Structural Studies of the Human Polymeric Immunoglobulin Receptor

by

Agnes Eva Hamburger

Submitted to the Department of Biology in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biology
at the

Massachