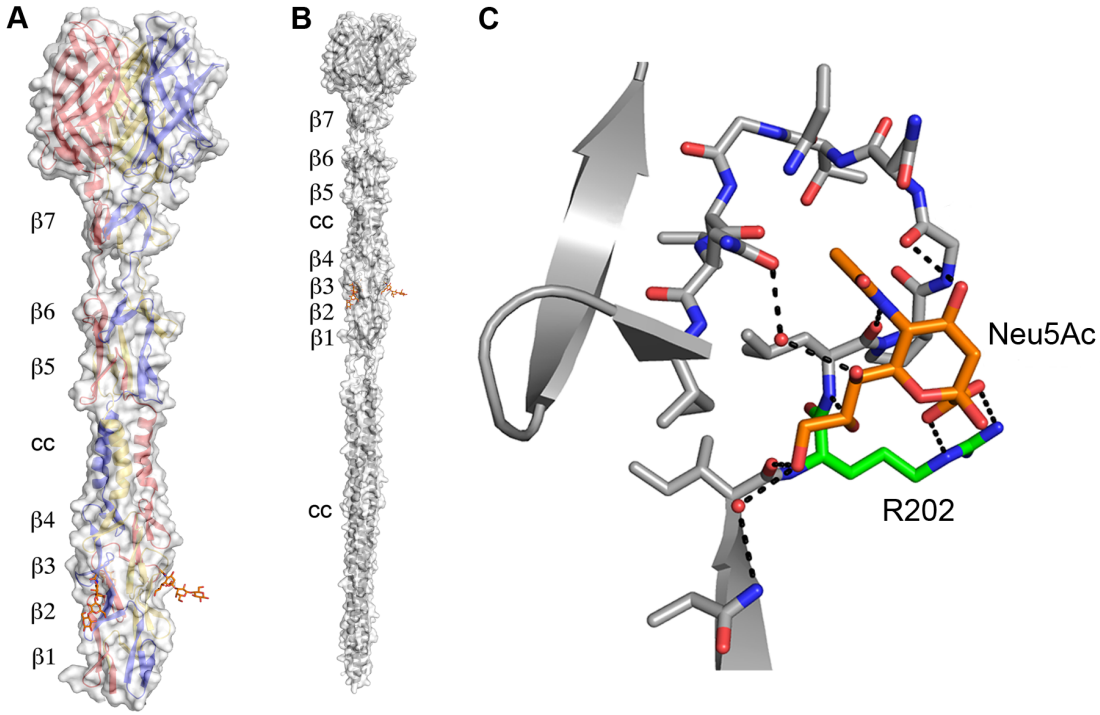

Figure 6: Crystal structure of T3D σ1 in complex with α2,3-sialyllactose. A: The σ1 trimer is shown as ribbon drawing in red, blue, and yellow; the protein surface is depicted in gray. β-spiral repeats β1-β7 and the coiled coil region (cc) between β4 and β5 are labeled. The carbohydrate is shown as sticks in orange. **B**: Surface representation of T3D σ1 including a model for the coiled coil tail domain. Modified from Reiter et al, 2011 (PDB code 3S6X). **C**: Close-up view of the carbohydrate binding site. Neu5Ac (orange) and residues that contact the glycan are shown as sticks. Residue R202, which forms two salt bridges with Neu5Ac, is highlighted in green (Panel C provided by Dirk M. Reiter, University of Tübingen, 2012).

According to the current model for reovirus attachment, the virus specifically binds with low affinity to its carbohydrate receptor, which is distributed on the cell surface. The virus then diffuses laterally until it meets its high-affinity receptor, JAM-A, which is located at tight junctions and thus less accessible for the virus. T3D reovirus attaches to host cells by an adhesion strengthening mechanism as carbohydrate engagement increases the binding of the virus to its target cell (Barton et al, 2001a). Hence, the interaction with carbohydrate receptors makes attachment more efficient.

The carbohydrate binding properties of σ1 also might contribute to the oncolytic potential of reoviruses: the Neu5Ac-binding Type 3 strain has a 100-fold greater binding capacity to human cholangiocarcinoma cells compared with a Type 3 strain that does not bind Neu5Ac (Barton et al, 2003).

The mechanism of Neu5Ac engagement and its effect on reovirus pathogenesis have at least partially been determined for Type 3 reovirus. However, much less

is known about the attachment mechanism of Type 1 reovirus. Neither the nature of the carbohydrate receptor for T1L and the structural basis for carbohydrate engagement, nor the influence of glycan binding on Type 1 reovirus spread and pathogenesis have been analyzed.

1.1.3 **The σ1 tail: a coiled coil motif?**

α-helical coiled coils are very common structural motifs composed of at least two α-helices winding around each other. Coiled coil domains occur in many proteins such as kinesin motor proteins, apolipoprotein E, DNA binding proteins, and response regulators of two-component signaling systems (Lupas & Gruber, 2005). In addition to these examples, coiled coil segments are often part of bacterial and viral attachment proteins, such as bacterial adhesins (Linke et al, 2006), influenza virus HA (Wilson et al, 1981), human immunodeficiency virus glycoprotein 41 (HIV gp41) (Weissenhorn et al, 1997), Moloney murine leukemia virus transmembrane protein (Fass et al, 1996), and reovirus σ1 (sections 1.1.2.1, 1.3.3). Coiled coils often induce formation of the correct oligomeric state of proteins (Kammerer et al, 1998; Kammerer et al, 2005). The architecture of whole proteins can be stabilized or modified by insertion of coiled coil domains with different oligomeric states, which is a useful tool for protein engineering purposes.

The assembly of α-helices into coiled coils and basic rules for this process were first described by Francis Crick (Crick, 1952; Crick, 1953a; Crick, 1953b). Amino acid sequences are predicted to form an α-helical coiled coil if they show a characteristic pattern of seven amino acid repeats (amino acids *a, b, c, d, e, f,* and *g)* called "heptad repeat". If positions *a* and *d* are occupied by hydrophobic residues it is likely that the sequence forms an α-helical coiled coil (McLachlan & Stewart, 1975; Parry, 1982; Lupas et al, 1991). The remaining positions usually carry polar amino acids. This principle is schematically illustrated in Figure 7 for a trimeric α-helical coiled coil. To predict if a certain amino acid sequence comprises a coiled coil motif, programs such as parcoil (Berger et al, 1995) or

coils (Parry, 1982; Lupas et al, 1991; Lupas, 1996) can be used to analyze amino acid sequences for the presence of heptad repeat patterns.

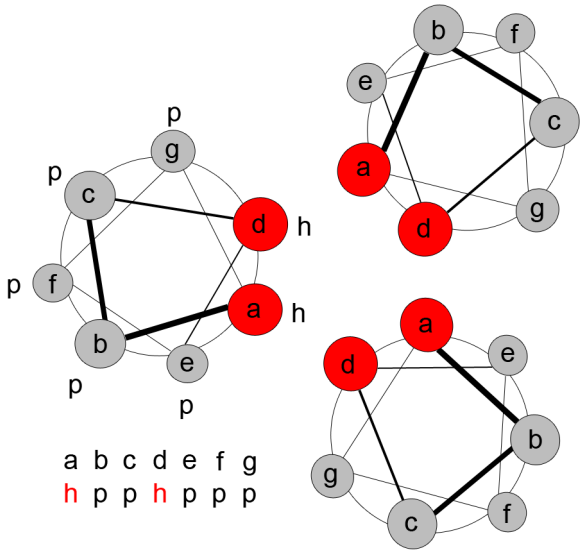

Figure 7: Helical wheel representation of a trimic α-helical coiled coil, viewed from the N-terminus of the helices. Hydrophobic residues (h) at positions *a* and *d* pointing to the inside of the coiled coil are highlighted in red. Polar amino acids (p) that occupy the remaining positions *b*, *c*, *e*, *f*, and *g* are shown in gray.

The oligomerization of coiled coils can be dimeric, trimeric, tetrameric, or even of higher order (Lupas & Gruber, 2005). The correct oligomeric state of coiled coils is often difficult to predict from the amino acid sequence. In order to understand the factors that determine the oligomerization of coiled coils, Harbury et al systematically exchanged the amino acids at positions *a* and *d* of the GCN4 leucine zipper, a native dimer (Harbury et al, 1993). They found that the protein forms a dimeric coiled coil, when β-branched residues (such as isoleucine or valine) are present at position *a* and γ-branched amino acids (i.e. leucine) at position *d*. When the positions of β- and γ-branched residues are reversed, the preferred oligomeric state is a tetramer. When all positions are occupied with β-branched residues, a trimeric α-helical coiled coil is formed. For instance, a trimeric version of GCN4, $(GCN4)_3$, was generated by introducing isoleucines at each *a* and *d* position. Both native and mutant GCN4 variants are commonly used to stabilize the native oligomeric state of proteins that are difficult to express and purify. In such fusion proteins, the GCN4 motif can be connected via a linker to the protein of interest (Pack et al, 1995; Wang et al, 2002; Reiter et al, 2011). If an α-helical coiled coil domain is present in the

protein of interest, the GCN4 domain can be attached in frame with the coiled coil register of this protein (Wolfe et al, 2003; Hernandez Alvarez et al, 2008).
Protein sequences that strictly obey the heptad repeat pattern, known as canonical coiled coils, form a left-handed supercoil. However, there are proteins with periodic insertions of one, three, or four additional amino acids into the heptad register. These modifications, called skips, stammers, and stutters, lead to the formation of straight and right-handed supercoils or to an increase or decrease of supercoiling. The basic rules for coiled coil formation have been expanded to reliably predict the structures from such sequences (Gruber & Lupas, 2003).
Sequence analyses of reovirus σ1 suggest that the tail domain forms a canonical trimeric α-helical coiled coil. However, no structural information has been available for this part of the protein.

1.1.4 **Helical elements in influenza virus hemagglutinin**

Influenza virus HA is a glycoprotein that mediates viral attachment by binding to Neu5Ac (section 1.1.2.4). Additionally, HA triggers membrane fusion when exposed to low pH after endocytosis of the virus. HA is a homotrimeric protein. Each subunit is composed of two chains, HA1 and HA2, which are derived from a common precursor protein HA0 by proteolytic cleavage (Klenk et al, 1975; Lazarowitz & Choppin, 1975).
HA1 comprises the globular receptor-binding head domain, while HA2 forms a slender stalk region (Figure 8A). The stalk comprises a trimeric α-helical coiled coil (the CD helix) that stabilizes the trimer by extensive interactions (Wilson et al, 1981). The coiled coil does not carry an uninterrupted heptad repeat pattern, but includes skip residues, which reduce the supercoiling of the helices compared to canonical coiled coils (Seo & Cohen, 1993).

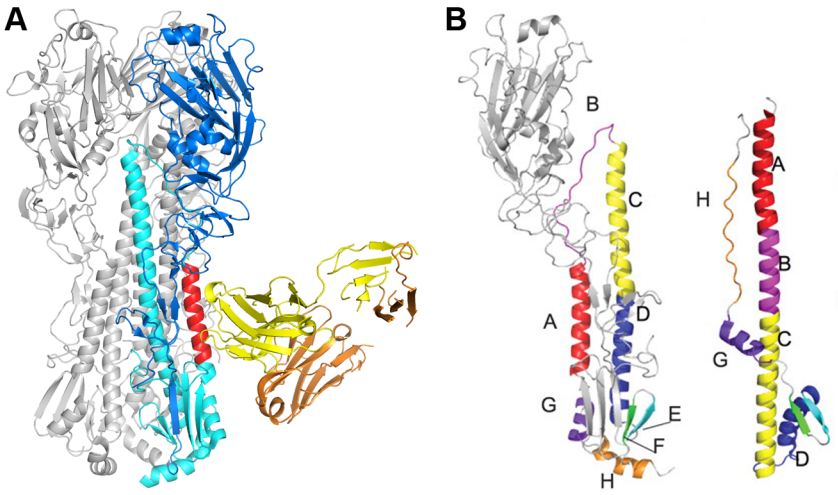

Figure 8: Structure of influenza virus HA. A: Ribbon tracing of trimeric HA in its pre-fusion state in complex with the CR6261 antibody (PDB code 3GBN). The antibody heavy and light chains are colored in yellow and orange; the HA1 and HA2 chains of one monomer are depicted in blue and cyan and the two other HA monomers are shown in gray. The A helix of HA2, to which the CR6261 antibody binds, is highlighted in red. **B:** Ribbon tracing of one HA subunit. Structural elements of the HA2 domain are labeled. Upon exposure to low pH and subsequent fusion, HA2 undergoes a drastic conformational change. Left: Pre-fusion state; right: post-fusion conformation. Modified from Ekiert et al, 2009.

The central CD helix of HA2 that forms a coiled coil is surrounded by several structural elements. One of them is the A helix, which changes its position drastically upon fusion (Figure 8B) (Bullough et al, 1994). In the post-fusion conformation, the A and CD helices no longer exist as separate structures, but they form one continuous helix. Additionally, the loop connecting helices A and CD becomes helical. The A helix sequence is highly conserved among influenza strains. In 2009, it was found that broadly neutralizing monoclonal antibodies such as CR6261 bind to the A helix of the HA2 domain in its pre-fusion state (Ekiert et al, 2009; Lingwood et al, 2012). This was surprising as most potent neutralizing influenza virus antibodies are directed against the receptor binding site of HA1, an area that is less conserved. Mutations frequently occur in this region, facilitating evasion from the host immune system (Wiley et al, 1981). Consequently, vaccines currently in use mainly protect against closely related influenza strains (Salzberg, 2008). A promising strategy in rational vaccine design therefore seeks to trigger the production of antibodies capable of broadly neutralizing influenza viruses by targeting conserved regions such as the A helix of HA2. In order to improve the accessibility of the A helix for the immune system, the epitope needs to be presented in a different context. The main

challenge of this approach is that the helix has to assemble correctly and that side chains must be oriented as in the native context so that the epitope is not distorted. Schneemann et al followed this strategy by placing the A helix into the capsid of Flock House virus (Schneemann et al, 2012). They were able to raise antibodies against the A helix. However, the antibodies were not able to neutralize influenza virus and to protect mice against influenza infection. This example illustrates the challenges of construct design for such purposes, and demonstrates that further optimization is necessary to successfully implement this strategy for vaccine development.

As the tail domain of the reovirus outer capsid protein σ1 was predicted to form an α-helical coiled coil, replacement of parts of the reovirus σ1 tail with the influenza virus A helix might serve as an alternative system to Flock House virus in rational vaccine design. To realize such a strategy, accurate structures of the σ1 coiled coil are needed.

1.2 Objectives

At the beginning of this study, it was known that the σ1 protein is a determinant of viral tropism. The distinct hemagglutination patterns of T1L and T3D reoviruses suggest that the two serotypes bind to different glycans. However, it was not known whether the differences in tropism are attributable to specific interactions of σ1 with carbohydrates on the host cell surface. T3D σ1 was known to bind sialyllactose. Although previous experiments suggested that T1L σ1 binds Neu5Ac at least in some contexts, the exact nature of the carbohydrate receptor remained undefined. There also was no experimental information on the structure of the σ1 tail region. The major aims of this study were therefore to:

- identify the carbohydrate receptor of T1L σ1
- determine the crystal structure of T1L σ1 at high resolution
- solve the crystal structure of a complex of T1L σ1 with its carbohydrate receptor
- define residues required for functional carbohydrate engagement
- determine whether specific interactions between σ1 and carbohydrates dictate viral tropism
- obtain information on the mechanism of viral attachment
- determine the structure of the σ1 tail
- establish the σ1 tail domain as a platform for development of vaccine vectors

1.3 Results and Discussion

1.3.1 The GM2 Glycan Serves as a Functional Co-Receptor for Serotype 1 Reovirus

Reiss K*, Stencel JE*, Liu Y, Blaum BS, Reiter DM, Feizi T, Dermody TS, Stehle T (2012). The GM2 Glycan Serves as a Functional Co-Receptor for Serotype 1 Reovirus. PLoS Pathogens, 8(12): e1003078.

*These authors contributed equally to this work.

Two constructs were used for structural and functional analyses of the T1L attachment protein. The first construct, T1L $\sigma1_{long}$ (aa 261-470), was predicted to form the three most C-terminal β-spiral repeats of the body and the C-terminal head domain. The second construct, T1L $\sigma1_{short}$ (aa 300-470), contains only the most C-terminal β-spiral repeat and the head domain. In addition to σ1 sequences, both constructs contained an N-terminal $(His)_6$-tag to facilitate protein purification and a $(GCN4)_3$ motif to stabilize the σ1 trimer (Harbury et al, 1994). Construct design, protein expression, and purification conditions are described in the publication.

Purified protein was used by Yan Liu and Ten Feizi (Imperial College London, UK) for glycan array analyses. These studies identified the GM2 glycan as a carbohydrate ligand for T1L σ1. The GM2 tetrasaccharide is a branched compound terminating in Neu5Ac and *N*-acetylgalactosamine (GalNAc) moieties. These residues are linked to a Gal and a Glc molecule. In the case of the GM2 ganglioside, which was present on the glycan array, Glc is further connected to a ceramide anchor. Soluble GM2 glycan, which lacks the lipid, was used in all other experiments. To validate the interaction between T1L σ1 and GM2 in solution, Saturation Transfer Difference (STD) NMR spectroscopy of T1L σ1 and the GM2 glycan was performed by Bärbel S. Blaum (University of Tübingen). This method is especially useful for detection of low affinity interactions between a small ligand, such as a carbohydrate, and a larger molecule. The STD NMR data demonstrated that T1L σ1 binds the GM2 oligosaccharide in solution and that the two terminal sugar moieties Neu5Ac and GalNAc contact the protein. Furthermore, the binding epitope of the carbohydrate could be determined. This

epitope comprises protons H5, H6, H7, and one of the two H9 protons of Neu5Ac and H1 through H4 of GalNAc. The methyl group of the *N*-acetyl chain of Neu5Ac gives the most prominent signal in the STD NMR spectrum, showing that this group is in close contact with the protein in the T1L σ1-GM2 glycan complex.

The crystal structure of T1L σ1 in complex with the GM2 oligosaccharide was determined at 3.6 Å resolution. Despite the intermediate resolution, the location of the carbohydrate binding site and the orientation of the sugar moieties in the T1L σ1-GM2 complex could be unambiguously defined in electron density maps. The GM2 glycan binds to the side of the T1L σ1 head domain and thus not near the predicted carbohydrate binding site within the body region. In agreement with the STD NMR data, the two terminal sugar moieties Neu5Ac and GalNAc contact σ1, while Gal and Glc point away from the protein. Neu5Ac inserts into a cleft between β-strands B and C, forming hydrogen bonds with backbone atoms of T1L σ1 and a hydrogen bond with the side chain of Q371. GalNAc attaches to the protein surface, forming van der Waals interactions with the protein.

In order to complement the structural data with functional studies, Jennifer E. Stencel and Terence S. Dermody (Vanderbilt University, Nashville, USA) performed infectivity studies. They showed that pre-incubation of T1L reovirus with the GM2 glycan blocks infectivity in a dose-dependent manner, while pre-incubation of T3D with this compound has no effect on infectivity. These data demonstrated that the GM2 oligosaccharide serves as a serotype-specific receptor for T1L reovirus. To validate the location of the carbohydrate binding site and to define residues involved in carbohydrate engagement, they further generated reoviruses carrying point mutations along the carbohydrate binding site of T1L σ1. The mutant viruses were tested for their carbohydrate binding capacity. These studies revealed that residues V354, S370, and Q371 are especially important for functional glycan engagement. T3D σ1 does not bind carbohydrates with its head domain, but instead uses its body region to bind glycans. Structural and sequence comparisons of the T1L σ1 carbohydrate binding site with T3D σ1 reveal that T3D carries bulkier residues along β-strands B and C. These bulky residues likely block the carbohydrate binding site

and therefore interfere with carbohydrate engagement. This conclusion was confirmed by mutational analyses.

The crystal structure of T1L σ1 in complex with the GM2 glycan demonstrates that Neu5Ac accounts for most of the interactions with the protein, while GalNAc contributes fewer interactions. We therefore tested whether the sugar compound lacking the GalNAc moiety also can serve as a receptor for T1L reovirus. This linear compound is the GM3 glycan, or α(2,3)-linked sialyllactose, which was shown to be the minimal receptor for T3D reovirus. We demonstrated that the GM3 glycan also binds to the T1L σ1 protein by STD NMR spectroscopy. The crystal structure of T1L σ1 in complex with this sugar shows that the GM3 glycan binds at the same site and in the same orientation as the branched GM2 glycan. However, infectivity assays suggest that T1L σ1 binds with higher affinity to GM2 than to GM3. Therefore, GM2 probably serves as the preferred co-receptor for Type 1 reoviruses *in vivo*.

From these studies, we conclude that T1L and T3D reovirus utilize different carbohydrates as co-receptors and possess very different carbohydrate binding sites. In contrast, JAM-A serves as a high affinity receptor for both serotypes, and the JAM-A binding site is well conserved between T1L and T3D σ1. As both serotypes make use of a common proteinaceous receptor and different carbohydrate receptors, it is possible that the mechanism of cell attachment varies between different serotypes. Thus, a major aim of future research in this field is to define the combined usage of proteinaceous and carbohydrate receptors and elucidate the contribution of the individual receptors to tropism and pathogenicity.

1.3.2 **Crystal structure of the T1L σ1 head domain**

To obtain a higher resolution structure of T1L σ1, T1L $\sigma1_{short}$ was crystallized. The structure of this protein was solved by molecular replacement and refined to 2.2 Å resolution (Table 1). The final model includes aa 307-469. Additional residues at the N- and C-termini are poorly ordered. $(His)_6$-tag and $(GCN4)_3$ motif, which were not removed during protein purification, are not visible in the

electron density map. As there is enough space in the crystal lattice it is likely that this part is not well ordered.

Table 1: Data collection and refinement statistics of the T1L σ1$_{short}$ crystal structure.

Data collection	
Resolution (Å)	50-2.20 (2.26-2.20)
Space group	$I2_12_12_1$
a, b, c (Å)	112.9, 113.0, 113.2
α, β, γ (°)	90, 90, 90
R_{meas} (%)	13.0 (131.8)
$CC_{1/2}$ (%)*	99.8 (56.1)
λ(Å)	1.0
I/σ(I)	11.7 (1.6)
Completeness (%)	98.7 (98.9)
Total reflections	188295 (13812)
Unique reflections	36975 (2690)
Redundancy	5.1
Wilson B factor (Å^2)	41.8
Refinement	
R_{work}/R_{free} (%)**	17.9/21.1
Protein atoms	3869
Number of molecules	
Water	106
Imidazole	2
Mg^{2+}	3
Glycerol	12
Ethylene glycol	23
Diethylene glycol	8
B factors (Å^2)	
Chain A	37.9
Chain B	37.7
Chain C	38.0
Water	48.7
Imidazole	74.2
Mg^{2+}	50.8
Glycerol	59.7
Ethylene glycol	61.8
Diethylene glycol	67.3
r.m.s.d.	
Bond lengths (Å)	0.010
Bond angles (°)	1.10
Ramachandran plot	
Favored (%)	469 (97.1)
Allowed (%)	11 (2.3)
Outliers (%)	3 (0.6)

* $CC_{1/2}$ = correlation coefficient (Karplus & Diederichs, 2012).
** R_{free} was calculated with 10 % of the data.
The Ramachandran plot was calculated with Rampage (CCP4, Lovell et al, 2003).

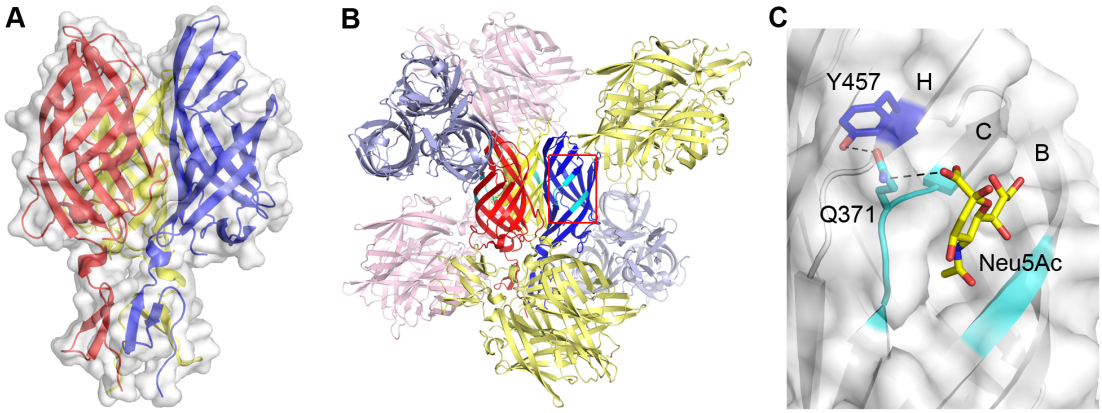

Figure 9: Crystal structure of the T1L σ1 head domain. A: Overall structure of the T1L σ1 trimer shown as ribbon representation. The σ1 monomers are colored in blue, red, and yellow. The protein surface is colored in gray. **B:** Crystal packing of T1L $\sigma1_{short}$. One trimer is present per asymmetric unit and is colored as in A. Symmetry-related molecules are shown in light blue, yellow, and pink. The carbohydrate binding site is highlighted in cyan and is surrounded with a red box. **C:** Close-up view of the carbohydrate binding site. A model for Neu5Ac engagement was generated by superimposing T1L $\sigma1_{short}$ with the crystal structure of T1L $\sigma1_{long}$ in complex with the GM2 glycan. The carbohydrate binding site, formed by β-strands B and C, is colored in cyan. Residue Q371, which is important for glycan engagement, is shown in stick representation. Residue Y457 from the neighboring β-strand H is shown as sticks in blue. Possible hydrogen bonds between Y457, Q371, and Neu5Ac are indicated with dashed lines.

The T1L $\sigma1_{short}$ protein folds into a β-spiral repeat and the compact head domain (Figure 9A). Due to the good resolution, main chain and most side chain atoms are clearly visible. The only exception is the loop connecting β-strands E and F (aa 419-423), which is not well defined in the electron density map due to flexibility.

The carbohydrate binding site is not blocked by crystal contacts and well accessible for the glycan (Figure 9B). However, a crystal structure of T1L $\sigma1_{short}$ in complex with its carbohydrate receptor at high resolution could not be obtained. The main problems were that only few crystals grown in this condition diffracted well and that crystals could only be obtained using the sitting drop vapor diffusion method. With this technique, many crystals stick to the surface of the well. These crystals are difficult to remove without being harmed or destroyed. The few well-diffracting crystals were damaged by the soaking procedure. To circumvent the problem of crystal soaking, co-crystals of T1L $\sigma1_{short}$ and the GM2 glycan were set up. The GM2 oligosaccharide was added to the protein solution to a final concentration of 20 mM prior to crystallization. However, no crystals appeared under this condition. Using a lower concentration

of GM2 for co-crystallization might be an option to obtain well-diffracting crystals of T1L σ1 in complex with the GM2 oligosaccharide.

Only poor electron density is visible for the side chain of Y457 in the T1L $\sigma1_{short}$ crystal structure. It is unusual that an aromatic residue is located on the protein surface. As shown in Figure 9C, Y457 is located on β-strand H, next to β-strand B, which is part of the carbohydrate binding site. In one of its three allowed conformations, Y457 might form a hydrogen bond with Q371, a residue that is important for functional carbohydrate engagement. Due to this arrangement, it is possible that Y457 undergoes an induced fit movement upon glycan binding and helps to orient and stabilize Q371 in the protein-carbohydrate complex. A high resolution structure of T1L σ1 in complex with the GM2 glycan would help to determine whether such an induced fit mechanism indeed takes place.

1.3.3 **The σ1 tail**

1.3.3.1 ***Coiled coil prediction and construct design***

In 1990, the domain organization and conserved regions of σ1 were predicted based on sequence analyses (Duncan et al, 1990; Nibert et al, 1990). Although the correct oligomeric state of σ1 was not known, it was thought that the σ1 tail forms an α-helical coiled coil as the sequence comprises the heptad repeat pattern characteristic of such structures (Bassel-Duby et al, 1985; Furlong et al, 1988). Today, software tools are available that facilitate such structural predictions. The result of a sequence analysis of the T1L σ1 tail (aa 1-180) calculated with COILS (Protein Bioinformatics & Computational Biology, Gene Center, LMU Munich) is given in Figure 10. The program takes windows of 14, 21, or 28 amino acids into account and calculates the probability (or score) of coiled coil formation at each position (Appendix, Table 4). For T1L σ1, the results are very similar for all windows. The T1L σ1 tail shows a high coiled coil score in two defined regions. The first region roughly comprises amino acids 25 to 80, and the second one includes residues 120 to 160. In contrast, the coiled coil probability is low between these two segments and within the 20 most N-terminal residues, which anchor the σ1 protein into the viral capsid. This

analysis suggests that the T1L σ1 tail consists of two coiled coil regions and an irregular structure in between.

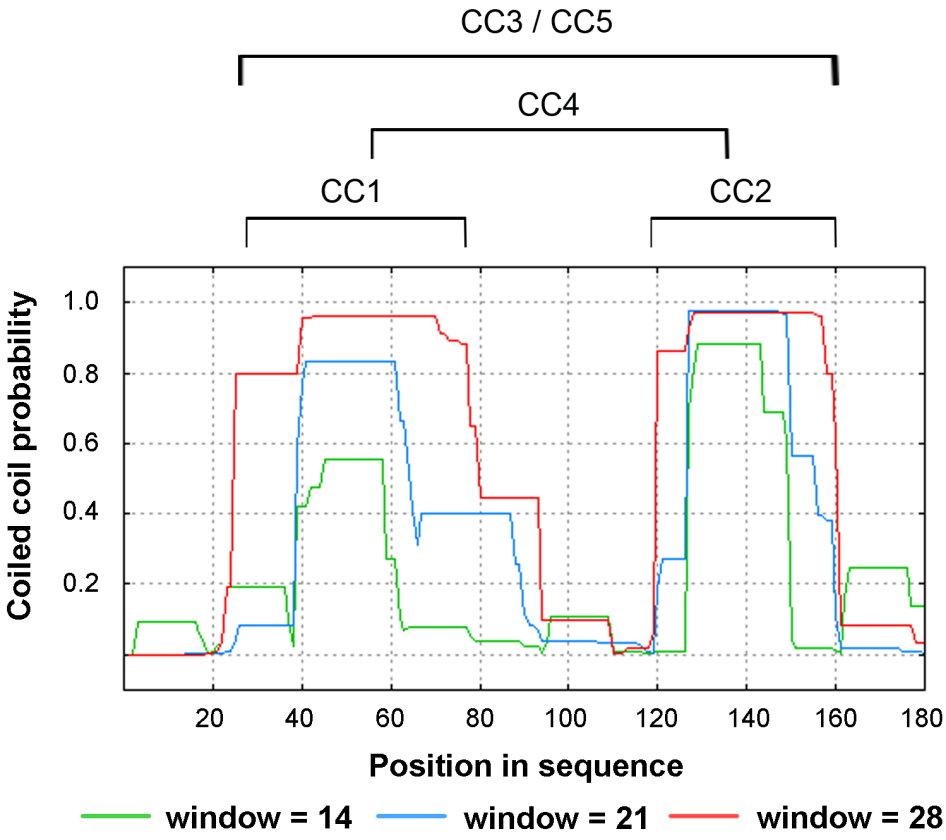

Figure 10: Coiled coil probabilities for the T1L σ1 tail (aa 1-180) were calculated with COILS (Protein Bioinformatics & Computational Biology, Gene Center, LMU Munich). Construct design for CC1-CC5 is indicated.

Based on this analysis, several constructs were designed that span different regions of the T1L σ1 tail. They are included in Figure 10. Construct CC1 comprises the more N-terminal area with high coiled coil scores (aa 29-76); CC2 spans the more C-terminal heptad region (aa 120-160). CC3 (aa 29-160) includes CC1, CC2, and the region with low scores. Construct CC4 (aa 57-139) represents a shortened version of CC3, which also includes the region with low scores, and starts and ends with residues that are predicted to form a coiled coil.

All constructs were cloned into the vector pIBA-GCN4tri-His, which is depicted in Figure 11 (provided by Dirk Linke, MPI for Developmental Biology, Tübingen) (Hernandez Alvarez et al, 2008). This vector attaches a $(GCN4)_3$ motif N- and C-terminally to the σ1 segment to stabilize the coiled coil protein. To generate a

continuous coiled coil structure, the predicted coiled coil sequence has to be in frame with the heptad repeat pattern of $(GCN4)_3$. Therefore, the Type IIS restriction enzyme BsaI was used. This enzyme does not cleave the DNA within its recognition sequence but a defined number of bases next to it. Thus, the resulting construct does not contain any additional residues from the cloning procedure and solely comprises $(GCN4)_3$ and σ1 sequences. CC1-CC4 furthermore carry C-terminal $(His)_6$-tags to facilitate protein purification.

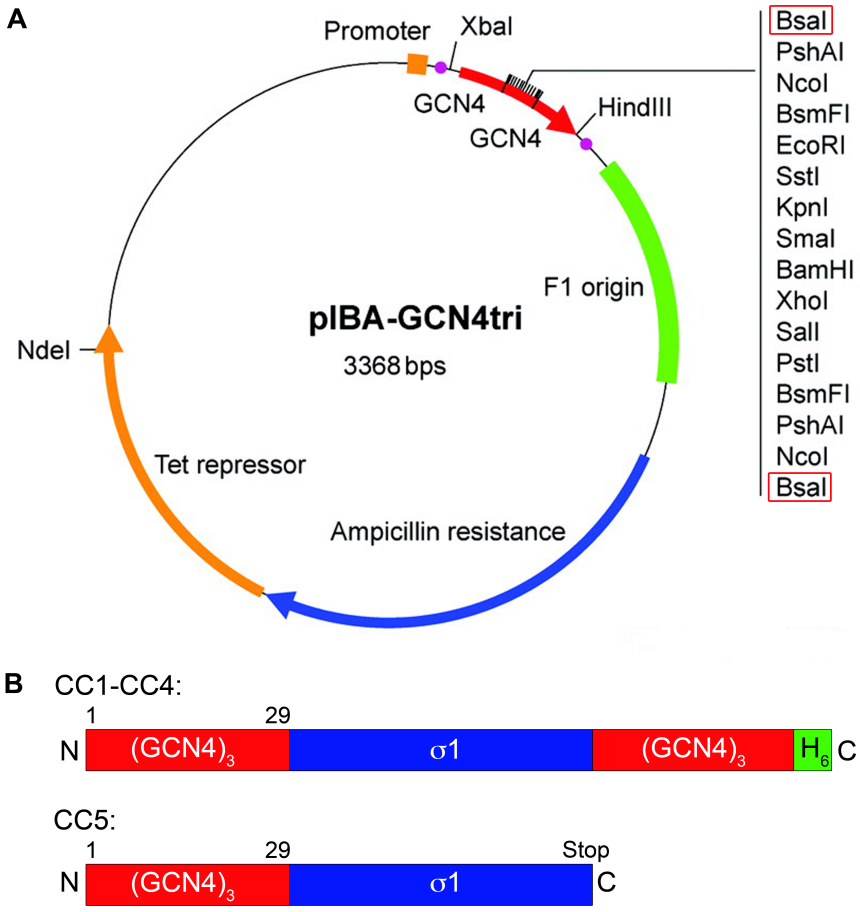

Figure 11: Expression system for σ1 coiled coil constructs. A: The σ1 sequences were cloned into the vector pIBA-GCN4tri-His. This vector adds the 29 amino acid long $(GCN4)_3$ motif N- and C-terminally to the σ1 sequence. The restriction enzyme BsaI was used. **B**: Schematic representation of coiled coil constructs: in CC1, CC2, CC3, and CC4, the σ1 sequence is flanked by two $(GCN4)_3$ domains, and a C-terminal $(His)_6$-tag is present. In construct CC5, a stop codon was introduced at the C-terminus of σ1. Therefore, the C-terminal $(GCN4)_3$ motif and the (His_6)-tag are absent. Modified from Hernandez Alvarez et al, 2008.

A possible interpretation of the coiled coil prediction pattern is that CC1 and CC2 each form a coiled coil region and that the region in between is flexible or has an irregular structure. As the σ1 molecule undergoes conformational changes

during attachment and internalization (Dryden et al, 1993), it is conceivable that when the flexible region kinks, the two trimeric coiled coils CC1 and CC2 could bind to each other (Figure 12A). In this situation, the two trimeric coiled coils would stay intact, and would run in opposite directions. An alternative possibility is that, under certain conditions, a structure similar to HIV fusion protein gp41 in its post-fusion state is formed (Chan et al, 1997). The sequence of HIV gp41 comprises two regions with heptad repeats that are connected by a linker. In the post-fusion conformation, the more N-terminal heptad of HIV gp41 forms a central trimeric α-helical coiled coil and the C-terminal heptad forms three individual α-helices that arrange around the N-terminal coiled coil (Figure 12B). In the resulting six-helix bundle, the C-terminal helices run in antiparallel direction to the central coiled coil. Hydrophobic residues at positions *a* and *d* of the C-terminal helices point to the inside of the six-helix bundle.

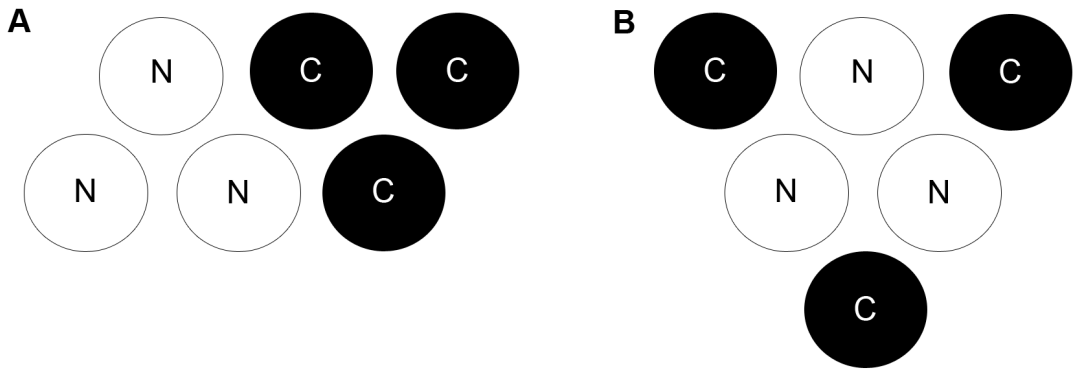

Figure 12: Possible conformations of the T1L σ1 tail domain. The more N-terminal helices are shown as white circles; the C-terminal helices are depicted as black circles. N- and C-termini are labeled. **A:** The two trimeric coiled coils bind to each other. They run in antiparallel directions. **B:** The N-terminal coiled coil stays intact, while the C-terminal trimer dissociates into three individual helices. Each helix binds in the gap between two helices of the N-terminal coiled coil. N- and C-terminal helices run in opposite directions. This conformation is analogous to HIV gp41 in its post-fusion state.

Both conceivable conformations of the σ1 tail could be unstable in the context of CC3: CC3 spans CC1 through CC2 and carries $(GCN4)_3$ motifs both N- and C-terminally to the σ1 sequence. If CC1 and CC2 interact with each other as assumed in the first scenario, the two $(GCN4)_3$ domains will also closely approach each other, which might not be favorable. If structural rearrangements similar to the six-helix bundle formation in HIV gp41 take place in σ1, both the

CC2 trimer and the C-terminal $(GCN4)_3$ motif would have to dissociate into three monomers. It is unlikely that the very stable $(GCN4)_3$ domain would undergo this change. Therefore, construct CC5 was designed. It spans residues 29-160 and carries a $(GCN4)_3$ domain at the N- but not the C-terminus. However, this construct has no C-terminal $(His)_6$-tag, which complicates the protein purification process.

1.3.3.2 *Expression and purification of coiled coil segments*

Expression and purification of CC1 and CC2 were performed as described by Hernandez Alvarez et al (construct SadAK3His) (Hernandez Alvarez et al, 2008). The protein was expressed in *E. coli* BL21 (DE3) cells at 37°C and purified using denaturing conditions. After a Ni affinity chromatography step, the denatured protein was refolded by removing the denaturant guanidinium hydrochloride via dialysis. The protein was subjected to size-exclusion chromatography. CC1 and CC2 eluted as a single peak from the gel filtration column. The elution time suggested that the oligomeric state of the protein was a trimer. The protein was concentrated and used for crystallization experiments.

CC3 was expressed and purified using the same conditions employed for CC1 and CC2. However, most of the protein precipitated during the refolding step, and the final amount of protein was too small to allow further characterization or crystallization.

Construct CC5 does not contain a $(His)_6$-tag. Therefore, the protein had to be purified from inclusion bodies. In the end, yield and purity of CC5 were too low for crystallization purposes. Construct CC4 is a shortened version of CC3 and CC5. It might therefore lead to higher protein yields that allow structural studies. However, experiments have yet to be performed with CC4.

1.3.3.3 *Crystallization of coiled coil segments*

Crystals of CC1 were obtained with the sitting drop vapor diffusion method using protein at a concentration of 6-8 mg/ml. Crystals grew at 4°C in 0.5 M $(NH_4)_2SO_4$, 0.1 M HEPES/NaOH (pH 7.5), 30 % v/v (+/-)-2-methyl-2,4-pentanediol. No additional cryoprotectant was necessary. The crystals diffracted to 2.2 Å

resolution. CC2 was crystallized with the hanging drop vapor diffusion method at 4°C in 0.1 M Tris/HCl (pH 8.5), 20 % ethanol at a concentration of 5-7 mg/ml. Glycerol 20 % served as cryoprotectant. The crystals diffracted to 2.5 Å resolution. In both cases, complete datasets were collected at the Swiss Light Source beamline X06SA and processed with XDS and XSCALE from the XDS program package (Kabsch, 2010). The CC1 and CC2 crystal structures were solved by molecular replacement using MOLREP (CCP4, Vagin & Teplyakov, 2000; Vagin & Isupov, 2001; Vagin & Teplyakov, 2010) and the coordinates of the trimeric $(GCN4)_3$ leucine zipper (PDB code 1GCM, Harbury et al, 1994) as a search model. Structural refinement was performed using Refmac5 (CCP4, Murshudov et al, 1997), Phenix (Adams et al, 2002), and autoBUSTER (Bricogne et al, 2011; Smart et al, 2012).

1.3.3.4 *Crystal structure of CC1*

The CC1 data were processed in space group P3 (Table 2). Two copies of CC1, chain A and B, are present per asymmetric unit. The crystal structure of the CC1 segment includes residues 2-107 in chain A and 2-109 in chain B. The two most N-terminal residues and part of the C-terminal $(His)_6$-tag are not visible in the electron density map. The threefold axis runs along the trimer axis of the α-helical coiled coil (Figure 13A). Thus, the two trimeric coiled coils of chain A and B are completed by symmetry-related molecules.

Table 2: Data collection and refinement statistics of the CC1 crystal structure.

Data collection	
Resolution (Å)	50-2.20 (2.26-2.20)
Space group	P3
a, b, c (Å)	39.72, 39.72, 169.82
α, β, γ (°)	90.0, 90.0, 120.0
R_{meas} (%)	13.1 (83.3)
$CC_{1/2}$ (%)*	99.9 (78.3)
λ(Å)	0.98
I/σ(I)	12.4 (2.5)
Completeness (%)	99.9 (99.9)
Total reflections	98152 (6914)
Unique reflections	15242 (1076)
Redundancy	6.4
Wilson B factor (Å^2)	37.7

Refinement	
R_{work}/R_{free} (%)**	20.8/24.9
Protein atoms	1719
Number of molecules	
Water	173
Cl^-	2
2-Methyl-2,4-pentanediol	2
Glycerol	2
B factors (Å²)	
Chain A	34.3
Chain B	34.4
Water	38.8
Cl^-	21.8
2-Methyl-2,4-pentanediol	57.1
Glycerol	53.4
r.m.s.d.	
Bond lengths (Å)	0.010
Bond angles (°)	0.97
Ramachandran plot	
Favored (%)	209 (99.5)
Allowed (%)	1 (0.5)
Outliers (%)	0

* $CC_{1/2}$ = correlation coefficient (Karplus & Diederichs, 2012).
** R_{free} was calculated with 10 % of the data.
The Ramachandran plot was calculated with Rampage (CCP4, Lovell et al, 2003).

The crystal structure is composed of an uninterrupted α-helical coiled coil (Figure 13). Due to the good resolution of 2.2 Å, side chains are clearly visible in the electron density map and the correct amino acid position of the helix could be assigned unambiguously. Chains A and B run in opposite directions (Figure 13A). In addition to van der Waals interactions, crystal contacts include two salt bridges between residues D50 and R54 from each chain (Figure 13C). The structure of the two $(GCN4)_3$ motifs are very similar to the previously solved structure of the individual $(GCN4)_3$ domain (PBD code 1GCM, Harbury et al, 1994). Their main chain atoms and 1GCM superpose with r.m.s.d. values of 0.43 and 0.74 Å, respectively. Within the σ1 segment, each individual helix is stabilized by a salt bridge between D40 and R43 and by a water-mediated hydrogen bond between D41 and N38. The composition of the trimer interface of both $(GCN4)_3$ and the σ1 segment fulfill the coiled coil prediction: hydrophobic residues at positions *a* and *d* of the heptad repeat form the trimer interface, while polar residues face solvent. Within the σ1 segment, amino acids V, L, A, and N occupy positions *a* and *d* and are therefore in frame with the coiled coil

prediction (Appendix, Table 4). The trimer interface also comprises a chloride ion, which probably bound to CC1 during protein purification or crystallization. The chloride ion is flanked by the side chains of V35 and N38. The temperature factors of the chloride ion and the neighboring amino acids V35 and N38 are in the same range indicating full occupancy of the ion (B-factors: Cl^-: 23.4 Å², 20.2 Å², V35: 26.1 Å². 25.2 Å², N38: 23.6 Å², 22.8 Å², in chain A and B, respectively). The σ1 trimer is further stabilized by salt bridges between the R69 side chain and the side chain of D74 from the neighboring helix (Figure 13D). Such interhelical salt bridges of charged side chains at position *g* with position *e* of the following heptad on the neighboring helix frequently occur in trimeric coiled coil structures (Harbury et al, 1994). Due to these stabilizing interactions within the σ1 segment, it is likely that this part of the σ1 tail forms the same structure without the flanking $(GCN4)_3$ motifs.

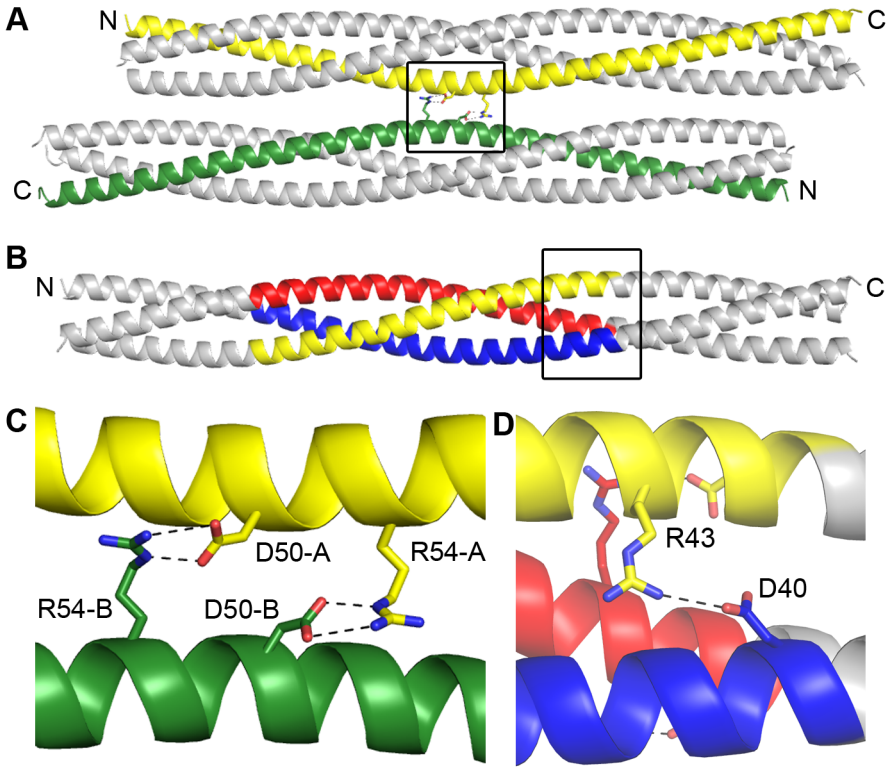

Figure 13: Crystal structure of CC1. A: Crystal packing of CC1 shown as ribbon representation. The two chains A and B that form the asymmetric unit are colored in yellow and green. Symmetry-related molecules that complete the two σ1 trimers are depicted in gray**. B:** Overall structure of CC1. The σ1 segment (aa 29-76) is colored in blue, red, and yellow; the $(GCN4)_3$ motifs attached N- and C-terminally to the σ1 sequence are depicted in gray. N- and C-termini are labeled. **C:** Close-up view of a salt bridge between chain A and B that stabilizes the crystal. **D:** close-up view of interhelical salt bridges between R69 and D74 that stabilizes the trimeric α-helical coiled coil.

1.3.3.5 *Crystal structure of CC2*

The data were processed in space group C2 (Table 3). The three chains of CC2 in the asymmetric unit form a coiled coil trimer. The final model comprises all amino acids present in the fusion protein, including the N- and C-terminal $(GCN4)_3$ motifs and the C-terminal $(His)_6$-tag.

Table 3: Data collection and refinement statistics of the CC2 crystal structure.

Data collection	
Resolution (Å)	50-2.50 (2.57-2.50)
Space group	C2
a, b, c (Å)	71.9, 41.5, 157.7
α, β, γ (°)	90.0, 98.7, 90.0
R_{meas} (%)	10.4 (123.5)
$CC_{1/2}$ (%)*	100 (85.6)
λ(Å)	0.98
I/σ(I)	12.9 (1.6)
Completeness (%)	99.5 (99.8)
Total reflections	67454 (5135)
Unique reflections	16233 (1224)
Redundancy	4.2
Wilson B factor (Å^2)	46.3
Refinement	
R_{work}/R_{free} (%)**	24.8/29.3
Protein atoms	2612
Number of molecules	
Ca^{2+}	1
Water	97
Ethanol	5
Glycerol	1
B factors (Å^2)	
Chain A	54.5
Chain B	53.0
Chain C	53.2
Ca^{2+}	41.9
Water	49.5
Ethanol	54.1
Glycerol	54.7
r.m.s.d.	
Bond lengths (Å)	0.010
Bond angles (°)	1.13
Ramachandran plot	
Favored (%)	312 (100)

* $CC_{1/2}$ = correlation coefficient (Karplus & Diederichs, 2012).
** R_{free} was calculated with 10 % of the data.
The Ramachandran plot was calculated with Rampage (CCP4, Lovell et al, 2003).

The R-factors of the CC2 model, 24.8 (R_{work}) and 29.3 % (R_{free}), are rather high. This is probably caused by minor pseudo-translational symmetry or twinning of the crystals. The problem of twinned crystals has been described for other proteins expressed with this vector system (Hernandez Alvarez et al, 2008).

As with CC1, the CC2 sequence forms an uninterrupted trimeric α-helical coiled coil (Figure 14A). As predicted, hydrophobic residues at positions *a* and *d* form the hydrophobic trimer interface, and polar residues face solvent. The correct assignment of the sequence within the helix could be unambiguously carried out by location of characteristic glycine residues, which do not show any electron density for side chains.

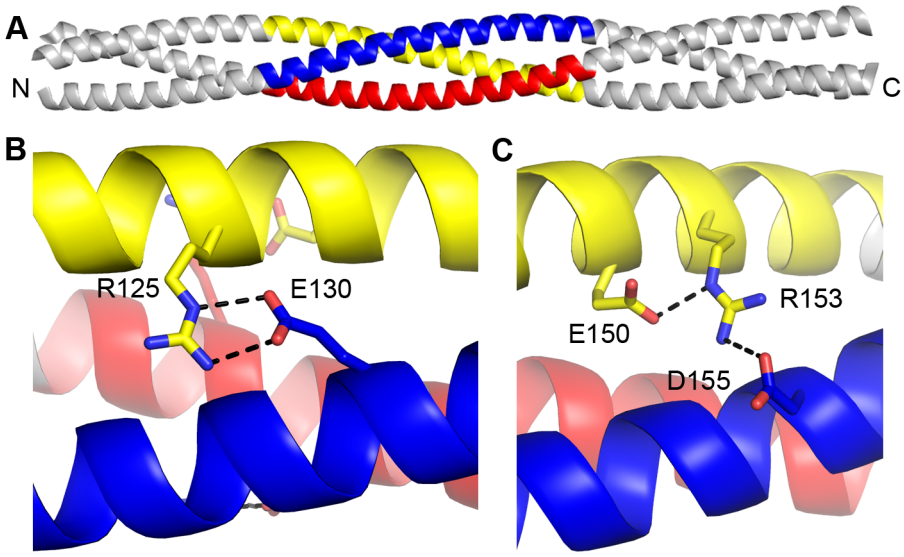

Figure 14: Crystal structure of CC2. A: A portion of the T1L σ1 tail domain (residues 120-160) is shown as a ribbon tracing in blue, red, and yellow. The two $(GCN4)_3$ motifs attached N- and C-terminally to the σ1 segment are depicted in gray. Interhelical salt bridges between R125 and E130 (**B**) and between R153 and D155, and a salt bridge between E150 and R153 on the same helix (**C**) are given as close-up views. Residues involved in these interactions are shown as sticks; contacts are shown with black dashes.

Several interhelical salt bridges stabilize the trimeric structure. The salt bridges between E130 and R125 (Figure 14B) and between R137 and D132 each connect charged residues at positions *e* and *g*, which is typical for trimeric α-helical coiled coils. R137 also forms a salt bridge with E139, again a residue at position *g*. An exception from this rule is the salt bridge between R153 at position *g* and D155 at position *b* from the neighboring helix. R153 also forms a salt bridge with E150 at position *d* in the same chain (Figure 14C). A charged residue at position

d is not favored. Hence, the C_β and C_γ atom of the glutamate side chain contribute hydrophobic interactions at the coiled coil interface, while the carboxyl group of the glutamate side chain points to the outside, where it forms a salt bridge with R153 (Figure 14C). Additionally, a hydrogen bond is present between the hydroxyl group of T148 (*b*) and the R146 (*g*) side chain of the neighboring helix. Interactions within the σ1 segment of the coiled coil are summarized in Figure 15. Moreover, a bivalent ion, probably Ca^{2+}, is coordinated by six histidine residues from the $(His)_6$-tag at the threefold axis of the trimer.

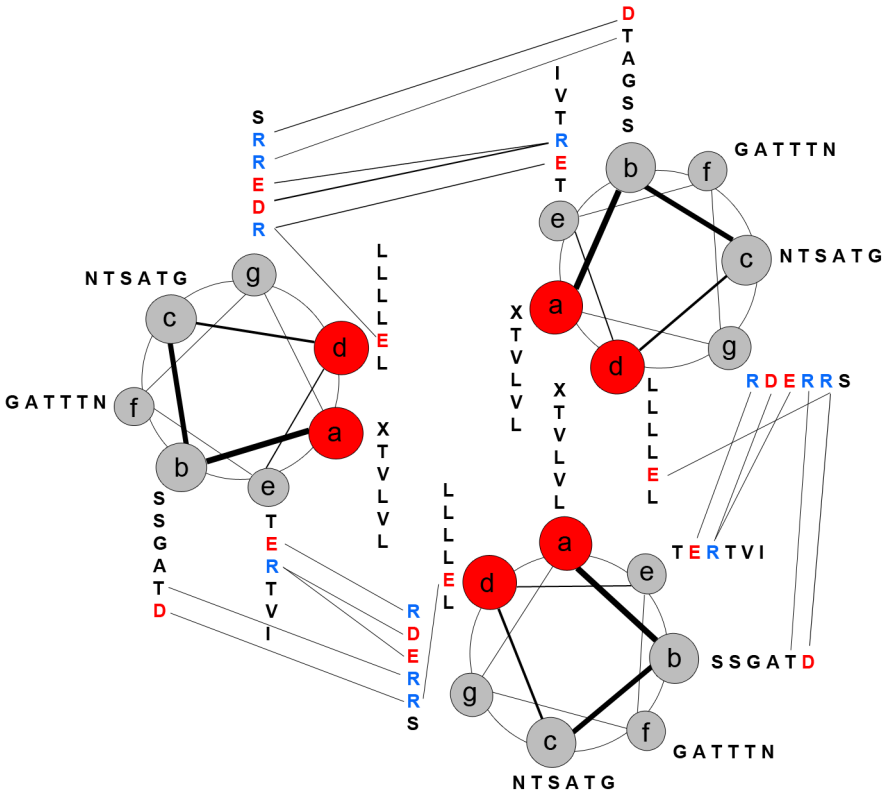

Figure 15: Stabilizing interactions within the σ1 segment of CC2. Helical wheel projection of T1L σ1 (aa 120-160), viewed from the N-terminus of the coiled coil. Heptad positions *a*-*g* are labeled. Positions *a* and *d* that form the hydrophobic trimer interface are highlighted in red. The amino acid composition at each position is included. Basic and acidic residues that are involved in interactions are colored in blue and red, respectively. Salt bridges and hydrogen bonds are shown with black lines.

1.3.3.6 *Outlook*

The fusion proteins CC1 and CC2 both form uninterrupted α-helical coiled coil structures. However, at this point it cannot be excluded that coiled coil formation

of the σ1 segments is artificially induced by the flanking $(GCN4)_3$ motifs and that σ1 does not form a coiled coil on its own. The stabilizing interactions within the hydrophobic σ1 trimer core and the numerous interhelical salt bridges of CC1 and CC2 suggest that these regions of the σ1 tail form an α-helical coiled coil, but to confirm the existence of this structure in the virus, the σ1 segments need to be analyzed without the $(GCN4)_3$ domains. The secondary structure composition, α-helical content, and oligomeric state could be determined by CD spectroscopy and analytical gel filtration. If these data show that the σ1 tail segments form trimeric α-helical coiled coils on their own, it will be interesting to define the conditions under which this structure is formed and how stable it is.

The expression and crystallization of CC3, CC4, or CC5, which include the region with low coiled coil scores between CC1 and CC2, would reveal if the σ1 tail comprises a continuous coiled coil, or if a structurally distinct area is inserted. It would be interesting to determine the conditions under which the structure is formed as this information could give insight into structural rearrangements taking place in this region during viral attachment, entry, and uncoating. A major aim of future research will be to link protein structures to σ1 dynamics and conformational changes during viral uptake. An interesting method to detect and characterize conformational changes and different states of a protein is small angle X-ray scattering. An advantage of this technique over X-ray crystallography is that the protein is analyzed in solution. Nevertheless, low resolution structures of proteins and information on flexibility and domain organization can be obtained.

1.3.3.7 *The σ1 tail: a platform for vaccine development*

As described in section 1.1.4, the A helix of influenza virus HA2 is a conserved part of the virus and therefore a target for rational vaccine design. However, in its natural context the A helix is not well exposed to the immune system. The A helix therefore needs to be presented in a different system to induce the production of antibodies against this region. The σ1 tail domain could serve as a platform to present epitopes like the A helix to the immune system. An advantage of reovirus σ1 is that the A helix would be flanked by trimeric α-

helical coiled coils, which would stabilize the secondary structure of this element. Due to the trimeric nature of σ1, three epitopes of the A helix would be presented per σ1 molecule. Further, the σ1 tail probably consists predominantly of coiled coils. Thus, several A helix "building blocks" could be inserted sequentially into the σ1 tail without distorting the overall structure of the protein. By displaying several A helices per molecule, it should be possible to enhance the antibody response against this epitope. In addition, reoviruses are only rarely associated with disease in humans. Compared to other viral delivery or vaccination systems, the administration of reoviruses would likely be associated with few side effects.

A

Residue		Heptad position
38	Q	a
39	K	b
40	S	c
41	T	d
42	Q	e
43	N	f
44	A	g
45	I	a
46	D	b
47	G	c
48	I	d
49	T	e
50	N	f
51	K	g
52	V	a
53	N	b
54	S	c
55	V	d
56	I	e
57	E	f
58	K	g

B

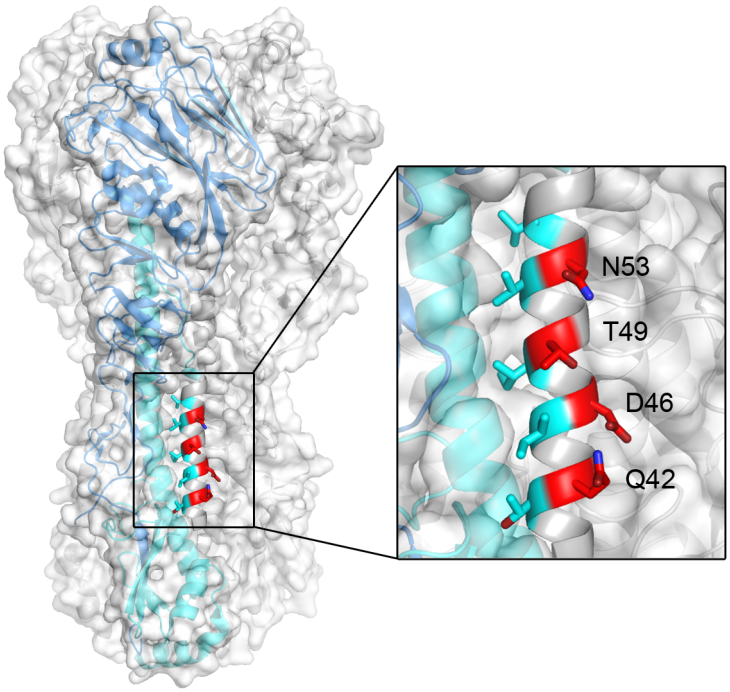

Figure 16: The A helix of influenza virus HA2. A: Amino acid composition and heptad register of the A helix. Positions *a* and *d* that point to the inside of HA are colored in cyan. Residues at positions *b* and *e* forming the epitope in the HA-CR6261 complex are highlighted in red. **B:** Crystal structure of HA with a close-up view of the A helix. In one monomer, HA1 and HA2 are colored in blue and cyan, respectively. The two other HA subunits and the protein surface are shown in gray. Residues that are highlighted in (A) are shown as sticks using the same color code (PDB code 3GBN).

The crystal structures of T1L σ1 tail segments are very useful for the design of σ1 proteins with influenza virus A helix insertions or substitutions. These chimeric proteins can be evaluated as immunogens in animal models. The A helix

is a straight helix composed of only 22 amino acids. CC1 and CC2 possess some left-handed supercoiling, but within such a short segment the supercoiling is negligible. For this reason, the main chain atoms of CC1 and the A helix superpose well, with an r.m.s.d. value of 0.77 Å.

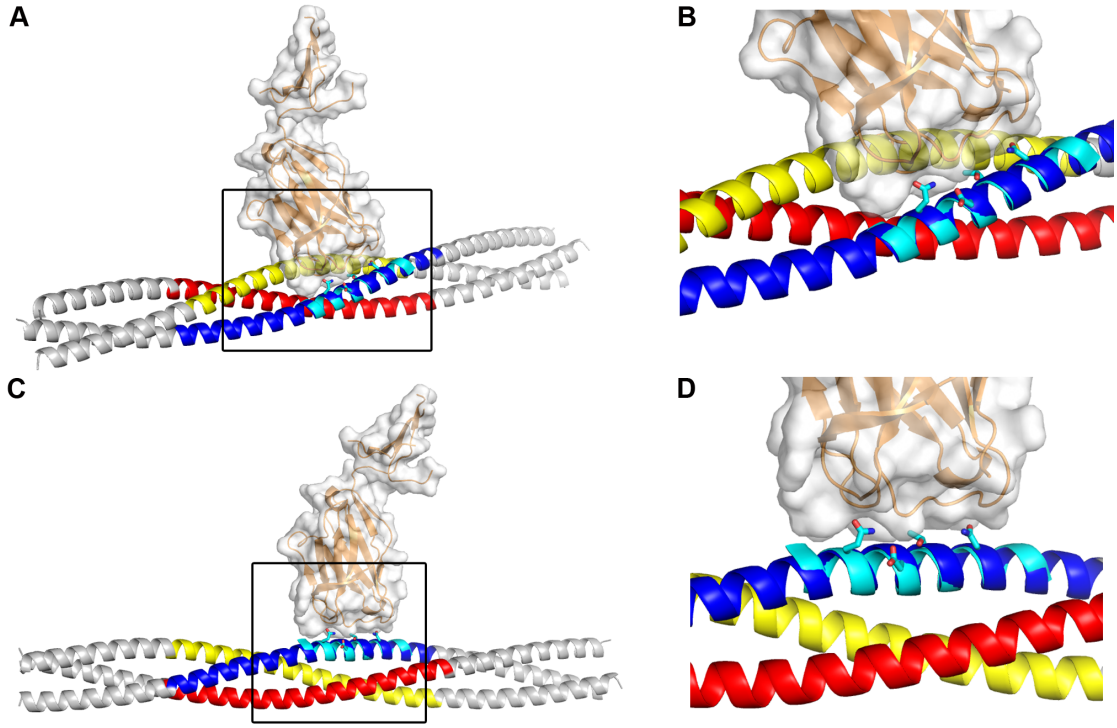

Figure 17: Possibilities to insert the A helix into the coiled coil σ1 tail. The crystal structure of CC1 is shown as ribbon drawing with the $(GCN4)_3$ motifs colored in gray and the σ1 segment in red, blue, and yellow. The A helix is colored in cyan. Residues that build the epitope are shown as sticks. The heavy chain of antibody CR6261 bound to the A helix is colored in orange with its surface depicted in gray. **A, B:** Superposition of the A helix (aa 37-58) with CC1 (residues 48-69). The A helix is inserted in frame with the σ1 heptad repeat pattern with positions *a* and *d* pointing to the trimer interface. The A helix epitope is presented at positions *e* and *b*. In this construct, the CR6261 antibody would clash with the neighboring helix. **C, D:** Superposition of the A helix (residues 38-58) with CC1 (residues 50-70). The A helix is rotated by one position in the heptad register compared to panels A and B. Thus, the A helix epitope is presented at positions *f* and *c*. There is no clash between the CR6261 antibody and the coiled coil.

The A helix sequence comprises a heptad repeat pattern with hydrophobic residues at positions *a* and *d* pointing to the inside of the HA protein, and polar residues facing the outside (Figure 16A, B). The conserved epitope bound by the antibody CR6261 is composed of amino acids Q42, D46, T49, and N53 (PDB code 3GBN, Ekiert et al, 2009). These residues alternately occupy positions *e* and *b*. The most obvious approach to insert the A helix sequence into the σ1 tail is to replace σ1 residues (for example aa 48-69 of CC1) with A helix sequences (aa 37-

58). Thereby, the A helix would be introduced in frame with the σ1 heptad register and positions *a* and *d* would be kept inside the coiled coil. The A helix epitope would be presented at positions e and b. However, the superposition of the A helix-CR6261 complex with CC1 reveals that this orientation would lead to clashes between the antibody's heavy chain and the neighboring helix within the CC1 coiled coil (Figure 17A, B). To substitute a σ1 segment with the A helix in frame with the σ1 heptad register is therefore not a promising strategy. Instead, the epitope residues have to be realigned within the heptad register to a more exposed position that avoids clashes between the antibody and neighboring coiled coil helices. This can be achieved by introducing point mutations into the σ1 tail, which generate the A helix epitope. This situation was simulated by the superposition of main chain atoms of the A helix (aa 38-58) with CC1 residues 50-70 (Figure 17C, D). In this scenario, the A helix is turned by one position in the heptad repeat pattern. Thus, residues not at positions *e* and *b* but at *f* and *c* are replaced with HA epitope sequences. In this arrangement, there are no clashes between the CR6261 antibody and the coiled coil.

To trigger the production of antibodies directed against the A helix, Schneeman et al followed a similar approach by replacing a helical segment within the B2 capsid protein of Flock House virus with the influenza virus A helix (Schneemann et al, 2012). After immunization with virus-like particles presenting these chimeric proteins, mice produced antibodies against the A helix. However, these antibodies were not able to neutralize influenza viruses and protect against influenza virus infection. The main problem was that the orientation of the A helix in influenza HA2 is reversed compared to the composition on Flock House virus. As illustrated in Figure 18A and B, the N-terminus of the A helix in influenza virus HA points towards the viral capsid and the C-terminus protrudes from the virus. In Flock House virus, N- and C-termini are oriented in the opposite direction. Therefore, CR6261-like antibodies could not be generated, as they would probably clash with the Flock House virus capsid or with neighboring B2 proteins (Figure 18C). On the other hand, antibodies that recognize the A helix in this reversed orientation would probably clash with the influenza virus capsid and therefore likely did not protect against influenza virus (Schneemann et al, 2012). In contrast, the orientation of the A helix will be kept in reovirus

chimera (σ1-A), because as in influenza virus HA, the N-terminus of reovirus σ1 points to the viral capsid and the C-terminus projects from it (Figure 18D). Thus, antibodies that are triggered by a reovirus σ1-A helix vector are probably also able to bind to the A helix in its natural context.

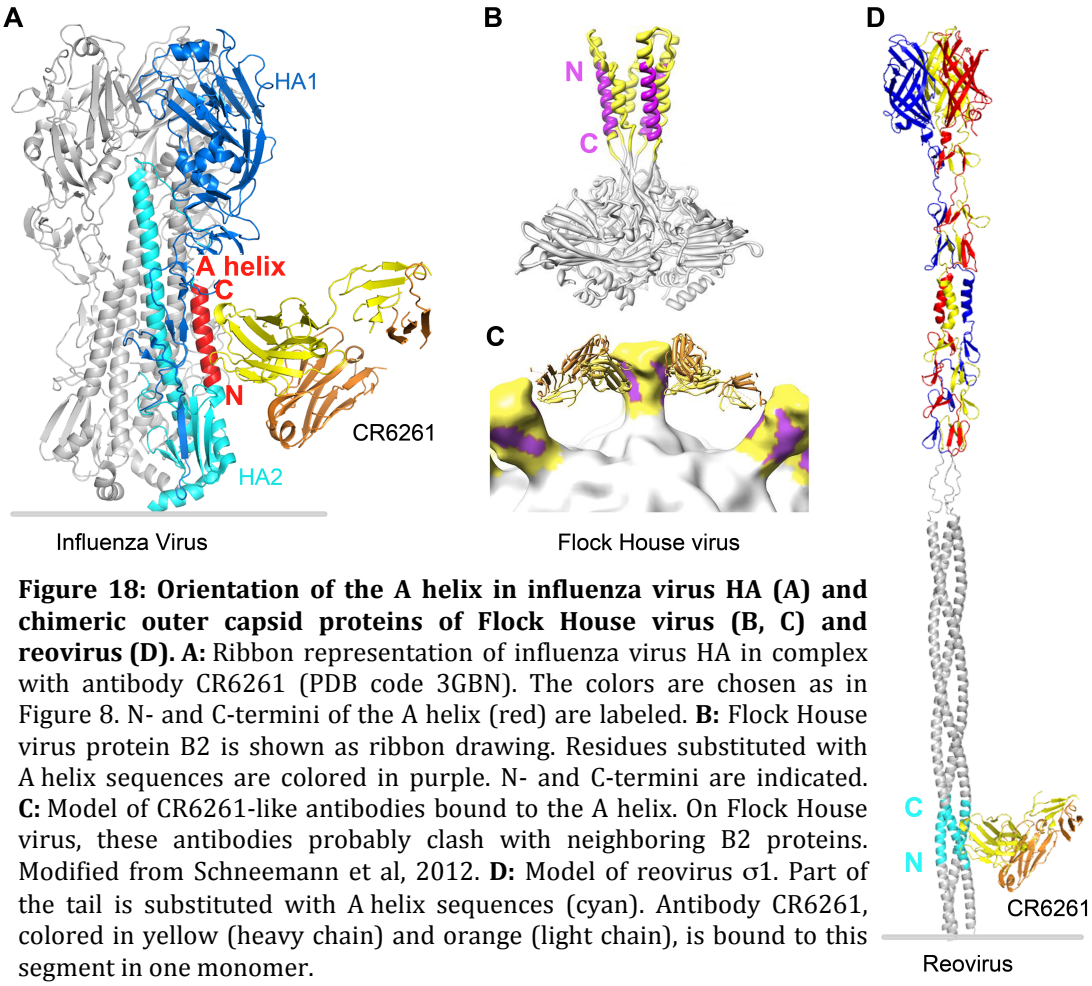

Figure 18: Orientation of the A helix in influenza virus HA (A) and chimeric outer capsid proteins of Flock House virus (B, C) and reovirus (D). A: Ribbon representation of influenza virus HA in complex with antibody CR6261 (PDB code 3GBN). The colors are chosen as in Figure 8. N- and C-termini of the A helix (red) are labeled. **B:** Flock House virus protein B2 is shown as ribbon drawing. Residues substituted with A helix sequences are colored in purple. N- and C-termini are indicated. **C:** Model of CR6261-like antibodies bound to the A helix. On Flock House virus, these antibodies probably clash with neighboring B2 proteins. Modified from Schneemann et al, 2012. **D:** Model of reovirus σ1. Part of the tail is substituted with A helix sequences (cyan). Antibody CR6261, colored in yellow (heavy chain) and orange (light chain), is bound to this segment in one monomer.

To foster development of reovirus-based vaccines, fragments of σ1-A will have to be expressed, purified, and characterized. Analytical gel filtration and SPR analyses could be carried out to determine whether the chimeric proteins are recognized by CR6261-like antibodies and to quantify the strength of the interaction. The next step would be to crystallize and determine the structures of σ1-A fragments in complex with the CR6261 antibody. With the crystal structures in hand, the binding mode could be described on an atomic level and the optimization of chimeric constructs could be guided. In addition, σ1-A will be introduced into reoviruses with the well-established reversed genetics system

(Kobayashi et al, 2007; Kobayashi et al, 2010; Boehme et al, 2011b). Enzyme-linked immunosorbent assays could be performed to detect whether the epitope is displayed correctly in the context of reovirus. The next steps would be to immunize mice or rabbits with these genetically manipulated reoviruses to determine whether the animals produce antibodies against σ1-A and to test whether these antibodies can neutralize influenza virus and protect against influenza virus infection. Several generations of chimeric constructs may have to be generated, tested in vaccination studies, and optimized further. It will be necessary to combine structural information and expert knowledge in virology and immunology to successfully establish the reovirus σ1 tail as a tool for vaccine design and development.

2 The *Phytophthora sojae* transglutaminase GP42

2.1 Introduction

2.1.1 Plant defense mechanisms

Plants are targeted by various pathogens and parasites, such as viruses, insects, fungi, and nematodes. Generally, plant pathogens can be divided into biotrophs, which need a living host to grow, and necrotrophs, which destroy the plant in order to live on its dead tissue (Dangl & Jones, 2001). Plants have evolved a variety of defense mechanisms against pest attack. In plants, neither a mechanism of somatic recombination nor a circulating system of specialized immune cells analogous to the animal immune system exists. Instead, microbe detection and antimicrobial defense take place in every single cell (Nurnberger et al, 2004). Despite this very different mechanism of immunity compared with animals, plants can trigger specific immune responses to distinguish between self and nonself molecules and establish long-lasting memory (Spoel & Dong, 2012).

2.1.1.1 *Plant immune response*

Upon pathogen detection, plants generate a complex immune response to deter or eliminate the assailant. At the site of pathogen attack, the cell wall is reinforced by deposition of callose and lignins and by oxidative cross-linking of cell wall components. The cross-linking reaction requires reactive oxygen species, which are also produced and released upon microbe attack. Reactive oxygen species are toxic for microbes and contribute to plant signaling (Grant & Loake, 2000; Huckelhoven & Kogel, 2003). The plant immune response also includes synthesis of hydrolytic enzymes that target the microbial cell wall, such as chitinases, that degrade the chitin shield of fungi. In addition, plants synthesize anti-microbial proteins (defensins) and anti-microbial secondary metabolites (phytoalexins) (Schafer et al, 1989; Vanetten et al, 1989) to constrain microbial growth. Furthermore, pathogen detection also leads to induction of apoptosis-like cell death at the site of infection and in surrounding tissue. This process, known as the hypersensitive response, confines growth of

biotrophic pathogens such as viruses (Lam et al, 2001). When a plant is locally infected by a pathogen, the entire plant establishes long-lasting resistance to a broad spectrum of pathogens (Ross, 1961).

2.1.1.2 ***Pathogen detection***

Before an effective immune response can be triggered, the plant has to specifically detect the presence of a pathogen. There are two branches of microbe detection, which are summarized in Figure 19. The first branch is activated by the detection of pathogen- or microbe-associated molecular patterns, known as PAMPs and MAMPs, respectively. PAMPs include cell wall components such as chitin for fungi and flagellin and lipopolysaccharide for bacteria. In animal immunology, PAMPs are highly conserved throughout classes of pathogens and functionally important for the pathogen, but absent in the host (Medzhitov & Janeway, 1997). These characteristics of PAMP recognition are probably similar in plant and animal immunity (Nurnberger & Brunner, 2002; Nurnberger et al, 2004).

PAMPs are detected by plant pattern recognition receptors (PRRs), which are localized at the plasma membrane. PRRs are composed of an extracellular leucine-rich repeat domain (LRR), a transmembrane region, and a cytoplasmic kinase domain (Figure 19-1). The detection of pathogens via PRRs leads to broad resistance to (non-adapted) microbe pathogens, a process referred to as PAMP-triggered immunity (PTI) (Figure 19-2). A well-studied example of PAMP detection is sensing of bacterial flagellin in *A. thaliana* by the LRR-receptor kinase flagellin-sensitive 2 (Gomez-Gomez & Boller, 2000). The most conserved epitope of flagellin is recognized (Felix et al, 1999), illustrating that PTI serves as a defense mechanism against a broad range of microbes rather than a specific immune response.

In non-adapted pathogens (Figure 19-2), PAMP detection by the plant triggers an immune response that abolishes colonization of the host by the pathogen. In order to suppress plant defense mechanisms induced by PAMP/MAMP detection, adapted pathogens express avirulence genes (Avr genes) encoding effector

molecules that interfere with host signaling and immune responses (Figure 19-3).

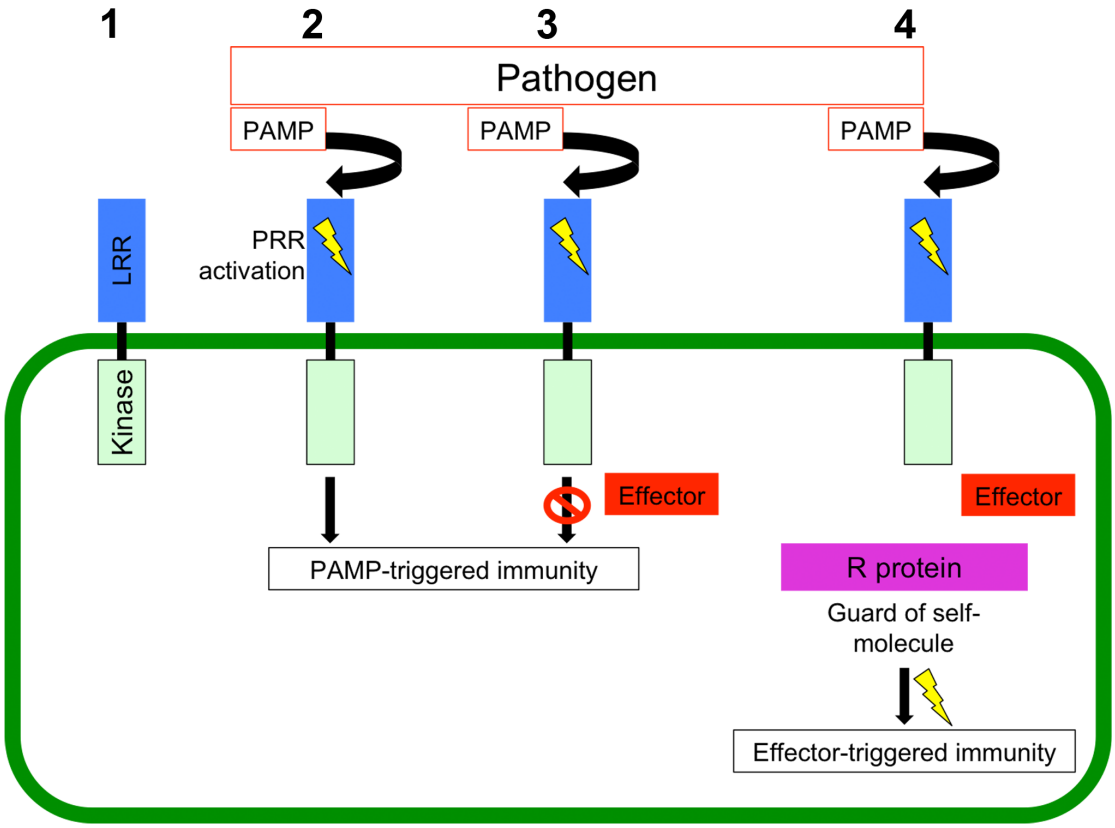

Figure 19: Pathogen detection in plants. PRRs are composed of an extracellular ligand-binding LRR domain, a transmembrane region, and a cytoplasmic kinase domain (**1**). Activation of PRRs leads to a PAMP-triggered immune response (**2**). Some pathogens synthesize effector molecules that suppress PTI mechanisms by interfering with host metabolism and signaling (**3**). Plants that are resistant to such pathogens express R genes matching specific effectors. The resulting R proteins (purple) "guard" hubs that are often targeted by effector molecules. Modification of the guarded protein by the pathogen results in activation of R proteins and effector-triggered immunity (**4**).

As a result of coevolution, plant cultivars that have acquired resistance to these microbes express resistance genes (R genes) (Figure 19-4). The corresponding gene products, the R proteins, represent the second branch of pathogen detection and plant immunity. The mechanism of pathogen detection by R proteins is known as effector-triggered immunity (ETI). In contrast to PTI, the expression of R genes induces a more specific mechanism that does not mediate broad immunity. Even closely related pathogen strains are not neutralized when the related strain does not carry a certain effector molecule. However, R genes confer durable resistance to pathogens (Bonas & Lahaye, 2002). Pathogen-synthesized compounds that raise an immune response in plants, for example by

triggering the production of phytoalexins, are termed “elicitors” (Ayers et al, 1976). As effectors are more species-specific than PAMPs, they are referred to as “specific elicitors” of immune response in contrast to PAMPs, which act as “general elicitors” (Chisholm et al, 2006; Jones & Dangl, 2006; Bent & Mackey, 2007).

In the 1950s, Harold Henry Flor proposed the “gene-for-gene model,” which postulates that a plant cultivar acquires resistance against an Avr gene-expressing pathogen strain by expressing the matching R gene (Flor, 1956). The mechanism was later renamed as effector-triggered immunity (ETI), which implies that the interactions take place on the protein level. The detection of a pathogen by an R protein induces a hypersensitive response in the surrounding plant tissue and the expression of hydrolytic enzymes that for example degrade cell wall components of the pathogen.

In most cases, R proteins and effectors do not interact directly with each other as in animal immunity, where binding of antibodies to nonself molecules is a key component of an immune response. Instead, R proteins sense the activity of effector molecules. Pathogen detection via R proteins can be described by the guard hypothesis (Van der Biezen & Jones, 1998). According to this model, R proteins detect perturbations of cellular functions by “guarding” self molecules that are often targeted by effectors. The R protein is activated when the self molecule is manipulated by a pathogen. A well-characterized example for the guard model is RPM1‐interacting protein 4 (RIN4) in *A. thaliana.* RIN4 serves as a target for the pathogen effectors AvrRpm1, AvrB, and AvrRpt2 from *Pseudomonas syringe*, probably due to its function in plant defense (Mackey et al, 2002; Kim et al, 2005). The two R proteins RPM1 and RPS2 guard RIN4. Bacterial AvrRpt2 is a cysteine protease (Axtell et al, 2003) that cleaves RIN4. Degradation of RIN4 activates the R protein RPS2 (Axtell & Staskawicz, 2003). Additionally, the presence of the bacterial effectors AvrB and AvrRpt2 leads to phosphorylation of RIN4 by the plant kinase RPM1‐induced protein kinase. This modification of RIN4 also activates the R protein RPM1 resulting in an immune response (Liu et al, 2011).

The animal immune system generates a huge repertoire of nonself detecting motifs via somatic recombination. In plants, such a system is absent. Therefore, it

remained unclear for a long time how a limited number of R proteins confers resistance to a huge variety of effectors from various attackers. The current explanation is that R proteins guard self molecules that are hubs of cell signaling and metabolism. These hubs are often targets of various effector molecules. Thus, a single R protein can respond to many pathogens that use the same strategy of host manipulation (Mukhtar et al, 2011).

2.1.2 **The plant pathogen *Phytophthora***

2.1.2.1 ***Classification and characterization***

Oomycetes are water molds that belong to the kingdom Stramenopiles, which includes brown algae and diatoms (Figure 20) (Forster et al, 1990; Harper et al, 2005). While brown algae and diatoms are photosynthetic organisms, oomycetes are nonphotosynthetic (Yoon et al, 2002). The oomycete phylum predominantly comprises plant pathogens, such as *Peronospora parasitica* and members of the genera *Phytophthora* and *Pythium*. *Peronospora parasitica* species cause downy mildew in Cruciferae, a family of plants with four-petaled flowers (Koch & Slusarenko, 1990). However, oomycetes also include some animal and human pathogens, such as *Pythium insidiosum,* which occurs in tropical and subtropical countries (Kaufman, 1998). Although the morphology and propagation of oomycetes resemble fungi, they are phylogenetically distinct (Forster et al, 1990; Harper et al, 2005). Similarities between oomycetes and fungi, including hyphal growth and spore formation, might have arisen through convergent evolution to facilitate the colonization of similar ecological niches. However, oomycetes and fungi differ in their biochemical processes and their physiology. Thus, oomycetes are poorly susceptible to anti-fungal strategies, which for example complicates management of *Phytophthora* pests (Tyler, 2001).

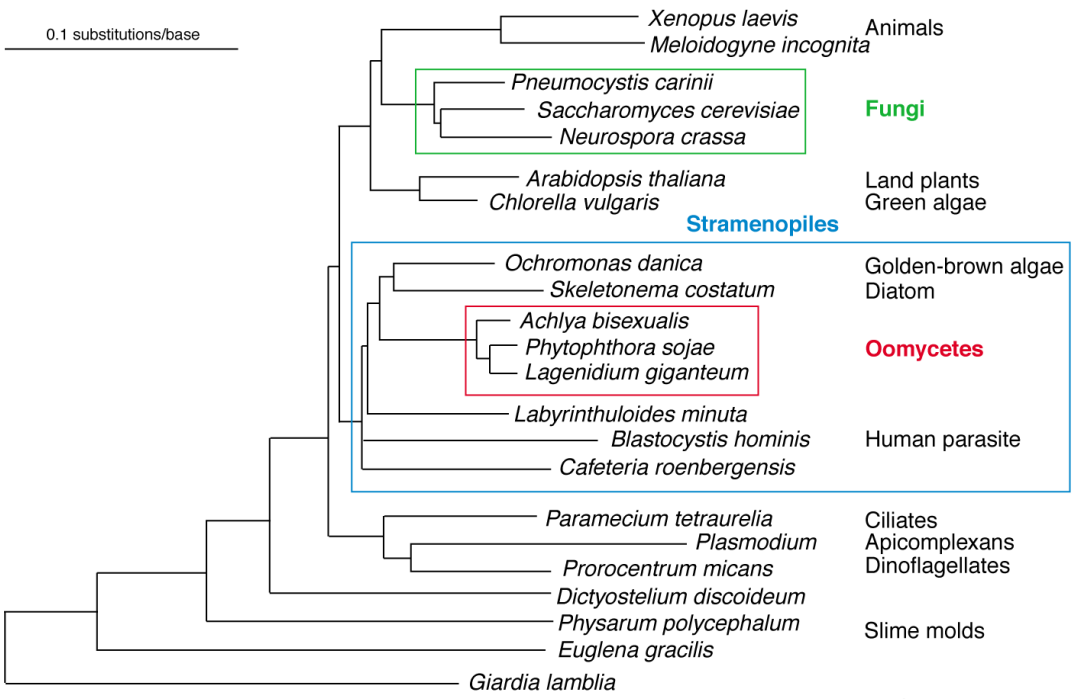

Figure 20: Eukaryotic phylogeny based on 18S ribosomal RNA sequences. The appearance of oomycetes resembles fungi. However, oomycetes are closer related to algae and diatoms. Modified from Tyler, 2001 and Sogin & Silberman, 1998.

Phytophthora, which literally means "plant destroyer," are parasitic plant pathogens comprising more than 80 species. To date, *Phytophthora* pests are difficult to control and cause huge economic losses. The three most prominent species are *Phytophthora (P.) infestans, P. ramorum,* and *P. sojae. P. infestans* is the causative agent of late blight in potatoes and tomatoes and was responsible for the Irish potato famine in the 19th century. *P. ramorum* emerged more recently causing sudden oak death in coastal forests of California and in Germany in the 1990s (Werres et al, 2001; Rizzo et al, 2005; Grunwald et al, 2008). However, this species has a broad host range and can infect a wide range of bushes and trees (Savidor et al, 2008). In the last two decades, *P. ramorum* spread to other parts of North America and Europe. The species *P. sojae* infects soybean, leading to root and stem rot (Kaufmann & Gerdemann, 1958). Other *Phytophthora* species with global impact include *P. palmivora* and *P. megakarya*, which infect cacao and thereby cause high cocoa yield losses worldwide (Guest, 2007). The genomes of *P. sojae, P. ramorum*, and *P. infestans* have been sequenced (oomycete genome sequencing project; Govers & Gijzen, 2006; Tyler et al, 2006). Comparison of the genomes, which represent three major clades of

Phytophthora, revealed that all three species express a variety of effector genes. These effector genes are localized in dynamic regions of the respective genome and are characterized by extensive genetic expansion, especially in the case of *P. infestans* (Haas et al, 2009). These findings might explain why *Phytophthora* species are able to rapidly adapt to chemical control strategies and genetically resistant cultivars (Tyler et al, 2006).

2.1.2.2 ***Life cycle***

Phytophthora species exhibit a complex life cycle that includes stages of both sexual and asexual reproduction (Figure 21). *Phytophthora* develops three kinds of asexual spores: sporangia, zoospores, and clamydospores. Sporangia (Figure 21-1) either germinate to produce hyphae that grow on host tissue or differentiate into 10-30 zoospores (2). Although zoospores are short-lived spores surviving only for hours, they represent the major route of infection. Zoospore formation is induced by flooding of a poorly drained field but can also be artificially triggered in the laboratory. Zoospores are motile and can swim with their two flagella until they meet their host. Zoospores from several *Phytophthora* species including *P. sojae* are chemotactically attracted to their hosts (Morris & Ward, 1992; Tyler, 2002). To infect a plant, zoospores attach to the host tissue and differentiate to adhesive cysts (3). The cysts germinate (4), penetrate the plant tissue with their hyphae, and initiate vegetative growth within the plant (5). In this phase, *Phytophthora* colonizes the plant tissue as mycelium (6). Sexual reproduction is carried out through meiosis generating the oogonium, which can differentiate into long-lived oospores (7). Oospores are more robust than zoospores, and can survive in soil over years in a dormant state until they initiate reinfection of plants. *P. sojae* is a homothallic *Phytophthora* species, whereas *P. infestans* and *P. ramorum* are heterothallic species, in which the presence of both mating types A1 and A2 is required for sexual reproduction (Tyler, 2007; Xavier et al, 2010). Another form of spores, the clamydospore, is a long-lived asexual spore that is present on dead plants or in late stages of the disease. Clamydospores are formed to allow survival under unfavorable conditions. *Phytophthora* initially live as biotrophs to grow on the

plant's nutrients. Later they switch to necrotrophic infection resulting in destruction of the plant.

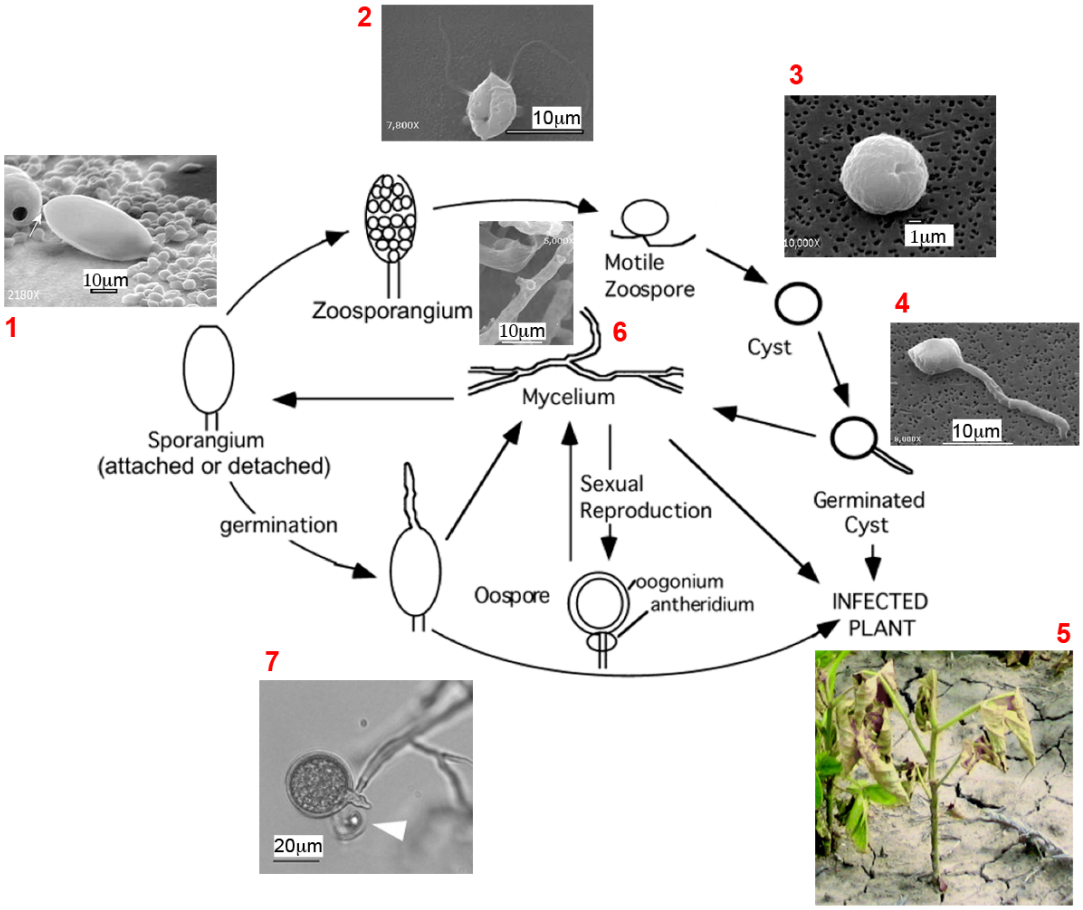

Figure 21: *Phytophthora* life cycle. The morphology of *Phytophthora* at different stages is shown as schematic drawing and as scanning electron micrograph (1-4 and 6) or light microscopy images (7) for *P. ramorum*. A picture of a plant infected by *P. sojae* is included (5). Modified from Tyler, 2007, Savidor et al, 2008, and Xavier et al, 2010.

2.1.2.3 *The GP42 protein from P. sojae*

P. sojae synthesizes a cell wall-associated glycoprotein (GP) with a molecular weight of 42 kDa (Hahlbrock et al, 1995). Due to these properties, the protein was named GP42. The GP42 sequence is conserved among *Phytophthora* species, and all *Phytophthora* species tested so far possess a GP42 homolog (Brunner et al, 2002). GP42 is a transglutaminase (TGase) (section 2.1.3).

GP42 contains a peptide of 13 amino acids, Pep-13, which acts as an elicitor of immune response both in host and nonhost plants (Nurnberger et al, 1994; Hahlbrock et al, 1995; Jabs et al, 1997; Ligterink et al, 1997; Brunner et al, 2002).

Defense mechanisms triggered by the detection of Pep-13 include phytoalexin and H_2O_2 production (Hahlbrock et al, 1995) and Ca^{2+} influx via the large conductance elicitor-activated ion channel (Zimmermann et al, 1997). In parsley, Pep-13 is recognized by a 91 kDa plasma membrane protein, which binds Pep-13 with high affinity (Nurnberger et al, 1994; Nennstiel et al, 1998). Introduction of point mutations that diminish elicitor activity also decrease TGase activity, demonstrating a functional relevance of this segment (Brunner et al, 2002).
Hydrophobicity analyses suggest that Pep-13 is located within a hydrophilic region of GP42. Additionally, sequence analyses predict that Pep-13 forms a surface-exposed loop (Rost, 1996). These predictions are reasonable as recognition of the Pep-13 motif by the plant immune system can only take place when the epitope is presented on the protein surface. Taken together, Pep-13 is an elicitor of immune response in plants that is highly conserved among *Phytophthora* species but absent in its host. Furthermore, Pep-13 is indispensable for enzymatic activity. Due to these criteria, Pep-13 is regarded as a PAMP (Brunner et al, 2002). However, it is unknown why the Pep-13 motif is essential for enzymatic activity and what functions GP42 fulfills in the pathogen.

2.1.3 **Transglutaminases**

The GP42 protein possesses TGase activity (Brunner et al, 2002). TGases comprise the enzyme family of R-glutaminyl-peptide:amine-γ-glutamyltransferases (EC 2.3.2.13). TGases catalyze the cross-linking between side chains of peptide-bound glutamine (Gln) and primary amines such as the side chain of lysine, resulting in the formation of peptide bonds (Figure 22) (Folk, 1980; Aeschlimann & Paulsson, 1994). As TGases synthesize peptide bonds, they catalyze the reversed reaction of proteases, which cleave peptide bonds. Free Gln is not recognized as a substrate by TGases. Only peptide- or protein-bound Gln in the context of certain sequences is incorporated into the product. In contrast, a broader range of primary amines is thought to function as the second substrate (Greenberg et al, 1991). In addition to the primary

sequence of the substrate, a certain flexibility of this sequence also seems to be required for substrate incorporation (Fontana et al, 2008).

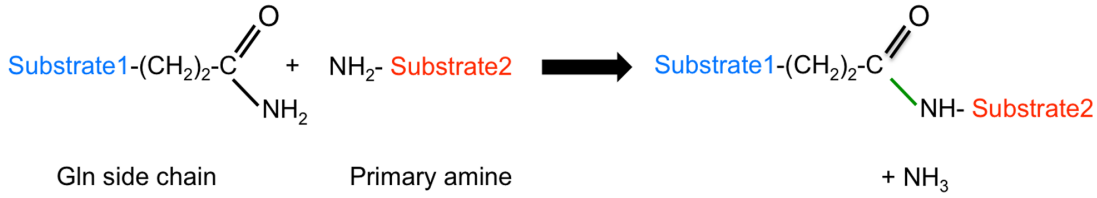

Figure 22: Cross-linking reaction catalyzed by TGases. TGases cross-link the side chains of peptide-bound glutamine residues (Gln, blue) and primary amines (red). Ammonia (NH_3) is released during the reaction. The newly formed peptide bond is shown in green.

2.1.3.1 *Structure and domain organization*

TGases are involved in many vital processes, such as blood coagulation: the TGase factor XIII stabilizes blood clots by cross-linking fibrin molecules or fibrin and other plasma molecules (Folk, 1980; Lorand et al, 1980; Shafer & Higgins, 1988). Two forms of factor XIII exist. The form that is found in plasma is composed of two *a* subunits with TGase activity and two inhibitory *b* subunits. In certain cells, such as platelets and monocytes/macrophages, factor XIII only consists of a dimer of two *a* subunits (Schwartz et al, 1971). In plasma, the *a* subunit is activated by thrombin cleavage of factor XIII in the presence of Ca^{2+}. The inhibitory *b* subunit dissociates from the factor XIII*a* dimer, which then assumes its catalytically active conformation (Curtis et al, 1973; Chung et al, 1974; Lorand et al, 1974).

Factor XIII*a* is one of nine members of the TGase family in humans, from which eight are catalytically active. The crystal structures of factor XIIIa (Yee et al, 1994), tissue transglutaminase TG2 (Liu et al, 2002), and TG3 (Ahvazi et al, 2002) have been determined. The three structures share a common fold. The crystal structure of the well-characterized TGase factor XIIIa is shown in Figure 23A. As seen in other TGases, factor XIIIa comprises an N-terminal β-sandwich domain, a catalytically active core, and two β-barrel domains. Some TGases, including factor XIIIa, carry an N-terminal activation peptide that is removed upon activation.

A sequence alignment of 15 members of the TGase family showed that these TGases have only modest sequence identity with factor XIIIa, ranging between 25 and 45%. However, the conserved residues cluster in the catalytic core domain, suggesting that members of the TGase family share a common reaction mechanism (Yee et al, 1994).

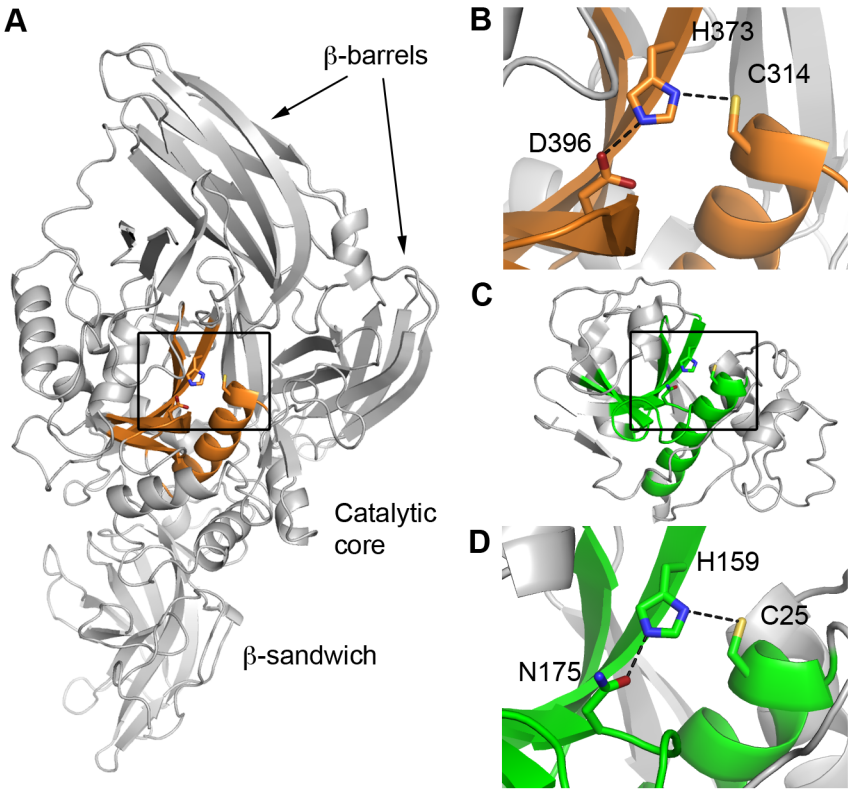

Figure 23: Structural similarities between TGases and cysteine proteases. A: Crystal structure of one factor XIII*a* subunit (PDB code 1GGT). The four domains of the TGase are labeled. **B:** Close-up view of the factor XIII*a* active site. The catalytic triad is composed of C314, H373, and D396. **C:** Crystal structure of papain (PDB code 1POP)**. D:** Close-up view of the catalytic triad of papain formed by C25, H159, and N175. The crystal structures are shown in ribbon representation and the active site residues as sticks. The central α–helix and the four neighboring β-strands that are conserved between TGases and cysteine proteases are highlighted in orange (factor XIII*a*) and green (papain).

The TGase active site is composed of a catalytic triad formed by a cysteine, a histidine, and an aspartate (Figure 23). The factor XIII*a* crystal structure revealed that mammalian TGases possess structural similarities with proteases (Pedersen et al, 1994; Yee et al, 1994). As the two enzyme classes catalyze reversed reactions, it is not unexpected that they share structural similarities. As with TGases, cysteine proteases such as papain also contain a Cys-His-Asp (or

Asn) catalytic triad. In both enzyme classes, the catalytically active cysteine is located at the N-terminus of a central α-helix. An adjacent β-sheet composed of four β-strands comprises the remaining catalytic residues (Figure 23).

In most TGase structures, the enzymes were crystallized in a conformation in which the active site is not accessible for the substrate. In 2007, the crystal structure of TG2 in its active conformation was reported (Pinkas et al, 2007). TG2 is a Ca^{2+}-dependent enzyme that can be allosterically inhibited with GDP. To trap the enzyme in its active conformation, TG2 was crystallized in the presence of a substrate mimic. Comparison of the structures of the activated and the GDP-bound enzyme revealed that the relative domain orientations change drastically upon TGase activation (Figure 24). However, the individual domains do not undergo significant structural changes.

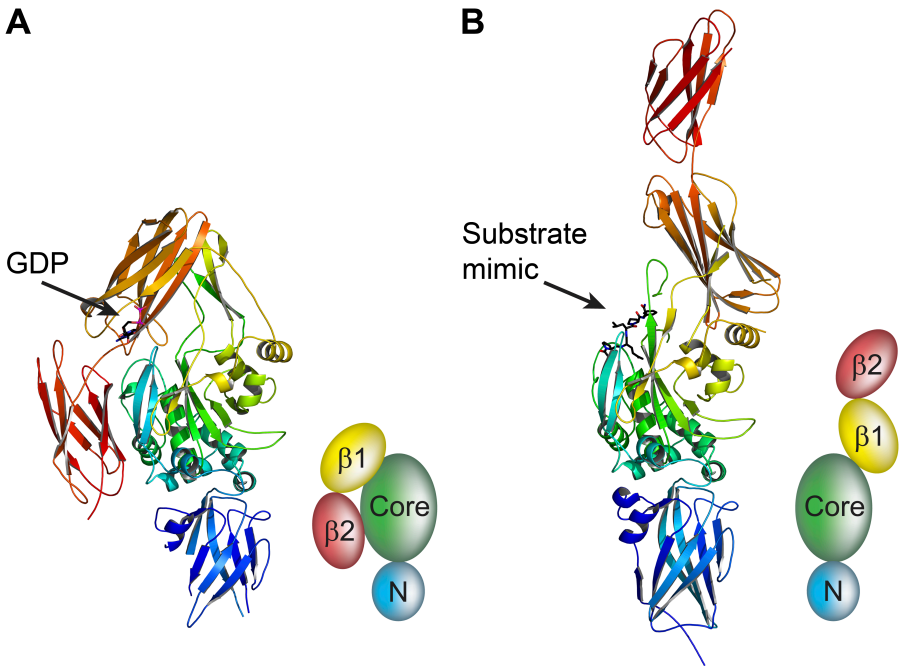

Figure 24: Conformational changes upon activation of TG2. A: GDP-bound, inactive conformation of TG2. **B:** Substrate mimic-bound, active conformation of TG2. The protein and the ligand are shown as ribbons and sticks, respectively. Schematics of the domain organization are included in both panels. The N-terminal β-sandwich domain, the catalytically active core, and the two β-barrel domains (β1 and β2) are colored in blue, green, yellow, and red. TG2 undergoes large conformational changes upon activation, but the structure within the individual domains does not change significantly. Modified from Pinkas et al, 2007.

2.1.3.2 *Reaction mechanism*

Due to the structural similarities between TGases and proteases, a reaction mechanism for TGases was proposed that resembles the reversed hydrolysis of papain (Schroder et al, 1993; Pedersen et al, 1994). An advanced reaction mechanism proposed for TG2 is schematically shown in Figure 25 (Iismaa et al, 2003).

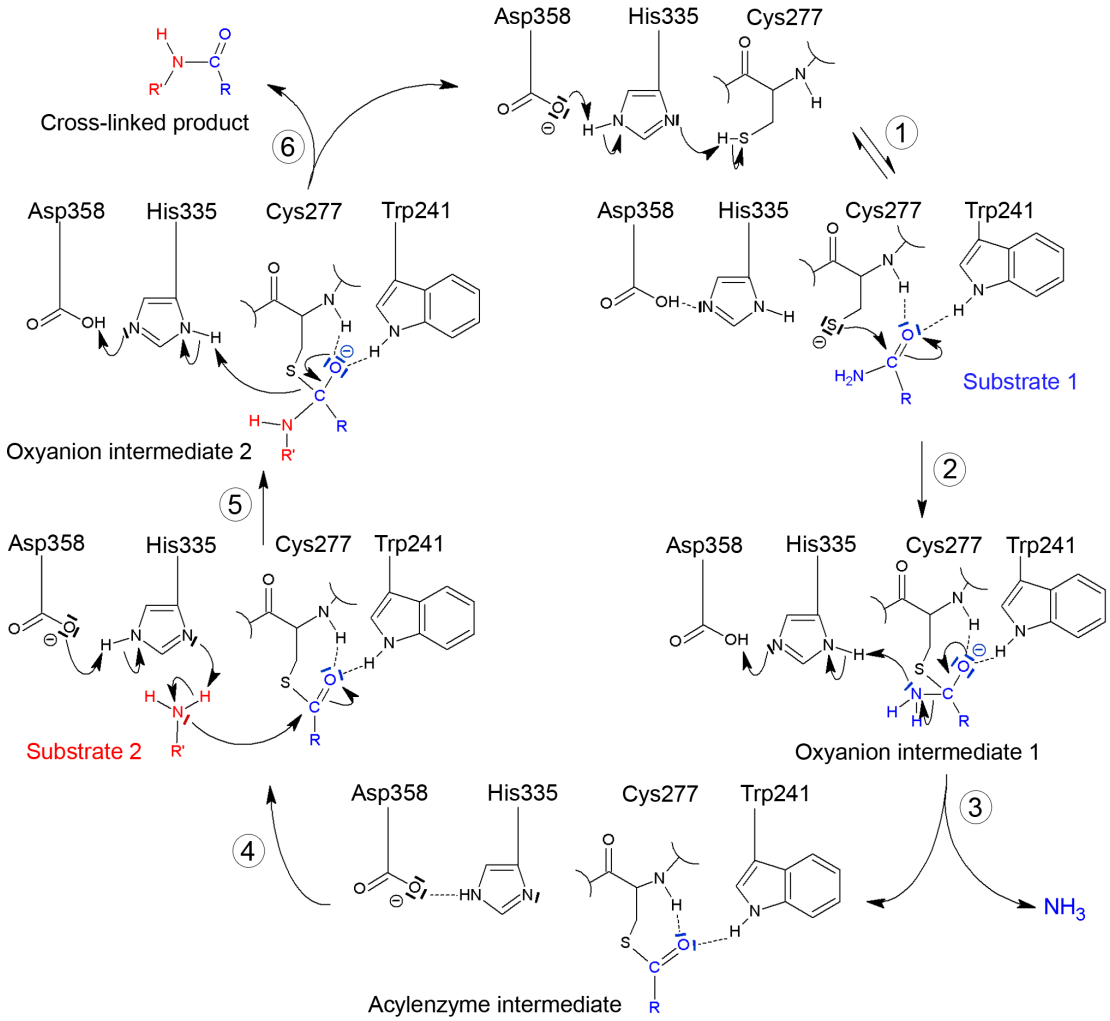

Figure 25: TGase reaction mechanism proposed for TG2. The catalytic triad of TG2 is composed of Cys277, His335, and Asp358 (step 1). In the thiol-imidazolium pair the acidity of the Cys277 sulfhydryl group is increased facilitating nucleophilic attack of the carboxyl group of substrate 1 (step 2). The resulting oxyanion intermediate 1 is stabilized by the backbone amide of Cys277 and the Nε1 nitrogen of Trp241. Ammonia release from substrate 1 leads to the formation of an acylenzyme intermediate (step 3), to which substrate 2 binds (step 4). Nucleophilic attack of the acylenzyme complex by the amine group of substrate 2 generates the second oxyanion intermediate that is again stabilized by the backbone amide of Cys277 and the side chain of Trp241 (step 5). At the end of the reaction cycle, the cross-linked product is released (Step 6). This image was created using ADC/ChemSketch (ACD/ChemSketch Freeware, version 12.0, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2012).

According to this mechanism, the active site residues, Cys277 and His335, form a thiolate-imidazolium pair (Figure 25, step 1). The polarized thiol group of Cys277 attacks the γ-carboxamide group of the glutamine substrate, generating oxyanion intermediate 1 (step 2). This first reaction intermediate is stabilized by the backbone amide proton of Cys277 and by the Nε1 nitrogen of the adjacent side chain of Trp241. An ammonium molecule is released from oxyanion intermediate 1, leading to the formation of an acylenzyme complex (step 3). Next, the second substrate, i.e. the primary amine, binds to the protein (step 4). The catalytic His335 stabilizes the deprotonated amine, facilitating nucleophilic attack of the carboxyl group of the acylenzyme complex by the primary amine (step 5). The resulting oxyanion intermediate 2 is again stabilized by the backbone amide group of Cys277 and the Nε1 nitrogen of Trp241. Finally, the cross-linked product is released from the enzyme (step 6) and the next reaction cycle is initiated.

In addition to the location in a catalytic triad, the acidity of the catalytic cysteine is increased by a second mechanism: in TGases and cysteine proteases the active site cysteine is situated at the N-terminus of an α-helix. Due to the dipole moment of α-helices, the deprotonation of the thiol group is favored and the nucleophilicity of the cysteine is increased (Hol et al, 1978). An analogous arrangement is present in the beta-ketoacyl-acyl carrier protein synthase III FabH. In this enzyme, the location of the catalytic cysteine at the N-terminus of an α-helix is sufficient to polarize the sulfhydryl group and to allow nucleophilic attack. The main function of the two remaining residues of the catalytic triad (His and Asn) is not to increase the acidity of the thiol but to stabilize a reaction intermediate (Davies et al, 2000).

2.1.3.3 *The GP42 transglutaminase*

The protein GP42 from *P. sojae* is a TGase as it catalyzes the incorporation of 5-(biotinamido)pentylamine into N,N'-dimethylcasein (Brunner et al, 2002), a standard reaction for the detection and quantification of TGase activity (Slaughter et al, 1992). A GP42 precursor is composed of 529 amino acids. The 162 N-terminal residues correspond to a pro-peptide, which is removed during enzyme maturation. The mature GP42 enzyme comprises amino acids 163-529

and possesses one putative glycosylation site at Asn379 (Sacks et al, 1995). As with many other TGases, GP42 activity depends on the presence of Ca^{2+} ions (Brunner et al, 2002). GP42 is well conserved among *Phytophthora* species, but it does not share any sequence similarities with any other characterized TGase (Sacks et al, 1995; Brunner et al, 2002). To identify the catalytic cysteine of GP42, the GP42-analog in *P. sanctorum* was analyzed. This enzyme was incubated with [^{14}C]-iodoacetamide to selectively label the active site cysteine. The protein was subsequently denatured, including disulfide bond reduction, followed by proteolytic digestion. Analysis of the resulting peptides by mass spectrometry then identified C296 as the catalytic cysteine (Dr. Sakari Kauppinen, Novo Nordisk, Copenhagen, Denmark, unpublished data). Sequence alignment revealed that this residue corresponds to C290 in *P. sojae*. To verify that C290 represents the catalytic cysteine of GP42, C290 was mutated to a serine, which resulted in a complete loss of TGase activity (Brunner et al, 2002; Brunner et al, 2006). Although C290 was identified as the active site cysteine, the two remaining residues of the supposed catalytic triad were unknown.

2.2 Objectives

The GP42 protein is a TGase from the plant pathogen *P. sojae*. The enzyme is well conserved among certain oomycete species, but it does not possess any homologs outside of this taxonomic group. The *in vivo* function of GP42 is unclear. A conserved peptide of 13 amino acids of GP42, Pep-13, is recognized by the plant immune system and elicits a defense response. Introduction of point mutations in this segment decreases both enzymatic activity and elicitor capacity. Little information about the enzymatic mechanism and residues involved in catalysis was available at the outset of this study. The major aims of this project were therefore to:

- determine the crystal structure of the GP42 TGase
- identify residues that form the GP42 active site
- determine the location of Pep-13 and elucidate its role in GP42 function
- gain insight into the catalytic mechanism and define possible substrates of the enzyme
- compare the GP42 structure with other protein structures to determine whether an evolutionary relationship with other proteins, particularly TGases, exists

2.3 Results and Discussion

Structural and Phylogenetic Analyses of the GP42 Transglutaminase from *Phytophthora sojae* Reveal an Evolutionary Relationship between Oomycetes and Marine *Vibrio* Bacteria

Reiss K, Kirchner E, Gijzen M, Zocher G, Löffelhardt B, Nürnberger T, Stehle T, Brunner F (2011). Structural and Phylogenetic Analyses of the GP42 Transglutaminase from Phytophthora sojae Reveal an Evolutionary Relationship between Oomycetes and Marine *Vibrio* Bacteria. J Biol Chem.; 286(49): 42585-93

GP42 is a cell wall-associated TGase from the parasitic plant pathogen *P. sojae*. The protein is well conserved among *Phytophthora* species, but it possesses no sequence homology with other TGases characterized to date. Prior to this study, C290 was identified as the catalytically active cysteine residue. As the wild type enzyme catalyzes disordered intermolecular cross-linking *in vitro* resulting in the formation of insoluble aggregates, the catalytically inactive point mutant C290S was used for structural analysis. This conservative mutation is unlikely to affect the protein fold. Therefore, serine and cysteine are expected to occupy equivalent positions in the enzyme. GP42 (C290S) was recombinantly expressed in *Pichia pastoris*, purified, and crystallized. Due to the lack of structural homologs, the crystal structure of GP42 was solved with a SIRAS experiment (Single Isomorphous Replacement with Anomalous Scattering) using a Ta_6Br_{12} derivative and refined to 2.95 Å resolution.

The protein adopts a seahorse-shaped fold that can be divided into three domains: the head, body, and tail. Details on domain organization and fold are described in the publication. The active site residue 290 is located at the N-terminus of an α-helix in the center of the enzyme at the interface between the body and tail domain. The side chain of the neighboring H291 is within a 5 Å distance of C290. These two residues could form the thiolate-imidazolium pair required for TGase catalysis. Residue D328 is within hydrogen bonding distance to H291 and could complete the catalytic triad.

The Pep-13 peptide, which triggers a defense response in plants, is located within the tail domain. Pep-13 is part of a β-strand at the surface of the protein and not a loop as predicted. In this partially surface-exposed position, some amino acids of Pep-13 can easily be detected by the plant's immune system. The N-terminal part of Pep-13 lies in close proximity to the active site. Thus, Pep-13 might be involved in substrate binding, which would explain its functional relevance for GP42 and the high degree of conservation among *Phytophthora* species.

A search for structural homologs of GP42 revealed that the protein has structural similarities to cysteine proteases. The central α-helix that harbors the active site cysteine and the neighboring β-sheet are structurally conserved between GP42 and cysteine proteases. These structural elements are generally present in papain-like cysteine proteases and TGases. However, in the case of cysteine proteases, the two remaining catalytic residues are located on the conserved β-sheet, whereas in GP42, His291 and Asp328 are not part of this β-sheet but are situated on the opposite side. Thus, GP42 and cysteine proteases (and other TGases) have common structural elements, but the compositions of their active sites differ.

A sequence alignment of GP42 and homologous proteins from various oomycete species revealed that the conserved residues cluster around the active site and include six N-terminal residues of the immunogenic Pep-13 fragment. Thus, these enzymes probably fulfill analogous functions and possess similar catalytic mechanisms. The crystal structure of GP42 from *P. sojae* can therefore serve as a model for a variety of homologous enzymes from other pathogenic oomycete species.

To identify GP42-like proteins outside the oomycete phylum, a BLAST search for homologous sequences was carried out. This search revealed that GP42 shares sequence similarities with certain proteins from marine *vibrio* bacteria. *Vibrio* bacteria, such as *Vibrio harveyi*, include several pathogenic strains that infect fish and marine invertebrates. Although the sequence similarities between GP42 and *vibrio* bacteria proteins are only modest, the conserved sequences cluster around the active site of the TGase. An evolutionary relationship between marine bacteria and oomycetes is therefore likely.

2.4 Outlook

The GP42 crystal structure allowed us to establish the overall fold of the enzyme and to identify catalytically important residues. However, there is still little knowledge on the catalytic mechanism and the *in vivo* function of the enzyme available. An aim of future research will therefore be to define minimal substrates and substrate range of GP42. Various substrates will have to be screened to identify motifs that are recognized by the TGase and incorporated into the product. The GP42 structure shows that the surface of the enzyme comprises an acidic patch in close proximity to the active site. Binding of peptides that contain basic amino acids is therefore a possibility. Such peptides should be considered first to determine the GP42 minimal substrate.

With the minimal substrate of GP42 in hand, the crystal structure of GP42 in complex with bound substrate could be determined to gain insight into the catalytic mechanism of GP42. To achieve this goal, the GP42 (C290S) crystals could either be soaked with substrate, or enzyme and peptide could be co-crystallized. An alternative approach to gain information on GP42 catalysis is to determine the structure of GP42 in complex with substrate mimics. In case of TG2, substrate mimics were used to trap the enzyme in its active conformation. The crystal structure of GP42 in its active conformation might differ from the structure of GP42 (C290S), as in the known structure the orientation of the (presumed) catalytic triad is not optimal for catalysis. Structural information on GP42 in its active conformation could give insight into enzyme flexibility and induced fit mechanisms: the loop connecting amino acids 339 and 360 is not defined in the structure of the inactive enzyme, but it may be better ordered in the active TGase. This segment could conceivably function as a lid that closes the active site to exclude water from the reaction, as water would abolish peptide cross-linking by hydrolyzing the Gln side chain.

In order to deduce the GP42 reaction mechanism from the structure of an enzyme-substrate complex, a higher resolution structure of the protein will be necessary. Mature GP42 possesses one glycosylation site. Flexible sugar chains of glycosylated proteins often interfere with tight crystal packing. A promising strategy to obtain better diffracting crystals of GP42 would therefore be to

crystallize the deglycosylated protein. A high resolution structure of the TGase could provide information on the enzymatic mechanism. Insight into GP42 catalysis and substrate range would help to elucidate the *in vivo* function of the TGase and foster development of specific GP42 inhibitors in order to combat *Phytophthora* pests.

3 References

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) Phenix: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58**:** 1948-1954

Aeschlimann D, Paulsson M (1994) Transglutaminases - protein cross-linking enzymes in tissues and body-fluids. *Thrombosis and Haemostasis* 71**:** 402-415

Ahvazi B, Kim HC, Kee SH, Nemes Z, Steinert PM (2002) Three-dimensional structure of the human transglutaminase 3 enzyme: Binding of calcium ions changes structure for activation. *EMBO J* 21**:** 2055-2067

Antar AA, Konopka JL, Campbell JA, Henry RA, Perdigoto AL, Carter BD, Pozzi A, Abel TW, Dermody TS (2009) Junctional adhesion molecule-a is required for hematogenous dissemination of reovirus. *Cell Host Microbe* 5**:** 59-71

Axtell MJ, Chisholm ST, Dahlbeck D, Staskawicz BJ (2003) Genetic and molecular evidence that the pseudomonas syringae type iii effector protein avrrpt2 is a cysteine protease. *Mol Microbiol* 49**:** 1537-1546

Axtell MJ, Staskawicz BJ (2003) Initiation of rps2-specified disease resistance in arabidopsis is coupled to the avrrpt2-directed elimination of rin4. *Cell* 112**:** 369-377

Ayers AR, Ebel J, Finelli F, Berger N, Albersheim P (1976) Host-pathogen interactions: Ix. Quantitative assays of elicitor activity and characterization of the elicitor present in the extracellular medium of cultures of phytophthora megasperma var. Sojae. *Plant Physiol* 57**:** 751-759

Bartlett NM, Gillies SC, Bullivant S, Bellamy AR (1974) Electron microscopy study of reovirus reaction cores. *J Virol* 14**:** 315-326

Barton ES, Connolly JL, Forrest JC, Chappell JD, Dermody TS (2001a) Utilization of sialic acid as a coreceptor enhances reovirus attachment by multistep adhesion strengthening. *J Biol Chem* 276**:** 2200-2211

Barton ES, Forrest JC, Connolly JL, Chappell JD, Liu Y, Schnell FJ, Nusrat A, Parkos CA, Dermody TS (2001b) Junction adhesion molecule is a receptor for reovirus. *Cell* 104**:** 441-451

Barton ES, Youree BE, Ebert DH, Forrest JC, Connolly JL, Valyi-Nagy T, Washington K, Wetzel JD, Dermody TS (2003) Utilization of sialic acid as a coreceptor is required for reovirus-induced biliary disease. *J Clin Invest* 111**:** 1823-1833

Bassel-Duby R, Jayasuriya A, Chatterjee D, Sonenberg N, Maizel JV, Jr., Fields BN (1985) Sequence of reovirus haemagglutinin predicts a coiled-coil structure. *Nature* 315**:** 421-423

Bassel-Duby R, Nibert ML, Homcy CJ, Fields BN, Sawutz DG (1987) Evidence that the sigma 1 protein of reovirus serotype 3 is a multimer. *J Virol* 61**:** 1834-1841

Baum LG, Paulson JC (1990) Sialyloligosaccharides of the respiratory epithelium in the selection of human influenza virus receptor specificity. *Acta Histochem Suppl* 40**:** 35-38

Bent AF, Mackey D (2007) Elicitors, effectors, and r genes: The new paradigm and a lifetime supply of questions. *Annu Rev Phytopathol* 45**:** 399-436

Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci U S A* 92**:** 8259-8263

Boehme KW, Frierson JM, Konopka JL, Kobayashi T, Dermody TS (2011a) The reovirus sigma1s protein is a determinant of hematogenous but not neural virus dissemination in mice. *J Virol* 85**:** 11781-11790

Boehme KW, Ikizler M, Kobayashi T, Dermody TS (2011b) Reverse genetics for mammalian reovirus. *Methods* 55**:** 109-113

Bokiej M, Ogden KM, Ikizler M, Reiter DM, Stehle T, Dermody TS (2012) Optimum length and flexibility of reovirus attachment protein sigma1 are required for efficient viral infection. *J Virol* 86**:** 10270-10280

Bonas U, Lahaye T (2002) Plant disease resistance triggered by pathogen-derived molecules: Refined models of specific recognition. *Curr Opin Microbiol* 5**:** 44-50

Borsa J, Morash BD, Sargent MD, Copps TP, Lievaart PA, Szekely JG (1979) Two modes of entry of reovirus particles into l cells. *J Gen Virol* 45**:** 161-170

Borsa J, Sargent MD, Lievaart PA, Copps TP (1981) Reovirus: Evidence for a second step in the intracellular uncoating and transcriptase activation process. *Virology* 111**:** 191-200

Bovarnick M, de Burgh PM (1947) Virus hemagglutination. *Science* 105**:** 550-552

Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart O, Vonrhein C, Womack T. (2011) Buster version 2.10.0.

Brunner F, Löffelhardt B, Müller C, Gijzen M, Nurnberger T (2006) Identification of the catalytic triad of the phytophthora sojae 42 kda transglutaminase by site-directed mutagenesis. In *Abstract book: annual meeting Oomycete Molecular Genetics Network*. De Wageningse Berg, Wageningen, The Netherlands

Brunner F, Rosahl S, Lee J, Rudd JJ, Geiler C, Kauppinen S, Rasmussen G, Scheel D, Nurnberger T (2002) Pep-13, a plant defense-inducing pathogen-associated pattern from phytophthora transglutaminases. *EMBO J* 21**:** 6681-6688

Bullough PA, Hughson FM, Skehel JJ, Wiley DC (1994) Structure of influenza haemagglutinin at the ph of membrane fusion. *Nature* 371**:** 37-43

Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N (2004) Crystal structure of species d adenovirus fiber knobs and their sialic acid binding sites. *J Virol* 78**:** 7727-7736

Burnet FM, Mc CJ, Stone JD (1946) Modification of human red cells by virus action; the receptor gradient for virus action in human red cells. *Br J Exp Pathol* 27**:** 228-236

Campbell JA, Schelling P, Wetzel JD, Johnson EM, Forrest JC, Wilson GA, Aurrand-Lions M, Imhof BA, Stehle T, Dermody TS (2005) Junctional adhesion molecule a serves as a receptor for prototype and field-isolate strains of mammalian reovirus. *J Virol* 79**:** 7967-7978

Carlin AF, Uchiyama S, Chang YC, Lewis AL, Nizet V, Varki A (2009) Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil siglec-9 and dampen the innate immune response. *Blood* 113**:** 3333-3336

Chan DC, Fass D, Berger JM, Kim PS (1997) Core structure of gp41 from the hiv envelope glycoprotein. *Cell* 89**:** 263-273

Chandran K, Farsetta DL, Nibert ML (2002) Strategy for nonenveloped virus entry: A hydrophobic conformer of the reovirus membrane penetration protein micro 1 mediates membrane disruption. *J Virol* 76**:** 9920-9933

Chandran K, Nibert ML (1998) Protease cleavage of reovirus capsid protein mu1/mu1c is blocked by alkyl sulfate detergents, yielding a new type of infectious subvirion particle. *J Virol* 72**:** 467-475

Chandrasekaran A, Srinivasan A, Raman R, Viswanathan K, Raguram S, Tumpey TM, Sasisekharan V, Sasisekharan R (2008) Glycan topology determines human adaptation of avian h5n1 virus hemagglutinin. *Nat Biotechnol* 26**:** 107-113

Chappell JD, Duong JL, Wright BW, Dermody TS (2000) Identification of carbohydrate-binding domains in the attachment proteins of type 1 and type 3 reoviruses. *J Virol* 74**:** 8472-8479

Chappell JD, Prota AE, Dermody TS, Stehle T (2002) Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. *EMBO J* 21**:** 1-11

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* 124**:** 803-814

Chou HH, Hayakawa T, Diaz S, Krings M, Indriati E, Leakey M, Paabo S, Satta Y, Takahata N, Varki A (2002) Inactivation of cmp-n-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc Natl Acad Sci U S A* 99**:** 11736-11741

Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, Muchmore EA, Nelson DL, Warren ST, Varki A (1998) A mutation in human cmp-sialic acid hydroxylase occurred after the homo-pan divergence. *Proc Natl Acad Sci U S A* 95**:** 11751-11756

Chung SI, Lewis MS, Folk JE (1974) Relationships of the catalytic properties of human plasma and platelet transglutaminases (activated blood coagulation factor xiii) to their subunit structures. *J Biol Chem* 249**:** 940-950

Cleveland DR, Zarbl H, Millward S (1986) Reovirus guanylyltransferase is l2 gene product lambda 2. *J Virol* 60**:** 307-311

Coffey MC, Strong JE, Forsyth PA, Lee PW (1998) Reovirus therapy of tumors with activated ras pathway. *Science* 282**:** 1332-1334

Crick F (1953a) The fourier transform of a coiled-coil. *Acta Crystallographica* 6**:** 685-689

Crick F (1953b) The packing of [alpha]-helices: Simple coiled-coils. *Acta Crystallographica* 6**:** 689-697

Crick FH (1952) Is alpha-keratin a coiled coil? *Nature* 170**:** 882-883

Curtis CG, Stenberg P, Chou CH, Gray A, Brown KL, Lorand L (1973) Titration and subunit localization of active center cysteine in fibrinoligase (thrombin-activated fibrin stabilizing fector). *Biochem Biophys Res Commun* 52**:** 51-56

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. *Nature* 411**:** 826-833

Daniels RS, Douglas AR, Skehel JJ, Wiley DC, Naeve CW, Webster RG, Rogers GN, Paulson JC (1984) Antigenic analyses of influenza virus haemagglutinins with different receptor-binding specificities. *Virology* 138**:** 174-177

Davies C, Heath RJ, White SW, Rock CO (2000) The 1.8 a crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase iii (fabh) from escherichia coli. *Structure* 8**:** 185-195

Dermody TS, Parker J, Sherry B (2012) Orthoreovirus. In *Fields virology*, Knipe DM, Howley PM (eds), 6 edn.: Philadelphia: Lippincott Williams & Wilkins

Dryden KA, Wang G, Yeager M, Nibert ML, Coombs KM, Furlong DB, Fields BN, Baker TS (1993) Early steps in reovirus infection are associated with dramatic changes in supramolecular structure and protein conformation: Analysis of virions and subviral particles by cryoelectron microscopy and image reconstruction. *J Cell Biol* 122**:** 1023-1041

Duncan R, Horne D, Cashdollar LW, Joklik WK, Lee PW (1990) Identification of conserved domains in the cell attachment proteins of the three serotypes of reovirus. *Virology* 174**:** 399-409

Ebert DH, Deussing J, Peters C, Dermody TS (2002) Cathepsin l and cathepsin b mediate reovirus disassembly in murine fibroblast cells. *J Biol Chem* 277**:** 24609-24617

Ehrlich M, Boll W, Van Oijen A, Hariharan R, Chandran K, Nibert ML, Kirchhausen T (2004) Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* 118**:** 591-605

Ekiert DC, Bhabha G, Elsliger MA, Friesen RH, Jongeneelen M, Throsby M, Goudsmit J, Wilson IA (2009) Antibody recognition of a highly conserved influenza virus epitope. *Science* 324**:** 246-251

Fass D, Harrison SC, Kim PS (1996) Retrovirus envelope domain at 1.7 angstrom resolution. *Nat Struct Biol* 3**:** 465-469

Fausnaugh J, Shatkin AJ (1990) Active site localization in a viral mrna capping enzyme. *J Biol Chem* 265**:** 7669-7672

Felix G, Duran JD, Volko S, Boller T (1999) Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant Journal* 18**:** 265-276

Flor HH (1956) Mutations in flax rust induced by ultraviolet radiation. *Science* 124**:** 888-889

Folk JE (1980) Transglutaminases. *Annu Rev Biochem* 49**:** 517-531

Fontana A, Spolaore B, Mero A, Veronese FM (2008) Site-specific modification and pegylation of pharmaceutical proteins mediated by transglutaminase. *Adv Drug Deliv Rev* 60**:** 13-28

Forster H, Coffey MD, Elwood H, Sogin ML (1990) Sequence-analysis of the small subunit ribosomal-rnas of 3 zoosporic fungi and implications for fungal evolution. *Mycologia* 82**:** 306-312

Forsyth P, Roldan G, George D, Wallace C, Palmer CA, Morris D, Cairncross G, Matthews MV, Markert J, Gillespie Y, Coffey M, Thompson B, Hamilton M (2008) A phase i trial of intratumoral administration of reovirus in patients with histologically confirmed recurrent malignant gliomas. *Mol Ther* 16**:** 627-632

Fraser RD, Furlong DB, Trus BL, Nibert ML, Fields BN, Steven AC (1990) Molecular structure of the cell-attachment protein of reovirus: Correlation of computer-processed electron micrographs with sequence-based predictions. *J Virol* 64**:** 2990-3000

Frierson JM, Pruijssers AJ, Konopka JL, Reiter DM, Abel TW, Stehle T, Dermody TS (2012) Utilization of sialylated glycans as coreceptors enhances the neurovirulence of serotype 3 reovirus. *J Virol*

Furlong DB, Nibert ML, Fields BN (1988) Sigma 1 protein of mammalian reoviruses extends from the surfaces of viral particles. *J Virol* 62**:** 246-256

Fuster MM, Esko JD (2005) The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nat Rev Cancer* 5**:** 526-542

Gentsch JR, Pacitti AF (1987) Differential interaction of reovirus type 3 with sialylated receptor components on animal cells. *Virology* 161**:** 245-248

Gomatos PJ, Tamm I (1962) Reactive sites of reovirus type 3 and their interaction with receptor substances. *Virology* 17**:** 455-461

Gomez-Gomez L, Boller T (2000) Fls2: An lrr receptor-like kinase involved in the perception of the bacterial elicitor flagellin in arabidopsis. *Mol Cell* 5**:** 1003-1011

Govers F, Gijzen M (2006) Phytophthora genomics: The plant destroyers' genome decoded. *Mol Plant Microbe Interact* 19**:** 1295-1301

Grant JJ, Loake GJ (2000) Role of reactive oxygen intermediates and cognate redox signaling in disease resistance. *Plant Physiol* 124**:** 21-29

Green RH, Woolley DW (1947) Inhibition by certain polysaccharides of hemagglutination and of multiplication of influenza virus. *J Exp Med* 86**:** 55-64

Greenberg CS, Birckbichler PJ, Rice RH (1991) Transglutaminases: Multifunctional cross-linking enzymes that stabilize tissues. *FASEB J* 5**:** 3071-3077

Gruber M, Lupas AN (2003) Historical review: Another 50th anniversary--new periodicities in coiled coils. *Trends Biochem Sci* 28**:** 679-685

Grunwald NJ, Goss EM, Press CM (2008) Phytophthora ramorum: A pathogen with a remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Mol Plant Pathol* 9**:** 729-740

Guest D (2007) Black pod: Diverse pathogens with a global impact on cocoa yield. *Phytopathology* 97**:** 1650-1653

Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AM, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JI, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J, Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs LJ, Griffith J, Grunwald NJ, Horn K, Horner NR, Hu CH, Huitema E, Jeong DH, Jones AM, Jones JD, Jones RW, Karlsson EK, Kunjeti SG, Lamour K, Liu Z, Ma L, Maclean D, Chibucos MC, McDonald H, McWalters J, Meijer HJ, Morgan W, Morris PF, Munro CA, O'Neill K, Ospina-Giraldo M, Pinzon A, Pritchard L, Ramsahoye B, Ren Q, Restrepo S, Roy S, Sadanandom A, Savidor A, Schornack S, Schwartz DC, Schumann UD, Schwessinger B, Seyer L, Sharpe T, Silvar C, Song J, Studholme DJ, Sykes S, Thines M, van de Vondervoort PJ, Phuntumart V, Wawra S, Weide R, Win J, Young C, Zhou S, Fry W, Meyers BC, van West P, Ristaino J, Govers F, Birch PR, Whisson SC, Judelson HS, Nusbaum C (2009) Genome sequence and analysis of the irish potato famine pathogen phytophthora infestans. *Nature* 461**:** 393-398

Hahlbrock K, Scheel D, Logemann E, Nurnberger T, Parniske M, Reinold S, Sacks WR, Schmelzer E (1995) Oligopeptide elicitor-mediated defense gene activation in cultured parsley cells. *Proc Natl Acad Sci U S A* 92**:** 4150-4157

Harbury PB, Kim PS, Alber T (1994) Crystal structure of an isoleucine-zipper trimer. *Nature* 371**:** 80-83

Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in gcn4 leucine zipper mutants. *Science* 262**:** 1401-1407

Harper JT, Waanders E, Keeling PJ (2005) On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *International Journal of Systematic and Evolutionary Microbiology* 55**:** 487-496

Hashiro G, Loh PC, Yau JT (1977) The preferential cytotoxicity of reovirus for certain transformed cell lines. *Arch Virol* 54**:** 307-315

Hayakawa T, Satta Y, Gagneux P, Varki A, Takahata N (2001) Alu-mediated inactivation of the human cmp- n-acetylneuraminic acid hydroxylase gene. *Proc Natl Acad Sci U S A* 98**:** 11399-11404

Helander A, Silvey KJ, Mantis NJ, Hutchings AB, Chandran K, Lucas WT, Nibert ML, Neutra MR (2003) The viral sigma1 protein and glycoconjugates containing alpha2-3-linked sialic acid are involved in type 1 reovirus adherence to m cell apical surfaces. *J Virol* 77**:** 7964-7977

Hernandez Alvarez B, Hartmann MD, Albrecht R, Lupas AN, Zeth K, Linke D (2008) A new expression system for protein crystallization using trimeric coiled-coil adaptors. *Protein Eng Des Sel* 21**:** 11-18

Higashi H, Naiki M, Matuo S, Okouchi K (1977) Antigen of "serum sickness" type of heterophile antibodies in human sera: Indentification as gangliosides with n-glycolylneuraminic acid. *Biochem Biophys Res Commun* 79**:** 388-395

Hirst GK (1941) The agglutination of red cells by allantoic fluid of chick embryos infected with influenza virus. *Science* 94**:** 22-23

Hirst GK (1942) Adsorption of influenza hemagglutinins and virus by red blood cells. *J Exp Med* 76**:** 195-209

Hol WG, van Duijnen PT, Berendsen HJ (1978) The alpha-helix dipole and the properties of proteins. *Nature* 273**:** 443-446

Huckelhoven R, Kogel KH (2003) Reactive oxygen intermediates in plant-microbe interactions: Who is who in powdery mildew resistance? *Planta* 216**:** 891-902

Iismaa SE, Holman S, Wouters MA, Lorand L, Graham RM, Husain A (2003) Evolutionary specialization of a tryptophan indole group for transition-state stabilization by eukaryotic transglutaminases. *Proc Natl Acad Sci U S A* 100**:** 12636-12641

Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A (1998) The molecular basis for the absence of n-glycolylneuraminic acid in humans. *J Biol Chem* 273**:** 15866-15871

Jabs T, Tschope M, Colling C, Hahlbrock K, Scheel D (1997) Elicitor-stimulated ion fluxes and o2- from the oxidative burst are essential components in triggering defense gene activation and phytoalexin synthesis in parsley. *Proc Natl Acad Sci U S A* 94**:** 4800-4805

Jones JD, Dangl JL (2006) The plant immune system. *Nature* 444**:** 323-329

Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66**:** 125-132

Kammerer RA, Kostrewa D, Progias P, Honnappa S, Avila D, Lustig A, Winkler FK, Pieters J, Steinmetz MO (2005) A conserved trimerization motif controls the topology of short coiled coils. *Proc Natl Acad Sci U S A* 102**:** 13891-13896

Kammerer RA, Schulthess T, Landwehr R, Lustig A, Engel J, Aebi U, Steinmetz MO (1998) An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc Natl Acad Sci U S A* 95**:** 13419-13424

Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336**:** 1030-1033

Kathan RH, Winzler RJ, Johnsom CA (1961) Preparation of an inhibitor of viral hemagglutination from human erythrocytes. *J Exp Med* 113**:** 37-45

Kaufman L (1998) Penicilliosis marneffei and pythiosis: Emerging tropical diseases. *Mycopathologia* 143**:** 3-7

Kaufmann MJ, Gerdemann JW (1958) Root and stem rot of soybeans caused by *phytophthora sojae* n. Sp. *Phytopathology* 48**:** 201-208

Khatua B, Ghoshal A, Bhattacharya K, Mandal C, Saha B, Crocker PR, Mandal C (2010) Sialic acids acquired by pseudomonas aeruginosa are involved in reduced complement deposition and siglec mediated host-cell recognition. *FEBS Lett* 584**:** 555-561

Kim MG, da Cunha L, McFall AJ, Belkhadir Y, DebRoy S, Dangl JL, Mackey D (2005) Two pseudomonas syringae type iii effectors inhibit rin4-regulated basal defense in arabidopsis. *Cell* 121**:** 749-759

Kirchner E. (2009) Structural and functional studies of the reovirus attachment protein sigma1 and its interaction with the receptor jam-a elektronische ressource. pp. Online-Ressource.

Kirchner E, Guglielmi KM, Strauss HM, Dermody TS, Stehle T (2008) Structure of reovirus sigma1 in complex with its receptor junctional adhesion molecule-a. *PLoS Pathog* 4**:** e1000235

Klenk HD, Rott R, Orlich M, Blodorn J (1975) Activation of influenza a viruses by trypsin treatment. *Virology* 68**:** 426-439

Kobayashi T, Antar AA, Boehme KW, Danthi P, Eby EA, Guglielmi KM, Holm GH, Johnson EM, Maginnis MS, Naik S, Skelton WB, Wetzel JD, Wilson GJ, Chappell JD, Dermody TS (2007) A plasmid-based reverse genetics system for animal double-stranded rna viruses. *Cell Host Microbe* 1**:** 147-157

Kobayashi T, Ooms LS, Ikizler M, Chappell JD, Dermody TS (2010) An improved reverse genetics system for mammalian orthoreoviruses. *Virology* 398**:** 194-200

Koch E, Slusarenko A (1990) Arabidopsis is susceptible to infection by a downy mildew fungus. *Plant Cell* 2**:** 437-445

Kyula JN, Roulstone V, Karapanagiotou EM, Melcher AA, Harrington KJ (2012) Oncolytic reovirus type 3 (dearing) as a novel therapy in head and neck cancer. *Expert Opin Biol Ther* 12**:** 1669-1678

Lajaunias F, Dayer JM, Chizzolini C (2005) Constitutive repressor activity of cd33 on human monocytes requires sialic acid recognition and phosphoinositide 3-kinase-mediated intracellular signaling. *Eur J Immunol* 35**:** 243-251

Lam E, Kato N, Lawton M (2001) Programmed cell death, mitochondria and the plant hypersensitive response. *Nature* 411**:** 848-853

Lazarowitz SG, Choppin PW (1975) Enhancement of the infectivity of influenza a and b viruses by proteolytic cleavage of the hemagglutinin polypeptide. *Virology* 68**:** 440-454

Lee PW, Hayes EC, Joklik WK (1981) Protein sigma 1 is the reovirus cell attachment protein. *Virology* 108**:** 156-163

Lerner AM, Cherry JD, Finland M (1963) Hemagglutination with reoviruses. *Virology* 19**:** 58-65

Ligterink W, Kroj T, zur Nieden U, Hirt H, Scheel D (1997) Receptor-mediated activation of a map kinase in pathogen defense of plants. *Science* 276**:** 2054-2057

Lingwood D, McTamney PM, Yassine HM, Whittle JR, Guo X, Boyington JC, Wei CJ, Nabel GJ (2012) Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* 489**:** 566-570

Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VA (2006) Trimeric autotransporter adhesins: Variable structure, common function. *Trends Microbiol* 14**:** 264-270

Liu J, Elmore JM, Lin ZJ, Coaker G (2011) A receptor-like cytoplasmic kinase phosphorylates the host target rin4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe* 9**:** 137-146

Liu S, Cerione RA, Clardy J (2002) Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. *Proc Natl Acad Sci U S A* 99**:** 2743-2747

Liu Y, Nusrat A, Schnell FJ, Reaves TA, Walsh S, Pochet M, Parkos CA (2000) Human junction adhesion molecule regulates tight junction resealing in epithelia. *J Cell Sci* 113 (Pt 13)**:** 2363-2374

Lorand L, Gray AJ, Brown K, Credo RB, Curtis CG, Domanik RA, Stenberg P (1974) Dissociation of the subunit structure of fibrin stabilizing factor during activation of the zymogen. *Biochem Biophys Res Commun* 56**:** 914-922

Lorand L, Losowsky MS, Miloszewski KJ (1980) Human factor xiii: Fibrin-stabilizing factor. *Prog Hemost Thromb* 5**:** 245-290

Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by calpha geometry: Phi,psi and cbeta deviation. *Proteins* 50**:** 437-450

Lupas A (1996) Prediction and analysis of coiled-coil structures. *Methods Enzymol* 266**:** 513-525

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252**:** 1162-1164

Lupas AN, Gruber M (2005) The structure of alpha-helical coiled coils. *Adv Protein Chem* 70**:** 37-78

Mackey D, Holt BF, 3rd, Wiig A, Dangl JL (2002) Rin4 interacts with pseudomonas syringae type iii effector molecules and is required for rpm1-mediated resistance in arabidopsis. *Cell* 108**:** 743-754

Maginnis MS, Forrest JC, Kopecky-Bromberg SA, Dickeson SK, Santoro SA, Zutter MM, Nemerow GR, Bergelson JM, Dermody TS (2006) Beta1 integrin mediates internalization of mammalian reovirus. *J Virol* 80**:** 2760-2770

Maginnis MS, Mainou BA, Derdowski A, Johnson EM, Zent R, Dermody TS (2008) Npxy motifs in the beta1 integrin cytoplasmic tail are required for functional reovirus entry. *J Virol* 82**:** 3181-3191

Mainou BA, Dermody TS (2012) Transport to late endosomes is required for efficient reovirus infection. *J Virol* 86**:** 8346-8358

Maitra R, Ghalib MH, Goel S (2012) Reovirus: A targeted therapeutic - progress and potential. *Mol Cancer Res*

Malykh YN, Schauer R, Shaw L (2001) N-glycolylneuraminic acid in human tumours. *Biochimie* 83**:** 623-634

Mao ZX, Joklik WK (1991) Isolation and enzymatic characterization of protein lambda 2, the reovirus guanylyltransferase. *Virology* 185**:** 377-386

Martin MJ, Rayner JC, Gagneux P, Barnwell JW, Varki A (2005) Evolution of human-chimpanzee differences in malaria susceptibility: Relationship to human genetic loss of n-glycolylneuraminic acid. *Proc Natl Acad Sci U S A* 102**:** 12819-12824

Martin-Padura I, Lostaglio S, Schneemann M, Williams L, Romano M, Fruscella P, Panzeri C, Stoppacciaro A, Ruco L, Villa A, Simmons D, Dejana E (1998) Junctional adhesion molecule, a novel member of the immunoglobulin superfamily that distributes at intercellular junctions and modulates monocyte transmigration. *J Cell Biol* 142**:** 117-127

McLachlan AD, Stewart M (1975) Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. *J Mol Biol* 98**:** 293-304

Medzhitov R, Janeway CA, Jr. (1997) Innate immunity: Impact on the adaptive immune response. *Curr Opin Immunol* 9**:** 4-9

Merckel MC, Huiskonen JT, Bamford DH, Goldman A, Tuma R (2005) The structure of the bacteriophage prd1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* 18**:** 161-170

Merrick JM, Zadarlik K, Milgrom F (1978) Characterization of the hanganutziu-deicher (serum-sickness) antigen as gangliosides containing n-glycolylneuraminic acid. *Int Arch Allergy Appl Immunol* 57**:** 477-480

Morozov SY (1989) A possible relationship of reovirus putative rna polymerase to polymerases of positive-strand rna viruses. *Nucleic Acids Res* 17**:** 5394

Morris DG, Feng X, Difrancesco LM, Fonseca K, Forsyth PA, Paterson AH, Coffey MC, Thompson B (2012) Reo-001: A phase i trial of percutaneous intralesional administration of reovirus type 3 dearing (reolysin(r)) in patients with advanced solid tumors. *Invest New Drugs*

Morris PF, Ward EWB (1992) Chemoattraction of zoospores of the soybean pathogen, phytophthora-sojae, by isoflavones. *Physiological and Molecular Plant Pathology* 40**:** 17-22

Morrison LA, Sidman RL, Fields BN (1991) Direct spread of reovirus from the intestinal lumen to the central nervous system through vagal autonomic nerve fibers. *Proc Natl Acad Sci U S A* 88**:** 3852-3856

Muchmore EA, Diaz S, Varki A (1998) A structural difference between the cell surfaces of humans and the great apes. *Am J Phys Anthropol* 107**:** 187-198

Mukhtar MS, Carvunis AR, Dreze M, Epple P, Steinbrenner J, Moore J, Tasan M, Galli M, Hao T, Nishimura MT, Pevzner SJ, Donovan SE, Ghamsari L, Santhanam B, Romero V, Poulin MM, Gebreab F, Gutierrez BJ, Tam S, Monachello D, Boxem M, Harbort CJ, McDonald N, Gai L, Chen H, He Y, Vandenhaute J, Roth FP, Hill DE, Ecker JR, Vidal M, Beynon J, Braun P, Dangl JL (2011) Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333**:** 596-601

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53**:** 240-255

Nennstiel D, Scheel D, Nurnberger T (1998) Characterization and partial purification of an oligopeptide elicitor receptor from parsley (petroselinum crispum). *FEBS Lett* 431**:** 405-410

Neu U, Bauer J, Stehle T (2011) Viruses and sialic acids: Rules of engagement. *Curr Opin Struct Biol* 21**:** 610-618

Neu U, Hengel H, Blaum BS, Schowalter RM, Macejak D, Gilbert M, Wakarchuk WW, Imamura A, Ando H, Kiso M, Arnberg N, Garcea RL, Peters T, Buck CB, Stehle T (2012) Structures of merkel cell polyomavirus vp1 complexes define a sialic acid binding site required for infection. *PLoS Pathog* 8**:** e1002738

Neu U, Maginnis MS, Palma AS, Stroh LJ, Nelson CD, Feizi T, Atwood WJ, Stehle T (2010) Structure-function analysis of the human jc polyomavirus establishes the lstc pentasaccharide as a functional receptor motif. *Cell Host Microbe* 8**:** 309-319

Neu U, Woellner K, Gauglitz G, Stehle T (2008) Structural basis of gm1 ganglioside recognition by simian virus 40. *Proc Natl Acad Sci U S A* 105**:** 5219-5224

Nibert ML, Dermody TS, Fields BN (1990) Structure of the reovirus cell-attachment protein: A model for the domain organization of sigma 1. *J Virol* 64**:** 2976-2989

Nibert ML, Fields BN (1992) A carboxy-terminal fragment of protein mu 1/mu 1c is present in infectious subvirion particles of mammalian reoviruses and is proposed to have a role in penetration. *J Virol* 66**:** 6408-6418

Nurnberger T, Brunner F (2002) Innate immunity in plants and animals: Emerging parallels between the recognition of general elicitors and pathogen-associated molecular patterns. *Curr Opin Plant Biol* 5**:** 318-324

Nurnberger T, Brunner F, Kemmerling B, Piater L (2004) Innate immunity in plants and animals: Striking similarities and obvious differences. *Immunol Rev* 198**:** 249-266

Nurnberger T, Nennstiel D, Jabs T, Sacks WR, Hahlbrock K, Scheel D (1994) High affinity binding of a fungal oligopeptide elicitor to parsley plasma membranes triggers multiple defense responses. *Cell* 78**:** 449-460

Odegard AL, Chandran K, Zhang X, Parker JS, Baker TS, Nibert ML (2004) Putative autocleavage of outer capsid protein micro1, allowing release of myristoylated peptide micro1n during particle uncoating, is critical for cell entry by reovirus. *J Virol* 78**:** 8732-8745

Pacitti AF, Gentsch JR (1987) Inhibition of reovirus type 3 binding to host cells by sialylated glycoproteins is mediated through the viral attachment protein. *J Virol* 61**:** 1407-1415

Pack P, Muller K, Zahn R, Pluckthun A (1995) Tetravalent miniantibodies with high avidity assembling in escherichia coli. *J Mol Biol* 246**:** 28-34

Parry DA (1982) Coiled-coils in alpha-helix-containing proteins: Analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. *Biosci Rep* 2**:** 1017-1024

Pedersen LC, Yee VC, Bishop PD, Le Trong I, Teller DC, Stenkamp RE (1994) Transglutaminase factor xiii uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Sci* 3**:** 1131-1135

Pinkas DM, Strop P, Brunger AT, Khosla C (2007) Transglutaminase 2 undergoes a large conformational change upon activation. *PLoS Biol* 5**:** e327

Prota AE, Campbell JA, Schelling P, Forrest JC, Watson MJ, Peters TR, Aurrand-Lions M, Imhof BA, Dermody TS, Stehle T (2003) Crystal structure of human junctional adhesion molecule 1: Implications for reovirus binding. *Proc Natl Acad Sci U S A* 100**:** 5366-5371

Reinisch KM, Nibert ML, Harrison SC (2000) Structure of the reovirus core at 3.6 a resolution. *Nature* 404**:** 960-967

Reiter DM, Frierson JM, Halvorson EE, Kobayashi T, Dermody TS, Stehle T (2011) Crystal structure of reovirus attachment protein sigma1 in complex with sialylated oligosaccharides. *PLoS Pathog* 7**:** e1002166

Reuter G, Gabius HJ (1996) Sialic acids structure-analysis-metabolism-occurrence-recognition. *Biol Chem Hoppe Seyler* 377**:** 325-342

Rizzo DM, Garbelotto M, Hansen EM (2005) Phytophthora ramorum: Integrative research and management of an emerging pathogen in california and oregon forests. *Annu Rev Phytopathol* 43**:** 309-335

Rogers GN, Paulson JC (1983) Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the h3 hemagglutinin based on species of origin. *Virology* 127**:** 361-373

Rogers GN, Paulson JC, Daniels RS, Skehel JJ, Wilson IA, Wiley DC (1983) Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* 304**:** 76-78

Ross AF (1961) Systemic acquired resistance induced by localized virus infections in plants. *Virology* 14**:** 340-358

Rost B (1996) Phd: Predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol* 266**:** 525-539

Rubin DH, Weiner DB, Dworkin C, Greene MI, Maul GG, Williams WV (1992) Receptor utilization by reovirus type 3: Distinct binding sites on thymoma and fibroblast cell lines result in differential compartmentalization of virions. *Microb Pathog* 12**:** 351-365

Sacks W, Nurnberger T, Hahlbrock K, Scheel D (1995) Molecular characterization of nucleotide sequences encoding the extracellular glycoprotein elicitor from phytophthora megasperma. *Mol Gen Genet* 246**:** 45-55

Salzberg S (2008) The contents of the syringe. *Nature* 454**:** 160-161

Sauter NK, Bednarski MD, Wurzburg BA, Hanson JE, Whitesides GM, Skehel JJ, Wiley DC (1989) Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: A 500-mhz proton nuclear magnetic resonance study. *Biochemistry* 28**:** 8388-8396

Savidor A, Donahoo RS, Hurtado-Gonzales O, Land ML, Shah MB, Lamour KH, McDonald WH (2008) Cross-species global proteomics reveals conserved and unique processes in phytophthora sojae and phytophthora ramorum. *Mol Cell Proteomics* 7**:** 1501-1516

Schafer W, Straney D, Ciuffetti L, HD VANE, Yoder OC (1989) One enzyme makes a fungal pathogen, but not a saprophyte, virulent on a new host plant. *Science* 246**:** 247-249

Schauer R (1991) Biosynthesis and function of n- and o-substituted sialic acids. *Glycobiology* 1**:** 449-452

Schelling P, Guglielmi KM, Kirchner E, Paetzold B, Dermody TS, Stehle T (2007) The reovirus sigma1 aspartic acid sandwich: A trimerization motif poised for conformational change. *J Biol Chem* 282**:** 11582-11589

Schiff LA, Nibert M, Tyler KL (2007) Orthoreoviruses and their replication. In *Fields virology*, Knipe DM, Howley PM (eds), Vol. 2, 5 edn, pp 1853–1915. Lippincott Williams & Wilkins, Philadelphia, PA

Schneemann A, Speir JA, Tan GS, Khayat R, Ekiert DC, Matsuoka Y, Wilson IA (2012) A virus-like particle that elicits cross-reactive antibodies to the conserved stem of influenza virus hemagglutinin. *J Virol* 86**:** 11686-11697

Schroder E, Phillips C, Garman E, Harlos K, Crawford C (1993) X-ray crystallographic structure of a papain-leupeptin complex. *FEBS Lett* 315**:** 38-42

Schwartz ML, Pizzo SV, Hill RL, McKee PA (1971) The subunit structures of human plasma and platelet factor xiii (fibrin-stabilizing factor). *J Biol Chem* 246**:** 5851-5854

Schwarzkopf M, Knobeloch KP, Rohde E, Hinderlich S, Wiechens N, Lucka L, Horak I, Reutter W, Horstkorte R (2002) Sialylation is essential for early development in mice. *Proc Natl Acad Sci U S A* 99**:** 5267-5270

Seliger LS, Zheng K, Shatkin AJ (1987) Complete nucleotide sequence of reovirus l2 gene and deduced amino acid sequence of viral mrna guanylyltransferase. *J Biol Chem* 262**:** 16289-16293

Seo J, Cohen C (1993) Pitch diversity in alpha-helical coiled coils. *Proteins* 15**:** 223-234

Shafer JA, Higgins DL (1988) Human fibrinogen. *Crit Rev Clin Lab Sci* 26**:** 1-41

Shatkin AJ, Sipe JD, Loh P (1968) Separation of ten reovirus genome segments by polyacrylamide gel electrophoresis. *J Virol* 2**:** 986-991

Shaw L, Schauer R (1989) Detection of cmp-n-acetylneuraminic acid hydroxylase activity in fractionated mouse liver. *Biochem J* 263**:** 355-363

Slaughter TF, Achyuthan KE, Lai TS, Greenberg CS (1992) A microtiter plate transglutaminase assay utilizing 5-(biotinamido)pentylamine as substrate. *Anal Biochem* 205**:** 166-171

Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G (2012) Exploiting structure similarity in refinement: Automated ncs and target-structure restraints in buster. *Acta Crystallogr D Biol Crystallogr* 68**:** 368-380

Sogin ML, Silberman JD (1998) Evolution of the protists and protistan parasites from the perspective of molecular systematics. *Int J Parasitol* 28**:** 11-20

Spoel SH, Dong XN (2012) How do plants achieve immunity? Defence without specialized immune cells. *Nature Reviews Immunology* 12**:** 89-100

Stehle T, Harrison SC (1996) Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure* 4**:** 183-194

Stehle T, Yan Y, Benjamin TL, Harrison SC (1994) Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature* 369**:** 160-163

Strong JE, Coffey MC, Tang D, Sabinin P, Lee PW (1998) The molecular basis of viral oncolysis: Usurpation of the ras signaling pathway by reovirus. *EMBO J* 17**:** 3351-3362

Sturzenbecker LJ, Nibert M, Furlong D, Fields BN (1987) Intracellular digestion of reovirus particles requires a low ph and is an essential step in the viral infectious cycle. *J Virol* 61**:** 2351-2361

Tai JH, Williams JV, Edwards KM, Wright PF, Crowe JE, Jr., Dermody TS (2005) Prevalence of reovirus-specific antibodies in young children in nashville, tennessee. *J Infect Dis* 191**:** 1221-1224

Takematsu H, Kawano T, Koyama S, Kozutsumi Y, Suzuki A, Kawasaki T (1994) Reaction mechanism underlying cmp-n-acetylneuraminic acid hydroxylation in mouse liver: Formation of a ternary complex of cytochrome b5, cmp-n-acetylneuraminic acid, and a hydroxylation enzyme. *J Biochem* 115**:** 381-386

Traving C, Schauer R (1998) Structure, function and metabolism of sialic acids. *Cell Mol Life Sci* 54**:** 1330-1349

Tyler BM (2001) Genetics and genomics of the oomycete-host interface. *Trends Genet* 17**:** 611-614

Tyler BM (2002) Molecular basis of recognition between phytophthora pathogens and their hosts. *Annu Rev Phytopathol* 40**:** 137-167

Tyler BM (2007) Phytophthora sojae: Root rot pathogen of soybean and model oomycete. *Mol Plant Pathol* 8**:** 1-8

Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CM, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee MK, McDonald WH, Medina M, Meijer HJ, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phuntumart V, Putnam NH, Rash S, Rose JK, Sakihama Y, Salamov AA, Savidor A, Scheuring CF, Smith BM, Sobral BW, Terry A, Torto-Alalibo TA, Win J, Xu Z, Zhang H, Grigoriev IV, Rokhsar DS, Boore JL (2006) Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313**:** 1261-1266

Tyler KL, McPhee DA, Fields BN (1986) Distinct pathways of viral spread in the host determined by reovirus s1 gene segment. *Science* 233**:** 770-774

Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* 56**:** 1622-1624

Vagin A, Teplyakov A (2010) Molecular replacement with molrep. *Acta Crystallogr D Biol Crystallogr* 66**:** 22-25

Vagin AA, Isupov MN (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr* 57**:** 1451-1456

Van der Biezen EA, Jones JD (1998) Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem Sci* 23**:** 454-456

van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* 401**:** 935-938

Vanetten HD, Matthews DE, Matthews PS (1989) Phytoalexin detoxification: Importance for pathogenicity and practical implications. *Annu Rev Phytopathol* 27**:** 143-164

Varki A (1992) Diversity in the sialic acids. *Glycobiology* 2**:** 25-40

Varki A (1993) Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology* 3**:** 97-130

Varki A (2007) Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* 446**:** 1023-1029

Varki A (2010) Colloquium paper: Uniquely human evolution of sialic acid genetics and biology. *Proc Natl Acad Sci U S A* 107 Suppl 2**:** 8939-8946

Varki A, Gagneux P (2009) Human-specific evolution of sialic acid targets: Explaining the malignant malaria mystery? *Proc Natl Acad Sci U S A* 106**:** 14739-14740

Vedula SR, Lim TS, Kirchner E, Guglielmi KM, Dermody TS, Stehle T, Hunziker W, Lim CT (2008) A comparative molecular force spectroscopy study of homophilic jam-a interactions and jam-a interactions with reovirus attachment protein sigma1. *J Mol Recognit* 21**:** 210-216

Wang Y, Gao R, Lynn DG (2002) Ratcheting up vir gene expression in agrobacterium tumefaciens: Coiled coils in histidine kinase signal transduction. *Chembiochem* 3**:** 311-317

Weiner HL, Drayna D, Averill DR, Jr., Fields BN (1977) Molecular basis of reovirus virulence: Role of the s1 gene. *Proc Natl Acad Sci U S A* 74**:** 5744-5748

Weiner HL, Powers ML, Fields BN (1980) Absolute linkage of virulence and central nervous system cell tropism of reoviruses to viral hemagglutinin. *J Infect Dis* 141**:** 609-616

Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC (1988) Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* 333**:** 426-431

Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC (1997) Atomic structure of the ectodomain from hiv-1 gp41. *Nature* 387**:** 426-430

Werres S, Marwitz R, Veld WAMI, De Cock AWAM, Bonants PJM, De Weerdt M, Themann K, Ilieva E, Baayen RP (2001) Phytophthora ramorum sp nov., a new pathogen on rhododendron and viburnum. *Mycological Research* 105**:** 1155-1165

White CL, Twigger KR, Vidal L, De Bono JS, Coffey M, Heinemann L, Morgan R, Merrick A, Errington F, Vile RG, Melcher AA, Pandha HS, Harrington KJ (2008) Characterization of the adaptive and innate immune response to intravenous oncolytic reovirus (dearing type 3) during a phase i clinical trial. *Gene Ther* 15**:** 911-920

Wiley DC, Wilson IA, Skehel JJ (1981) Structural identification of the antibody-binding sites of hong kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* 289**:** 373-378

Williams LA, Martin-Padura I, Dejana E, Hogg N, Simmons DL (1999) Identification and characterisation of human junctional adhesion molecule (jam). *Mol Immunol* 36**:** 1175-1188

Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 a resolution. *Nature* 289**:** 366-373

Wolfe SA, Grant RA, Pabo CO (2003) Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry* 42**:** 13401-13409

Xavier B, Annelies V, Kurt H, Frederic L, Anne C (2010) Oospores progenies from phytophthora ramorum. *Fungal Biol* 114**:** 369-378

Yap TA, Brunetto A, Pandha H, Harrington K, Debono JS (2008) Reovirus therapy in cancer: Has the orphan virus found a home? *Expert Opin Investig Drugs* 17**:** 1925-1935

Yee VC, Pedersen LC, Le Trong I, Bishop PD, Stenkamp RE, Teller DC (1994) Three-dimensional structure of a transglutaminase: Human blood coagulation factor xiii. *Proc Natl Acad Sci U S A* 91**:** 7296-7300

Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002) The single, ancient origin of chromist plastids. *Proc Natl Acad Sci U S A* 99**:** 15507-15512

Zhang X, Tang J, Walker SB, O'Hara D, Nibert ML, Duncan R, Baker TS (2005) Structure of avian orthoreovirus virion by electron cryomicroscopy and image reconstruction. *Virology* 343**:** 25-35

Zimmermann S, Nurnberger T, Frachisse JM, Wirtz W, Guern J, Hedrich R, Scheel D (1997) Receptor-mediated activation of a plant ca2+-permeable ion channel involved in pathogen defense. *Proc Natl Acad Sci U S A* 94**:** 2751-2755

4 Appendix

4.1 Acknowledgments

Ein großes Dankeschön geht an…

… Prof. Dr. Thilo Stehle für die freundliche und kompetente Betreuung meiner Promotion und für die Möglichkeiten, auch über den Tellerrand des eigenen Projekts hinaus zu schauen. Seine engagierte und von einem unerschütterlichen Optimismus getragene Projektleitung sorgte auch in schwierigen Phasen für das nötige Durchhaltevermögen.

… Prof. Dr. Terence S. Dermody, my second supervisor, for examination of my dissertation, helpful discussions on our regular telephone conferences, and his positive and encouraging attitude during our pleasant collaboration.

… Dr. Frédéric Brunner, Birgit Löffelhardt und Prof. Dr. Thorsten Nürnberger für die gute Zusammenarbeit im Rahmen des GP42-Projekts.

… das Reovirus-Team, vor allem Dirk, Jen, Eva und Melanie, mit denen ich Rückschläge, Herausforderungen und Fortschritte des σ1-Projektes teilen durfte.

… die gesamte Arbeitsgruppe, insbesondere Georg, Ulla, Johannes und Bärbel, für Hilfestellungen und Projektdiskussionen bei Kaffee-, Schoki- oder Obstpausen.

… Susi und Tina, die mich als Hiwis bei der Laborarbeit unterstützt und auch außerhalb des Labors immer wieder für Abwechslung gesorgt haben.

… die „Girls Gang & Nici“, sowie Riki, Vera, Yvonne, Denise, Kathrin, Kristin, Ancilla und Ulla für „Frauenabende“ und diverse andere Unternehmungen.

… die Sportfreunde, allen voran Volker, Johannes, Dirk, Sebastian, Manuel, Riki und Moritz, mit denen ich die beiden schönsten Sportarten der Welt (Beachvolleyball und Fußball) bei vielen Gelegenheiten erleben konnte.

… alle Kollegen des AK Stehle für die gute Arbeitsatmosphäre und die vielen positiven Erinnerungen und lustigen Begebenheiten.

… meine Eltern und meine Schwester, die meine Arbeit mit Interesse verfolgt und mich immer und in jeder Hinsicht unterstützt haben.

4.2 COILS/PCOILS results for the T1L σ1 tail

Table 4: COILS/PCOILS numerical results for T1L σ1 (aa 1-180). Column 1: residue number and amino acid; columns 2-4: positions in heptad repeat and coiled coil scores for windows of 14, 21, and 28 residues.

		14 aa window		21 aa window		28 aa window	
Residue		Position	Score	Position	Score	Position	Score
1	M	d	0.001	g	0	g	0
2	D	e	0.001	a	0	a	0
3	A	b	0.093	b	0	b	0
4	S	c	0.093	c	0	c	0
5	L	d	0.093	d	0	d	0
6	I	e	0.093	e	0	e	0
7	T	f	0.093	f	0	d	0
8	E	g	0.093	g	0	e	0
9	I	a	0.093	a	0	f	0
10	R	b	0.093	b	0	g	0
11	K	c	0.093	e	0	e	0
12	I	d	0.093	f	0	f	0
13	V	e	0.093	g	0.001	c	0
14	L	f	0.093	a	0.003	d	0
15	Q	g	0.093	b	0.003	e	0
16	L	a	0.093	c	0.003	f	0
17	S	b	0.06	d	0.003	g	0
18	V	c	0.03	e	0.003	e	0.001
19	S	g	0.005	f	0.003	f	0.004
20	S	g	0.008	g	0.003	g	0.004
21	N	a	0.023	a	0.003	a	0.007
22	G	b	0.034	b	0.003	b	0.036
23	S	c	0.192	c	0.007	c	0.194
24	Q	d	0.192	g	0.008	d	0.194
25	S	e	0.192	e	0.046	e	0.798
26	K	f	0.192	f	0.082	f	0.798
27	E	g	0.192	g	0.082	g	0.798
28	I	a	0.192	a	0.082	a	0.798

29	E	b	0.192	b	0.082	b	0.798
30	E	c	0.192	c	0.082	c	0.798
31	I	d	0.192	d	0.082	d	0.798
32	K	e	0.192	e	0.082	e	0.798
33	K	f	0.192	f	0.082	f	0.798
34	Q	g	0.192	g	0.082	g	0.798
35	V	a	0.192	a	0.082	a	0.798
36	Q	b	0.192	b	0.082	b	0.798
37	V	c	0.094	c	0.082	c	0.798
38	N	g	0.025	d	0.082	d	0.798
39	V	e	0.42	e	0.585	e	0.798
40	D	f	0.42	f	0.768	f	0.957
41	D	g	0.42	g	0.832	g	0.957
42	I	a	0.474	a	0.832	a	0.957
43	R	b	0.474	b	0.832	b	0.963
44	A	c	0.474	c	0.832	c	0.963
45	A	d	0.557	d	0.832	d	0.963
46	N	e	0.557	e	0.832	e	0.963
47	I	f	0.557	f	0.832	f	0.963
48	K	g	0.557	g	0.832	g	0.963
49	L	a	0.557	a	0.832	a	0.963
50	D	b	0.557	b	0.832	b	0.963
51	G	c	0.557	c	0.832	c	0.963
52	L	d	0.557	d	0.832	d	0.963
53	G	e	0.557	e	0.832	e	0.963
54	R	f	0.557	f	0.832	f	0.963
55	Q	g	0.557	g	0.832	g	0.963
56	I	a	0.557	a	0.832	a	0.963
57	A	b	0.557	b	0.832	b	0.963
58	D	c	0.557	c	0.832	c	0.963
59	I	d	0.274	d	0.832	d	0.963
60	S	e	0.274	e	0.832	e	0.963
61	N	f	0.274	f	0.832	f	0.963
62	S	g	0.104	g	0.664	g	0.963

63	I	a	0.07	a	0.664	a	0.963
64	S	b	0.077	b	0.52	b	0.963
65	T	c	0.077	c	0.41	c	0.963
66	I	d	0.077	d	0.314	d	0.963
67	E	e	0.077	e	0.402	e	0.963
68	S	f	0.077	f	0.402	f	0.963
69	R	g	0.077	g	0.402	g	0.963
70	L	a	0.077	a	0.402	a	0.963
71	G	b	0.077	b	0.402	b	0.914
72	E	c	0.077	c	0.402	c	0.914
73	M	d	0.077	d	0.402	d	0.89
74	D	e	0.077	e	0.402	e	0.89
75	N	f	0.077	f	0.402	f	0.89
76	R	g	0.077	g	0.402	g	0.881
77	L	a	0.077	a	0.402	a	0.881
78	V	b	0.055	b	0.402	b	0.648
79	G	c	0.041	c	0.402	c	0.648
80	I	d	0.041	d	0.402	d	0.443
81	S	e	0.041	e	0.402	e	0.443
82	S	f	0.041	f	0.402	f	0.443
83	Q	g	0.041	g	0.402	g	0.443
84	V	a	0.041	a	0.402	a	0.443
85	T	b	0.041	b	0.402	b	0.443
86	Q	c	0.041	c	0.402	c	0.443
87	L	d	0.041	d	0.402	d	0.443
88	S	e	0.041	e	0.258	e	0.443
89	N	f	0.041	f	0.258	f	0.443
90	S	g	0.023	g	0.111	g	0.443
91	V	a	0.023	a	0.081	a	0.443
92	S	b	0.023	b	0.081	b	0.443
93	Q	c	0.023	c	0.068	c	0.443
94	N	d	0.004	d	0.039	d	0.096
95	T	e	0.032	e	0.039	e	0.096
96	Q	f	0.109	f	0.039	f	0.096

97	S	g	0.109	g	0.039	g	0.096
98	I	a	0.109	a	0.039	a	0.096
99	S	b	0.109	b	0.039	b	0.096
100	S	c	0.109	c	0.039	c	0.096
101	L	d	0.109	d	0.039	d	0.096
102	G	e	0.109	e	0.039	e	0.096
103	D	f	0.109	f	0.039	f	0.096
104	R	g	0.109	g	0.039	g	0.096
105	I	a	0.109	a	0.039	a	0.096
106	N	b	0.109	b	0.039	b	0.096
107	A	c	0.109	c	0.035	c	0.096
108	V	d	0.109	d	0.035	d	0.096
109	E	e	0.109	e	0.035	e	0.096
110	P	f	0.008	f	0.035	f	0.004
111	R	g	0.008	g	0.035	g	0.004
112	V	a	0.008	a	0.035	a	0.005
113	D	b	0.008	b	0.035	b	0.019
114	S	c	0.008	c	0.035	c	0.019
115	L	d	0.008	d	0.035	d	0.019
116	D	e	0.008	e	0.028	e	0.019
117	T	f	0.005	f	0.012	f	0.019
118	V	g	0.002	g	0.01	g	0.019
119	T	a	0.002	a	0.002	a	0.066
120	S	b	0.01	b	0.188	b	0.864
121	N	c	0.01	c	0.274	c	0.864
122	L	d	0.01	d	0.274	d	0.864
123	T	e	0.01	e	0.274	e	0.864
124	G	f	0.01	f	0.274	f	0.864
125	R	g	0.01	g	0.274	g	0.864
126	T	a	0.01	a	0.274	a	0.864
127	S	b	0.698	b	0.975	b	0.927
128	T	c	0.812	c	0.975	c	0.973
129	L	d	0.883	d	0.975	d	0.973
130	E	e	0.883	e	0.975	e	0.973

131	A	f	0.883	f	0.975	f	0.973
132	D	g	0.883	g	0.975	g	0.973
133	V	a	0.883	a	0.975	a	0.973
134	G	b	0.883	b	0.975	b	0.973
135	S	c	0.883	c	0.975	c	0.973
136	L	d	0.883	d	0.975	d	0.973
137	R	e	0.883	e	0.975	e	0.973
138	T	f	0.883	f	0.975	f	0.973
139	E	g	0.883	g	0.975	g	0.973
140	L	a	0.883	a	0.975	a	0.973
141	A	b	0.883	b	0.975	b	0.973
142	A	c	0.883	c	0.975	c	0.973
143	L	d	0.883	d	0.975	d	0.973
144	T	e	0.686	e	0.975	e	0.973
145	T	f	0.686	f	0.975	f	0.973
146	R	g	0.686	g	0.975	g	0.973
147	V	a	0.686	a	0.975	a	0.973
148	T	b	0.686	b	0.965	b	0.973
149	T	c	0.602	c	0.965	c	0.973
150	E	d	0.045	d	0.563	d	0.973
151	V	a	0.019	e	0.563	e	0.973
152	T	b	0.019	f	0.563	f	0.973
153	R	c	0.019	g	0.563	g	0.973
154	L	d	0.019	a	0.563	a	0.973
155	D	e	0.019	b	0.563	b	0.973
156	G	f	0.019	c	0.398	c	0.959
157	L	g	0.019	d	0.398	d	0.959
158	I	a	0.019	e	0.379	e	0.798
159	N	b	0.019	f	0.379	f	0.798
160	S	c	0.012	g	0.106	g	0.612
161	G	d	0.004	d	0.021	d	0.081
162	Q	e	0.158	e	0.021	e	0.081
163	N	f	0.245	f	0.021	f	0.081
164	S	g	0.245	g	0.021	g	0.081

165	I	a	0.245	a	0.021	a	0.081
166	G	b	0.245	b	0.021	b	0.081
167	E	c	0.245	c	0.021	c	0.081
168	L	d	0.245	d	0.021	d	0.081
169	S	e	0.245	e	0.021	e	0.081
170	T	f	0.245	f	0.021	f	0.081
171	R	g	0.245	g	0.021	g	0.081
172	L	a	0.245	a	0.021	a	0.081
173	S	b	0.245	b	0.021	b	0.081
174	N	c	0.245	c	0.021	c	0.081
175	V	d	0.245	d	0.01	d	0.081
176	E	e	0.245	e	0.01	e	0.081
177	T	f	0.14	f	0.01	f	0.081
178	S	g	0.14	g	0.009	g	0.035
179	M	a	0.14	a	0.008	a	0.035
180	V	b	0.14	b	0.003	b	0.032

4.3 Publications

Reiss K, Stencel JE, Liu Y, Blaum BS, Reiter DM, Feizi T, Dermody TS, Stehle T (2012). The GM2 Glycan Serves as a Functional Co-Receptor for Serotype 1 Reovirus. PLoS Pathogens 8(12): e1003078.

Reiss K, Kirchner E, Gijzen M, Zocher G, Löffelhardt B, Nürnberger T, Stehle T, Brunner F (2011). Structural and Phylogenetic Analyses of the GP42 Transglutaminase from *Phytophthora sojae* Reveal an Evolutionary Relationship between Oomycetes and Marine *Vibrio* Bacteria. J Biol Chem. 286(49): 42585-93.

Bauer J, **Reiss K**, Veerabagu M, Heunemann M, Harter K, and Stehle T (2012). Structure-function analysis of Arabidopsis thaliana histidine kinase AHK5 bound to its cognate phosphotransfer protein AHP1. Mol Plant, doi: 10.1093/mp/sss126.

OPEN ACCESS Freely available online

PLOS | PATHOGENS

The GM2 Glycan Serves as a Functional Coreceptor for Serotype 1 Reovirus

Kerstin Reiss[1☯], Jennifer E. Stencel[2,3☯], Yan Liu[4], Bärbel S. Blaum[1], Dirk M. Reiter[1], Ten Feizi[4], Terence S. Dermody[2,3,5*], Thilo Stehle[1,5*]

1 Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany, **2** Department of Pathology, Microbiology, and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **3** Elizabeth B. Lamb Center for Pediatric Research, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **4** Glycosciences Laboratory, Department of Medicine, Imperial College London, London, United Kingdom, **5** Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America

Abstract

Viral attachment to target cells is the first step in infection and also serves as a determinant of tropism. Like many viruses, mammalian reoviruses bind with low affinity to cell-surface carbohydrate receptors to initiate the infectious process. Reoviruses disseminate with serotype-specific tropism in the host, which may be explained by differential glycan utilization. Although α2,3-linked sialylated oligosaccharides serve as carbohydrate receptors for type 3 reoviruses, neither a specific glycan bound by any reovirus serotype nor the function of glycan binding in type 1 reovirus infection was known. We have identified the oligosaccharide portion of ganglioside GM2 (the GM2 glycan) as a receptor for the attachment protein σ1 of reovirus strain type 1 Lang (T1L) using glycan array screening. The interaction of T1L σ1 with GM2 in solution was confirmed using NMR spectroscopy. We established that GM2 glycan engagement is required for optimal infection of mouse embryonic fibroblasts (MEFs) by T1L. Preincubation with GM2 specifically inhibited type 1 but not type 3 reovirus infection of MEFs. To provide a structural basis for these observations, we defined the mode of receptor recognition by determining the crystal structure of T1L σ1 in complex with the GM2 glycan. GM2 binds in a shallow groove in the globular head domain of T1L σ1. Both terminal sugar moieties of the GM2 glycan, *N*-acetylneuraminic acid and *N*-acetylgalactosamine, form contacts with the protein, providing an explanation for the observed specificity for GM2. Viruses with mutations in the glycan-binding domain display diminished hemagglutination capacity, a property dependent on glycan binding, and reduced capacity to infect MEFs. Our results define a novel mode of virus-glycan engagement and provide a mechanistic explanation for the serotype-dependent differences in glycan utilization by reovirus.

Citation: Reiss K, Stencel JE, Liu Y, Blaum BS, Reiter DM, et al. (2012) The GM2 Glycan Serves as a Functional Coreceptor for Serotype 1 Reovirus. PLoS Pathog 8(12): e1003078. doi:10.1371/journal.ppat.1003078

Editor: Billy Tsai, University of Michigan, United States of America

Received July 13, 2012; **Accepted** October 23, 2012; **Published** December 6, 2012

Funding: This work was supported by US Public Health Service award R01 AI76983, the Elizabeth B. Lamb Center for Pediatric Research, the UK Research Councils' Basic Technology Initiative 'Glycoarrays' (GRS/79268), EPSRC Translational Grant (EP/G037604/1), the Wellcome Trust (093378MA), and the National Cancer Institute Alliance of Glycobiologists (U01CA128416). Additional support was provided by US Public Health Service awards P30 CA68485 for the Vanderbilt-Ingram Cancer Center and P60 DK20593 for the Vanderbilt Diabetes Research and Training Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: terry.dermody@vanderbilt.edu (TSD); thilo.stehle@uni-tuebingen.de (TS)

☯ These authors contributed equally to this work.

Introduction

Virus infections are initiated by attachment of the virus to target cells of susceptible hosts. Receptors facilitate attachment, determine host range, and govern susceptibility of particular cells to infection. While viral attachment can be a monophasic event, this process frequently involves multiple receptors, and adhesion strengthening is a common mechanism that facilitates virus entry [1]. Thus, a virus may interact with an attachment factor, commonly a carbohydrate, to adhere via low-affinity interaction to the cell-surface, where it then binds to an additional receptor with high affinity that leads to viral entry. The identities of the low-affinity attachment factors are not known for many viruses.

Mammalian orthoreoviruses (reoviruses) serve as highly tractable models to study virus-receptor interactions. These viruses replicate to high titer, facilitating biochemical and biophysical studies, and both the virus and host can be manipulated genetically. Reoviruses contain ten segments of double-stranded RNA (dsRNA) encapsidated within two protein shells. Reoviruses can infect the gastrointestinal and respiratory tracts of a variety of mammals but rarely cause systemic disease outside of the immediate newborn period [2]. Most children are seropositive for reovirus by the age of 5 years [3]. Reoviruses preferentially infect tumor cells and are being tested in clinical trials for the treatment of a variety of cancers [4–6]. It is not yet clear why reoviruses infect tumor cells more efficiently than untransformed cells, but it is likely that distribution, accessibility, and density of cellular receptors contribute to this process.

The three known reovirus serotypes are represented by the prototype strains type 1 Lang (T1L), type 2 Jones (T2J), and type 3 Dearing (T3D). These three strains differ markedly in cell tropism and viral spread, and these properties have been studied extensively using newborn mice [7]. T1L spreads hematogenously and infects ependymal cells, leading to non-lethal hydrocephalus

Author Summary

Receptor utilization plays an important role in viral disease. Viruses must recognize a receptor or sometimes multiple receptors to infect a cell. Mammalian orthoreoviruses (reoviruses) serve as useful models for studies of viral receptor binding and pathogenesis. The reovirus experimental system allows manipulation of both the virus and the host to define mechanisms of viral attachment and disease. Like many viruses, reoviruses engage carbohydrate molecules on the cell-surface, but the oligosaccharide sequences bound and the function of glycan binding in infection were not known prior to this study. We used glycan array screening to determine that serotype 1 reoviruses bind ganglioside GM2 and found that this interaction is required for efficient infection of some types of cells. To better understand how reovirus engages GM2, we determined the structure of the reovirus attachment protein σ1 in complex with the GM2 glycan and defined residues that are required for functional receptor binding. Reoviruses are being tested in clinical trials for efficacy in the treatment of cancer. Cancer cells commonly have altered glycan profiles. Therefore, understanding how reoviruses engage cell-surface glycans might lead to improvements in oncolytic therapy.

[8,9]. In contrast, T3D disseminates hematogenously and neurally and infects neurons, causing lethal encephalitis [7–12]. These serotype-dependent differences are linked to sequence variations in the σ1 outer-capsid protein [7,9].

The σ1 protein mediates the attachment of the virus to target cells [9,13]. It is a 150 kDa homotrimeric protein that assembles into a long fiber that protrudes from the virion surface [14]. The σ1 protein can be partitioned into three functionally and structurally distinct domains: the tail, body, and head. The N-terminal tail spans about 170 residues and is predicted to form an α-helical coiled coil [15–17]. The body domain comprises approximately 100 residues and primarily consists of β-spiral repeats [18,19]. The C-terminal 150 residues fold into the compact head domain composed of eight antiparallel β-strands that assemble into a jelly-roll [18]. The head binds with high affinity to junctional adhesion molecule-A (JAM-A) [20], which serves as a receptor for all known reovirus serotypes [21]. JAM-A is a homodimeric member of the immunoglobulin superfamily [22] located in tight junctions [23].

The structure and receptor-binding properties of reovirus T3D σ1 have been studied most extensively [18,19,24,25]. Interactions of T3D σ1 and JAM-A exclusively involve the σ1 head, which binds the N-terminal D1 domain of JAM-A [25,26]. JAM-A binds with higher affinity to σ1 than to itself; thus, the engagement of σ1 to JAM-A disrupts the JAM-A homodimer. The JAM-A-binding site is highly conserved among the three reovirus serotypes; thus, it is predicted that the T1L, T2J, and T3D reovirus σ1 proteins engage JAM-A in a similar manner and with similar affinities. Although binding to JAM-A is required for hematogenous dissemination, differences in target cell selection within the CNS displayed by T1L and T3D are retained in JAM-A deficient mice inoculated with the viruses intracranially [11]. Therefore, interactions with JAM-A are unlikely to dictate the serotype-specific differences in cell tropism in the nervous system. Instead, these differences in tropism are likely a consequence of virus binding to serotype-specific receptors.

In addition to JAM-A, reoviruses bind to cell-surface glycans. However, the limited knowledge of glycan coreceptors for reovirus is an obstacle to a precise understanding of the contribution of individual receptors to viral tropism and disease. While there is considerable information about carbohydrate-mediated interactions of T3D with host cells, the role of glycan binding in other reovirus serotypes is not known. T3D σ1 interacts with α-linked 5-*N*-acetyl neuraminic acid (Neu5Ac) [19,27], and crystal structures of T3D σ1 in complex with sialyllactose-based compounds terminating in α2,3-, α2,6-, and α2,8-linked Neu5Ac have identified the glycan-binding site [19]. The N-terminal portion of the T3D σ1 body, which lies close to the mid-point of the molecule, engages Neu5Ac via a complex network of interactions that are identical for the three linkages tested. Contacts include a bidentate salt bridge, which connects arginine 202 with the Neu5Ac carboxylate, and a number of augmenting hydrogen bonds and non-polar interactions. The additional sugar rings of the lactose backbone make minimal contacts with T3D σ1, suggesting that T3D σ1 recognizes a different carbohydrate sequence on the cell-surface [19].

Much less is known about the interaction of type 1 reovirus with cell-surface glycans. Hemagglutination is dependent on glycan-engagement, and serotypes 1 and 3 display differences in hemagglutination profiles, suggesting that they differentially engage cell-surface glycans [28]. Type 1 reoviruses agglutinate human and not bovine red blood cells, whereas type 3 reoviruses agglutinate bovine erythrocytes well and human erythrocytes less efficiently than type 1 strains [29]. Hemagglutination studies using chimeric and truncated σ1 proteins expressed in insect cells using baculovirus vectors suggest that the carbohydrate-binding site of T1L σ1 resides just beneath the head domain [27]. Additionally, neuraminidase treatment diminishes infection of intestinal M cells by T1L, suggesting that type 1 reoviruses can engage sialic acid at least in some contexts [30]. T1L reovirus does not bind to sialylated glycophorin, whereas T3D reovirus does [27,31]. Therefore, the glycan recognized by type 1 reoviruses differs from that recognized by type 3 strains.

In this study, we employed glycan microarray analyses to identify ganglioside GM2 as a glycan receptor for reovirus T1L, and we used structural and infectivity data to define the glycan-protein interaction and the biological relevance of glycan binding to infection of host cells. Taken together, our structure-function data provide insight into how the GM2 glycan is specifically recognized by type 1 reovirus and explain the serotype-specific nature of reovirus glycan utilization.

Results

Infectivity of T1L reovirus is dependent on Neu5Ac

To investigate glycan engagement by T1L, we established a cell-culture system in which glycan binding could be evaluated. Binding to sialic acid is dispensable for infection of murine L929 (L) fibroblast cells by either type 1 or type 3 reovirus [27,32,33]. However, sialic acid engagement is required for optimal infection of MEFs [11,33] and HeLa cells by type 3 reoviruses [25,27,33]. To determine whether sialylated glycan engagement is required for efficient infection by T1L, we pretreated L cells (Figure 1A) and MEFs (Figure 1B) with *Arthrobacter ureafaciens* neuraminidase to remove cell-surface sialic acid. Neuraminidase treatment did not impair the capacity of T1L to infect L cells, as previously shown [32]. In contrast, neuraminidase treatment reduced T1L infectivity of MEFs (Figure 1B) and also HeLa cells (data not shown), suggesting that sialic acid engagement by T1L is required for optimal infection of some cell types. Of note, GM2 is expressed on MEFs [34], which display glycan-dependent infection, and L cells [35], which do not require glycan-binding for infection. While

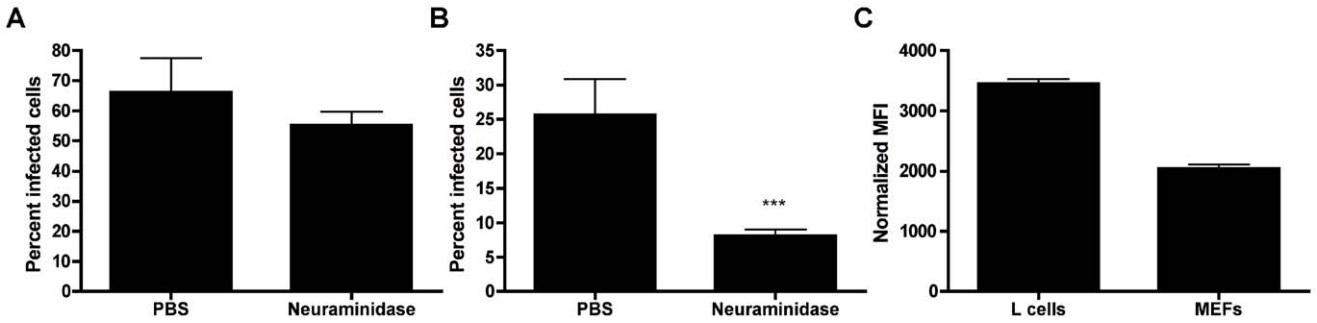

Figure 1. The effect of neuraminidase treatment on T1L infectivity in L cells and MEFs. (**A**) L cells or (**B**) MEFs were treated with *A. ureafaciens* neuraminidase for 1 h, followed by adsorption of T1L at an MOI of 10 or 100 PFU/cell, respectively. Cells were washed twice with PBS, and fresh medium was added. After incubation at 37°C for 20 h, cells were fixed, and reovirus antigen was detected by indirect immunofluorescence. Nuclei were stained with DAPI. The percentage of infected cells in three fields of view per well was determined. The results are expressed as the mean percent infected cells per well in triplicate wells for two independent experiments. Error bars represent standard deviations. (**A**) n.s., (**B**) ***, $P<0.0001$, as determined by two-tailed Student's *t* test. (**C**) L cells or MEFs were stained with anti-JAM-A antibody followed by Alexa-488 labeled secondary antibody to measure cell-surface JAM-A expression. Fluorescence was detected by flow cytometry. Cells were gated on forward and side scatter and the mean fluorescence intensity (MFI) of Alexa-488 was quantified. Results shown are from a representative of three experiments each done with duplicate samples.
doi:10.1371/journal.ppat.1003078.g001

both L cells and MEFs are of murine origin, differences in sialic acid requirements are likely accounted for by differences in the expression on these cells of the known proteinaceous reovirus receptor, JAM-A. L cells, which do not require sialic acid for efficient entry, express higher levels of cell-surface JAM-A than do MEFs (Figure 1C). Thus, T1L may infect MEFs using an adhesion-strengthening mechanism in which binding to glycans must precede engagement of the relatively low abundance JAM-A receptor.

Glycan array screening identifies GM2 as a preferred ligand for T1L σ1

To assess the carbohydrate-binding specificity of T1L reovirus, we expressed and purified recombinant hexahistidine-tagged T1L σ1 protein for binding analyses in neoglycolipid-based glycan microarrays. Based on sequence alignment with T3D σ1, for which several crystal structures exist [18,24,25], two constructs were designed. The first construct, $\sigma1_{long}$, comprised amino acids 261–470, which were predicted to fold into three β-spiral repeats and the C-terminal head domain. The second construct, $\sigma1_{short}$, comprised amino acids 300–470, which were predicted to form only the most C-terminal β-spiral and the head domain. Both σ1 constructs included the predicted carbohydrate-binding site, which was reported to lie in close proximity to the head domain [27].

Glycan microarray analyses were carried out initially with $\sigma1_{long}$ using an array composed of 124 lipid-linked oligosaccharide probes. Among these are 119 sialylated probes with differing sialic acid linkages, backbone sequences, chain lengths, and branching patterns; five non-sialylated probes were included as negative controls (Table S1). The results from the glycan array screening showed a signal for the ganglioside GM2 that, despite its low intensity, was significantly stronger than the other signals (Figure S1). The GM2 glycan sequence contains two terminal sugars, Neu5Ac and *N*-acetylgalactosamine (GalNAc), that are both linked to a central galactose (Gal) via α2,3 and β1,4 linkages, respectively. The Gal is connected, via a β1,4 linkage, to a glucose (Glc), which is attached to a ceramide anchor.

Additional analyses were carried out with the $\sigma1_{short}$ construct, which was predicted to have less steric hindrance imposed by the long body domain and, therefore, to perhaps yield clearer results. Since the initial screen with $\sigma1_{long}$ revealed GM2 as a likely carbohydrate receptor, the second array was comprised of 21 ganglioside-related saccharide probes that included GM2 (Table S2). The results from this screen confirmed binding of the protein to GM2 and yielded a higher signal-to-noise ratio than the initial screen (Figure 2A). GM2 clearly exhibited the highest signal among the probes investigated, whereas several other structurally closely related probes (Figure 2B), e.g., the "a series" gangliosides GM3, GM1, and GD1a (sequences in Table S2), elicited marginally detectable low signals. The overall binding intensity of the σ1 protein, even with the short construct, is lower than that of other proteins tested in the same arrays, e.g., the VP1 proteins of polyoma viruses JCV and SV40, and the fiber knobs of adenovirus Ad37 (data not shown).

T1L σ1 interacts with the GM2 glycan in solution

To verify that T1L σ1 binds specifically to the GM2 glycan, we performed STD NMR spectroscopy experiments with σ1 and the glycan. This method is especially well suited to detect low-affinity binding between a large molecule, such as σ1, and a small oligosaccharide [36–38]. In an STD NMR experiment, the protein is selectively excited, and magnetization transfer to the ligand is observed if complex formation and rapid release of the ligand take place. If these conditions are fulfilled, the STD spectrum contains ligand resonances belonging to the binding epitope. A control experiment without protein serves to exclude direct excitation of the ligand. Using STD NMR, we found that T1L σ1 binds to the GM2 oligosaccharide in solution. Moreover, the STD analysis identified the protons of the carbohydrate that lie in close proximity (about 5 Å) to σ1 in the complex (Figure 2C, Figure S2A). All of the GM2 protons in the σ1-GM2 complex are part of the terminal Neu5Ac or the GalNAc moieties. The most prominent peak in the STD NMR spectrum belongs to the Neu5Ac methyl group, which receives considerably more saturation than the GalNAc methyl group. Protons H5, H6, H7, and one of the two H9 protons of Neu5Ac also are readily identified in the STD NMR spectrum, while the axial and equatorial H3 protons of this moiety receive little, if any, magnetization from the protein. Saturation transfer to the Neu5Ac protons H4 and H8 cannot be evaluated unambiguously because the resonances of both overlap with each other and with the GalNAc H6 resonance. Protons H1 through H4 of the GalNAc ring also are seen in the difference

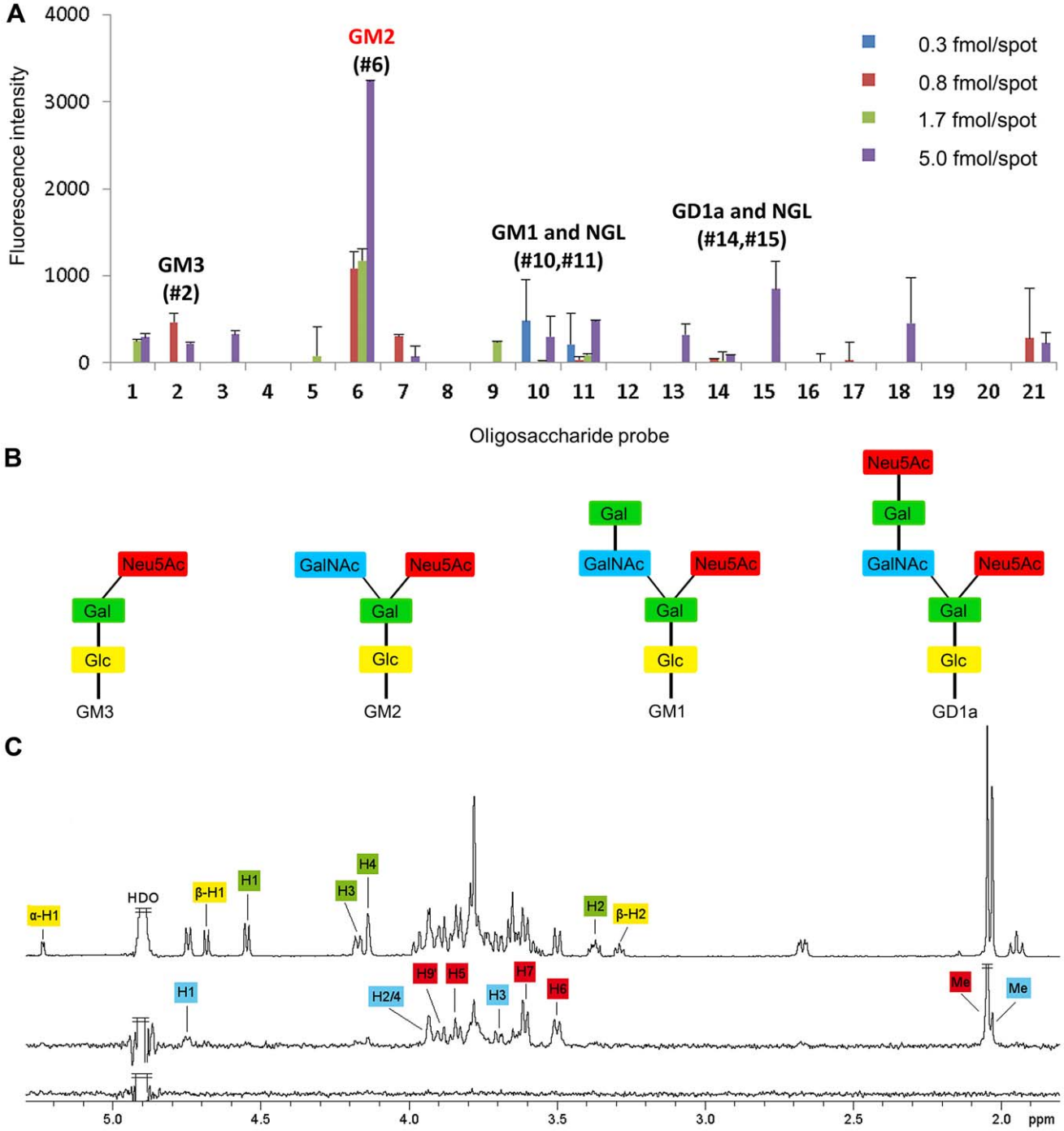

Figure 2. T1L reovirus uses GM2 as a coreceptor. (**A**) Glycan microarray analysis of recombinant T1L $\sigma1_{short}$ using 21 lipid-linked oligosaccharide probes. Each oligosaccharide probe was arrayed at four levels (as indicated) in duplicate. Numerical scores of the binding signals are means of duplicate spots (with error bars). The complete list of probes and their sequences are provided in Table S2. (**B**) Diagrams of "a series" gangliosides GM3, GM2, GM1, and GD1a present in the glycan array. Ceramide, glucose (Glc), galactose (Gal), *N*-5-acetyl neuraminic acid (Neu5Ac), and *N*-acetylgalactosamine (GalNAc) moieties are indicated. (**C**) STD NMR spectroscopy demonstrates that T1L σ1 binds to the GM2 glycan in solution. Upper spectrum: ^{1}H spectrum of the GM2 oligosaccharide alone; middle: STD spectrum of T1L σ1 and the GM2 glycan; and lower spectrum: STD spectrum of the GM2 glycan alone as a control for direct excitation of the ligand. The protons are labeled and color-coded according to the sugar moieties within the GM2 oligosaccharide. The large peak just below 3.8 ppm unites the Neu5Ac H4 and H8 and the GalNAc H6 resonances.
doi:10.1371/journal.ppat.1003078.g002

spectrum, although they are generally less prominent than the Neu5Ac protons. No noteworthy transfer was observed for the GM2 galactose and glucose rings. Thus, the STD NMR spectroscopy data show that the T1L σ1-GM2 glycan interaction is based on contacts with ring atoms and the glycerol side chain of Neu5Ac, with additional contacts contributed by GalNAc ring atoms. The STD NMR experiment was repeated with the linear GM3 glycan (Figure S2B), which lacks the terminal GalNAc

present on GM2. The difference spectrum demonstrates that the GM3 trisaccharide interacts with T1L σ1 and that saturation transfer is observed to Neu5Ac protons only. The STD NMR experiments allow no direct estimate of relative affinities for GM2 and GM3, but it is likely that T1L σ1 binds with greater affinity to the GM2 glycan because of the additional contacts with the terminal GalNAc of this compound. This assumption is consistent with our observation that the GM2 binding signal on the microglycan array is much higher compared with the GM3 signal (Figure 2A).

Infection of MEF cells with T1L reovirus is blocked by preincubation with the GM2 glycan

To investigate whether GM2 serves as a functional receptor for T1L reovirus, we tested the soluble GM2 glycan for the capacity to inhibit T1L infection of MEFs. Preincubation of the GM2 glycan with T1L resulted in a dose-dependent decrease in T1L infectivity (Figure 3A). However, preincubation of T1L with the GM3 glycan diminished infectivity to a lesser extent and was not dose-dependent (Figure 3B). As a specificity control, incubation of reovirus T3D with the GM2 glycan did not diminish the capacity of T3D to infect MEFs (Figure 3C). These findings demonstrate that the GM2 glycan is specifically recognized by T1L and serves as a physiologically relevant coreceptor.

Crystal structure of T1L σ1 in complex with the GM2 glycan

To visualize interactions between T1L σ1 and its coreceptor, we determined the crystal structure of the $\sigma1_{long}$ construct in complex with the GM2 glycan. The overall structure of the monomer and the organization of the trimer are similar to the T3D σ1 structure [18]. The crystallized T1L σ1 protein folds into three β-spiral repeats and a globular C-terminal head domain (Figure 4A–C). The head domain, comprising amino acids 327–470, is constructed from two Greek-key motifs, each consisting of four β-strands (β-strands A–D and E-H). β-spiral repeats 1 (amino acids 310–326) and 3 (residues 268 to 287) form proline-type β-turns, with both prolines being in the cis-configuration, again similar to T3D σ1. β-spiral repeat 2 (amino acids 288–305) is initiated by a serine residue (S291). In T3D σ1, threonine 278 occupies an analogous position. Both residues are non-standard, as normally only glycines or prolines are tolerated at this position [18,39].

Although the structure has only intermediate resolution, it has good refinement statistics (Table 1). The unbiased electron density map shown in Figure 4 was determined prior to inclusion of the glycan in the refinement and therefore does not contain any information about GM2. The map has interpretable electron density for all four sugar moieties of GM2, including the unique features of Neu5Ac, in all three T1L σ1 monomers. The three copies of the glycan are crystallographically independent but nevertheless make nearly identical contacts with their respective binding pockets, providing additional support for the validity of the observed interactions. The GM2 glycan binds to the upper region of the T1L σ1 head and thus not near the β-spiral region as predicted earlier [18]. A schematic representation of the σ1 domain organization is shown in Figure 4D, including the localization of the respective binding sites for carbohydrate and JAM-A in T1L and T3D σ1.

The Neu5Ac residue contributes the majority of the contacts between GM2 and T1L σ1 and is wedged into a shallow groove bordered on each side by β-strands B and C. Additional contacts involve the GalNAc moiety. The lactose component, which forms the backbone of the branched glycan and would be linked to the ceramide anchor in the GM2 ganglioside, points away from the protein. The mobilities of the sugar moieties are reflected in their thermal factors (B-factors). The average B-factors of Neu5Ac and GalNAc are in the same range as those of the neighboring protein residues, indicating nearly complete occupancy of the glycan-binding pockets (Table 1). The remaining two sugars, and especially the glucose moiety, have elevated B-factors, in agreement with their lack of contacts to protein residues and resultant higher mobility (Table 1).

The Neu5Ac residue can be unambiguously placed in the electron density map due to unique identifying features of this sugar compound (Figure 5A, B). The *N*-acetyl and glycerol chains of Neu5Ac insert between β-strands B and C, where they form hydrogen bonds with backbone atoms of both β-strands (Figure 5A, B). Additionally, the methyl group of the Neu5Ac *N*-acetyl chain inserts into a hydrophobic pocket flanked by V354, F369, and M372, consistent with the dominance of this group in the STD NMR spectrum. The side chain of Q371 likely forms a hydrogen bond with the Neu5Ac carboxylate. However, at 3.6 Å resolution, the conformations of protein side chains cannot be unambiguously determined.

There are two possible orientations for the GalNAc group as a result of the electron density. For our crystallographic model, we selected the sugar conformation that is favored according to the

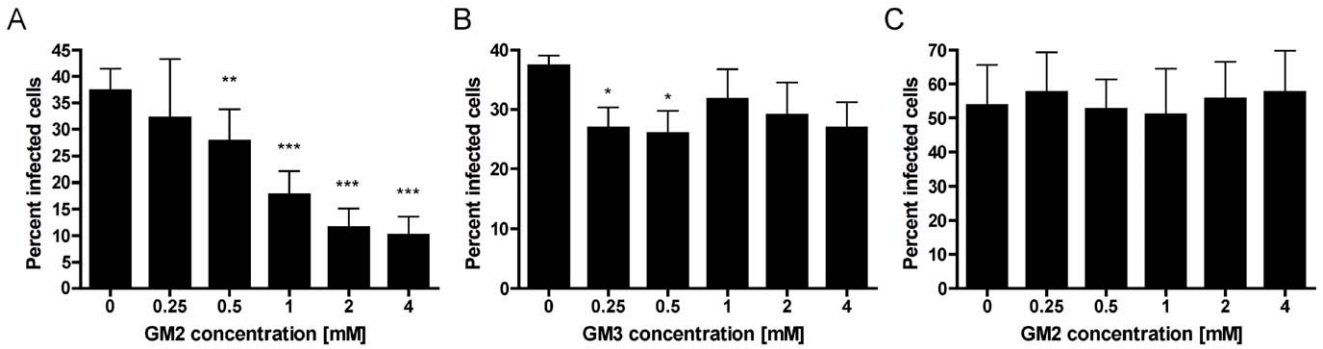

Figure 3. The effect of soluble glycans on T1L infectivity of MEFs. (**A,B**) T1L or (**C**) T3D (10^7 PFU/well) were pre-incubated with the GM2 (**A,C**) or GM3 (**B**) glycan at the concentrations shown for 1 h prior to adsorption to MEFs at a final MOI of 100 PFU/cell. Cells were washed twice with PBS, and fresh medium was added. After incubation at 37°C for 20 h, cells were fixed and reovirus antigen was detected by indirect immunofluorescence. Nuclei were quantified by DAPI staining. The results are expressed as the mean percent infected cells per field in triplicate wells for two independent experiments. Error bars represent standard deviations. *, $P<0.05$; **, $P<0.01$; ***, $P<0.0001$, as determined by two-tailed Student's *t* test.
doi:10.1371/journal.ppat.1003078.g003

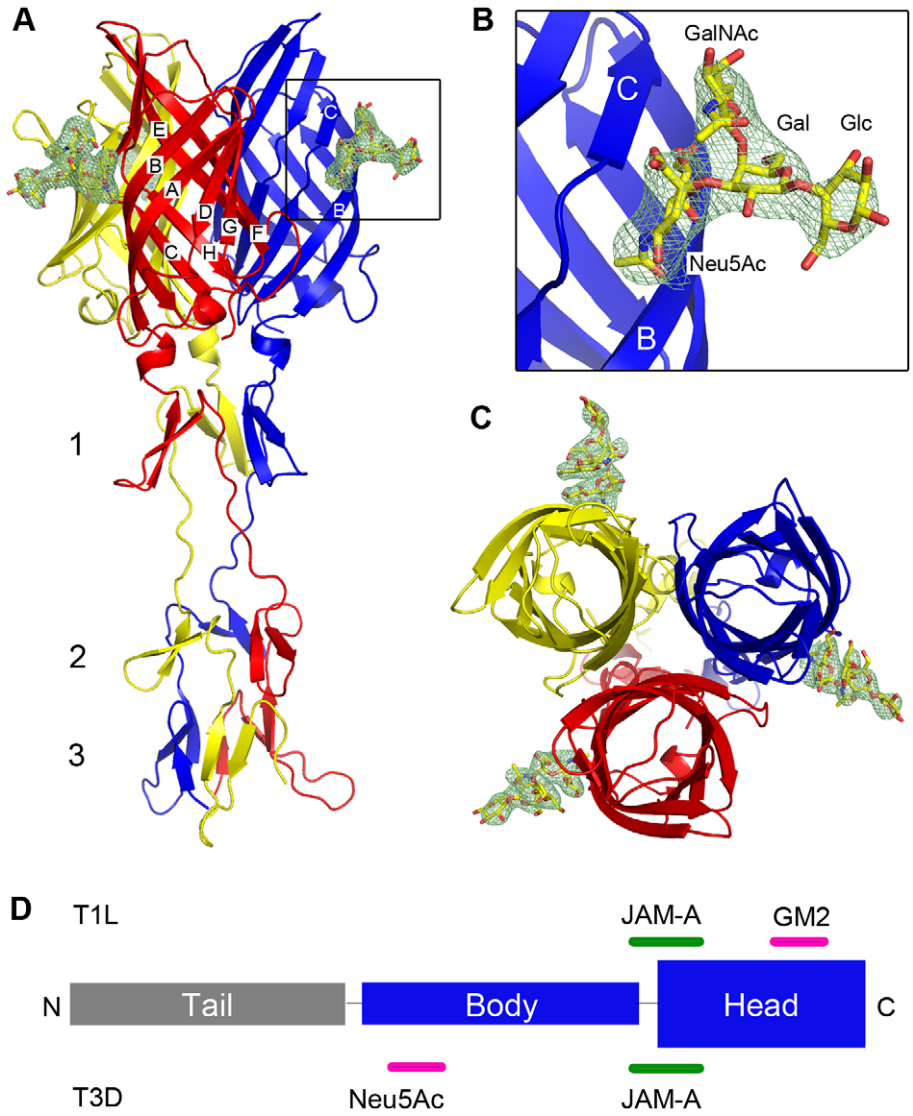

Figure 4. Crystal structure of T1L σ1 in complex with the GM2 glycan. Ribbon tracing of the complex viewed from the side (**A**) with a close-up of the carbohydrate-binding site (**B**) and top-view of the complex (**C**). The three T1L σ1 monomers are depicted in blue, red, and yellow. β-spiral repeats 1, 2, and 3 and β-strands A-H are labeled. The GM2 oligosaccharide is shown in stick representation, with carbons, oxygens, and nitrogens colored yellow, red, and blue, respectively. An unbiased F_o-F_c map of the carbohydrate is shown at a contour level of 3 σ for 2.0 Å around the GM2 glycan (see Materials and Methods section). (**D**) Schematic representation of the σ1 domain organization. Binding sites in T1L and T3D σ1 for JAM-A and carbohydrate are depicted in green and pink, respectively.
doi:10.1371/journal.ppat.1003078.g004

corresponding *Car*bohydrate *R*amachandran *p*lot (CaRp) (Figure S3, Table S3) [40]. This orientation of GalNAc also is preferred by GM2 in solution as assessed by NMR spectroscopy [41]. The GalNAc moiety does not form any hydrogen bonds with T1L σ1, but it clearly interacts with the protein through van der Waals contacts (Figure 5A). Similar contacts are made for each of the two possible orientations of the GalNAc ring.

Crystal structure of T1L σ1 in complex with the GM3 glycan

The GM3 glycan differs from the GM2 oligosaccharide in lacking the GalNAc moiety (Figure 2B). Although GM3 exhibited only very weak binding to T1L σ1 in the glycan arrays (Figure 2A), the structure of T1L σ1 in complex with the GM2 glycan indicated that GM3 contains most of the essential features for complex formation and could potentially engage T1L σ1, albeit with lower affinity compared to GM2. We therefore determined a crystal structure of T1L σ1 in complex with the GM3 glycan at 3.5 Å resolution (Table 2). The structure shows that T1L σ1 binds to the GM3 glycan at the same site as the GM2 glycan, using identical contacts for the Neu5Ac group (Figure 6). The Neu5Ac residues of the T1L σ1-GM3 and T1L σ1-GM2 complex structures superimpose with an r.m.s.d. value of 0.76 Å (Figure S4). As is the case for the T1L σ1-GM2 complex, the lactose moiety of the GM3 glycan points away from the protein.

Residues in T1L reovirus required for carbohydrate engagement

To identify residues in T1L σ1 required for glycan binding, we generated T1L reoviruses carrying point mutations in the GM2-

Table 1. Data collection and refinement statistics for the T1L σ1-GM2 complex.

Data collection	
Resolution (Å)	50-3.60 (3.69-3.60)
Space group	$P3_221$
a, c (Å)	147.5, 164.5
α, β, γ (°)	90, 90, 120
R_{meas} (%)	11.5 (61.9)
$CC_{1/2}$ (%)*	99.8 (88.1)
λ(Å)	1.0
I/σ(I)	17.2 (3.1)
Completeness (%)	99.9 (99.8)
Total reflections	151484 (11137)
Unique reflections	24422 (1782)
Redundancy	6.2
Refinement	
R_{work}/R_{free} (%)**	18.5/20.4
B-factors	
Chain A (Å^2)	86.6
Chain B (Å^2)	86.9
Chain C (Å^2)	98.7
GM2-A (complete) (Å^2)	99.4
Neu5Ac/GalNAc-A (Å^2)	85.9/89.0
GM2-B (complete) (Å^2)	101.2
Neu5Ac/GalNAc-B (Å^2)	87.9/94.3
GM2-C (complete) (Å^2)	111.4
Neu5Ac/GalNAc-C (Å^2)	95.2/108.3
Number of atoms	
Protein	4776
GM2 glycan	171
r.m.s.d.	
Bond lengths (Å)	0.01
Bond angles (°)	1.11
Ramachandran Plot	
Favored (%)	593 (97.5)
Allowed (%)	15 (2.5)
Outliers (%)	0

r.m.s.d. = root-mean-square deviation.
*$CC_{1/2}$ = correlation coefficient ([90]).
**R_{free} was calculated with 10% of the data.
doi:10.1371/journal.ppat.1003078.t001

binding site using plasmid-based reverse genetics [42]. Residues V354, S370, Q371, and M372 were chosen for mutational analysis, as inspection of the T1L σ1-GM2 complex structure showed that each of these residues is in close proximity to the bound glycan (Figure 5B). For point mutants V354F, V354L, and M372L, the amino acids present in T1L σ1 were replaced with residues predicted to partially block the putative Neu5Ac-binding pocket. Residue Q371 was replaced with an acidic residue to introduce a negative charge that was expected to repel the Neu5Ac moiety and interfere with binding to the GM2 glycan (Figure 5B). Point mutants S370P, Q371A, and M372F were generated to replace a T1L σ1 residue with the corresponding residue in T3D σ1, which does not bind a carbohydrate receptor via its head domain [19] (Figure 5C). The S1 genes of all mutant viruses were sequenced to confirm the fidelity of mutagenesis.

We thought it possible that mutations within the putative carbohydrate-binding site might result in diminished infectivity in MEFs due to impaired glycan engagement or some other impairment in viral fitness. To eliminate the latter possibility and normalize infectious units for the virus strains tested, we used L cells, which do not require sialylated glycan engagement to support infection, likely due to an abundance of JAM-A on the cell surface. Unlike our findings with MEFs, neither neuraminidase treatment of cells (Figure 1) nor pretreatment of virus with GM2 (data not shown) altered T1L infectivity in L cells. To determine whether the mutant σ1 proteins are properly folded, we tested the conformation-sensitive monoclonal antibody (mAb) 5C6 for the capacity to inhibit mutant virus infection of L cells. Neutralization-resistant T1L mutants selected by mAb 5C6 have alterations at Q417 and G447 in T1L σ1 [43]. These residues are located at the upper part of the T1L σ1 head domain, close to the intersubunit interface (Figure 7A). An antibody that recognizes these residues likely binds a trimeric conformer of the T1L σ1 head and thus indicates the presence of properly folded and assembled σ1 trimers. Preincubation with mAb 5C6 significantly diminished the capacity of wildtype and mutant T1L viruses to infect L cells (Figure 7B), suggesting that the σ1 head domain of the mutants is recognized by mAb 5C6 and not grossly misfolded.

To test whether the σ1 point mutants have impaired glycan binding, we quantified the capacity of wildtype and mutant viruses to agglutinate human erythrocytes (Figure 8), a property linked to carbohydrate binding [28]. All of the mutants had a significant defect in hemagglutination, with alterations of V354, S370, and Q371 showing the greatest impairment. To determine whether the point mutants have an altered capacity to infect cells in a carbohydrate-dependent fashion, we quantified infectivity in MEFs, which require carbohydrate binding for optimal infection (Figure 1). MEFs were inoculated with wildtype and mutant viruses at an MOI of 1 FFU/cell for each virus as equilibrated in assays using L cells. The V354F, S370P, Q371A, and Q371E mutants displayed a significant defect in infectivity in MEFs (Figure 9). Taken together, these data suggest that residues V354, S370, and Q371, which flank the carbohydrate-binding site of T1L σ1, are required for functional engagement of the GM2 glycan.

Discussion

Although all known reovirus serotypes utilize JAM-A as a receptor, they display striking differences in viral tropism and spread. These differences segregate with the S1 gene, which encodes the σ1 attachment protein [7]. The σ1 residues that interact with JAM-A are conserved among the serotypes [25], and serotype-dependent tropism in the CNS is observed in JAM-A-null mice [11]. These observations suggest that serotype-dependent differences in host disease are attributable to σ1 engagement of cell-surface receptors other than JAM-A.

T3D σ1 binds to sialic acid using residues in its body domain, interacting with α2,3, α2,6, and α2,8-linked sialic acid in a similar manner [19,27]. Although hemagglutination data [28] and lectin-based studies [30] demonstrate that T1L interacts with α2,3-linked sialic acid, neither the identity of the specific glycan nor the molecular basis of T1L-glycan interactions was known. In this study, we found that T1L uses the GM2 glycan as a functional receptor, which is the first identification of a specific glycan recognized by any reovirus serotype.

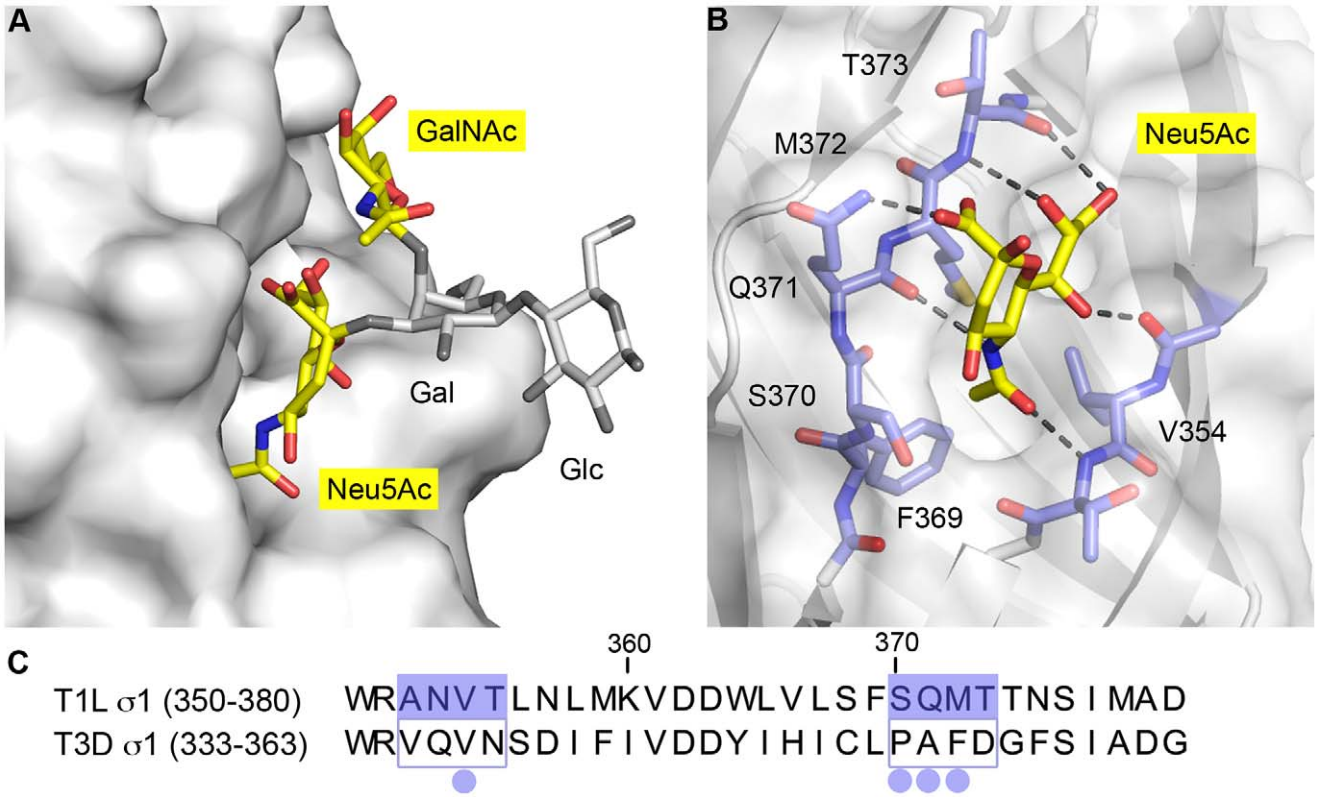

Figure 5. The carbohydrate-binding site of T1L σ1. (**A**) Surface representation of T1L σ1 shown in light gray. The GM2 glycan is depicted in stick representation with the two terminal sugars, Neu5Ac and GalNAc, that contact T1L σ1 shown in color, and the Gal and Glc residues shown in gray. (**B**) Close-up view of the Neu5Ac-binding pocket, with contacting residues shown in stick representation in blue (carbons) and the protein surface shown in light gray. Neu5Ac is depicted in stick representation and colored as in Figure 4. Hydrogen bonds between T1L σ1 and Neu5Ac are represented with black dashes. The methyl group of the *N*-acetyl chain of Neu5Ac inserts into a hydrophobic pocket formed by residues V354, F369, and M372. (**C**) Sequence alignment of the carbohydrate-binding site of T1L σ1 (amino acids 350–380) with the corresponding region of T3D σ1 (residues 333–363). The two β-strands forming the carbohydrate-binding site of T1L σ1 are highlighted in blue. The four residues included in the mutational analyses are marked with blue dots.
doi:10.1371/journal.ppat.1003078.g005

Hemagglutination assays have been used in many previous studies of reovirus-glycan interactions [27–29]. Reovirus displays serotype-dependent hemagglutination profiles. Type 1 reoviruses agglutinate human but not bovine erythrocytes, whereas type 3 reoviruses preferentially agglutinate bovine erythrocytes and agglutinate human erythrocytes less efficiently [29]. These observations suggest that the glycan-binding sites of type 1 and type 3 reovirus are distinct, a hypothesis that is now confirmed by this study and that of Reiter, et al [19]. Analysis of the respective crystal structures sheds light on the potential species differences in hemagglutination behavior. Whereas human erythrocytes express the Neu5Ac form of sialic acid [44], bovine cells express mostly Neu5Gc and less Neu5Ac [45]. The additional hydroxyl group of Neu5Gc would face a hydrophobic pocket in the type 1 σ1 glycan-binding site, making a favorable interaction unlikely. In contrast, the type 3 σ1 binding site likely could accommodate either Neu5Ac or Neu5Gc (D.M. Reiter and T. Stehle, unpublished data).

The GM2 glycan binds to the head domain of T1L σ1 and not, as predicted earlier, to the body region of the protein [27]. It is possible that cell-surface structures in addition to glycans contribute to hemagglutination by type 1 reovirus and this may explain why the chimeric σ1 proteins used in the earlier study had diminished, but not abolished, hemagglutination capacity. Alternatively, disruption of the neck domain of σ1 in the chimeric proteins used in the previous study [27] might have altered the conformation of the glycan-binding domain in the head.

Inspection of the carbohydrate-binding site reveals that the two terminal sugar moieties of the branched GM2 glycan, Neu5Ac and GalNAc, contact the protein, explaining the observed specificity of T1L σ1 for this receptor. Most of the contacts are contributed by Neu5Ac, which is wedged into a cleft between β-strands B and C at the side of the σ1 head, while the GalNAc docks onto a shallow protein surface using van der Waals interactions.

Although the GM3 oligosaccharide is also able to bind T1L σ1 in solution, infectivity studies indicate that GM2 is the preferred glycan receptor for T1L reovirus. While preincubation with either GM2 or GM3 oligosaccharides resulted in diminished infectivity of MEFs, the GM2 glycan blocked infectivity more efficiently and in a dose-dependent fashion. The "extra" GalNAc moiety of GM2 is likely responsible for the selectivity of T1L σ1 for this glycan. At only 41 $Å^2$, the surface area in T1L σ1 buried by interactions with GalNAc is very small compared to the 284 $Å^2$ surface buried by contacts with Neu5Ac in the same complex (Table S4), but the small additional interactions are nevertheless expected to mediate higher-affinity binding of the GM2 glycan compared with GM3, which lacks GalNAc. In addition, due to its branched structure, the GM2 glycan has less conformational freedom in solution than the linear GM3 molecule [41], which may also facilitate interactions with the virus. Entropy furthermore favors binding

Table 2. Data collection and refinement statistics for the T1L σ1-GM3 complex.

Data collection	
Resolution (Å)	50-3.50 (3.59-3.50)
Space group	$P3_221$
a, c (Å)	149.4, 165.2
α, β, γ (°)	90, 90, 120
R_{meas} (%)	12.7 (64.3)
$CC_{1/2}$ (%)*	99.6 (86.8)
λ(Å)	1.0
I/σ(I)	11.77 (3.34)
Completeness (%)	97.7 (98.7)
Total reflections	107527 (8217)
Unique reflections	26751 (1984)
Redundancy	4.0
Refinement	
R_{work}/R_{free} (%)**	18.6/19.7
B-factors	
Chain A (Å^2)	81.4
Chain B (Å^2)	83.4
Chain C (Å^2)	90.5
GM3-A (complete) (Å^2)	104.1
Neu5Ac-A (Å^2)	91.4
GM3-B (complete) (Å^2)	101.4
Neu5Ac-B (Å^2)	81.3
GM3-C (complete) (Å^2)	103.0
Neu5Ac-C (Å^2)	91.4
Number of atoms	
Protein	4794
GM3	118
r.m.s.d.	
Bond lengths (Å)	1.09
Bond angles (°)	0.01
Ramachandran plot	
Favored (%)	602 (98.9)
Allowed (%)	6 (1.0)
Outliers (%)	1 (0.2)

r.m.s.d. = root-mean-square deviation.
*$CC_{1/2}$ = correlation coefficient ([90]).
**R_{free} was calculated with 10% of the data.
doi:10.1371/journal.ppat.1003078.t002

of the branched GM2 glycan over the linear GM3 molecule. In support of this idea, limited conformational freedom of the branched glycan GM1 is essential for its selective engagement by cholera toxin over related compounds [46]. Therefore, the branched sequence of the GM2 glycan sequence is preferred over the linear sequence of GM3.

Interactions between T1L σ1 and GM2 are primarily comprised of hydrogen bonds between the sugar molecule and backbone atoms of the protein. Nevertheless, we were able to identify residues required for functional glycan engagement by introducing mutations into the glycan-binding site. All mutants displayed impaired hemagglutination capacity, with mutations altering V354, S370, and Q371 having the greatest effect (Figure 8). Mutations affecting these same residues resulted in the greatest defect in infectivity of MEFs (Figure 9). Residue V354 flanks a hydrophobic pocket into which the methyl group of the *N*-acetyl chain of Neu5Ac inserts. Mutation of V354 to phenylalanine impairs infectivity of MEFs, while mutating the residue to leucine had a less dramatic effect. Changing S370 to proline introduces a protruding and rigid ring structure, which is expected to create steric hindrance within the glycan-binding pocket (Figure 5). Q371 likely forms a hydrogen bond with the carboxyl group of Neu5Ac. In the point mutants Q371E and Q371A, this hydrogen bond would be lost, which would lead to reduced ligand binding and, in the case of Q371E, electrostatic repulsion.

Interestingly, for the mutants S370P, and Q371A, the residue in T1L σ1 was changed to the corresponding residue in T3D σ1. Structural data suggest that the T1L glycan-binding pocket does not exist in T3D (Figure 10), which likely explains the serotype-dependent inhibition of infection by GM2 (Figure 3). Collectively, these data suggest that residues V354, S370, and Q371, which flank the carbohydrate-binding site of T1L σ1, are important for recognition and engagement of the GM2 glycan despite the predominant role of main-chain interactions in the crystallographic model.

The GM2-binding site in T1L σ1 is distinct from the site of JAM-A binding, and we think that T1L σ1 can bind both receptors, perhaps in a sequential manner (Figure 11A, B). The N-terminal D1 domain of human JAM-A is not glycosylated [47]. Therefore, the glycan receptor must be an independent entity. Reovirus engagement of host cells is likely a multistep process in which interactions with glycans function in adhesion strengthening [33]. We anticipate that the virus first encounters cell-surface GM2 and binds with relatively low affinity (in line with the NMR data) and then binds JAM-A with high affinity [20,48], followed by integrin-mediated uptake [49]. This model is supported by the finding that glycan binding is required for T1L infection of MEFs, which express modest levels of JAM-A, and dispensable in L cells, which display significantly higher levels of JAM-A expression. Glycan binding also can function independently of JAM-A engagement, as the relatively modest infectivity of JAM-A-null MEFs can be further reduced by neuraminidase treatment (data not shown). Furthermore, it is possible that the glycan functions with unknown receptors in the host or serves as the sole cell-surface molecule used by T1L in some tissues. The function of adhesion-strengthening and the interactions or lack thereof between GM2 and other reovirus receptors is an important topic for future research.

The precise tissue distribution of GM2 is not completely understood, but the glycan is a component of the mammalian nervous system [50–52]. In mice, T1L reovirus infects ependymal cells and causes hydrocephalus [8,53]. The presence of GM2 in the brain provides an attractive explanation for the use of this coreceptor by T1L. Because ganglioside expression may differ in cell types that serve as targets for reovirus infection in vivo, there may be cells in which one glycan or another predominates as a T1L coreceptor. Type 3 reoviruses differing only in the capacity to engage cell-surface glycans display marked differences in tropism [54,55]. We anticipate that glycan binding also functions in the pathogenesis of type 1 reovirus infections, which is an area of current investigation in our laboratories.

Reovirus is being tested in clinical trials as an oncolytic adjunct to conventional cancer therapy. Some tumor cells have altered ganglioside expression compared with untransformed cells, and some overexpress GM2 [56–58]. Humanized antibodies directed against GM2 prevent the formation of organ metastases in mice with small-cell lung cancer [59]. It is possible that ganglioside overexpression in

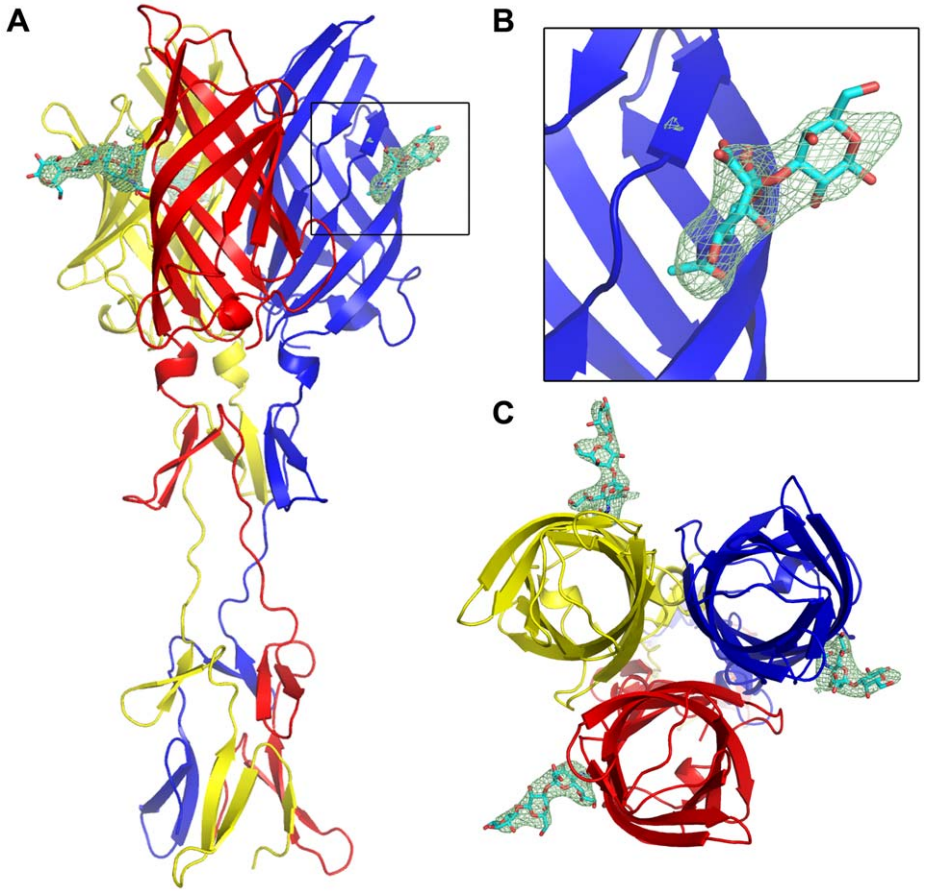

Figure 6. Crystal structure of T1L σ1 in complex with the GM3 glycan. Ribbon tracing of the complex viewed from the side (**A**) with a close-up of the carbohydrate-binding site (**B**) and top-view of the complex (**C**). The three σ1 monomers are depicted in blue, red, and yellow. The ligand is depicted in stick representation in cyan. An unbiased F_o-F_c electron density map is shown at 3.0 σ contour level for 3 Å around the GM3 glycan.
doi:10.1371/journal.ppat.1003078.g006

tumor cells alters the susceptibility of certain cancers to reovirus infection. Understanding the molecular basis of reovirus-glycan interactions might improve the design of effective oncolytics.

Although T1L and T3D reoviruses bind sialylated glycans as receptors using their σ1 proteins, the locations of the respective carbohydrate-binding sites differ substantially (Figure 11A, B). The T1L σ1 glycan-binding site resides in the head domain. In contrast, the T3D σ1 glycan-binding site is in the N-terminal part of the body domain, close to the midpoint of the σ1 molecule. Structure and sequence comparisons show that the head of T3D σ1 would not be capable of engaging Neu5Ac-based receptors because the carbohydrate-binding site of the T1L σ1 head is blocked in T3D σ1 (Figures 5C, 10). It also is unlikely that the region of T1L σ1 corresponding to the T3D σ1 glycan-binding site would interact with sialic acid. T3D σ1 residue Arg202 forms critical interactions with Neu5Ac and, in T1L σ1, there is an aspartate instead of an arginine at the equivalent position. The negatively charged aspartate side chain would probably repel Neu5Ac and, thus, carbohydrate engagement at this site is impeded (Figure 11C). The different locations of the carbohydrate-binding sites contrast with the conserved interactions of both σ1 proteins with JAM-A. The JAM-A-binding sites of both T1L and T3D σ1 proteins are located at the base of the head domain, and interactions between σ1 and JAM-A are similar in both serotypes [25,26]. Assuming that both protein- and carbohydrate-binding sites are accessible for both serotype 1 and serotype 3 reoviruses, it is possible that the mechanisms of attachment are not conserved between the reovirus serotypes, which may contribute to the observed differences in viral tropism and spread.

Materials and Methods

T1L σ1 protein expression and purification

Construct $\sigma1_{long}$ comprises the three most C-terminal predicted β-spirals of T1L σ1 and the head domain (amino acids 261–470). Construct $\sigma1_{short}$ comprises the most C-terminal predicted β-spiral of T1L σ1 and the head domain (amino acids 300–470). Expression and purification of T1L $\sigma1_{long}$ and T1L $\sigma1_{short}$ were facilitated by attaching a trimeric version of the GCN4 leucine zipper [60,61] to the N-terminus of the σ1 sequence, similar to the strategy we used to express T3D σ1 [19]. The σ1 construct was cloned into the pQE-80L expression vector (Qiagen), which includes a non-cleavable N-terminal His_6-tag. The protein was expressed in *E. coli* Rosetta 2 (DE3) (Novagen) by autoinduction at 20°C for 48 to 72 h. Bacteria were lysed using an EmulsiFlex (Avestin) homogenizer and purified via Ni-affinity chromatography (His-Trap FF column, GE Healthcare). The fusion protein was eluted from the column, and the protein solution was desalted using a PD10 desalting column (GE Healthcare). The GCN4 domain and the His_6-tag were removed from the fusion protein using 1 μg trypsin per mg protein at 20°C for 4 h. The resultant products were subjected to size-exclusion chromatography (Superdex 200) to remove the tags, trypsin, and other minor impurities.

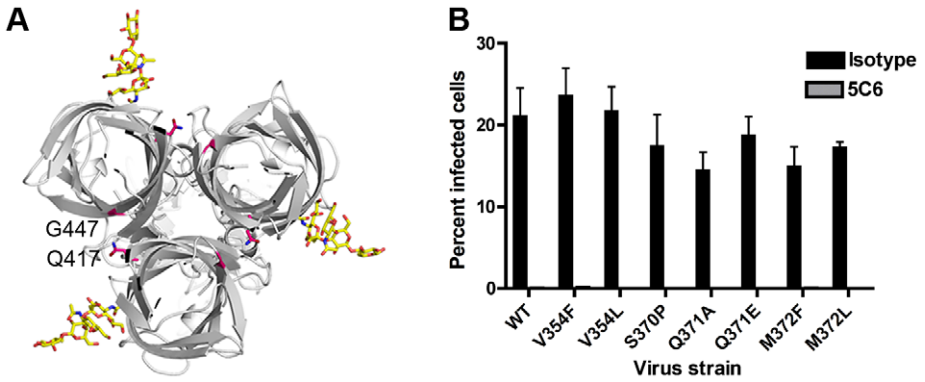

Figure 7. T1L σ1 point-mutant viruses are neutralized by mAb 5C6. (**A**) Top-view of T1L σ1 (gray) in complex with the GM2 glycan (yellow). Residues Q417 and G447, which are altered in mAb 5C6-resistant mutants and likely form part of the 5C6 epitope, are shown in stick representation in pink. (**B**) Wildtype and mutant viruses were incubated with conformation-specific T1L σ1-specific mAb 5C6 for 1 h, and the virus-antibody mixture was adsorbed to L cells for 1 h. Cells were washed twice with PBS, and fresh medium was added. After incubation at 37°C for 20 h, cells were fixed, and reovirus antigen was detected by indirect immunofluorescence. Nuclei were stained with DAPI. The percentage of infected cells in three fields of view per well was determined. The data shown are the mean infectivity per well from three independent experiments each performed in triplicate. Error bars represent S.E.M. $P<0.001$, as determined by two-tailed Student's *t* test for all virus strains.
doi:10.1371/journal.ppat.1003078.g007

Undigested versions of both constructs were used for glycan array screening. STD NMR experiments were performed using $\sigma1_{long}$. Both constructs were used for structural analysis. Uncleaved $\sigma1_{short}$ yielded crystals diffracting to 2.6 Å resolution. This higher resolution structure was used as a reference model for refinement of the lower-resolution structures of cleaved $\sigma1_{long}$ in complex with the GM2 or GM3 glycan.

Glycan microarray analyses

Microarrays were composed of lipid-linked oligosaccharide probes, neoglycolipids (NGLs) and glycolipids, robotically printed on nitrocellulose-coated glass slides at 2 and 7 fmol per spot using a non-contact instrument, and analyses were performed as described [62,63]. For analysis of T1L $\sigma1_{long}$, the results of 124 oligosaccharide probes (5 non-sialylated and 119 sialylated, Glycosciences Array Set 40–41), at 5 fmol per spot are shown in Figure S1 and Table S1. For the analysis of T1L $\sigma1_{short}$, a different version of the microarray (in house designation Ganglioside Dose Response Array set 1) was used; results of the 21 ganglioside-related probes (Table S2) each arrayed at four levels: 0.3, 0.8, 1.7 and 5.0 fmol/spot, are shown in Figure 2A.

For the initial analysis of His-tagged T1L $\sigma1_{long}$, the protein was incubated with mouse monoclonal anti-poly-histidine (Ab1) and biotinylated anti-mouse IgG antibodies (Ab2) (both antibodies

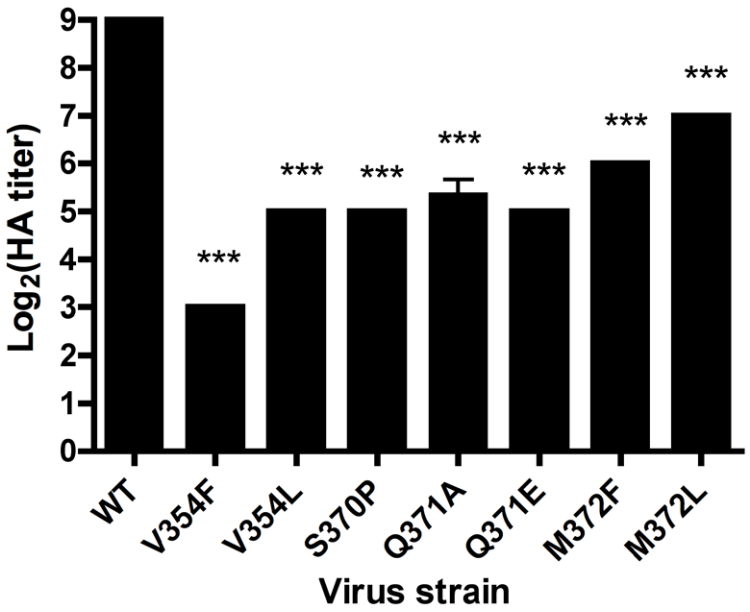

Figure 8. Hemagglutination by σ1 mutant viruses. Purified virions of the strains shown (10^{11} particles/well) were serially diluted 1:2 in PBS in 96-well U-bottom plates. Human erythrocytes were washed several times with PBS, resuspended to a concentration of 1% (vol/vol) in PBS, added to virus-containing wells, and incubated at 4°C for 3 h. Results are expressed as $\log_2$ (HA titer). HA titer is defined as 10^{11} particles divided by the number of particles/HA unit. One HA unit is the particle number sufficient to produce hemagglutination. *** $P<0.001$, as determined by one-way Anova followed by Bonferroni's multiple comparison test.
doi:10.1371/journal.ppat.1003078.g008

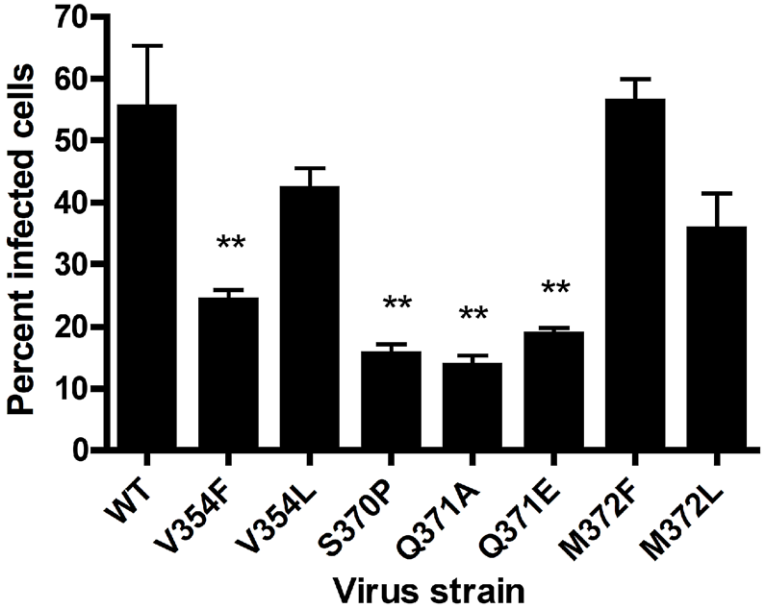

Figure 9. Infectivity of σ1 mutant viruses in MEFs. Monolayers of MEFs were adsorbed with the strains shown at an MOI of 1 FFU/field (as titered in L cells) at room temperature for 1 h. Cells were washed twice with PBS, and fresh medium was added. After incubation at 37°C for 20 h, cells were fixed, and reovirus antigen was detected by indirect immunofluorescence. Nuclei were stained with DAPI. The percentage of infected cells in three fields of view per well was determined. The results are from a representative experiment of three experiments performed with triplicate wells. Error bars represent standard deviations. **, $P<0.01$, as determined by two-tailed Student's *t* test.
doi:10.1371/journal.ppat.1003078.g009

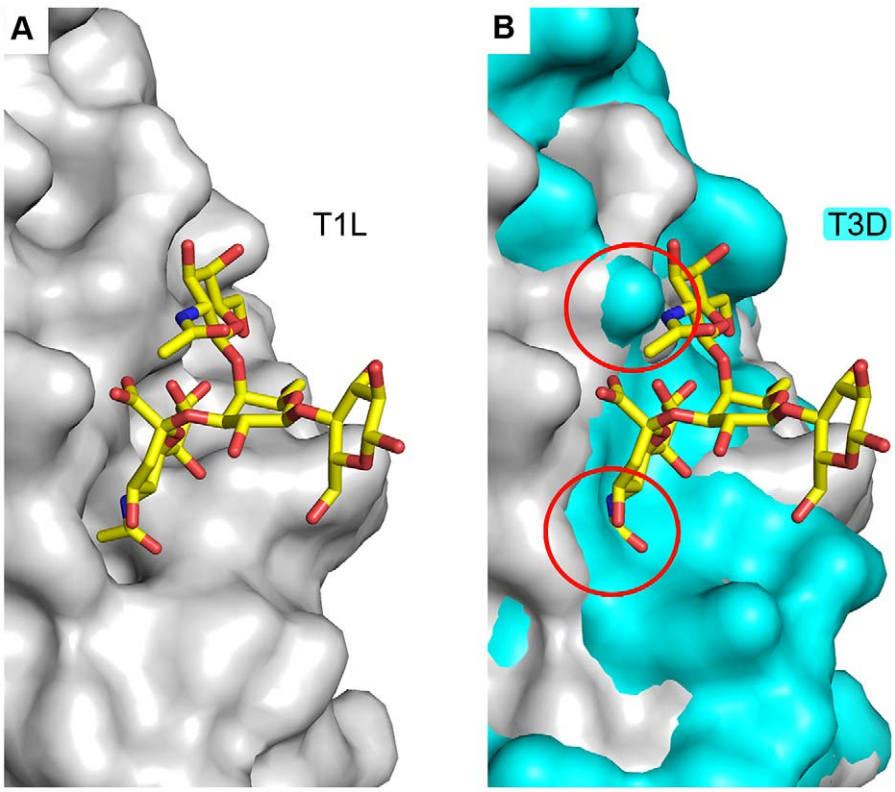

Figure 10. The head domain of T3D σ1 does not bind Neu5Ac. (**A**) Surface representation of T1L σ1 depicted in gray. (**B**) SSM superposition of T1L (gray) and T3D (cyan) σ1. The GM2 glycan is shown in stick representation (colors as in Figure 4) in both panels. Clashes between the carbohydrate and T3D σ1 are highlighted with red circles in panel B. Both the Neu5Ac and GalNAc moieties of the GM2 oligosaccharide would clash with T3D σ1 residues.
doi:10.1371/journal.ppat.1003078.g010

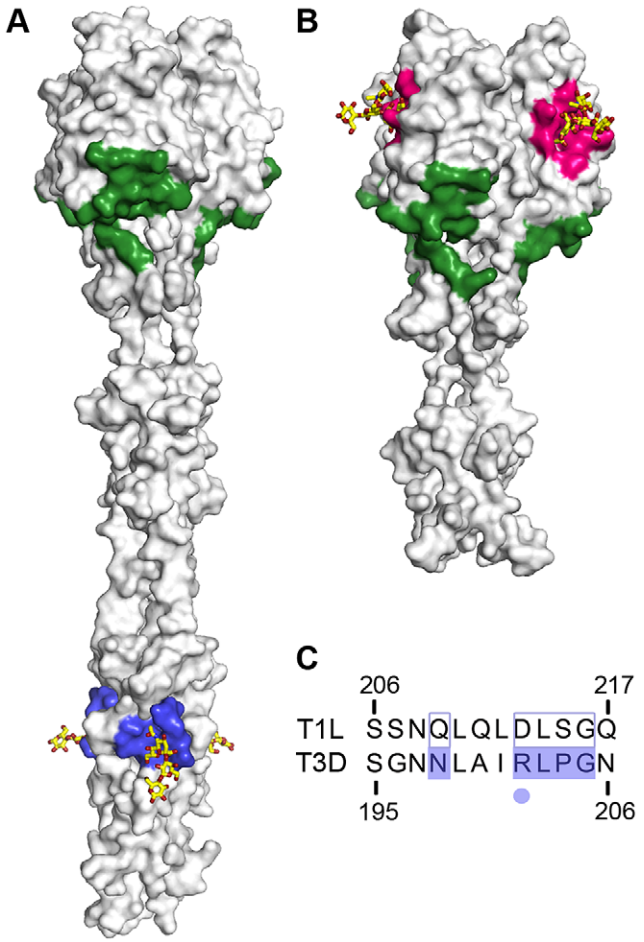

Figure 11. Comparison of the receptor-binding sites of T1L and T3D σ1. Surface representations of (**A**) T3D σ1 in complex with the GM3 glycan (PDB accession code 3S6X) and (**B**) T1L σ1 in complex with the GM2 glycan. The carbohydrates are shown in stick representation and colored as in Figure 4. The JAM-A-binding sites are highlighted in green, and the carbohydrate-binding sites in T1L and T3D σ1 are depicted in pink and blue, respectively. (**C**) Sequence alignment of the carbohydrate-binding site in T1L and T3D σ1. Residues required for carbohydrate engagement in T3D σ1 are highlighted in blue. Residue R202, which forms a central interaction with Neu5Ac in T3D σ1, is marked with a blue dot.
doi:10.1371/journal.ppat.1003078.g011

from Sigma) at a ratio of 4:2:1 (by weight). The $\sigma1_{long}$-antibody complexes were prepared by preincubating Ab1 with Ab2 at ambient temperature for 15 min, followed by addition of His-tagged T1L $\sigma1_{long}$ and incubation on ice for 15 min. The $\sigma1_{long}$-antibody complexes were diluted in 5 mM HEPES (pH 7.4), 150 mM NaCl, 0.3% (v/v) Blocker Casein (Pierce), 0.3% (w/v) bovine serum albumin (Sigma), 5 mM $CaCl_2$ and 40 mM imidazole (referred to as HBS-Casein/BSA-imidazole), to provide a final $\sigma1_{long}$ concentration of 150 μg/ml, and overlaid onto the arrays at 20 °C for 2 h. Binding was detected using Alexa Fluor 647-labeled streptavidin (Molecular Probes) at 1 μg/ml. Microarray data analyses and presentation were facilitated using dedicated software [64].

For the analyses of His-tagged T1L $\sigma1_{short}$, different assay conditions were evaluated with and without complexation (not shown). The condition selected as optimal was without precomplexation. His-tagged $\sigma1_{short}$ was diluted in HBS-Casein/BSA-imidazole, overlaid at 300 μg/ml, followed by incubation with Ab1 and Ab2 (each at 10 mg/ml, precomplexed at ambient temperature for 15 min). Binding was detected using Alexa Fluor 647-labeled streptavidin.

Crystallization, x-ray structure determination, and refinement

Crystals of uncleaved $\sigma1_{short}$ formed in 0.1 M MES/imidazole (pH 6.5), 10% PEG 4000, 20% glycerol, 0.02 M sodium formate, 0.02 M ammonium acetate, 0.02 M trisodium citrate, 0.02 M sodium potassium L-tartrate, 0.02 M sodium oxamate at 4°C using the sitting-drop-vapor-diffusion method. No additional cryoprotection was necessary. Crystals of $\sigma1_{long}$ formed in 0.1 M Na cacodylate (pH 6.0–6.6), 1.2–1.5 M $(NH_4)_2SO_4$ at 4°C using the sitting-drop-vapor-diffusion method. For preparation of complexes, these crystals were transferred to 20 mM GM2 or GM3 oligosaccharide (Elicityl) in the crystallization solution for 5–10 min. Prior to flash-freezing, the crystals were transferred to a solution containing 0.1 M Na cacodylate, 1.34 M $(NH_4)_2SO_4$, 25% glycerol, and 20 mM GM2 or GM3 glycan.

The crystals belonged to space group $P3_221$ and contained one trimer in the asymmetric unit. A complete data set was collected at the Swiss Light Source, beamline X06SA. XDS was used to index and scale the reflection data [65]. The structure was determined by molecular replacement with Phaser (CCP4) [66],[67] using the coordinates of T1L σ1 derived from the previously determined T1L σ1-JAM-A complex structure as a search model [26]. Manual model building was carried out using coot [68]. Structural refinement was performed using Refmac5 (CCP4) [69], Phenix [70], and autoBUSTER [71],[72].

Inspection of the $2F_o$-F_c maps for the structures of the T1L σ1-glycan complexes revealed clear, unambiguous electron density for most of the GM2 and GM3 oligosaccharides at a 1.5 σ contour level. The glycans also were visible in difference electron density maps. The unbiased electron density maps in Figures 4, 6, and S3 show the initial F_o-F_c maps of the T1L σ1-GM2 and T1L σ1-GM3 glycan complexes obtained after molecular replacement using the previously solved structure of unliganded T1L σ1. The carbohydrates were included in the model at this point. Refinement of the ligands was performed using the CCP4 library and user-defined constraints. Structure images were created using PyMOL [73]. Coordinates and structure factors of both complexes have been deposited in the Protein Data Bank with accession codes 4GU3 (T1L-σ1-GM2 glycan complex) and 4GU4 (T1L σ1-GM3 glycan complex).

Sequence and structural analysis

Sequence alignments were performed using T-Coffee [74] and analyzed using Jalview [75,76]. Structure alignments were calculated by secondary-structure matching (SSM) superposition in coot [77]. The Ramachandran plot was generated with Rampage (CCP4) [78]. Buried surface areas were calculated using AreaImol (CCP4) [79,80].

STD NMR spectroscopy

NMR spectra were recorded using 3 mm tubes and a Bruker AVIII-600 spectrometer equipped with a room temperature probe head at 283 K and processed with TOPSPIN 3.0 (Bruker). Samples containing 1 mM GM2 or GM3 glycan (Elicityl), 20 mM potassium phosphate (pH 7.4), and 150 mM NaCl with and without 20 μM T1L σ1 were used for the STD NMR measurements and the frequency control, respectively. Samples were prepared in D_2O, and no additional water suppression was used to preserve the anomeric proton signals. The sample without

protein also was used for spectral assignment. The off- and on-resonance irradiation frequencies were set to −30 ppm and 7.3 ppm, respectively. The irradiation power of the selective pulses was 57 Hz, the saturation time was 2 s, and the total relaxation delay was 3 s. A 50 ms continuous-wave spin-lock pulse with a strength of 3.2 kHz was employed to suppress residual protein signals. A total number of 512 scans were recorded. A total of 10,000 points were collected, and spectra were multiplied with a Gaussian window function prior to Fourier transformation. Spectra were referenced using HDO as an internal standard [81].

Cells

Spinner adapted murine L cells were grown in suspension culture in Joklik's minimum essential medium (Lonza) supplemented to contain 5% fetal bovine serum (FBS) (Gibco), 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin (Invitrogen), and 25 ng/ml amphotericin B (Sigma-Aldrich). MEFs were generated from C57/BL6 mice at embryonic day 13.5 as described [82]. MEFs were maintained in Dulbecco's modified Eagle's minimum essential medium (DMEM) (Gibco) supplemented to contain 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin, 1X MEM nonessential amino acids (Sigma-Aldrich), 20 mM HEPES, and 0.1 mM 2-mercaptoethanol (Sigma-Aldrich). Cells at passages 3–6 were used in this study.

Viruses and plasmid-based reovirus rescue

Viruses were generated using plasmid-based reverse genetics [42,83]. BHK-T7 cells (5×10^5) were seeded in 60 mm tissue-culture dishes (Corning) and allowed to incubate at 37°C overnight. OptiMEM (Invitrogen) (0.75 ml) was mixed with 53.25 μl TransIT-LT1 transfection reagent (Mirus) and incubated at RT for 20 min. Plasmid constructs representing cloned gene segments from the T1L genome, pT7S1 T1L, pT7S2 T1L, pT7L3S3 T1L, pT7S4 T1L, pT7M1 T1L, pT7L1M2 T1L, and pT7L2M3 T1L were mixed into the OptiMEM/TransIT-LT solution. Equal amounts of each plasmid were added for a total of 17.75 μg DNA. The plasmid-transfection solution was added to BHK-T7 cells and incubated for 3–5 days. Following two freeze-thaw cycles, recombinant viruses were isolated by plaque purification using L-cell monolayers [84]. Purified virions were generated using second-passage L cell-lysate stocks. Viral particles were Freon-extracted from infected cell lysates and layered onto 1.2 to 1.4 g/cm^3 CsCl gradients and centrifuged at 62,000×g for 18 h. Bands were collected and dialyzed exhaustively in virion-storage buffer as described [12,85]. To generate mutant viruses, resides V354, S370, Q371, and M372 in the S1 gene plasmid were altered by QuickChange (Stratagene) site-directed mutagenesis. S1 gene sequences were confirmed using the OneStep RTPCR kit (Qiagen), gene-specific primers, and viral dsRNA extracted from infected L cells (RNAeasy, Qiagen). Primer sequences for mutagenesis and sequencing are available from the corresponding authors by request. Sanger sequencing was performed using purified PCR products (Gene Hunter and Vanderbilt Sequencing Core). Genotypes were confirmed by electrophoresis of viral particles in 4-to-20% gradient sodium dodecyl sulfate polyacrylamide gels stained with ethidium bromide and visualized by UV illumination [86]. Particle concentrations were determined using the conversion 1 $AU_{260} = 2.1\times10^{12}$ particles [85]. Viral titers were quantified by plaque assay [84] or fluorescent focus assay [33].

Antibodies

Reovirus polyclonal immunoglobulin G (IgG) raised against T1L and T3D was used to stain for reovirus antigen [87]. Alexa-488 conjugated goat anti-rabbit antibody (Invitrogen) was used as a secondary antibody. Monoclonal rat anti-mouse JAM-A (Abcam, clone H202-106) was used to stain for JAM-A expression followed by goat anti-rat secondary antibody conjugated to Alexa-488 (Invitrogen). Conformation-sensitive neutralizing mAb 5C6 specific for T1L [43,88] was used in neutralization assays as described [89].

Infectivity studies

L cells (10^5) or MEFs (5×10^4) were incubated in 24-well plates (Costar) at 37°C overnight. To evaluate the importance of sialic acid engagement in T1L infection, cell monolayers were treated with 100 mU/ml of *A. ureafaciens* neuraminidase diluted in PBS (MP Biomedicals, LLC) or PBS alone (mock) at RT for 1 h prior to virus adsorption at an MOI of 1 PFU/cell in L cells or 100 PFU/cell (as titered in L cells) in MEFs. Following incubation at RT for 1 h, the inoculum was removed, and cells were washed twice with PBS and incubated at 37°C for 20 h. Cells were fixed in methanol and visualized by indirect immunofluorescence [33] with the addition of a DAPI stain to quantify cell nuclei. Cells were blocked in PBS supplemented to contain 5% bovine serum albumin (BSA) (Sigma). Infected cells were detected by staining with reovirus polyclonal antiserum diluted 1:1000 and secondary Alexa-488 goat anti-rabbit Ig 1:1000 (Invitrogen). Nuclei were quantified using DAPI (1:1000). All antibodies were diluted in PBS supplemented to contain 0.5% Triton X-100. Infectivity studies were performed in triplicate wells. Three fields of view per well were quantified using the Axiovert 200 fluorescence microscope (Carl Zeiss).

To determine the effect of soluble glycans on viral infectivity, virus was incubated with various concentrations of GM2 or GM3 glycan (Elicityl) at room temperature for 1 h. The virus-glycan mixture was adsorbed to MEFs (MOI of 100 PFU/cell as titered on L cells) at room temperature for 1 h. The cells were washed twice, and infectivity was determined by immunofluorescence assay.

Flow cytometry

To determine the relative amount of JAM-A on L cells and MEFs, 5×10^5 cells were stained with rat anti-mouse JAM-A at a dilution of 1:200 followed by staining with Alexa-488 labeled goat anti-rat Ig at 1:1000. All staining was done in PBS supplemented to contain 2% FBS. Fluorescence was measured using an LSRII (BD, Vanderbilt University Flow Cytometry Shared Resource). Mean fluorescence intensity of a forward and side scatter gated population was determined using FlowJo software.

Hemagglutination assay

Purified reovirus virions (10^{11} particles) were distributed into 96-well U-bottom microtiter plates (Costar) and serially diluted twofold in 0.05 ml of PBS. Human type O erythrocytes (Vanderbilt Blood Bank) were washed twice with PBS and resuspended at a concentration of 1% (vol/vol). Erythrocytes (0.05 ml) were added to wells containing virus particles and incubated at 4°C for 3 h. A partial or complete shield of erythrocytes on the well bottom was interpreted as a positive HA result; a smooth, round button of erythrocytes was interpreted as a negative result. HA titer is expressed as 10^{11} particles divided by the number of particles/HA unit. One HA unit equals the number of particles sufficient to produce HA.

Statistical analysis

Statistical analysis was performed using Prism (Graphpad). Two-tailed Student's *t* tests were used for all infectivity studies.

The hemagglutination assays were analyzed using a one-way Anova followed by a Bonferroni's correction. *P* values of less than 0.05 were considered to be statistically significant.

Supporting Information

Figure S1 Glycan microarray analyses of T1L-σ1$_{long}$ using a microarray of 124 lipid-linked oligosaccharide probes. Numerical scores of the binding signals are means of duplicate spots at 5 fmol/spot (with error bars). The various types of terminal sialic acid linkage are indicated by the colored panels as defined at the bottom of the figure. Error bars are all relatively large due to the low fluorescent signals. The list of probes and their sequences and binding scores are provided in Table S1. The X indicates an artifact on the slide giving a false signal resulting in a large error bar.
(TIF)

Figure S2 STD NMR spectroscopy of T1L σ1 with GM2 and GM3 oligosaccharide. (**A**) Chemical structure of the GM2 glycan. Protons that receive saturation upon binding to T1L σ1 are color-coded according to the corresponding STD NMR spectrum in Figure 2C. (**B**) T1L σ1 binds to the GM3 glycan in solution. STD NMR experiment of T1L σ1 and the GM3 oligosaccharide. Upper spectrum: ^{1}H spectrum of the GM3 glycan alone; middle: STD spectrum of T1L σ1 and the GM3 glycan; and lower spectrum: STD spectrum of the GM3 oligosaccharide alone to ensure that no direct excitation of the glycan takes place. A schematic drawing of GM3 is provided in the upper left corner.
(TIF)

Figure S3 CaRp analysis of the T1L σ1-GM2 complex. CaRp analysis (Carbohydrate Ramachandran plot, www.glycosciences.de) of the three GM2 oligosaccharide molecules in the T1L σ1-GM2 complex. A schematic of the GM2 oligosaccharide is included with the three glycosidic bonds numbered. The structure of one GM2 glycan molecule and its unbiased F_o-F_c map at 3.0 σ contour level for 2.0 Å are shown at the bottom right.
(TIF)

Figure S4 T1L σ1 binds Neu5Ac of the GM2 glycan and the GM3 glycan at the same site. SSM superposition of the T1L σ1-GM2 complex (yellow) and the T1L σ1-GM3 complex (cyan). The protein chains are shown as ribbon tracings, and the Neu5Ac moieties of the GM2 and GM3 glycan are depicted in stick representation in yellow and cyan, respectively. They superimpose with an r.m.s.d. value of 0.76 Å.
(TIF)

Table S1 Oligosaccharide probes used in the initial glycan microarray analyses, sorted by sialyl linkage and backbone sequence, and the binding signals (means of the fluorescence intensity at ~5 fmol/probe spot) of T1L-σ1$_{long}$.
(DOC)

Table S2 List of probes and sequences included in the ganglioside dose-response array set.
(DOC)

Table S3 Dihedral angles of the glycosidic linkages of the three GM2 oligosaccharides bound to T1L σ1.
(DOC)

Table S4 T1L σ1 surface areas buried by GM2 and GM3 in the σ1-glycan complex structures.
(DOC)

Acknowledgments

We acknowledge the staff at the Swiss Light Source for beam time and technical support. We are grateful to Susanne Bruekner and Tina Wagner (Interfaculty Institute of Biochemistry, University of Tuebingen) for assistance with the expression and purification of T1L σ1 protein. We thank members of the Glycosciences Laboratory for their collaboration in the establishment of the neoglycolipid-based microarray System and Professor Makoto Kiso for the chemically synthesized GSC sialoglycolipids. We thank Dr. Remco Sprangers (Max Planck Institute for Developmental Biology, Tuebingen, Germany) for assistance in recording the NMR data. We acknowledge the staff in the Vanderbilt University Flow Cytometry Shared Resource for technical support.

Author Contributions

Conceived and designed the experiments: KR JES YL TSD TS. Performed the experiments: KR JES YL BSB. Analyzed the data: KR JES YL BSB DMR TF TSD TS. Wrote the paper: KR JES YL BSB TF TSD TS.

References

1. Haywood AM (1994) Virus receptors: binding, adhesion strengthening, and changes in viral structure. J Virol 68: 1–5.
2. Dermody TS, Parker J, Sherry B (2012) Orthoreovirus. In: D.M. K, P.M. H, editors. Fields Virology. 6th edition. Philadelphia: Lippincott Williams & Wilkins.
3. Tai JH, Williams JV, Edwards KM, Wright PF, Crowe JE, Jr., et al. (2005) Prevalence of reovirus-specific antibodies in young children in Nashville, Tennessee. J Infect Dis 191: 1221–1224.
4. Coffey MC, Strong JE, Forsyth PA, Lee PW (1998) Reovirus therapy of tumors with activated Ras pathway. Science 282: 1332–1334.
5. Strong JE, Coffey MC, Tang D, Sabinin P, Lee PW (1998) The molecular basis of viral oncolysis: usurpation of the Ras signaling pathway by reovirus. EMBO J 17: 3351–3362.
6. Adair RA, Roulstone V, Scott KJ, Morgan R, Nuovo GJ, et al. (2012) Cell carriage, delivery, and selective replication of an oncolytic virus in tumor in patients. Sci Transl Med 4: 138ra177.
7. Tyler KL, McPhee DA, Fields BN (1986) Distinct pathways of viral spread in the host determined by reovirus S1 gene segment. Science 233: 770–774.
8. Weiner HL, Drayna D, Averill DR, Jr., Fields BN (1977) Molecular basis of reovirus virulence: role of the S1 gene. Proc Natl Acad Sci U S A 74: 5744–5748.
9. Weiner HL, Powers ML, Fields BN (1980) Absolute linkage of virulence and central nervous system cell tropism of reoviruses to viral hemagglutinin. J Infect Dis 141: 609–616.
10. Morrison LA, Sidman RL, Fields BN (1991) Direct spread of reovirus from the intestinal lumen to the central nervous system through vagal autonomic nerve fibers. Proc Natl Acad Sci U S A 88: 3852–3856.
11. Antar AA, Konopka JL, Campbell JA, Henry RA, Perdigoto AL, et al. (2009) Junctional adhesion molecule-A is required for hematogenous dissemination of reovirus. Cell Host Microbe 5: 59–71.
12. Boehme KW, Frierson JM, Konopka JL, Kobayashi T, Dermody TS (2011) The reovirus sigma1s protein is a determinant of hematogenous but not neural virus dissemination in mice. J Virol 85: 11781–11790.
13. Lee PW, Hayes EC, Joklik WK (1981) Protein sigma 1 is the reovirus cell attachment protein. Virology 108: 156–163.
14. Furlong DB, Nibert ML, Fields BN (1988) Sigma 1 protein of mammalian reoviruses extends from the surfaces of viral particles. J Virol 62: 246–256.
15. Nibert ML, Dermody TS, Fields BN (1990) Structure of the reovirus cell-attachment protein: a model for the domain organization of sigma 1. J Virol 64: 2976–2989.
16. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252: 1162–1164.
17. Duncan R, Horne D, Cashdollar LW, Joklik WK, Lee PW (1990) Identification of conserved domains in the cell attachment proteins of the three serotypes of reovirus. Virology 174: 399–409.
18. Chappell JD, Prota AE, Dermody TS, Stehle T (2002) Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. EMBO J 21: 1–11.
19. Reiter DM, Frierson JM, Halvorson EE, Kobayashi T, Dermody TS, et al. (2011) Crystal structure of reovirus attachment protein sigma1 in complex with sialylated oligosaccharides. PLoS Pathog 7: e1002166.
20. Barton ES, Forrest JC, Connolly JL, Chappell JD, Liu Y, et al. (2001) Junction adhesion molecule is a receptor for reovirus. Cell 104: 441–451.

21. Campbell JA, Schelling P, Wetzel JD, Johnson EM, Forrest JC, et al. (2005) Junctional adhesion molecule a serves as a receptor for prototype and field-isolate strains of mammalian reovirus. J Virol 79: 7967–7978.
22. Prota AE, Campbell JA, Schelling P, Forrest JC, Watson MJ, et al. (2003) Crystal structure of human junctional adhesion molecule 1: implications for reovirus binding. Proc Natl Acad Sci U S A 100: 5366–5371.
23. Liu Y, Nusrat A, Schnell FJ, Reaves TA, Walsh S, et al. (2000) Human junction adhesion molecule regulates tight junction resealing in epithelia. J Cell Sci 113 (Pt 13): 2363–2374.
24. Schelling P, Guglielmi KM, Kirchner E, Paetzold B, Dermody TS, et al. (2007) The reovirus sigma1 aspartic acid sandwich: a trimerization motif poised for conformational change. J Biol Chem 282: 11582–11589.
25. Kirchner E, Guglielmi KM, Strauss HM, Dermody TS, Stehle T (2008) Structure of reovirus sigma1 in complex with its receptor junctional adhesion molecule-A. PLoS Pathog 4: e1000235.
26. Kirchner E (2009) Structural and functional studies of the reovirus attachment protein sigma1 and its interaction with the receptor JAM-A Elektronische Ressource. pp. Online-Resource.
27. Chappell JD, Duong JL, Wright BW, Dermody TS (2000) Identification of carbohydrate-binding domains in the attachment proteins of type 1 and type 3 reoviruses. J Virol 74: 8472–8479.
28. Lerner AM, Cherry JD, Finland M (1963) Hemagglutination with reoviruses. Virology 19: 58–65.
29. Gomatos PJ, Tamm I (1962) Reactive sites of reovirus type 3 and their interaction with receptor substances. Virology 17: 455–461.
30. Helander A, Silvey KJ, Mantis NJ, Hutchings AB, Chandran K, et al. (2003) The viral sigma1 protein and glycoconjugates containing alpha2-3-linked sialic acid are involved in type 1 reovirus adherence to M cell apical surfaces. J Virol 77: 7964–7977.
31. Paul RW, Lee PW (1987) Glycophorin is the reovirus receptor on human erythrocytes. Virology 159: 94–101.
32. Nibert ML, Chappell JD, Dermody TS (1995) Infectious subvirion particles of reovirus type 3 Dearing exhibit a loss in infectivity and contain a cleaved sigma 1 protein. J Virol 69: 5057–5067.
33. Barton ES, Connolly JL, Forrest JC, Chappell JD, Dermody TS (2001) Utilization of sialic acid as a coreceptor enhances reovirus attachment by multistep adhesion strengthening. J Biol Chem 276: 2200–2211.
34. Shevchuk NA, Hathout Y, Epifano O, Su Y, Liu Y, et al. (2007) Alteration of ganglioside synthesis by GM3 synthase knockout in murine embryonic fibroblasts. Biochim Biophys Acta 1771: 1226–1234.
35. Yogeeswaran G, Fujinami R, Kiessling R, Welsh RM (1982) Interferon-induced alterations in sialic acid and glycoconjugates of L-929 cells. Virology 121: 363–371.
36. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angewandte Chemie-International Edition 38: 1784–1788.
37. Meyer B, Peters T (2003) NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angewandte Chemie-International Edition 42: 864–890.
38. Neu U, Hengel H, Blaum BS, Schowalter RM, Macejak D, et al. (2012) Structures of Merkel cell polyomavirus VP1 complexes define a sialic acid binding site required for infection. PLoS Pathog 8: e1002738.
39. van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. Nature 401: 935–938.
40. Lutteke T, Frank M, von der Lieth CW (2005) Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. Nucleic Acids Res 33: D242–246.
41. Levery SB (1991) 1H-NMR study of GM2 ganglioside: evidence that an interresidue amide-carboxyl hydrogen bond contributes to stabilization of a preferred conformation. Glycoconj J 8: 484–492.
42. Kobayashi T, Ooms LS, Ikizler M, Chappell JD, Dermody TS (2010) An improved reverse genetics system for mammalian orthoreoviruses. Virology 398: 194–200.
43. Helander A, Miller CL, Myers KS, Neutra MR, Nibert ML (2004) Protective immunoglobulin A and G antibodies bind to overlapping intersubunit epitopes in the head domain of type 1 reovirus adhesin sigma1. J Virol 78: 10695–10705.
44. Muchmore EA, Diaz S, Varki A (1998) A structural difference between the cell surfaces of humans and the great apes. Am J Phys Anthropol 107: 187–198.
45. Musielak M (2004) Are there two functionally distinguished Neu5Gc pools with respect to rouleau formation on the bovine red blood cell? Clin Hemorheol Microcirc 30: 435–438.
46. Turnbull WB, Precious BL, Homans SW (2004) Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry. J Am Chem Soc 126: 1047–1054.
47. Chen R, Jiang X, Sun D, Han G, Wang F, et al. (2009) Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry. J Proteome Res 8: 651–661.
48. Guglielmi KM, Kirchner E, Holm GH, Stehle T, Dermody TS (2007) Reovirus binding determinants in junctional adhesion molecule-A. J Biol Chem 282: 17930–17940.
49. Maginnis MS, Forrest JC, Kopecky-Bromberg SA, Dickeson SK, Santoro SA, et al. (2006) Beta1 integrin mediates internalization of mammalian reovirus. J Virol 80: 2760–2770.
50. Mikami T, Kashiwagi M, Tsuchihashi K, Daino T, Akino T, et al. (1998) Further characterization of equine brain gangliosides: the presence of GM3 having N-glycolyl neuraminic acid in the central nervous system. J Biochem 123: 487–491.
51. Rodrig N, Osanai T, Iwamori M, Nagai Y (1988) Uncoupling of intracellular cyclic AMP and dome formation in cultured canine kidney epithelial cells: effects of gangliosides and vasopressin. J Biochem 104: 215–219.
52. Cochran FB, Jr., Yu RK, Ledeen RW (1982) Myelin gangliosides in vertebrates. J Neurochem 39: 773–779.
53. Phillips PA, Alpers MP, Stanley NF (1970) Hydrocephalus in mice inoculated neonatally by the oronasal route with reovirus type 1. Science 168: 858–859.
54. Barton ES, Youree BE, Ebert DH, Forrest JC, Connolly JL, et al. (2003) Utilization of sialic acid as a coreceptor is required for reovirus-induced biliary disease. J Clin Invest 111: 1823–1833.
55. Frierson JM, Pruijssers AJ, Konopka JL, Reiter DM, Abel TW, et al. (2012) Utilization of Sialylated Glycans as Coreceptors Enhances the Neurovirulence of Serotype 3 Reovirus. J Virol: in press.
56. Raval G, Biswas S, Rayman P, Biswas K, Sa G, et al. (2007) TNF-alpha induction of GM2 expression on renal cell carcinomas promotes T cell dysfunction. J Immunol 178: 6642–6652.
57. Watanabe T, Pukel CS, Takeyama H, Lloyd KO, Shiku H, et al. (1982) Human melanoma antigen AH is an autoantigenic ganglioside related to GD2. J Exp Med 156: 1884–1889.
58. Cahan LD, Irie RF, Singh R, Cassidenti A, Paulson JC (1982) Identification of a human neuroectodermal tumor antigen (OFA-I-2) as ganglioside GD2. Proc Natl Acad Sci U S A 79: 7629–7633.
59. Yamada T, Bando H, Takeuchi S, Kita K, Li Q, et al. (2011) Genetically engineered humanized anti-ganglioside GM2 antibody against multiple organ metastasis produced by GM2-expressing small-cell lung cancer cells. Cancer Sci 102: 2157–2163.
60. Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262: 1401–1407.
61. Harbury PB, Kim PS, Alber T (1994) Crystal structure of an isoleucine-zipper trimer. Nature 371: 80–83.
62. Palma AS, Feizi T, Zhang Y, Stoll MS, Lawson AM, et al. (2006) Ligands for the beta-glucan receptor, Dectin-1, assigned using ''designer'' microarrays of oligosaccharide probes (neoglycolipids) generated from glucan polysaccharides. J Biol Chem 281: 5771–5779.
63. Palma AS, Zhang Y, Childs RA, Campanero-Rhodes MA, Liu Y, et al. (2012) Neoglycolipid-based ''designer'' oligosaccharide microarrays to define beta-glucan ligands for Dectin-1. Methods Mol Biol 808: 337–359.
64. Stoll M, Feizi T. Software Tools for Storing, Processing and Displaying Carbohydrate Microarray Data. In: Kettner C, editor. 4–8 October, 2009. Potsdam, Germany. Frankfurt, Germany: Beilstein Institute for the Advancement of Chemical Sciences. pp. 123–140.
65. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66: 125–132.
66. (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50: 760–763.
67. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674.
68. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
69. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240–255.
70. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, et al. (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58: 1948–1954.
71. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, et al. (2011) BUSTER version 2.10.0.
72. Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, et al. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. Acta Crystallogr D Biol Crystallogr 68: 368–380.
73. DeLano WL (2010) The PyMOL Molecular Graphics System. Version 1.2r3pre. Schrödinger, LLC.
74. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302: 205–217.
75. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189–1191.
76. Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. Bioinformatics 20: 426–427.
77. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60: 2256–2268.
78. Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, et al. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50: 437–450.
79. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55: 379–400.
80. Saff EB, Kuijlaars ABJ (1997) Distributing many points on a sphere. Mathematical Intelligencer 19: 5–11.

81. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, et al. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 6: 135–140.
82. Danthi P, Pruijssers AJ, Berger AK, Holm GH, Zinkel SS, et al. (2010) Bid regulates the pathogenesis of neurotropic reovirus. PLoS Pathog 6: e1000980.
83. Kobayashi T, Antar AA, Boehme KW, Danthi P, Eby EA, et al. (2007) A plasmid-based reverse genetics system for animal double-stranded RNA viruses. Cell Host Microbe 1: 147–157.
84. Virgin HWt, Bassel-Duby R, Fields BN, Tyler KL (1988) Antibody protects against lethal infection with the neurally spreading reovirus type 3 (Dearing). J Virol 62: 4594–4604.
85. Smith RE, Zweerink HJ, Joklik WK (1969) Polypeptide components of virions, top component and cores of reovirus type 3. Virology 39: 791–810.
86. Wilson GJ, Wetzel JD, Puryear W, Bassel-Duby R, Dermody TS (1996) Persistent reovirus infections of L cells select mutations in viral attachment protein sigma1 that alter oligomer stability. J Virol 70: 6598–6606.
87. Wetzel JD, Chappell JD, Fogo AB, Dermody TS (1997) Efficiency of viral entry determines the capacity of murine erythroleukemia cells to support persistent infections by mammalian reoviruses. J Virol 71: 299–306.
88. Virgin HWt, Mann MA, Fields BN, Tyler KL (1991) Monoclonal antibodies to reovirus reveal structure/function relationships between capsid proteins and genetics of susceptibility to antibody action. J Virol 65: 6772–6781.
89. Iskarpatyoti JA, Willis JZ, Guan J, Ashley Morse E, Ikizler M, et al. (2012) A rapid, automated approach for quantitation of rotavirus and reovirus infectivity. J Virol Methods 184: 1–7.
90. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. Science 336: 1030–1033.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 49, pp. 42585–42593, December 9, 2011

Structural and Phylogenetic Analyses of the GP42 Transglutaminase from *Phytophthora sojae* Reveal an Evolutionary Relationship between Oomycetes and Marine *Vibrio* Bacteria*[S]

Received for publication, August 8, 2011, and in revised form, September 16, 2011 Published, JBC Papers in Press, October 12, 2011, DOI 10.1074/jbc.M111.290544

Kerstin Reiss[‡], Eva Kirchner[‡1], Mark Gijzen[§], Georg Zocher[‡], Birgit Löffelhardt[¶], Thorsten Nürnberger[¶], Thilo Stehle[‡‖2], and Frédéric Brunner[¶3]

From the [‡]Interfakultäres Institut für Biochemie and [¶]Zentrum für Molekularbiologie der Pflanzen, Universität Tübingen, 72076 Tübingen, Germany, [§]Agri-Food Canada, London, Ontario N5V 4T3, Canada, and the [‖]Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232

Background: The *Phytophthora sojae* GP42 transglutaminase induces defense responses in parsley and potato.
Results: GP42 folds into a novel structure that has arisen through convergent evolution.
Conclusion: GP42 has unique structural and enzymatic properties distinct from mammalian or bacterial transglutaminases.
Significance: This offers a basis to engineer durable broad spectrum resistance in plants with the design and use of GP42 inhibitors.

Transglutaminases (TGases) are ubiquitous enzymes that catalyze selective cross-linking between protein-bound glutamine and lysine residues; the resulting isopeptide bond confers high resistance to proteolysis. *Phytophthora sojae*, a pathogen of soybean, secretes a Ca^{2+}-dependent TGase (GP42) that is activating defense responses in both host and non-host plants. A GP42 fragment of 13 amino acids, termed Pep-13, was shown to be absolutely indispensable for both TGase and elicitor activity. GP42 does not share significant primary sequence similarity with known TGases from mammals or bacteria. This suggests that GP42 has evolved novel structural and catalytic features to support enzymatic activity. We have solved the crystal structure of the catalytically inactive point mutant GP42 (C290S) at 2.95 Å resolution and identified residues involved in catalysis by mutational analysis. The protein comprises three domains that assemble into an elongated structure. Although GP42 has no structural homolog, its core region displays significant similarity to the catalytic core of the Mac-1 cysteine protease from Group A *Streptococcus*, a member of the papain-like superfamily of cysteine proteases. Proteins that are taxonomically related to GP42 are only present in plant pathogenic oomycetes belonging to the order of the Peronosporales (*e.g. Phytophthora*, *Hyaloperonospora*, and *Pythium* spp.) and in marine *Vibrio* bacteria. This suggests that a lateral gene transfer event may have occurred between bacteria and oomycetes. Our results offer a basis to design and use highly specific inhibitors of the GP42-like TGase family that may impair the growth of important oomycete and bacterial pathogens.

Transglutaminases (TGases)[4] ((*R*)-glutamyl-peptide:amine-γ-glutamyltransferases, EC 2.3.2.13) catalyze an acyl transfer reaction between peptide-bound glutamine residues and primary amines, including the ϵ-amino group of peptide-bound lysine residues. TGase activity has been implicated in a multitude of physiological activities in animals, plants, and bacteria (1–3). Human TGases such as the blood coagulation factor XIIIa or the tissue TGase 2 have been structurally characterized (4, 5). Despite the low degree of homology at the amino acid sequence level, it appears that human TGases share a core structural fold with the papain-like cysteine proteases (1). The reaction mechanism of TGases and cysteine proteases is based on a catalytic triad formed by Cys, His, and Asp residues and proceeds via an acyl-enzyme intermediate in which the active site cysteine is covalently bound to the ligand (1). This is a striking example of the conservation in evolutionary distant taxa of a tertiary structure in proteins mechanistically related but with highly diverging primary sequences. Bioinformatics analyses performed by Makarova *et al.* (6) demonstrated a statistically significant sequence similarity between mammalian TGases and a group of microbial proteins of which the only functionally characterized member is a protease (7). Subsequently, it was suggested that animal TGases have evolved from

* This work was supported by German Research Foundation (DFG) Grants Br 3875/2-1 (to F. B.) and SFB766 (to T. S.)
[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1 and Figs. S1–S6.
The atomic coordinates and structure factors (code 3TW5) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).
[1] Present address: Institut Pasteur, Département de Virologie, Unité de Virologie Structurale and CNRS URA3015, 25 rue du Dr Roux, 75724 Paris Cedex 15, France.
[2] To whom correspondence may be addressed: Interfakultäres Institut für Biochemie, Universität Tübingen, Hoppe-Seyler-Str. 4, D-72076 Tübingen, Germany. Tel.: 49-7071-2973043; Fax: 49-7071-295565; E-mail: thilo.stehle@uni-tuebingen.de.
[3] To whom correspondence may be addressed: Frédéric Brunner, Zentrum für Molekularbiologie der Pflanzen, Universität Tübingen, Auf der Morgenstelle 5, D-72076 Tübingen, Germany. Tel.: 49-7071-2976667; Fax: 49-7071-295226; E-mail: frederic.brunner@zmbp.uni-tuebingen.de.
[4] The abbreviations used are: TGase, transglutaminase; VhTGase, *V. harveyi* TGase-like protein; r.m.s.d., root mean square deviation.

ancestral cysteine proteases, but to date, there is a lack of evidence to support such a model of divergent evolution.

Phytophthora species are pathogens that infect various plants, leading to major economic and environmental damage. *Phytophthora sojae* in particular causes root and stem rot of soybeans. Immunolocalization assays demonstrated that the GP42 TGase is associated with hyphal cell walls of *P. sojae* growing *in planta* (8). We have previously shown that extracellular TGase activity is detectable in at least 10 different *Phytophthora* spp. and that proteins serologically related to GP42 are expressed in all these species (9). The GP42 TGase is similar in both its biochemical and its enzymatic characteristics to mammalian TGases (9), but its biological function is yet to be elucidated. Notably, GP42 serves as a microbe-associated molecular pattern that triggers plant immunity in a plant pattern recognition receptor-dependent manner (9–11). A stretch of 13 amino acids (Pep-13; VWNQPVRGFKVYE), located in the C-terminal region of GP42, has been shown to be sufficient for receptor-mediated perception and plant immunogenic activity (11).

To provide a platform for understanding the molecular mechanism of the GP42-catalyzed transamidation reaction and to allow for its comparison with other known enzymes, we have determined the crystal structure of a catalytically inactive form of GP42 in which the active site cysteine residue is replaced with a serine. The core region of GP42 superposes well with the canonical domain characteristic of proteins belonging to the papain-like protein superfamily of cysteine proteases. However, the GP42 sequences surrounding the catalytic core fold into domains that are novel and have not been observed to date in other proteins. Moreover, the arrangement of the catalytic triad in GP42 differs from that seen in cysteine proteases. Hence, the structural analysis clearly suggests that GP42 has arisen through convergent evolution. GP42-related sequences are only present in closely related oomycetes belonging to the order of the Peronosporales and in a small group of marine *Vibrio* bacteria that can infect numerous fish species and marine invertebrates (12). We show that the TGase-like protein from *Vibrio harveyi* displays significant TGase activity *in vitro*. We therefore conclude that TGases from plant pathogenic oomycetes were acquired from bacteria by horizontal gene transfer, probably conferring a selective advantage to Peronosporales over other oomycetes.

EXPERIMENTAL PROCEDURES

Recombinant Protein Expression and Purification—The cDNA sequence encoding the mature *P. sojae* GP42 (GenBank accession AAA67875.1, amino acids 163–529) was cloned into the pPIC9K vector (Invitrogen) and subjected to site-directed mutagenesis using the GeneEditor *in vitro* site-directed mutagenesis system (Promega) to replace Cys-290 with serine. The expression of recombinant GP42 (C290S) was performed in *Pichia pastoris* GS115 according to the multi-copy Pichia expression kit manual (Invitrogen). The culture supernatant containing recombinant GP42 (C290S) was adjusted to 3.5 M $(NH_4)_2SO_4$ and incubated for 2 h at 4 °C prior to centrifugation at 15,000 × *g* for 20 min. The pellet was then resuspended in water and dialyzed against 0.1 M K_2HPO_4/KH_2PO_4, pH 7.9, 5 mM DTT, 5 mM $CaCl_2$ prior to loading onto a DEAE-Sepharose column (GE Healthcare) equilibrated with the same buffer. Bound protein was eluted with a gradient of 0–0.5 M KCl. GP42 (C290S)-containing fractions were pooled, concentrated, and subjected to gel filtration chromatography on a Superdex 75 16/60 column (GE Healthcare). The column was eluted with 0.1 M K_2HPO_4/KH_2PO_4, pH 7.9, 5 mM DTT, 5 mM $CaCl_2$, and 0.1 M KCl. Fractions containing pure recombinant GP42 (C290S) at 2–3 mg/ml were used for crystallization.

The cDNA sequence encoding the *V. harveyi* TGase-like protein (VhTGase, Zp_01984668) was synthesized (GeneArt) and cloned into the pDEST17 vector (Invitrogen). Expression and purification of recombinant VhTGase with an N-terminal His_6 tag was performed in *Escherichia coli* BL21-AI according to the *E. coli* expression systems with Gateway Technology manual (Invitrogen). Guinea pig liver TGase and trypsin were delivered by Sigma and Roche Diagnostics, respectively.

Site-directed Mutagenesis and in Vitro Enzyme Activity Assays—Site-directed mutagenesis to generate the GP42 TGase mutants listed in Table 2 and expression in *P. pastoris* were performed as described above. The TGase assay was directly conducted with the culture filtrate containing the recombinant proteins. Briefly, *P. pastoris* transformants were grown for 3 days at 30 °C in induction medium (buffered methanol complex medium) before harvesting the culture supernatant. Aliquots were then desalted by size exclusion chromatography on PD-10 columns (GE Healthcare) and freeze-dried. The proteins were resuspended in 0.1 M MES, pH 6.0, and separated by SDS-PAGE followed by Commassie Brilliant Blue staining or transfer onto nitrocellulose. Western blots were performed according to standard protocols. Both primary (anti-Pep-25) and secondary (ECL Plex goat anti-rabbit IgG-Cy3-conjugated, GE Healthcare) antibodies were used at 1:2000 dilution. Immunodetection was performed using an FMBIO III bio-imager (Hitachi), and protein quantification was performed with the Aida image analyzer software (Raytest). Protein concentration was calculated as the mean of three independent experimental replicates. Serial dilutions of the purified recombinant GP42 (C290S) protein preparation (1 μg/μl) were used for calibration. TGase activity was determined using the [^{3}H]putrescine assay (13), with minor modifications. The reaction was performed in 0.1 M MES, pH 6.0, 10 mM DTT, 5 mM $CaCl_2$, 10 μg/μl *N*,*N*′-dimethylcasein, 1 μM [2,3-^{3}H]putrescine (60–120 Ci/mmol; Hartmann Analytic), 10 mM putrescine. Protease activity was determined using the ENZCHEK® peptidase/protease assay kit (Invitrogen) following the instructions provided by the manufacturer.

Crystallization and Heavy Atom Derivatization—GP42 (C290S) crystals grew with the sitting drop vapor diffusion method. 450 μl of GP42 (C290S) at a concentration of 2–3 mg/ml in 0.1 M K_2HPO_4/KH_2PO_4, pH 7.9, 5 mM DTT, 5 mM $CaCl_2$, and 0.1 M KCl were mixed with 450 μl of reservoir solution: 0.1 M CAPS, pH 11, 0.2 M Li_2SO_4, 1.6 M $(NH_4)_2SO_4$. For cryoprotection, crystals were transferred into 3 M sodium malonate (pH 7.4) before freezing them in liquid nitrogen. Heavy atom soaks were carried out with 1 mM $[Ta_6Br_{12}]^{2+} \times 2\ Br^-$; (Jena Biosciences) dissolved in 3 M sodium malonate (pH 7.4). The crystals turned green and were transferred back into 3 M

sodium malonate (pH 7.4) prior to flash-freezing. The crystals belong to space group $P6_222$ and contain two molecules per asymmetric unit (see Table 1). Both native and derivative datasets were collected at beamline PXIII of the Swiss Light Source in Villigen, Switzerland. Diffraction data were processed with XDS (14).

Crystal Structure Determination and Structural Refinement—The structure was solved by single isomorphous replacement with anomalous scattering phasing using autoSHARP (15). As the phasing power was poor beyond 6 Å resolution, phase extension to the resolution of the native dataset was performed. To further improve the initial phases, the two-fold non-crystallographic symmetry and the high solvent content of the crystals (78%) were exploited. Solvent flattening was performed with RESOLVE (Los Alamos National Laboratory). The resulting electron density map was of good quality, allowing manual model building using Coot (16). Crystallographic refinement was carried out with Refmac5 (CCP4) (17). PyMOL (35) was used to create the structure figures. The topology diagram was created using topdraw (CCP4) (18).

Phylogenetic Analysis—To determine the extent and distribution of TGase-like sequences among organisms, known and predicted proteins within the National Center for Biotechnology Information (NCBI) non-redundant databases were searched by BLASTP using the *P. sojae* GP42 sequence as a query. Predicted proteins from whole genome sequencing of *Hyaloperonospora arabidopsidis* (VBI Microbial Database Version 5.0), *Phytophthora infestans* (available through the Broad Institute of MIT and Harvard), *Phytophthora sojae* and *Phytophthora ramorum* (www.jgi.doe.gov), and from *Pythium ultimum* (*Pythium* Genome Database) were retrieved and searched using BLASTP. Sequences were edited to remove redundant and fragmentary proteins, and low scoring sequences (expected value $>10^{-5}$). Protein sequence alignments were made using CLUSTALW, MUSCLE, and T-COFFEE, and phylogenic inferences and bootstrap values were calculated by maximum likelihood, neighbor joining, or the unweighted pair group method with arithmetic means, using the computer software PHYLIP and MEGA. Jalview was used to further analyze the alignments made with CLUSTALW. N- and C-terminal sequences that did not correspond to the mature sequence of *P. sojae* GP42 (amino acids 163–529) were removed. The alignment was colored according to percentage of sequence identity (19).

TABLE 1
X-ray data collection and structural refinement

	Native	Ta_6Br_{12} derivative
Data collection[a]		
Resolution (Å)	30–2.95 (3.03–2.95)	50–5.50 (5.64–5.50)
Space group	$P6_222$	$P6_222$
Cell dimensions		
a and *c* (Å)	195.5, 137.3	195.1, 138.4
α, β, and γ	90, 90, 120	90, 90, 120
$R_{observed}$	9.0 (58.3)	10.2 (27.2)
λ(Å)	1.0	1.255
$I/\sigma(I)$	15.51 (2.75)	15.93 (6.99)
Completeness (%)	96.4 (98.2)	99.8 (100)
Redundancy	3.7	6.5
R_{iso}(%)		36.5
Refinement		
R_{work}/R_{free} (%)	21.5/23.5	
Figure of merit	0.85	
B-factors		
Chain A (Å^2)	47.0	
Chain B (Å^2)	47.0	
No. of atoms		
Protein	5297	
CAPS	28	
r.m.s.d.		
Bond lengths (Å)	0.01	
Bond angles (°)	0.9	
Ramachandran plot[b]		
Favored (%)	656 (96.5)	
Allowed (%)	24 (3.5)	

[a] Data collection was at Beamline PXIII (SLS, Villigen, Switzerland). Values in parentheses are for highest resolution shell.
[b] Ramachandran plot generated with Rampage (CCP4) (34).

RESULTS AND DISCUSSION

Overall Structure of GP42—GP42 from *P. sojae* consists of an N-terminal prodomain region (amino acids 1–162) that is cleaved off during enzyme maturation and a region that comprises the TGase (amino acids 163–529). The TGase was recombinantly expressed in *P. pastoris* and purified from the culture medium using a two-step chromatography protocol (see "Experimental Procedures"). We solved the crystal structure of the catalytically inactive point mutant (C290S) at a resolution of 2.95 Å by single isomorphous replacement with anomalous scattering (Table 1). The structure of GP42 reveals a highly intertwined fold. Its overall appearance resembles a sea horse, with a head, a body, and a curved tail (Fig. 1, *A* and *B*). To facilitate discussion of its structural elements, GP42 can be divided into three domains: the head (domain I), body (domain II), and tail (domain III). The head, depicted in *yellow*, consists of a small β-sheet formed by two β-strands (β2 and β3) that faces a short 3_{10} helix. The body (shown in *cyan*) is compact, globular, and predominantly composed of α-helices (α2, α3, α4, α5, α8, and α9). The long α5 helix forms the core of the domain, whereas the other helices are arranged circularly around it. The body domain also contains a short 3_{10} helix following α8 and a small β-sheet (β4 and β5) between α4 and α5. The tail domain (depicted in *navy*) has an elongated, curved shape and features elaborate loops. The core of the tail domain is primarily formed by a long and highly twisted antiparallel β-sheet (β-strands β1, β8, β9, β10, β11) that wraps tightly around helix α6. Loops connecting the β-strands in some cases feature helices (α7, α8, and a 3_{10} helix). The tail domain is connected to the body domain via a small four-stranded antiparallel β-sheet (β6, β7, β8′, and β12). An additional link to the body domain is provided by the N-terminal helix α1 and strand β1. Although both are part of the tail domain, the chain enters the body domain with residues following β1. GP42 contains seven cysteine residues. Intramolecular disulfide bridges between Cys-256 and Cys-271 and between Cys-262 and Cys-282 stabilize the body domain, whereas a third bridge links Cys-305 with Cys-527 in the tail domain. The presence of disulfide bridges is quite unusual because they can interfere with conformational changes upon activation in mammalian TGases (1, 20). The seventh cysteine, Cys-290, is the active-site nucleophile for TGase activity. Replacement of Cys-290 with serine was necessary for crystallization as active TGases are able to catalyze *in vitro* disordered intermolecular cross-linking, resulting in insoluble aggregates (5, 21, 22). As replacement of cysteine with serine is unlikely to affect the fold of the protein, we expect that both side chains will be in equivalent positions in GP42. For

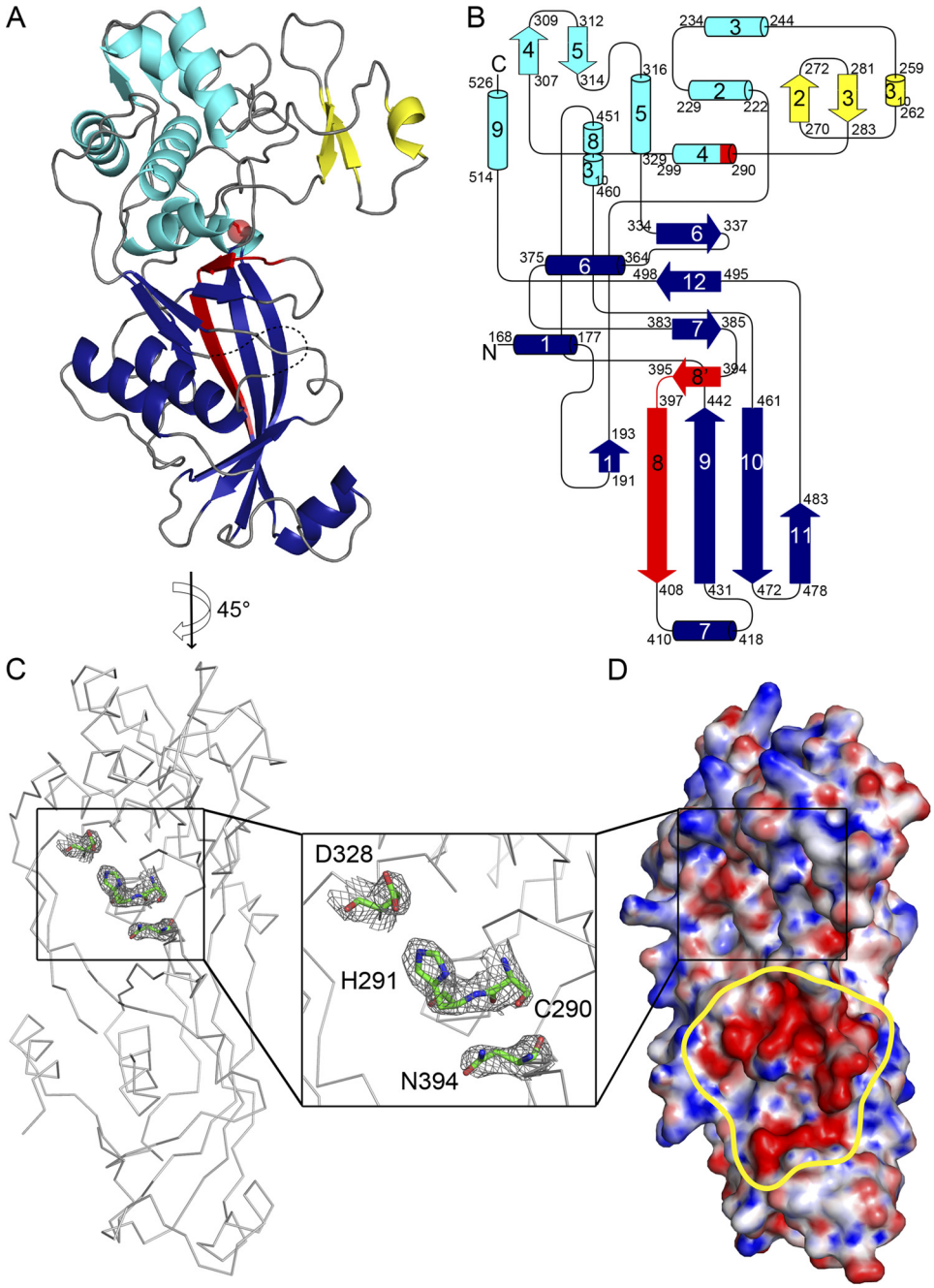

FIGURE 1. **Crystal structure of GP42 (C290S).** *A*, graphic representation. *B*, topology diagram of GP42. The head domain is depicted in *yellow*, the body is depicted in *cyan*, and the tail is depicted in *navy*. The catalytic residue Cys-290 and the Pep-13 fragment, which induces an immune response in plants, are highlighted in *red*. *C*, α-carbon backbone trace of GP42, rotated by 45° as compared with the view in *panel A*. The postulated catalytic triad, consisting of Cys-290, His-291, and Asp-328, and Asn-394, which forms a hydrogen bond with Cys-290, are shown in stick representation together with their ($2F_o - F_c$) electron density maps in the *inset*. The electron density map is contoured at 1.7 σ. *D*, electrostatic surface of GP42 with a color scale that varies from *blue* to *red*, representing positive and negative potential, respectively. The electrostatic surface of GP42 (C290S) was calculated using Adaptive Poisson-Boltzmann Solver (APBS (33)) with an electrostatic potential between −20 and +20 k_BT/e.

simplicity, we will from here on refer to the residue at position 290 as Cys-290.

The amino acid sequence corresponding to Pep-13 is located on β-strands β8 and 8′, near Cys-290 (Fig. 1, *A* and *B*). Belonging to two different sheets, these two strands are separated by a single proline residue (Pro-396) that causes the chain to bend at this position. One face of β8/8′ is fully exposed to solvent, rendering many of the residues of Pep-13 accessible to putative ligands. Moreover, because Pep-13 is part of β-strands 8 and 8′, many of its residues are conformationally restrained. The surface exposure of the Pep-13 sequence within GP42 is one feature that logically explains why this sequence constitutes an immunogenic motif that triggers plant immunity through activation of plant pattern recognition receptors.

Architecture of the Active Site—Inspection of the structure reveals that Cys-290 is located at the interface between the body and tail domain, at the N terminus of helix α4. One side of Cys-290 is covered by the loop connecting β-strands β8′ and

β9, but the other side is exposed to the surface and poised to interact with its substrate. The reaction mechanism of all known TGases is based on a Cys-His-Asp triad or, less frequently, a Cys-His dyad (1). Nucleophilic attack on the substrate is performed by the sulfhydryl group of the cysteine following activation by a thiolate-imidazolium ion pair involving the histidine side chain. Six histidines are present in the sequence of GP42, but only one of these, His-291, is in close proximity to Cys-290. The remaining histidines have distances of at least 17 Å from the Cys-290 sulfhydryl group and thus cannot participate in catalysis. His-291 is located immediately next in sequence to Cys-290, and it is part of the same α4 helix. Such an unusual arrangement of two neighboring amino acids forming a catalytic dyad or triad has so far not been described in the literature for any cysteine protease or TGase. Moreover, the distance between the imino nitrogen atom of His-291 and the sulfhydryl group of Cys-290 is, with 4.8 Å in chain A and 5.0 Å in chain B, rather large, and it is therefore unclear whether Cys-290 can be directly activated by the His-291 imidazole group. However, the His-291 side chain does form a salt bridge with Asp-328 (3.0 Å distance) located on the α5 helix, and therefore Cys-290, His-291, and Asp-328 could constitute a functional catalytic triad (Fig. 1*C*). Different rotamers of His-291 would bring its side chain closer to Cys-290. Such rotamers are not observed in the crystal structure but may occur in solution.

It is also conceivable that the main function of His-291 is not to lower the pK_a value of Cys-290 but to support catalysis by stabilizing an intermediate. Such a mechanism was shown for FabH, an enzyme that catalyzes a transacylase reaction in the biosynthesis of fatty acids (23). The active site of this enzyme is composed of Cys-112, His-244, and Asn-274. The distance between Cys-112 and His-244 is, with 3.6 and 4.1 Å in different protomers, also larger than a hydrogen bond. Thus, deprotonation of Cys-112 would not be supported by interaction with His-244. Instead, the pK_a of Cys-112 is lowered due to its localization at the N terminus of an α-helix and the resulting dipole effect (24). The GP42 residue Cys-290 is also located at the N terminus of an α-helix (α4), and a similar dipole effect could increase its acidity. It is therefore possible that the His-291-Asp-328 pair stabilizes a reaction intermediate by hydrogen bonding.

Mapping the electrostatic potential of GP42 onto its surface reveals that most of the protein is either electropositive or neutral (Fig. 1*D*). Strikingly, however, a strong negative potential delineates a groove adjacent to the active site. This region features several acidic residues projecting from different β-strands and loops. The presence of a highly negatively charged region in close proximity to the active site suggests that the groove may interact with a positively charged substrate. Although the identity of GP42 substrates is not yet known, it is likely that the enzyme catalyzes the cross-linking between peptides that contain basic amino acids.

Mutational Analyses of Residues Cys-290, His-291, and Asp-328—To determine the function of residues Cys-290, His-291, and Asp-328, which constitute the catalytic triad in GP42, point mutants of all three residues were generated, and the culture filtrates of *P. pastoris* expressing the recombinant proteins were tested for enzymatic activity (Table 2, supplemental Fig. S1). The advantage to using *P. pastoris* for protein expression is that there is no detectable TGase activity in the culture filtrate of the untransformed strain. In a standard assay, based on the incorporation of [^{3}H]putrescine into *N*,*N*′-dimethylcasein, the specific TGase activity of GP42 was 30.15 units/mg. This value was about 2 orders of magnitude above the value obtained for the guinea pig liver TGase (0.117 units/mg). As expected, the C290S mutant had no detectable activity. We also generated single mutants H291A, H291F, H291Y, and D328N as well as the double mutant H291A/D328N. The D328N mutation is a conservative exchange unlikely to cause structural changes. In line with our hypothesis that Asp-328 is part of the catalytic triad of GP42, the D328N mutant had only 2.6% of wild-type GP42 activity (Table 2). To determine how size and polarity of His-291 affects catalytic activity, it was first replaced with Phe or Tyr. The His-291 side chain is solvent-exposed, and the two substitutions are unlikely to produce steric clashes. Both mutants exhibit negligible activity (Table 2), suggesting that neither Tyr-291 nor Phe-291 is able to functionally replace His-291. However, the slightly larger side chains of Phe-291 and Tyr-291 may also make access of the substrate to the active site more difficult, thus affecting activity. To test this possibility, we mutated His-291 to Ala and found that the H291A mutant retained 19.1% of wild-type activity (Table 2). An explanation for this unexpected finding could be that due to the small size of the Ala side chain, the substrate can enter the active site more easily, perhaps facilitating cleavage in the absence of the proton-abstracting histidine side chain. A second possibility is that one or two water molecules could enter the active site in the H291A mutant, linking Cys-290 with Asp-328 and thus helping to deprotonate Cys-290 through a proton relay system. The additive effect of H291A and D328N in the double mutant, which has an activity of 0.3% of the wild type (Table 2), clearly demonstrates the crucial role of the two residues in catalysis. Finally, we also mutated Asn-394 to Ala (N394A) because this residue forms direct hydrogen bonds with Ser-290 (Fig. 1*C*) as well as with Trp-496. It is likely that Asn-394 also makes a similar interaction with Cys-290 in the wild-type protein, although the hydrogen bond would be somewhat weaker. Replacement with Ala reduced the activity to 0.02% (Table 2), indicating that Asn-394 may be responsible for the correct positioning of Cys-290.

TABLE 2
TGase activity of GP42 point mutants

GP42	TGase activity	Relative TGase activity
	Units/mg ± S.D.	*%*
Wild type	30.15 ± 1.80	100
C290S	ND[a]	ND[a]
D328N	0.77 ± 0.01	2.6
W393F	22.70 ± 3.72	75.3
N394A	0.007 ± 0.0001	0.02
H291A/D328N	0.088 ± 0.001	0.3
H291A	5.76 ± 0.58	19.1
H291F	ND[a]	ND[a]
H291Y	0.068 ± 0.001	0.2

[a] ND, not detectable

GP42 Exhibits Structural Homology to the Cysteine Protease Mac-1 from Group A Streptococcus—A search for tertiary structure homologs using the DALI server did not reveal any

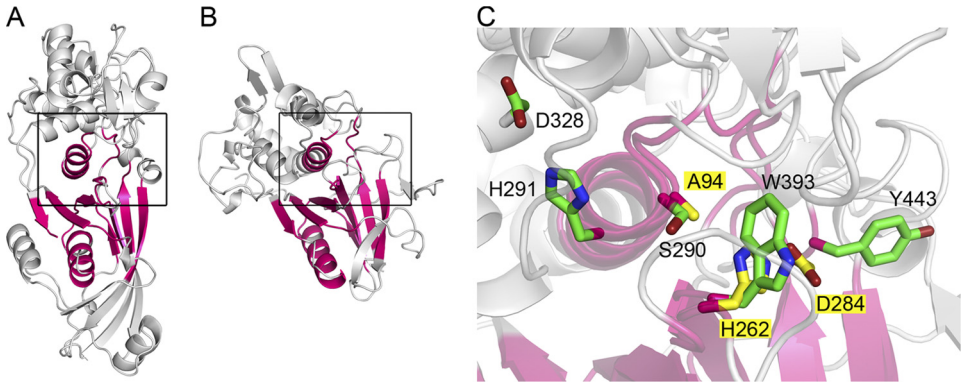

FIGURE 2. **The structure of the catalytic core domain is conserved between the *P. sojae* GP42 and the bacterial Mac-1 cysteine protease.** Structurally conserved regions in GP42 (*A*) and Mac-1 (*B*) are shown in *pink*. Alignment of the Cα atoms using lsqkab resulted in an r.m.s.d. value of 1.4 Å over 77 aligned residues (out of 367 GP42 residues). *C*, close-up view of the superposition of GP42 and Mac-1, centered at their active sites. The postulated catalytic triads of GP42 (Ser-290, His-291, and Asp-328) and Mac-1 (Ala-94, His-262, and Asp-284) and the GP42 residues Trp-393 and Tyr-443 that superpose with the catalytic residues His-262 and Asp-284 of Mac-1 are shown as stick representations and are highlighted in *green* (GP42) and *yellow* (Mac-1).

significant (based on *Z*-scores >6.0) homology of GP42 to other proteins when the individual domains are considered. However, searches with the complete structure of GP42 yielded modest hits with bacterial cysteine proteases, indicating the presence of some structural similarity. The highest score of $Z = 6.4$ was obtained with the IgG-degrading cysteine protease Mac-1 from Group A *Streptococcus*. Inspection of the crystal structure of the inactive (C94A) point mutant of Mac-1 (25) reveals that its core structure indeed possesses some similarity with GP42 (Fig. 2, *A* and *B*). Alignment of the Cα atoms of both structures using lsqkab (26) confirms this by yielding a root mean square deviation (r.m.s.d.) value of 1.4 Å for 77 superposed residues out of 367 GP42 residues. Regions of GP42 that can be superposed onto Mac-1 include elements of both the body and the tail domains. Most noticeably, a region comprising the active-site cysteine of GP42, including helix α4 and the two neighboring β-strands β8/8′ and β9, superposes well onto the corresponding helix and β-strands that form the catalytic core in Mac-1. These structural elements are shared by all members of the papain-like protein superfamily, including the papain-like cysteine proteases and peptide: *N*-glycanase or *N*-acetyl transferases. Interestingly, the superposition described above brings the two catalytically active cysteine residues (Cys-290 in GP42 and Cys-94 in Mac-1) in almost perfect alignment, with an r.m.s.d. value of 0.5 Å (Fig. 2*C*). However, the remaining residues of the catalytic triad of Mac-1 (His-262 and Asp-284) have no equivalent residues in GP42. Although His-262 and Asp-284 of Mac-1 align well with Trp-393 (r.m.s.d. = 1.0 Å) and Tyr-443 (r.m.s.d. = 2.0 Å) of GP42, neither Trp-393 nor Tyr-443 could function in a catalytic triad (Fig. 2*C*). The residues that most likely participate in the catalytic triad of GP42, His-291 and Asp-328, face into the opposite direction. Thus, although the central cysteine and key secondary structure elements are structurally conserved in the two proteins, the catalytic triads have apparently evolved differently.

Previous studies have shown that replacing Trp-393 and Pro-396 within the Pep-13 motif with alanines compromised the ability of the protein to induce defense responses in plants and also affected TGase activity to a similar extent (2–6% of wild-type activity) (9, 11). Interestingly, the structural comparison shows that Trp-393 aligns with the catalytically important His-262 of the bacterial Mac-1 protease (Fig. 2*C*). The W393F mutant retained 75% activity, indicating that the tryptophan is not directly involved in catalysis. Trp-393 may, however, play a structural role or be required for substrate binding. It was proposed for the human TGase 2 and red sea bream tissue TGase that highly conserved tryptophans close to the active site regulate substrate entry to the active site or are involved in substrate binding (27–29). We therefore think it likely that a rearrangement of the catalytic triad occurred during evolution of the *Phytophthora* TGases to accommodate aromatic residues (such as Trp-393 and Tyr-443) that were required for specificity. It is tempting to speculate that such a change was accompanied by a switch from hydrolytic to transamidating reactions.

Phylogenic Analysis of TGase Sequences Delineates Bacterial and Oomycete Sequences—Using the GP42 sequence as a query, we performed BLAST searches of public databases and sets of predicted proteins resulting from whole genome sequencing of selected organisms (as described under "Experimental Procedures"). More than 100 sequences were obtained, but these were reduced to a core set of 61 sequences after removing redundant, fragmentary, and low scoring hits (expected value $>10^{-5}$) (supplemental Fig. S2). This core set of 61 TGase-like sequences originated from a limited number of species, as shown in supplemental Table S1. Remarkably, all seven of the prokaryotic TGase-like sequences were from marine bacteria, whereas all 54 of the eukaryotic TGases were from plant pathogenic oomycetes belonging to the order of the Peronosporales.

To determine the relationships among the 61 TGase-like sequences retrieved from databases searches, we performed alignments and phylogenetic analysis using diverse programs. The TGase-like proteins consistently separate into bacterial and eukaryotic groups regardless of the methods used to make the alignments or construct the phylogeny. The branch separation of bacterial from oomycete TGase-like proteins always returns the maximum bootstrap value (equal to the number of replications), whereas other branches within the two groups vary in their placement and bootstrap values depending on the algorithm (Fig. 3*A*). The oomycete TGase-like proteins form a monophyletic group that is well separated from the bacterial proteins. Within the oomycete branch, TGase-like proteins from different species of *Phytophthora*, *Hyaloperonospora*, and

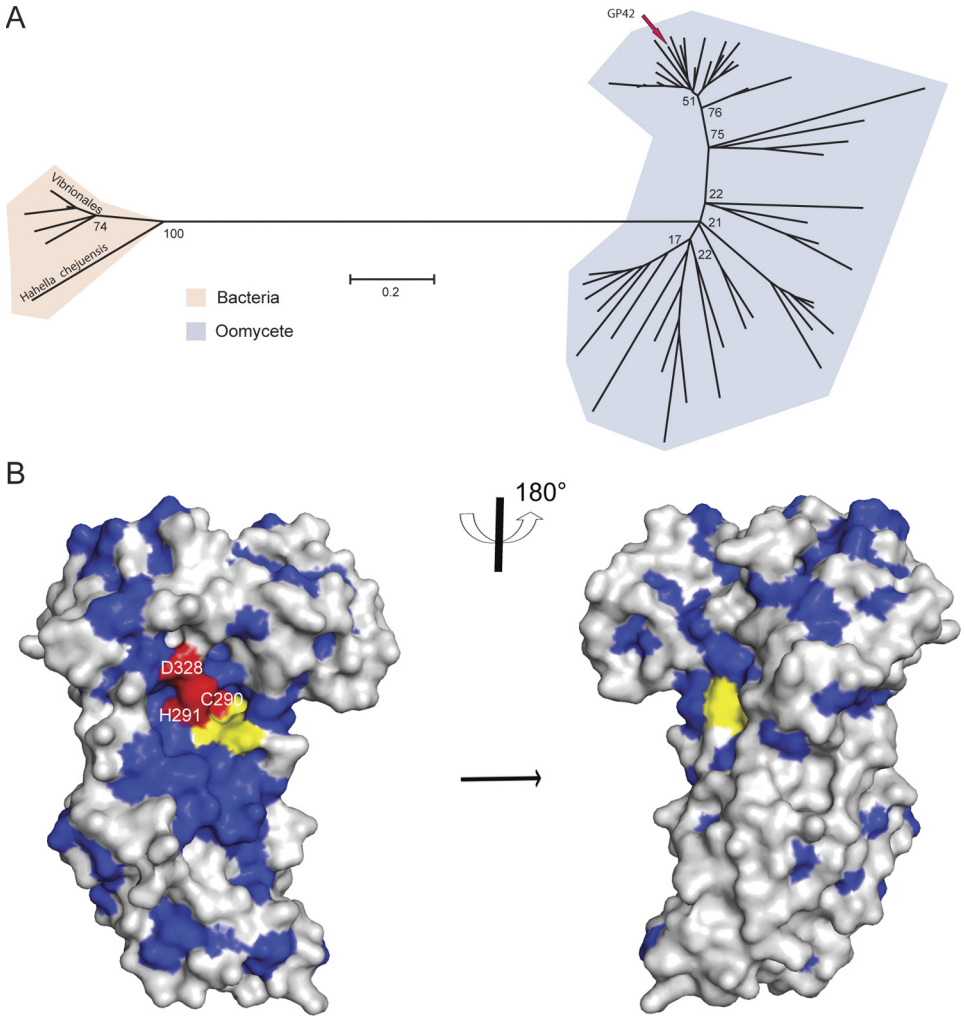

FIGURE 3. **Structurally related GP42 TGases are conserved in all plant pathogenic oomycetes.** *A*, phylogenetic analysis of TGase protein sequences. Shown is a hypothetical radial tree resulting from comparison of 61 TGase protein sequences. The unrooted, maximum likelihood tree was produced from a MUSCLE alignment. Bootstrap values from 100 replicates are shown for major branches. The *scale bar* represents 20% weighted sequence divergence. *B*, surface representation of conserved amino residues within the oomycete TGases. The program Jalview was used to identify residues that are identical in >75% (highlighted in *blue*) of the 54 GP42-related protein sequences from *Phytophthora*, *Pythium*, and *Hyaloperonospora*. These highly conserved residues include the active-site residues Cys-290, His-291, and Asp-328 (highlighted in *red*) and the N-terminal part of the Pep-13 motif (highlighted in *yellow*).

Pythium are variously interspersed (orthologous) and clustered (paralogous) among each other. Our expectation was to identify TGase-like sequences in marine oomycetes that are more likely to share the same habitat as compared with the *Vibrio* species, but there is no evidence of TGase-like sequences in the genome of *Saprolegnia* and *Aphanomyces* species. Our analysis therefore suggests that a single ancestral gene gave rise to all Peronosporales TGases. This hypothesis is further supported by the analysis of surface residue conservation among all members of the oomycete TGase-like proteins (Fig. 3*B* and supplemental Fig. S3). Amino acid residues that were identical in >75% of the sequences are predominantly concentrated at the active site and in the surrounding area, which is probably involved in substrate binding and recognition. The residues forming the catalytic triad and the first six amino acids of the Pep-13 fragment (VWNQPV), which are located in close proximity to the active site, are present in >90% of the analyzed sequences. The less conserved residues in the C-terminal portion of Pep-13 are not as relevant as Trp-393 and Pro-396 for elicitor activity. The heptamer VWNQPVR was found to be completely inactive, demonstrating that the C-terminal part is important for plant perception and induction of immune responses (11). The higher degree of variability might be a strategy of the pathogen to evade plant perception. Whether such TGase homologs with a degenerated Pep-13 motif still trigger plant immune responses and display TGase activity requires further experiments.

The high structural similarity among the oomycete TGase-like proteins suggests that they share similar substrates and a similar catalytic mechanism. The presence of functionally identical TGases, highly conserved within the order of the Peronosporales, is an important criterion of genuine microbial-associated molecular patterns that are by definition present in a broad range of microorganisms.

The Oomycete TGases Originate from Marine Vibrio Species—The phylogenetic analysis has shown that the bacterial TGase-like sequences also formed a monophyletic group, and within this, a subgroup of TGase-like proteins from *Vibrio* species clustered together. In fact, most (six out of seven) of the bacterial sequences originated from Vibrionales, the exception being one

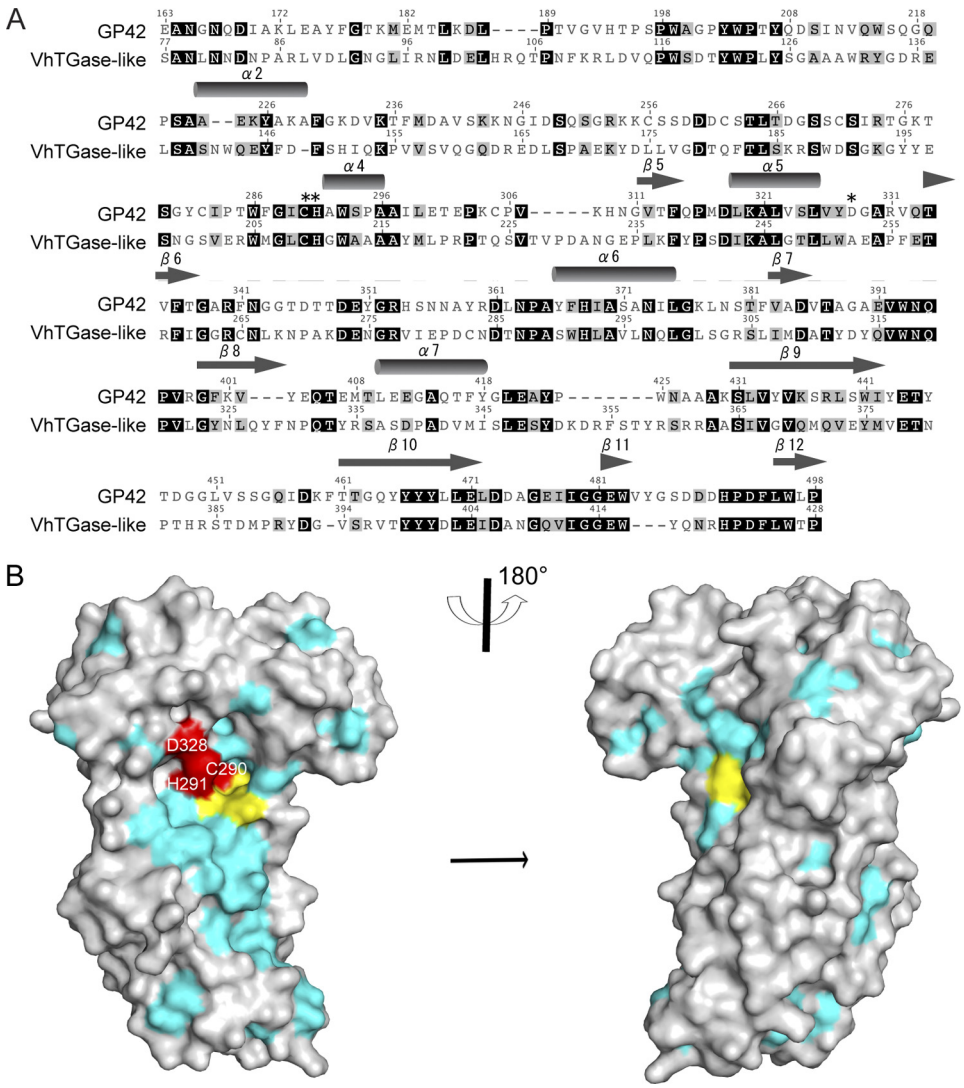

FIGURE 4. **A GP42-related TGase is present in marine bacteria.** *A*, sequence alignment of the *P. sojae* GP42 and *V. harveyi* TGase. Identical (*black*) and similar (gray) residues are highlighted. Based on the crystal structure of GP42 and the structure prediction of VhTGase, conserved secondary elements are indicated by *cylinders* (α-helices) and *arrows* (β-strands). The positions of the residues involved in the catalytic reaction (Cys-His-Asp) are marked with *asterisks*. *B*, surface representation of conserved amino residues between the *P. sojae* GP42 and bacterial TGases. The program Jalview was used to identify residues that are identical in >75% (highlighted in *cyan*) of the six GP42-related protein sequences from *Vibrio* spp. The conserved residues include the active-site residues Cys-290 and His-291 (highlighted in *red*) and the N-terminal part of the Pep-13 motif (highlighted in *yellow*). Asp-328 (highlighted in *red*), which is postulated to be part of the catalytic triad in GP42, is not conserved between the oomycete and the *Vibrio* sequences.

predicted protein from *Hahella chejuensis.* Alignment of the amino acid sequences of GP42 and its closest homolog from *V. harveyi* (VhTGase) revealed the presence of a highly conserved core sequence (31% identity; 45% similarity over 336 amino acids; Fig. 4*A*). In particular, the positions of the residues that were subjected to mutational analyses are conserved, with the noticeable exception of Asp-328, which is replaced with an alanine in the VhTGase sequence. Furthermore, the secondary structure of the VhTGase protein was predicted using the JPRED program. The high degree of secondary structure conservations with GP42, spanning helices α2 and α4–7 and sheets β5–12 (Fig. 4*A*), is another piece of evidence in support of functional homology between the two proteins. There was no significant similarity between the N-terminal prodomain sequence of GP42 and VhTGase. This lack of conservation may reflect the evolutionary divergence to achieve specific regulatory processes. In addition, GP42 is lacking the predicted C-terminal ricin B-lectin domain present in the *Vibrio* homolog but also in secreted bacterial peptidases. To determine whether VhTGase is a functional TGase, we expressed the protein in *E. coli* (supplemental Fig. S4). The purified recombinant protein catalyzes the incorporation of [^{3}H]putrescine into *N*,*N*′-dimethylcasein at very low rate (0.0038 units/mg). The activity is 4 orders of magnitude lower than the GP42 TGase activity measured in yeast culture filtrates, which could be partially explained by the aforementioned substitution of the catalytic Asp by an Ala residue in the *Vibrio* protein sequence. We do not know whether the activity reflects the physiological function of the protein *in vivo*. The presence of the predicted ricin B-lectin domain at the C terminus suggests rather a function in proteolysis. However, no activity was detected in standard *in vitro* assays for proteases (supplemental Fig. S5). Interestingly, surface residue representation showed that the pattern of conserved residues between bacterial and oomycete TGase-like

proteins is very similar to the pattern obtained for the oomycete TGases alone (Fig. 4*B* and supplemental Fig. S6). High sequence similarity in regions adjacent to the catalytic residues and Pep-13 motif favors the hypothesis that the Vibrionales and oomycete TGase-like proteins arise from a common ancestor rather than from convergent evolution. There is a high probability that a lateral gene transfer event with *Vibrio* spp. as the bacterial source occurred prior to speciation and radiation of the Peronosporales. Combined with the absence of TGase-related sequences in animals, plants, fungi, and other bacteria, it is unlikely that a common ancestral protein has been selectively retained or lost through evolution because this would have required a high number of independent gene losses. Interestingly, homologs of the necrosis- and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs), which represent a major class of toxin-like virulence factors in Peronosporales, are also present in the genomes of several Vibrionales (30). Thus, a comprehensive search of the *P. sojae* genome for bacterium-derived genes might reveal additional acquisitions and help to document at which frequency gene transfer occurred. Although the biological function of GP42 has not yet been elucidated, the protein is thought to play a crucial function for pathogen lifestyle or in the infection process on plants. This hypothesis is strongly supported by several lines of evidence that allowed us to consider the Pep-13 motif as a genuine microbe-associated molecular pattern (9–11). We conclude that the GP42-related TGase family probably confers an important gain of pathogenicity to all the Peronosporales that may have supported colonization of land plants. Several virulence-associated proteases have been identified from *V. harveyi* strains (31, 32). Whether the TGase-like proteins from marine *Vibrio* species are playing a crucial role for pathogenicity on crustacean and mollusks also needs to be demonstrated.

Acknowledgments——We thank Dr. Leighton Pritchard for helpful discussions about the evolutionary aspect of TGases, and we are grateful to the staff at beamline PXIII at the Swiss Light Source (SLS), Villigen, Switzerland, for technical support.

REFERENCES

1. Lorand, L., and Graham, R. M. (2003) *Nat. Rev. Mol. Cell Biol.* **4,** 140–156
2. Serafini-Fracassini, D., and Del Duca, S. (2008) *Ann. Bot.* **102,** 145–152
3. Yokoyama, K., Nio, N., and Kikuchi, Y. (2004) *Appl. Microbiol. Biotechnol.* **64,** 447–454
4. Liu, S., Cerione, R. A., and Clardy, J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99,** 2743–2747
5. Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E., and Teller, D. C. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91,** 7296–7300
6. Makarova, K. S., Aravind, L., and Koonin, E. V. (1999) *Protein Sci.* **8,** 1714–1719
7. Pfister, P., Wasserfallen, A., Stettler, R., and Leisinger, T. (1998) *Mol. Microbiol.* **30,** 233–244
8. Hahlbrock, K., Scheel, D., Logemann, E., Nürnberger, T., Parniske, M., Reinold, S., Sacks, W. R., and Schmelzer, E. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92,** 4150–4157
9. Brunner, F., Rosahl, S., Lee, J., Rudd, J. J., Geiler, C., Kauppinen, S., Rasmussen, G., Scheel, D., and Nürnberger, T. (2002) *EMBO J.* **21,** 6681–6688
10. Halim, V. A., Hunger, A., Macioszek, V., Landgraf, P., Nurnberger, T., Scheel, D., and Rosahl, S. (2004) *Physiol. Mol. Plant Pathol.* **64,** 311–318
11. Nürnberger, T., Nennstiel, D., Jabs, T., Sacks, W. R., Hahlbrock, K., and Scheel, D. (1994) *Cell* **78,** 449–460
12. Austin, B., and Zhang, X. H. (2006) *Lett. Appl. Microbiol.* **43,** 119–124
13. Lorand, L., Campbell-Wilkes, L. K., and Cooperstein, L. (1972) *Anal. Biochem.* **50,** 623–631
14. Kabsch, W. (2010) *Acta Crystallogr. D.* **66,** 125–132
15. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) *Methods Mol. Biol.* **364,** 215–230
16. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D.* **60,** 2126–2132
17. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D.* **53,** 240–255
18. Bond, C. S. (2003) *Bioinformatics* **19,** 311–312
19. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) *Bioinformatics* **25,** 1189–1191
20. Begg, G. E., Carrington, L., Stokes, P. H., Matthews, J. M., Wouters, M. A., Husain, A., Lorand, L., Iismaa, S. E., and Graham, R. M. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 19683–19688
21. Ahvazi, B., Boeshans, K. M., and Rastinejad, F. (2004) *J. Struct. Biol.* **147,** 200–207
22. Noguchi, K., Ishikawa, K., Yokoyama, Ki, Ohtsuka, T., Nio, N., and Suzuki, E. (2001) *J. Biol. Chem.* **276,** 12055–12059
23. Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) *Structure* **8,** 185–195
24. Hol, W. G., van Duijnen, P. T., and Berendsen, H. J. (1978) *Nature* **273,** 443–446
25. Agniswamy, J., Nagiec, M. J., Liu, M., Schuck, P., Musser, J. M., and Sun, P. D. (2006) *Structure* **14,** 225–235
26. Kabsch, W. (1976) *Acta Crystallogr. A* **32,** 922–923
27. Chica, R. A., Gagnon, P., Keillor, J. W., and Pelletier, J. N. (2004) *Protein Sci.* **13,** 979–991
28. Murthy, S. N., Iismaa, S., Begg, G., Freymann, D. M., Graham, R. M., and Lorand, L. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99,** 2738–2742
29. Pinkas, D. M., Strop, P., Brunger, A. T., and Khosla, C. (2007) *PLoS Biol.* **5,** e327
30. Ottmann, C., Luberacki, B., Küfner, I., Koch, W., Brunner, F., Weyand, M., Mattinen, L., Pirhonen, M., Anderluh, G., Seitz, H. U., Nürnberger, T., and Oecking, C. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 10359–10364
31. Cheng, S., Hu, Y. H., Zhang, M., and Sun, L. (2010) *Vaccine* **28,** 2716–2721
32. Liu, P. C., Lee, K. K., Tu, C. C., and Chen, S. N. (1997) *Curr. Microbiol.* **35,** 32–39
33. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10037–10041
34. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins Struct. Funct. Genet.* **50,** 437–450
35. DeLano, W. L. (2010) *The PyMOL Molecular Graphics System*, Version 1.3, Schrödinger, LLC, New York

Molecular Plant Advance Access published January 24, 2013

Structure-Function Analysis of *Arabidopsis thaliana* Histidine Kinase AHK5 Bound to Its Cognate Phosphotransfer Protein AHP1

Johannes Bauer[a], Kerstin Reiss[a], Manikandan Veerabagu[b], Michael Heunemann[b], Klaus Harter[b,1] and Thilo Stehle[a,c,1]

a Interfaculty Institute of Biochemistry, University of Tübingen, D-72076 Tübingen, Germany
b Center for Plant Molecular Biology, University of Tübingen, D-72076 Tübingen, Germany
c Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

ABSTRACT The multi-step phosphorelay (MSP) system defines a key signal transduction pathway in plants and many eukaryotes. In this system, external stimuli first lead to the activation of a histidine kinase, followed by transfer of a phosphoryl group from the receiver domain of the kinase (HK_{RD}) to downstream, cytosolic phosphotransfer proteins (HPs). In order to establish the determinants of specificity for this signaling relay system, we have solved the first crystal structure of a plant HK_{RD}, $AHK5_{RD}$, in complex with one of its cognate HPs, AHP1. AHP1 binds $AHK5_{RD}$ via a prominent hydrogen bond docking ridge and a hydrophobic patch. These features are conserved among all AHP proteins, but differ significantly from other structurally characterized prokaryotic and eukaryotic HPs. Surface plasmon resonance experiments show that $AHK5_{RD}$ binds to AHP1-3 with similar, micromolar affinity, consistent with the transient nature of this signaling complex. Our correlation of structural and functional data provide the first insight, at the atomic level as well as with quantitative affinity data, into the molecular recognition events governing the MSP in plants.

Key words: multi-step phosphorelay; phosphotransfer protein; plant signaling; sensor histidine kinase; two-component system.

INTRODUCTION

The multi-step phosphorelay (MSP) system of plants has evolved from the prokaryotic two-component system (TCS), which allows organisms to sense and respond to changes in environmental conditions. In its most basic form, the TCS consists of a membrane-bound histidine kinase (HK) that senses a specific environmental stimulus and a corresponding response regulator (RR) that mediates the cellular response, for example via the differential expression of its target genes (Laub and Goulian, 2007). Compared with most prokaryotic TCS, the MSP signaling system of plants employs a more sophisticated His-Asp-His-Asp phosphorelay system by incorporating additional signaling domains and phosphotransfer proteins (AHPs) (Figure 1A and B).

Recently, substrate specificity determining amino acid residues have been identified in the *Escherichia coli* TCS, namely for the HK-RR pairs of EnvZ–OmpR (Skerker et al., 2008; Capra et al., 2010) and $CheA_3$–$CheY_6$ (Bell et al., 2010). However, this phosphorelay in the prokaryotic TCS corresponds to an *intra*molecular phosphoryl-transfer reaction in the MSP system (Figure 1A and B). The determinants of the subsequent phosphorelay in MSP, the *inter*molecular phosphoryl transfer from the receiver domain of the kinase (HK_{RD}) to the cognate AHP, are therefore not known, and it is furthermore not understood whether and how this interaction confers specificity onto a signal.

The *Arabidopsis thaliana* genome encodes eight canonical HKs (AHK1–5, ETR1, ERS1, and CKI1), five canonical AHPs (AHP1–5), and one pseudo-AHP (AHP6) that carries an asparagine instead of the critical histidine residue (Grefen and Harter, 2004). MSP signaling regulates a wide variety of key processes in *A. thaliana*, including osmoregulation (AHK1) (Urao et al.,

[1] To whom correspondence should be addressed. (T.S.) E-mail thilo.stehle@uni-tuebingen.de, tel. +49–7071–2973043, fax +49–7071–295565. (K.H.) E-mail klaus.harter@zmbp.uni-tuebingen.de, tel. +49–7071–2972605, fax +49–7071–293287.

doi: 10.1093/mp/sss126

Received 21 September 2012; accepted 31 October 2012

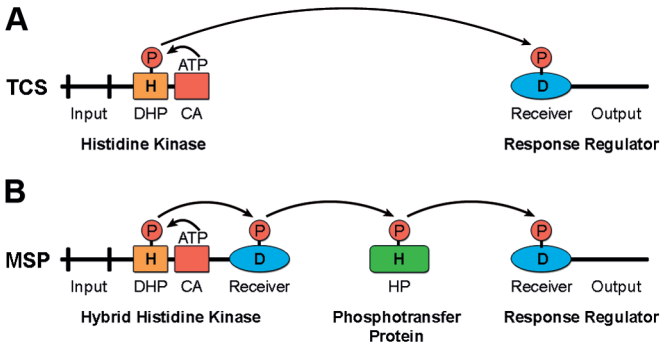

Figure 1. Schematic Overview of His–Asp Phosphorelay Systems.

(A) Overview of phosphorelay signaling in two-component systems (TCS). Signal perception induces ATP-binding to the catalytic domain (CA), followed by phosphorylation of a conserved histidine residue in the dimerization and phosphotransfer domain (DHP) of the kinase. Finally, the phosphoryl group (P) is relayed onto a conserved aspartic acid residue in the receiver domain of the cognate response regulator (RR).

(B) Overview of phosphorelay signaling in multi-step phosphorelay systems (MSP). MSP signaling involves a hybrid histidine kinase containing a fused receiver domain and an additional AHP protein.

1999; Tran et al., 2007), hormonal signaling (AHK2–4, ETR1, ERS1) (Hass et al., 2004; Nishimura et al., 2004), and megagametogenesis (CKI1) (Pischke et al., 2002; Deng et al., 2010; Horak et al., 2011). Recently, AHK5 was found to positively regulate salt sensitivity and to contribute to resistance to the bacterium *Pseudomonas syringae* and the fungal pathogen *Botrytis cinerea* (Pham et al., 2012). Additionally, AHK5 mediates stomatal responses to exogenous and endogenous signals such as flagellin, which alter both the intracellular level of reactive oxygen species (ROS) and redox homeostasis (Desikan et al., 2008).

AHPs can interact with several upstream HKs and downstream response regulators (ARRs) (Mahonen et al., 2006). Thus, AHPs confer an enormous flexibility onto the MSP system and, for this reason, they play a central role in signal integration (Horak et al., 2011). For example, in the well-investigated plant cytokinin signaling pathway, AHP1–3 and AHP5 act as redundant positive regulators, whereas AHP4 is thought to function in some cases as a negative regulator (Hutchison et al., 2006). Expression of the pseudo-AHP AHP6 was also shown to be induced by cytokinin (Mahonen et al., 2006). AHP6 inhibited signal transduction from a phosphorylated HK but also competed with AHP1 and blocked the cognate ARR, thus acting as a negative regulator (Mahonen et al., 2006). Notably, pseudo–AHP homologs also exist in other plant species, indicating that such negative regulation of the plant MSP system is commonly used (Mahonen et al., 2006). However, neither the structural basis of AHP-recognition nor to what extent this redundancy in AHP-recognition manifests in a similar mode of recognition have been defined.

Since the MSP system plays a central role in plant signaling, many MSP proteins have been functionally investigated in *Arabidopsis* and other plants. However, only a small number of components of the *Arabidopsis* MSP system have been structurally characterized to date: (1) the AHK4 sensor domain in complex with its cytokinin ligand (Hothorn et al., 2011), (2) the receiver domains of ETR1 ($ETR1_{RD}$) (Muller-Dieckmann et al., 1999) and CKI1 ($CKI1_{RD}$) (Pekarova et al., 2011), and (3) the DNA-binding motif of ARR10 (Hosoda et al., 2002). Importantly, structural information about the mode of transfer of the phosphoryl group from the HK to the AHP is lacking.

In order to define the specificity and mode of recognition for this interaction, we have determined the first crystal structure of a plant HK_{RD}, $AHK5_{RD}$, in complex with one of its cognate HPs, AHP1. This structure identifies key contacts, including a conserved AHP surface that is recognized by AHK5. In combination with *in vitro* binding studies and functional complementation assays, our analysis provides an excellent basis for understanding the molecular recognition events governing MSP in plants.

RESULTS

Crystal Structure of $AHK5_{RD}$ in Complex with AHP1

In order to reveal the structural features that underlie the AHK5–AHP interaction, we determined the crystal structure of a representative complex, between the receiver domain of the kinase, $AHK5_{RD}$, and AHP1 at 1.95-Å resolution (Table 1).

AHP1 folds into an elongated, all α-helical bundle formed by six α-helices (Figure 2A). Four of these (α3, α4, α5, and α6) form a central core structure that is augmented by helices α1 and α2. A search for structural homologs using the DALI server (Holm and Rosenstrom, 2010) revealed that the overall fold of AHP1 is most similar to those of the plant HPs MtHpt1 (pdb-code: 3US6), ZmHP2 (pdb-code: 1WN0 (Sugawara et al., 2005)), and AK104879 (pdb-code: 1YVI), with Z-scores of 23.9, 21.2, and 21.1 and r.m.s.d. values of 1.1 Å, 1.2 Å, and 1.1 Å, respectively. The conserved residue carrying the phosphoryl group, His79, protrudes from one side of the bundle, facing towards its interaction partner (Figure 2A and B).

The $AHK5_{RD}$ structure resembles other receiver domains of TCSs from different organisms (Volz and Matsumura, 1991; Volz, 1993; Muller-Dieckmann et al., 1999; Xu et al., 2003; Pekarova et al., 2011), all of which belong to the CheY-like protein superfamily (Wilson et al., 2009). Members of this superfamily possess a central five-stranded parallel β-sheet that is surrounded by five helices, giving rise to an $(\alpha/\beta)_5$ topology. However, $AHK5_{RD}$ contains six instead of the usually observed five α-helices. The extra helix α4 lies on the surface opposite to the binding interface, a region in which $AHK5_{RD}$ carries about 25 additional residues compared to $ETR1_{RD}$ and $CKI1_{RD}$ (Muller-Dieckmann et al., 1999; Pekarova et al., 2011), or the yeast osmoregulator $SLN1_{RD}$ (Posas et al., 1996; Xu et al., 2003; Zhao et al., 2008). The residue carrying the phosphoryl group, Asp828, is located at the tip of strand β3 (Figure 2A and B, and Supplemental Figure 1). Loop L5, which connects strand

Table 1. Data Collection and Refinement Statistics.

$AHK5_{RD}$–AHP1	Native	NaI-soak		
	PX III (SLS)	ID 14–4 (ESRF)		
Data collection				
Wavelength (Å)	1.0000	1.2000		
Space group	P 2_13	P 2_13		
Cell dimensions				
a=b=c (Å)	106.8	106.6		
$\alpha=\beta=\gamma$ (°)	90.00	90.00		
Resolution (Å)	47.76–1.95 (2.00–1.95)	47.67–2.90 (2.98–2.90)		
R_{meas} (%)	6.5 / (61.0)	11.8 (67.4)		
$I/\sigma I$	16.7 (2.8)	48.7 (11.5)		
Completeness (%)	99.9 (99.4)	100.0 (100.0)		
Redundancy	13.4 (11.2)	76.6 (76.7)		
Wilson factor (Å²)	40.6	52.4		
Phasing				
Sites per ASU	–	2		
Resolution bin (Å)	–	37.76–8.30	8.30–5.94	5.94–4.87	4.87–4.23
Figure of merit (%)	–	74.3	60.5	50.6	37.5
Phasing Power, isomorphous	–	1.74	1.38	(0.82)	(0.49)
Phasing Power, anomalous	–	1.87	1.64	1.21	(0.92)
R_{cullis} (%), isomorphous	–	75.4	(86.8)	(88.9)	(94.5)
R_{cullis} (%), anomalous	–	55.8	61.8	75.6	(82.0)
Refinement					
Resolution (Å)	1.95				
No. of unique reflections	28 327 (2063)				
R_{work}/R_{free} (%)	17.2/19.7				
No. of non-H atoms					
$AHK5_{RD}$/AHP1/water/Mg^{2+}	1052/1229/181/1				
B-factors (*Å²*)					
$AHK5_{RD}$/AHP1/water/Mg^{2+}	36.3/43.5/42.4/29.2				
R.m.s. deviations					
Bond lengths (Å)	0.013				
Bond angles (°)	1.248				

Phasing statistics were calculated before density modification as defined in autoSHARP (Vonrhein et al., 2007).
$R_{meas} = \Sigma_h (N/N-1)^{1/2} \Sigma^N_{i=1}|I_{h,i}-\bar{I}_h|/\Sigma_h \Sigma^N_{i=1} I_{h,i}$, where *N* is the redundancy.
$R_{work} = \Sigma_{h \notin free} |F_{obs}-kF_{calc}|/\Sigma_{h \notin free}$ Fobs
$R_{free} = \Sigma_{h \in free} |F_{obs}-kF_{calc}|/\Sigma_{h \in free}$ Fobs

β3 to helix α3, bears a γ-turn motif (residues 831–836) that is structurally conserved among receiver domains (Volz, 1993). The γ-turn is inserted into a longer loop, termed the γ-loop (Volz, 1993), which has the sequence MPVLDG in $AHK5_{RD}$. Loop L8, which connects strand β4 and helix α5 and can adopt different conformations depending on Mg^{2+}-binding and the state of phosphorylation (Guhaniyogi et al., 2006; Zhao et al., 2008), lies in close proximity to the γ-loop.

Architecture of the $AHK5_{RD}$–AHP1 Interface

$AHK5_{RD}$ and AHP1 form a 1:1 complex in solution, and this complex is also found in the crystals. Contacts between $AHK5_{RD}$ and AHP1 shield a surface of 787 Å² from solvent (Krissinel and Henrick, 2007). This surface is small in size compared with the contact areas of other protein–protein complexes (Bahadur and Zacharias, 2008). In the complex, $AHK5_{RD}$-helix α1 is inserted into a groove formed by three AHP1-helices (α2, α3, and α4) and one adjacent loop (L2) (Figure 2B). This results in a rectangular contact area, with one half primarily involving hydrophobic contacts and the other half mainly consisting of polar residues that line a ridge on AHP1 and engage in hydrogen bond formation with AHK5 (Figure 2C and Supplemental Table 1). The contact area is largely contiguous and devoid of solvent. The histidine carrying the

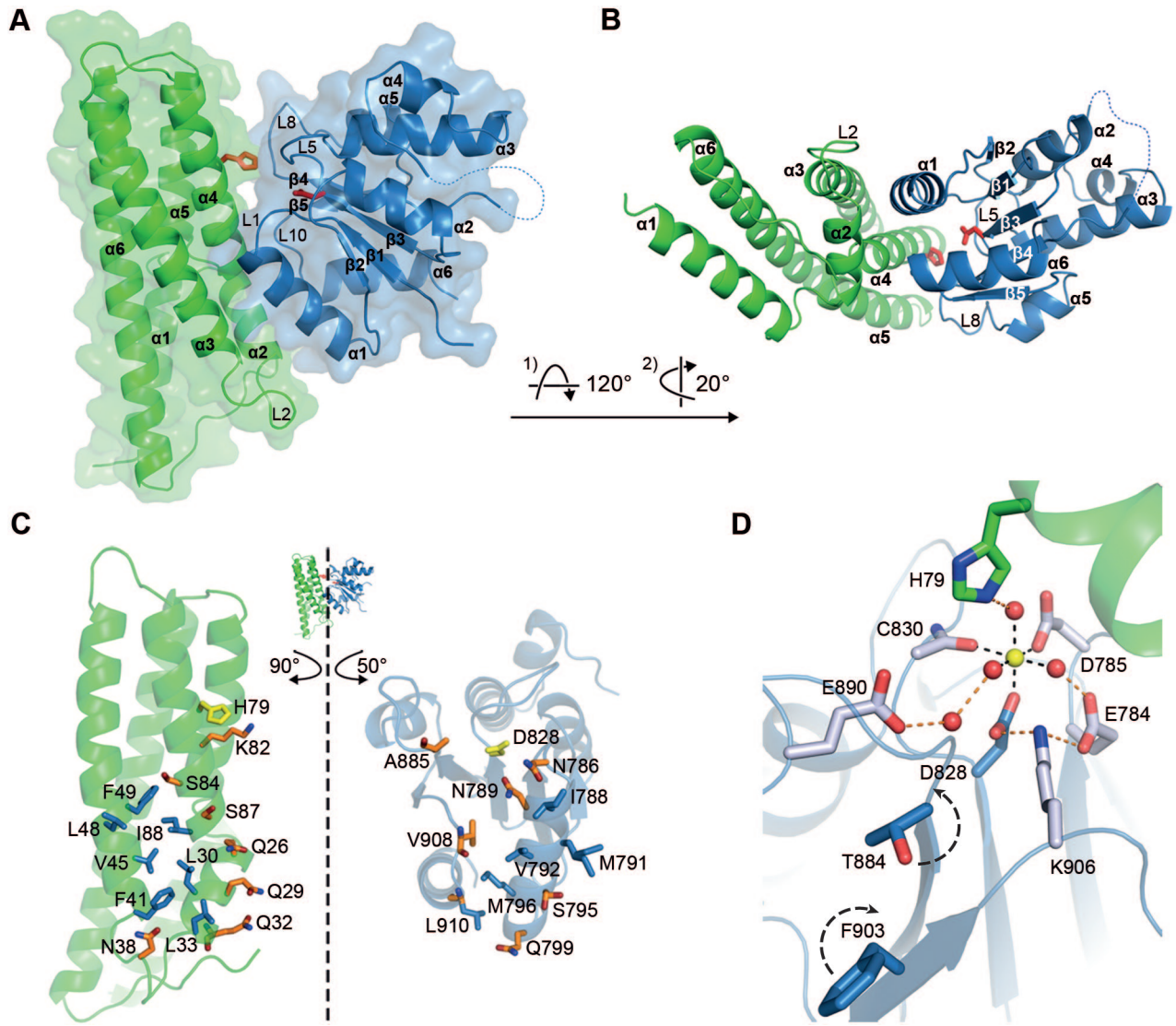

Figure 2. Structure of the AHK5$_{RD}$–AHP1 Complex.
AHP1 and AHK5$_{RD}$ are shown in all figures as ribbon tracings and colored green and blue, respectively.
(A, B) Overall structure. The conserved Asp and His residues are drawn as red sticks. In (A), AHP1 and AHK5$_{RD}$ are additionally shown in surface representation.
(C) Binding interface of AHP1 (left) and AHK5$_{RD}$ (right). Contacting residues are shown in stick representation. Residues involved in hydrogen bonds are highlighted orange, with oxygens in red and nitrogens in blue. Hydrophobic patch residues are highlighted blue. The histidine and aspartic residues that donate and accept the phosphoryl group are shown as yellow sticks.
(D) Mg^{2+}-coordination in the phosphorylation site. The Mg^{2+}-ion (yellow) and water molecules (red) are shown as spheres. The octahedral coordination geometry of Mg^{2+} is indicated by black broken lines; further hydrogen bonds of Mg^{2+}-coordinating water molecules are indicated by orange broken lines. The phosphoryl-transferring histidine and aspartic acid residues are shown as sticks, with oxygens in red and nitrogens in blue. All other Mg^{2+}-contacting residues are drawn in stick representation with carbons colored light blue. F903 and T884 of AHK5$_{RD}$ are involved in 'F-T'-coupling and are shown as blue sticks; the predicted reorientation upon phosphorylation is indicated by black arrows.

phosphoryl group, His79, lies at the very edge of the contact area, and is mostly accessible to solvent.

Phosphorylation Site of the AHK5$_{RD}$–AHP1 Complex

A Mg^{2+}-ion is located at the edge of the AHP1–AHK5$_{RD}$ interface, in close vicinity to the two residues that are involved in phosphoryl transfer (His79 and Asp828). The magnesium ion is located in an acidic pocket on AHK5$_{RD}$, and it is held in place by three amino acids (Asp828, Cys830, and Asp785) and three water molecules that together form an almost perfect octahedral coordination sphere (Figure 2D and Supplemental Figure 2). The positions of the three water molecules are secured by either a second hydrogen bond to neighboring AHK5$_{RD}$-residues or by interactions with an adjacent water molecule. One such water molecule forms a bridge to His79 of AHP1, which would accept the phosphoryl group during the phosphotransfer process. As

Table 2. Summary of the SPR Measurements and the Calculated Affinities.

	AHP1	AHP2	AHP3
Concentration range (μM)	0.05–110.60	0.08–39.30	0.05–96.40
R_{max} (RU)	69.3	26.5	82.4
K_D (μM)	4.1	2.7	4.4
X^2 $(RU)^2$	0.55	0.21	0.88

the complex structure does not include a phosphoryl group, the side chain of Lys906 coordinates the carboxyl oxygens of Asp828 and Glu784. Moreover, the side chain of Thr884 points away from the phosphorylation site, so that, in turn, Phe903 is forced to occupy a solvent-exposed 'out'-conformation. We expect these two interactions to be altered in a setting in which Asp828 would be phosphorylated. As five of the six Mg^{2+}-coordination sites and all direct protein-Mg^{2+} contacts are provided by $AHK5_{RD}$, conformational changes upon Mg^{2+}-binding would primarily affect residues in $AHK5_{RD}$.

Affinity Measurements Using Surface Plasmon Resonance

In order to obtain quantitative data for the affinity and kinetics of the interaction of $AHK5_{RD}$ with AHPs, surface plasmon resonance (SPR) experiments were conducted. We immobilized anti-His_6 monoclonal antibodies on CM5 sensor chips, and bound purified $AHK5_{RD}$ non-covalently via its hexahistidine tag to the antibodies. Then we applied purified, untagged full-length AHP1, AHP2, and AHP3 proteins as analytes to the chip. All three proteins bound to $AHK5_{RD}$ following a 1:1 binding model (Figure 3A–F), consistent with gel filtration experiments (unpublished observations). Interestingly, the dissociation constant (K_D) values for all three proteins are highly similar *in vitro*, ranging from 2.7 to 4.4 μM (Figure 3D–F and Table 2). This indicates that the three AHPs interact with $AHK5_{RD}$ using a conserved strategy. The observed intermediate affinity is in agreement with the size of the contact area in the structure of the $AHK5_{RD}$–AHP1 complex. The association and dissociation of all $AHK5_{RD}$–AHP complexes occurred very rapidly *in vitro* and prevented us from determining association/dissociation rate constants. Rapid association and dissociation of these complexes would be required for their physiological function of quickly transferring a phosphoryl group. The remaining three AHPs (AHP4, AHP5, and AHP6) could not be purified in soluble forms and were therefore not investigated.

Specificity of $AHK5_{RD}$ for Different AHPs

To substantiate the biochemical data and to determine whether the six known phosphotransfer proteins (AHP1–6) from *A. thaliana* differ in their binding specificity for $AHK5_{RD}$ *in planta*, we conducted bimolecular fluorescence complementation (BiFC) experiments (Walter et al., 2004; Schuetze et al., 2009). Therefore, the C-terminal YFP fragment (*YFP-C*) was fused to the N-terminus of $AHK5_{RD}$ generating *YFP-C::*$AHK5_{RD}$, whereas the N-terminal YFP fragment (*YFP–N*) was fused to the C-terminus of the different *AHPs* generating the *AHP::YFP–N* constructs. The combined expression of the BiFC constructs in transiently transformed tobacco epidermal leaf cells revealed an interaction of $AHK5_{RD}$ with AHP1, AHP2, AHP3, AHP5, and AHP6 but no interaction with AHP4 (Figure 3G and H). Although the expression levels of investigated AHP fusion proteins are similar by visual inspection, they are not identical, and thus the intensity of the fluorescence signals provides only *semi*-quantitative information about differences in their interaction with $AHK5_{RD}$. On this basis, the strongest interaction with $AHK5_{RD}$ was found for AHP2 and AHP5 followed by AHP6, AHP3, and AHP1.

Our SPR experiments show that the affinities of the investigated $AHK5_{RD}$–AHP interactions are highly similar. However, the SPR experiments were conducted *in vitro*, and thus they give only information about this particular interaction. In contrast, the BiFC assays were carried out *in planta*, and thus the $AHK5_{RD}$–AHP interaction is analyzed in the context of complete signaling cascades. Therefore, the interaction of AHPs with cognate RRs might modify the investigated $AHK5_{RD}$–AHP interaction. Furthermore, in order to maintain specificity during signaling, there are most likely (yet unknown) additional molecular mechanisms which affect and interfere with the $AHK5_{RD}$–AHP interaction *in planta*.

Conservation of AHP Binding Interface

Our SPR experiments demonstrate that recombinantly expressed, full-length AHP1, AHP2, and AHP3 engage $AHK5_{RD}$ with similar affinity. On the other hand, BiFC experiments identify differences among AHPs with regard to their *in vivo* interactions with $AHK5_{RD}$. In order to reconcile these findings with the structure of the complex, we examined the level of conservation of residues at the interface among all six AHPs proteins (Figure 4A and Supplemental Figure 3). The underlying assumptions for our analysis are (1) that the six AHPs are likely to have similar folds, which is a reasonable assumption given their high level of sequence identity (ranging from 36% to 81%), and (2) that they all form similar contacts with AHK5, which, at least in the cases of AHP1–AHP3, is likely since all three proteins have similar dissociation constants for their interactions with $AHK5_{RD}$.

Examination of the seven residues mediating hydrogen bonds showed that three of these residues, Thr26, Ser84, and Ser87, are strictly conserved in all AHPs. In addition, the hydrogen bonds mediated by Gln29 and Gln32 are functionally conserved. Although Gln29 is sometimes replaced with a glutamic acid, both residues can serve as a hydrogen acceptor for the backbone nitrogen of Leu910 of $AHK5_{RD}$. The hydrogen bond involving Gln32 is mediated by its backbone oxygen, and similar interactions can also be made by residues at the corresponding positions in the other AHPs. Thus, residues Thr26, Gln29, Gln32, Ser84, and Ser87 form a prominent docking ridge that should exist in very similar form in all six

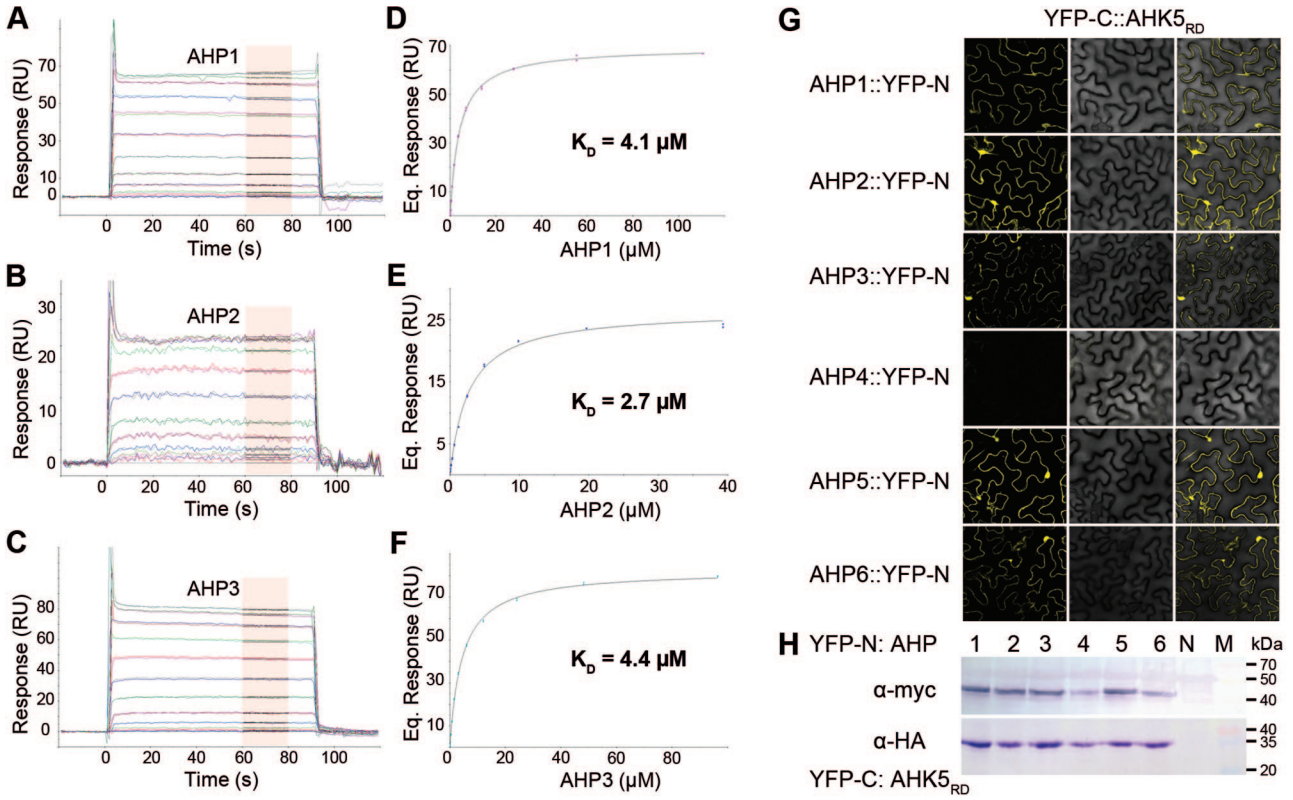

Figure 3. Interaction Studies of $AHK5_{RD}$–AHP Complexes.
(A–F) Surface plasmon resonance analyses. Double-referenced sensorgrams of the $AHK5_{RD}$-interaction with AHP1 (A), AHP2 (B), and AHP3 (C) are shown. Red boxes show the range used for calculation of averaged equilibrium response values. (D–F) Averaged equilibrium response values plotted against the injected concentrations of AHP1 (D), AHP2 (E), and AHP3 (F).
(G, H) BiFC analysis of the $AHK5_{RD}$-interaction with AHP1-6 *in planta*. (G) Confocal images of tobacco leaf cells expressing both the YFP-C::$AHK5_{RD}$ fusion protein and respective AHP::YFP–N fusion proteins. The fluorescence images (left column), corresponding bright field images (middle column), and respective overlays (right column) are shown. (H) Western blot analysis of protein extracts from tobacco leaves used for BiFC analysis. Immunodetection was carried out by α-myc antibody for AHP::YFP–N fusion proteins and by α-HA antibody for YFP::C–$AHK5_{RD}$. N, protein extract from non-transformed tobacco leaves; M, protein size standard.

AHPs, allowing them to engage $AHK5_{RD}$ using a conserved hydrogen binding pattern. Another hydrogen bond provided by the ε-amino-group of Lys82 is conserved in AHP1–5, but, as the corresponding residue in AHP6 is a valine, this hydrogen bond would not be present in an $AHK5_{RD}$–AHP6 complex. The seventh hydrogen bond is formed by the side chain amide group of Asn38 in AHP1. This asparagine is only conserved in AHP4, but is replaced with either a serine or a threonine in AHP2, AHP3, AHP5, and AHP6. Serines and threonines would not be able to form similar direct hydrogen bonds, but they would also not clash with $AHK5_{RD}$, and perhaps even allow for the formation of water-mediated, indirect hydrogen bonds.

Inspection of the hydrophobic patch revealed seven residues in AHP1 that form non-polar contacts with $AHK5_{RD}$: Leu30, Leu33, Phe41, Val45, Leu48, Phe49, and Ile88. Four of these residues (Leu30, Leu33, Phe41, and Val45) are strictly conserved in AHP1–6; the remaining three are functionally conserved with respect to their surface complementarity. Leu48 is conserved in AHP1–5, but is an isoleucine residue in AHP6. Phe49 is replaced with tyrosines in AHP4 and AHP6, but the tyrosine hydroxyl group would be expected to point towards the AHP1-core. Ile88 is sometimes replaced with a similarly hydrophobic valine. Due to these conservative mutations, the primary characteristics of the hydrophobic patch should be maintained throughout all AHPs.

DISCUSSION

Two-component signal–response systems serve as the foundation of intracellular communication in many biological systems. Our structure-function analysis provides an understanding, at the atomic level and with quantitative affinity data, of the specificity that underlies such signaling systems in plants. The $AHK5_{RD}$–AHP1 crystal structure demonstrates that the contact area between the two proteins consists of two prominent features: the hydrogen bond docking ridge and the hydrophobic patch. Both features are highly conserved in all AHP proteins, allowing conserved modes of engagement for all known AHP proteins in $AHK5_{RD}$–AHP complexes. These findings are in good agreement with our SPR experiments,

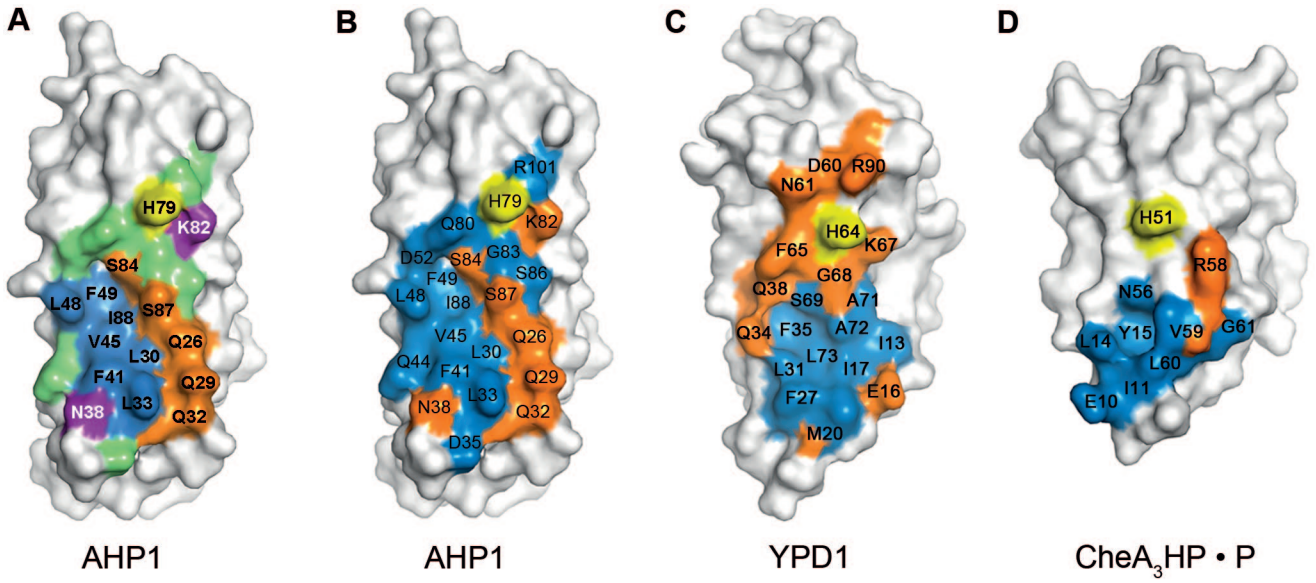

Figure 4. Interface Architectures of AHP1, YPD1, and $CheA_3HP$.
(A) Conservation of interface residues in AHP proteins based on the $AHK5_{RD}$–AHP1 complex structure. AHP1 is shown in surface representation with non-interface residues colored white. The phosphorylation site is shown in yellow. Functionally conserved hydrogen bonds in AHP1–6 are colored orange; not functionally conserved hydrogen bonds are colored purple. All hydrophobic residues in the interface are conserved and colored blue. All other AHP interface residues within a radius of 5 Å to $AHK5_{RD}$ are shown in green and they are all conserved in terms of their surface complementarity with respect to the interface.
(B–D) Comparison of interface architectures of AHP1 (B), YPD1 (C), and $CheA_3HP$ (D) based on the $AHK5_{RD}$–AHP1 complex structure, the phosphorylated $SLN1_{RD}$–YPD1 complex structure, and the phosphorylated $CheY_6$–$CheA_3HP$ ($CheY_6$–$CheA_3HP$•P) complex structure, respectively. All three proteins are shown in surface representation, with non-interface residues colored white. The phosphorylation site is shown in yellow. Amino acids that contribute hydrogen bonds are colored orange; residues contributing hydrophobic interactions with their respective complex partner are colored blue.

which show nearly identical binding affinities of $AHK5_{RD}$ for AHP1, AHP2, and AHP3 *in vitro*. BiFC experiments indicate that at least five of six AHP proteins can engage $AHK5_{RD}$ *in planta*, and we expect that they do so using the same contact surface and with similar affinities. The specificity of signal transduction from AHK5 to an AHP does therefore not depend on unique contacts with any one AHP, and more likely involves spatial distribution and local concentrations of an AHP. Interestingly, sequence comparison of all AHPs with other structurally characterized plant HPs from maize, rice, and clover and with numerous protein sequences found by a search in the UniProtKB database (Altschul et al., 1997) revealed that the residues involved in $AHK5_{RD}$ recognition are highly conserved across plant species (Supplemental Figure 4). Thus, the mode of AHP1 recognition serves as a template for understanding interactions in plant MSP systems.

Comparison with Non-Plant Signaling Complex Structures

Although several receiver domains ((Muller-Dieckmann et al., 1999; Pekarova et al., 2011) and pdb-codes: 3LUA, 3C97, 3GT7) and phosphotransfer proteins ((Sugawara et al., 2005) and pdb-codes: 1YVI, 3US6) have been crystallized in isolation, only two other structures of complexes between two such signal proteins have been solved to date: the osmosensing $SLN1_{RD}$–YPD1 complex from *S. cerevisiae* (Posas et al., 1996; Xu et al., 2003; Zhao et al., 2008) and the $CheY_6$–$CheA_3HP$ complex from *R. sphaeroides* (Bell et al., 2010). The receiver domains in these complexes share little sequence identity (ranging from 17% to 28%) but nevertheless possess the same fold (Figure 5A). $AHK5_{RD}$ has especially high structural similarity to $SLN1_{RD}$ (Xu et al., 2003; Zhao et al., 2008) (r.m.s.d. values of 1.0 Å), while its structural similarity to $CheY_6$ is less pronounced (r.m.s.d. value of 2.0 Å).

Likewise, the three HP proteins AHP1, YPD1, and $CheA_3HP$ share the four-helix bundle core, which is assembled from helices α3, α4, α5, and α6 in AHP1 (Figure 2A and B). In contrast to both eukaryotic HPs, however, the AHP1-helix α2, which contributes three hydrogen bonds and two hydrophobic residues to the interface, is absent in prokaryotic $CheA_3HP$ (Figure 5B). Thus, the $CheY_6$–$CheA_3HP$ contact area is decreased by about 200 Å² compared to the $AHK5_{RD}$–AHP1 complex, consistent with a significant reduction in affinity of $CheY_6$ for $CheA_3HP$ (K_D = 218 μM) (Bell et al., 2010). The observed differences in affinity and structure point to a central role of the AHP1-helix α2 in interface formation. As the α2 helix is also present in the $SLN1_{RD}$–YPD1 complex, this extra helix might be a common feature within the more highly evolved eukaryotic MSP systems. In addition to enlarging the contact surface, the α2 helix could also play a role in modulating interactions with selected ligands.

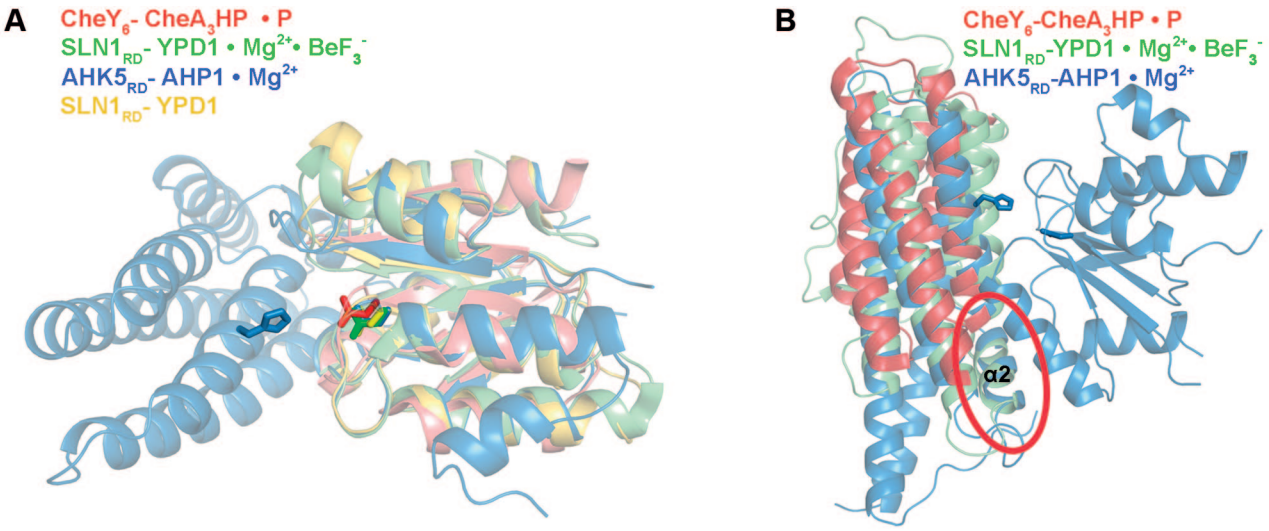

Figure 5. Superposition of $AHK5_{RD}$–AHP1 Complex with Structurally Similar Complexes.
(A) Superposition of the $AHK5_{RD}$–AHP1 complex (blue) with the phosphorylated $SLN1_{RD}$–YPD1 complex (green), the apo $SLN1_{RD}$–YPD1 complex (yellow), and the phosphorylated $CheY_6$–$CheA_3HP$ complex (red) by superposing the receiver domains only. The phosphoryl-transferring histidine and aspartic acid residues are highlighted in stick representation.
(B) Comparison of the $AHK5_{RD}$–AHP1 complex (blue) with the phosphorylated $SLN1_{RD}$–YPD1 complex (green) and the phosphorylated $CheY_6$–$CheA_3HP$ complex (red). The superpositions are based on the receiver domains only and, for reasons of clarity, only $AHK5_{RD}$–AHP1 is shown as a complete complex, with phosphoryl-transferring histidine and aspartic acid residues highlighted in stick representation. The extra helix of AHP1 and YPD1 is indicated.

The eukaryotic $AHK5_{RD}$–AHP1 and $SLN1_{RD}$–YPD1 complexes are well conserved with respect to the structures of their components and the overall organization of the complexes, including the sizes of the buried surface areas. It is therefore surprising that the two complexes use entirely unrelated strategies for complex formation. The amino acid residues contributing to hydrogen bonds and hydrophobic interactions are distributed differently across the interfaces (Figure 4B and C). In YPD1, nine of 11 hydrogen bond-forming residues cluster around the phosphorylatable histidine in a pattern that resembles the letter Y (Figure 4C). This Y-shaped distribution in YPD1 contrasts with the hydrogen bond docking ridge observed in AHP1 (Figure 4B). In addition, the interface in the $SLN1_{RD}$–YPD1 complex contains a hydrophobic patch that is enclosed by the hydrogen bond forming residues. Molecular recognition of $CheA_3HP$ is again organized in a different manner. Only two hydrogen bonds are present in the phosphorylated $CheY_6$–$CheA_3HP$ complex, and these residues bridge to a patch of residues involved in hydrophobic interactions (Figure 4D).

Structural Changes upon Phosphorylation of the $AHK5_{RD}$–AHP1 Complex

The impact of phosphorylation on the structures of HK or RR has been most extensively investigated for the bacterial chemotaxis protein CheY (Guhaniyogi et al., 2006). These studies revealed that phosphorylation of the conserved aspartate residue induces local and medium-distance changes of conserved amino acids within CheY, known as 'Y–T' coupling. In eukaryotes, this 'Y–T'-coupling was also observed in the crystal structures of the $SLN1_{RD}$–YPD1 complex in both phosphorylation states (Xu et al., 2003; Zhao et al., 2008).

Our efforts to produce complexes of $AHK5_{RD}$–AHP1 with ions such as BeF_3^- or AlF_4^-, two compounds that mimic the phosphorylated state of $AHK5_{RD}$, were not successful. We furthermore attempted to produce a complex with AlF_3, which mimics the transition state of the phosphoryltransfer reaction (Xu et al., 1997), also without success. In order to predict the structural changes that would be triggered by phosphorylation of Asp828, we therefore have to rely on comparisons with the homologous $SLN1_{RD}$–YPD1 structure (Xu et al., 2003; Zhao et al., 2008). All phosphoryl-coordinating residues are conserved between $AHK5_{RD}$ and $SLN1_{RD}$, so that the phosphoryl group would likely be coordinated by the same residues. Therefore, we predict that, upon phosphorylation, the side chain of Thr884 rotates into the active site to coordinate the first oxygen of the phosphoryl group. This would in turn make room for the Phe903 side chain, allowing it to flip from a solvent-exposed 'out'-conformation into a buried 'in'-conformation (Figure 2D).

The coordinated movement of this proposed 'F-T'-coupling in phosphorylated $AHK5_{RD}$ would result in optimal charge compensation and in a moderate rearrangement of the Thr884-containing loop L8, similar to what has been seen in the $SLN1_{RD}$–YPD1 complex structures.

Structural Implications for the Phosphoryl-Transfer Mechanism

In order to enable a phosphoryl transfer from a histidine to an aspartic acid in TCSs, two mechanisms have to be considered

(Knowles, 1980): an associative S_N2-mechanism and a dissociative S_N1-mechanism. Recently characterized TCS-complex structures favor the associative S_N2-mechanism (Varughese et al., 2006; Zhao et al., 2008). Both mechanisms require a linear O_{Asp}–P–N_{His} geometry, but they differ in their respective O_{Asp}–P and P–N_{His} distances. The dissociative mechanism involves the formation of a monomeric metaphosphate, which in turn requires that the O_{Asp}–P and P–N_{His} atom radii must not intersect, resulting in O_{Asp}–P and P–N_{His} distances of longer than 3.3 Å. In contrast, the associative mechanism requires a penta-coordinated transition state, with O_{Asp}–P and P–N_{His} distances shorter than 3.3 Å (Varughese et al., 2006).

In the non-phosphorylated $AHK5_{RD}$–AHP1 complex, the phosphoryl-donating oxygen atom of Asp828 (O_{D828}) and the phosphoryl-accepting nitrogen atom of His79 (N_{H79}) are 6.4 Å apart. The O–P–N angle α_{OPN} is 155°, indicating that O_{D828} and N_{H79} are not yet in an optimal in-line position for phosphoryl transfer. Upon phosphorylation, the Thr884-containing loop L8 in $AHK5_{RD}$ needs to be rearranged to accommodate the phosphoryl group and also to compensate for its charge. In order to provide a perfectly linear O–P–N geometry for phosphoryl transfer, His79 of AHP1 has to move about 2–3 Å closer towards $AHK5_{RD}$. One could envision that the AHK5-helix α1 serves as a hinge that immobilizes the bottom part of the complex but allows the upper part of AHP1 and especially AHP1-helix α4 to tilt towards $AHK5_{RD}$ for proper reception of the phosphoryl group. As a consequence, the O_{D828}–N_{H79} distance would decrease to about 5 Å, with optimal in-line geometry. In this scenario, the geometry in the $AHK5_{RD}$–AHP1 complex would support an associative mechanism.

Commutability of AHP Proteins

The structure-function analysis presented here extends our understanding of the parameters that underlie the specificity and affinity of plant MSP complexes. AHK5 is able to bind to a range of AHPs in a similar manner, consistent with the general commutability of AHP proteins shown for other HKs from *A. thaliana* (Mahonen et al., 2006; Deng et al., 2010). Therefore, our biochemical and structural findings are in good agreement with the hypothesis that AHP proteins can act as functionally redundant signaling hubs in the MSP system in *A. thaliana* (Schaller et al., 2008; Horak et al., 2011).

Notably, examination of conserved residues among all AHP proteins suggests that AHK5 can also bind the pseudo–AHP AHP6, but with somewhat weaker affinity. Thus, for AHK5, engagement of AHP6 might also contribute to fine-tuned signaling by blocking the MSP to certain extents depending on the cytoplasmic concentration of AHP6 relative to the concentration of canonical AHPs. Our structure and sequence analysis does not provide clues as to why AHP4 would not be able to engage $AHK5_{RD}$, nor why there are differences in the *in planta* interaction strength. We therefore think it possible that additional molecular mechanisms that interfere with the $AHK5_{RD}$–AHP complex formation, perhaps in order to fine-tune or regulate the interaction, are involved in some cases. Such mechanisms might provide additional specificity in the $AHK5_{RD}$–AHP interaction and, therefore, in MSP signaling processes in general. In combination with the established SPR protocol, the observed interactions in the $AHK5_{RD}$–AHP complex structure provide a firm platform for ongoing efforts to reengineer and evaluate MSP signaling cascades.

METHODS

Cloning

DNA coding for amino acid residues 774–922 of AHK5 ($AHK5_{RD}$) was cloned into the pET28a plasmid (Novagen) using *Nde*I and *BamH*I restriction sites. The plasmid contained an N-terminal His_6-tag and a thrombin protease cleavage site. Full-length AHP1–6 genes were cloned into pGEX–6PI (Amersham Biosciences) plasmids using *BamH*I and *Xho*I restriction sites. This plasmid contained an N-terminal GST-tag and a PreScission protease cleavage site.

Protein Expression and Purification

$AHK5_{RD}$ was expressed in *E. coli* Rosetta2 (DE3) cells (Novagen). Cells were grown to OD_{595} = 0.5 at 37°C, and the temperature was then reduced to 20°C. Gene expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cell pellets were harvested 17 h after induction. The pellet from a 2-L culture was re-suspended in 20 mM Tris (pH 8.4), 500 mM NaCl, and 20 mM imidazole. Cells were then lysed using an EmulsiFlex homogenizer (Avestin). After centrifugation (26 000 *g*, 15 min, 4°C), the supernatant was applied onto a 1-ml HisTrap™ HP column (GE Healthcare), followed by protein elution with 250 mM imidazole. $AHK5_{RD}$-containing fractions were purified over a Superdex 75 column (GE Healthcare) using 20 mM Tris (pH 8.4), 100 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, and 5% glycerol. Fractions containing monomeric $AHK5_{RD}$ were pooled, and the His-tag was then cleaved through a 40-h incubation with 15 U Thrombin protease/mg protein at 4°C. Cleaved $AHK5_{RD}$ was first applied onto a 1-ml HisTrap™ HP and then re-purified over a Superdex 75 column (GE Healthcare) using 20 mM Tris (pH 8.4), 100 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP and 5% (v/v) glycerol.

Expression of AHP proteins was carried out as described for $AHK5_{RD}$. In each case, the pellet from a 4-L culture was re-suspended in 50 mM Tris (pH 8.4), 150 mM NaCl, and 5% (v/v) glycerol. Cells were then lysed by EmulsiFlex homogenizer (Avestin). After centrifugation (26 000 *g*, 15 min, 4°C), the supernatant was applied onto a 5-ml GSTrap™ FF column (GE Healthcare) followed by protein elution with 10 mM reduced glutathione. The GST-tag was then cleaved through incubation for 16 h with 1 U PreScission protease/mg protein at 4°C. AHPs were further purified using a Superdex 75 column (GE Healthcare) in 20 mM Tris (pH 8.4), 100 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, and 5% (v/v) glycerol. For SPR experiments, a gel filtration buffer containing 20 mM

HEPES (pH 7.5), 100 mM NaCl, and 5 mM $MgCl_2$ was used. AHP-containing fractions were applied onto a 5-ml GSTrap™ FF column for re-purification. For crystallization, AHP1 and $AHK5_{RD}$ were mixed in an equimolar ratio and then concentrated to 9.2 mg ml^{-1}.

Crystallization and Structure Determination

Crystals of $AHK5_{RD}$–AHP1 complex were grown at 20°C with the hanging drop method using a reservoir solution of 10% PEG 20 000 (w/v), 20% PEG MME 550 (v/v), 100 mM MES/imidazole (pH 6.5), and 20 mM of each following alcohol: 1,6-hexanediol, 1-butanol, (RS)-1,2-propanediol, 2-propanol, 1,4-butanediol and 1,3-propanediol. Native crystals were frozen in liquid nitrogen and used to collect data at beamline X06DA at SLS (Swiss Light Source, Villigen, Switzerland). For experimental phasing, $AHK5_{RD}$–AHP1 complex crystals were soaked for 8 min in the crystallization condition complemented with 400 mM NaI. Data were collected at beamline ID 14–4 at ESRF (European Synchrotron Radiation Facility, Grenoble, France).

Diffraction data were processed with the XDS-package (Kabsch, 2010). The structure was solved with the autoSHARP pipeline (Vonrhein et al., 2007). For initial automated model building and refinement, the ARP/wARP package (Langer et al., 2008) in CCP4 (Winn et al., 2011) was used. Structural refinement was carried out using alternating rounds of model building in COOT (Emsley et al., 2010) and restrained refinement, including TLS refinement (Painter and Merritt, 2006), with Refmac5 (Murshudov et al., 1997). Simulated annealing was performed with PHENIX (Adams et al., 2010). The final model had excellent geometry. The topology plot was prepared with TopDraw (Bond, 2003); structure figures were prepared with PyMOL (www.pymol.org, last accessed 19 December 2012). Statistics on data collection and refinement are given in Table 1. Protein sequence alignments were performed using ClustalW 2.1 (Goujon et al., 2010). Structural superpositions were calculated by Superpose using the secondary structure matching function in CCP4 (Winn et al., 2011). For O–P–N angle calculation, the unphosphorylated $AHK5_{RD}$–AHP1 complex was superimposed with the phosphorylated $SLN1_{RD}$–YPD1 complex. The position of the P-site of the phosphorylated $SLN1_{RD}$–YPD1 complex was then used for calculating the O–P–N angle in the $AHK5_{RD}$–AHP1 complex. For structural alignment, all mainchain atoms of residues 824–828 in $AHK5_{RD}$ and 1140–1144 in $SLN1_{RD}$ (corresponding to the β-strand bearing at its C-terminal end the phosphorylatable Asp residues) were superimposed using the LSQ algorithm in COOT (Emsley et al., 2010).

Surface Plasmon Resonance

SPR measurements were performed on a Biacore 2000 instrument at 13°C. Using two consecutive flow cells, an anti-His_6 monoclonal antibody (Novagen, N-70796) was covalently bound to a CM5 sensorchip (GE Healthcare) with the amine coupling kit (GE Healthcare). 5000–7000 RU of antibody were immobilized in both the reference and the experimental flow cells. His_6-tagged $AHK5_{RD}$ protein was diluted in running buffer (10 mM HEPES pH = 7.4, 150 mM NaCl, 0.05 % P20) and non-covalently bound to the monoclonal antibodies in the experimental flow cell. For $AHK5_{RD}$-binding experiments, AHP1, AHP2, or AHP3 were serially diluted in running buffer and injected for 90 s in both the reference and the experimental cell at a flow rate of 20 µl min^{-1}.

In the case of AHP1 and AHP3, 12 different concentrations ranging from 0.05 to 111 µM (AHP1) or from 0.05 to 96 µM (AHP3) were measured; 11 of these conditions were run in duplicates. For AHP2, 10 different concentrations ranging from 0.08 to 39 µM were used, and each concentration was measured twice. The data points were fitted using a '1:1 Langmuir isotherm' binding model (BIAevaluation).

To remove $AHK5_{RD}$ after each cycle, 10 µl of regeneration solution (33 mM ethanolamine, 33 mM Na_3PO_4, 33 mM piperazine, 33 mM glycine, 77 mM KSCN, 31 mM $MgCl_2$, 15 mM urea, 31 mM guanidine-HCl in the case of AHP2; 77 mM KSCN, 31 mM $MgCl_2$, 15 mM urea, 31 mM guanidine-HCl, 3 mM EDTA in the cases of AHP1 and AHP3) were injected into the measurement cell. The optimal regeneration solutions were determined through a multivariate cocktail approach (Andersson et al., 1999).

Bimolecular Fluorescence Complementation

BiFC experiments were essentially carried out as described previously (Walter et al., 2004; Schuetze et al., 2009). The cDNAs of AHP1–6 were recombined via LR-reaction into pSPYNE–35S (Horak et al., 2008), and the $AHK5_{RD}$ cDNA was inserted into pSPYCE–35S. All binary vectors were transformed into *Agrobacterium tumefaciens* strain GV3101 pMP90 and infiltrated to tobacco leaves (Horak et al., 2008). The p19 protein from tomato bushy stunt virus cloned in pBIN61 (Voinnet et al., 2000) was used to suppress gene silencing in all infiltration experiments. Abaxial epidermis of infiltrated tobacco leaves was assayed for fluorescence by confocal laser scanning microscopy (CLSM) 2 d post infiltration. Expression of the BiFC fusion proteins was determined by Western blot analysis of the transfected leaf tissues using α-HA and α-myc antibodies (Walter et al., 2004; Schuetze et al., 2009).

Accession Numbers

Atomic coordinates and structure factors have been deposited with the Protein Data Bank (accession code 4EUK).

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online.*

FUNDING

This work was supported by Deutsche Forschungsgemeinschaft (grant HA 2146/16–1) (to K.H. and T.S.) and a student fellowship from the University of Tübingen (to J.B.).

ACKNOWLEDGMENTS

We highly appreciate the assistance of Stefan Schütz and Mayra Lorenz with cloning, expression, and purification of $AHK5_{RD}$ and AHP1–6 proteins. We are also grateful to Johannes Romir for providing the AHP2 gene cloned into the pGEX–6PI plasmid as well as AHP2 protein for initial gel filtration studies. We thank the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS) for beamtime and beamline support. No conflict of interest declared.

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 Crystallographica Section D–Biological Crystallography. **66**, 213–221.

Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389–3402.

Andersson, K., Hamalainen, M., and Malmqvist, M. (1999). Identification and optimization of regeneration conditions for affinity-based biosensor assays: a multivariate cocktail approach. Anal. Chem. **71**, 2475–2481.

Bahadur, R.P., and Zacharias, M. (2008). The interface of protein–protein complexes: analysis of contacts and prediction of interactions. Cellular Mol. Life Sci. **65**, 1059–1072.

Bell, C.H., Porter, S.L., Strawson, A., Stuart, D.I., and Armitage, J.P. (2010). Using structural information to change the phosphotransfer specificity of a two-component chemotaxis signalling complex. Plos Biol. **8**, e1000306.

Bond, C.S. (2003). TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics. **19**, 311–312.

Capra, E.J., Perchuk, B.S., Lubin, E.A., Ashenberg, O., Skerker, J.M., and Laub, M.T. (2010). Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. Plos Genetics. Plos Genetics, **6**, doi: e1001220.

Deng, Y., Dong, H., Mu, J., Ren, B., Zheng, B., Ji, Z., Yang, W.-C., Liang, Y., and Zuo, J. (2010). *Arabidopsis* histidine kinase CKI1 acts upstream of histidine phosphotransfer proteins to regulate female gametophyte development and vegetative growth. Plant Cell. **22**, 1232–1248.

Desikan, R., Horak, J., Chaban, C., Mira-Rodado, V., Witthoeft, J., Elgass, K., Grefen, C., Cheung, M.-K., Meixner, A.J., Hooley, R., et al. (2008). The histidine kinase AHK5 integrates endogenous and environmental signals in *Arabidopsis* guard cells. Plos One, **3**, e2491.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica Section D-Biological Crystallography. **66**, 486–501.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. **38**, W695–W699.

Grefen, C., and Harter, K. (2004). Plant two-component systems: principles, functions, complexity and cross talk. Planta. **219**, 733–742.

Guhaniyogi, J., Robinson, V.L., and Stock, A.M. (2006). Crystal structures of beryllium fluoride-free and beryllium fluoride-bound CheY in complex with the conserved C-terminal peptide of CheZ reveal dual binding modes specific to CheY conformation. J. Mol. Biol. **359**, 624–645.

Hass, C., Lohrmann, J., Albrecht, V., Sweere, U., Hummel, F., Yoo, S.D., Hwang, I., Zhu, T., Schafer, E., Kudla, J., et al. (2004). The response regulator 2 mediates ethylene signalling and hormone signal integration in *Arabidopsis*. EMBO J. **23**, 3290–3302.

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

Horak, J., Grefen, C., Berendzen, K.W., Hahn, A., Stierhof, Y.D., Stadelhofer, B., Stahl, M., Koncz, C., and Harter, K. (2008). The *Arabidopsis thaliana* response regulator ARR22 is a putative AHP phospho-histidine phosphatase expressed in the chalaza of developing seeds. BMC Plant Biol. **8**, 77.

Horak, J., Janda, L., Pekarova, B., and Hejatko, J. (2011). Molecular mechanisms of signalling specificity via phosphorelay pathways in *Arabidopsis*. Current Protein & Peptide Science. **12**, 126–136.

Hosoda, K., Imamura, A., Katoh, E., Hatta, T., Tachiki, M., Yamada, H., Mizuno, T., and Yamazaki, T. (2002). Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the *Arabidopsis* response regulators. Plant Cell. **14**, 2015–2029.

Hothorn, M., Dabi, T., and Chory, J. (2011). Structural basis for cytokinin recognition by *Arabidopsis thaliana* histidine kinase 4. Nature Chemical Biology. **7**, 766–768.

Hutchison, C.E., Li, J., Argueso, C., Gonzalez, M., Lee, E., Lewis, M.W., Maxwell, B.B., Perdue, T.D., Schaller, G.E., Alonso, J.M., et al. (2006). The *Arabidopsis* histidine phosphotransfer proteins are redundant positive regulators of cytokinin signaling. Plant Cell. **18**, 3073–3087.

Kabsch, W. (2010). XDS. Acta Crystallographica Section D-Biological Crystallography. **66**, 125–132.

Knowles, J.R. (1980). Enzyme-catalyzed phosphoryl transfer reactions. Ann. Rev. Biochem. **49**, 877–919.

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

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature Protocols. **3**, 1171–1179.

Laub, M.T., and Goulian, M. (2007). Specificity in two-component signal transduction pathways. In Annual Review of Genetics (Palo Alto: Annual Reviews), pp. 121–145.

Mahonen, A.P., Bishopp, A., Higuchi, M., Nieminen, K.M., Kinoshita, K., Tormakangas, K., Ikeda, Y., Oka, A., Kakimoto, T., Helariutta, Y., et al. (2006). Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. Science. **311**, 94–98.

Muller-Dieckmann, H.J., Grantz, A.A., and Kim, S.H. (1999). The structure of the signal receiver domain of the *Arabidopsis thaliana* ethylene receptor ETR1. Structure with Folding & Design. **7**, 1547–1556.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D-Biological Crystallography. **53**, 240–255.

Nishimura, C., Ohashi, Y., Sato, S., Kato, T., Tabata, S., and Ueguchi, C. (2004). Histidine kinase homologs that act as cytokinin receptors possess overlapping functions in the regulation of shoot and root growth in *Arabidopsis*. Plant Cell. **16**, 1365–1377.

Painter, J., and Merritt, E.A. (2006). TLSMD web server for the generation of multi-group TLS models. J. Applied Crystallography. **39**, 109–111.

Pekarova, B., Klumpler, T., Triskova, O., Horak, J., Jansen, S., Dopitova, R., Borkovcova, P., Papouskova, V., Nejedla, E., Sklenar, V., et al. (2011). Structure and binding specificity of the receiver domain of sensor histidine kinase CKI1 from *Arabidopsis thaliana*. Plant J. **67**, 827–839.

Pham, J., Liu, J., Bennett, M.H., Mansfield, J.W., and Desikan, R. (2012). *Arabidopsis* histidine kinase 5 regulates salt sensitivity and resistance against bacterial and fungal infection. New Phytologist. **194**, 168–180.

Pischke, M.S., Jones, L.G., Otsuga, D., Fernandez, D.E., Drews, G.N., and Sussman, M.R. (2002). An *Arabidopsis* histidine kinase is essential for megagametogenesis. Proc. Natl Acad. Sci. U S A. **99**, 15800–15805.

Posas, F., WurglerMurphy, S.M., Maeda, T., Witten, E.A., Thai, T.C., and Saito, H. (1996). Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1–YPD1–SSK1 'two-component' osmosensor. Cell. **86**, 865–875.

Schaller, G.E., Kieber, J.J., and Shiu, S.-H. (2008). Two-component signaling elements and histidyl-aspartyl phosphorelays. The Arabidopsis Book/American Society of Plant Biologists. **6**, e0112.

Schuetze, K., Harter, K., and Chaban, C. (2009). Bimolecular fluorescence complementation (BiFC) to study protein–protein interactions in living plant cells. In Methods in Molecular Biology, Pfannschmidt, T., ed. (Totowa: Human Press Inc), pp. 189–202.

Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M., and Laub, M.T. (2008). Rewiring the specificity of two-component signal transduction systems. Cell. **133**, 1043–1054.

Sugawara, H., Kawano, Y., Hatakeyama, T., Yamaya, T., Kamiya, N., and Sakakibara, H. (2005). Crystal structure of the histidine-containing phosphotransfer protein ZmHP2 from maize. Protein Sci. **14**, 202–208.

Tran, L.-S.P., Urao, T., Qin, F., Maruyama, K., Kakimoto, T., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in *Arabidopsis*. Proc. Natl Acad. Sci. U S A. **104**, 20623–20628.

Urao, T., Yakubov, B., Satoh, R., Yamaguchi-Shinozaki, K., Seki, M., Hirayama, T., and Shinozaki, K. (1999). A transmembrane hybrid-type histidine kinase in *Arabidopsis* functions as an osmosensor. Plant Cell. **11**, 1743–1754.

Varughese, K.I., Tsigelny, I., and Zhao, H.Y. (2006). The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. J. Bacteriol. **188**, 4970–4977.

Voinnet, O., Lederer, C., and Baulcombe, D.C. (2000). A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. Cell. **103**, 157–167.

Volz, K. (1993). Structural conservation in the CheY superfamily. Biochemistry. **32**, 11741–11753.

Volz, K., and Matsumura, P. (1991). Crystal structure of *Escherichia coli* CheY refined at 1.7-A resolution. J. Biol. Chem. **266**, 15511–15519.

Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with autoSHARP. Methods Mol. Biol. (Clifton, NJ). **364**, 215–230.

Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., et al. (2004). Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J. **40**, 428–438.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY-sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res. **37**, D380–D386.

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 Crystallographica Section D–Biological Crystallography. **67**, 235–242.

Xu, Q., Porter, S.W., and West, A.H. (2003). The yeast YPD1/SLN1 complex: insights into molecular recognition in two-component signaling systems. Structure (Cambridge). **11**, 1569–1581.

Xu, Y.W., Morera, S., Janin, J., and Cherfils, J. (1997). AlF3 mimics the transition state of protein phosphorylation in the crystal structure of nucleoside diphosphate kinase and MgADP. Proc. Natl Acad. Sci. U S A. **94**, 3579–3583.

Zhao, X., Copeland, D.M., Soares, A.S., and West, A.H. (2008). Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. J. Mol. Biol. **375**, 1141–1151.

SUMMARY

The crystal structure of the $AHK5_{RD}$–AHP1 complex reveals the determinants of specificity in plant two-component signaling. Our correlation of structural and quantitative affinity data provide the first insight into the molecular recognition events governing the multi-step phosphorelay system in plants.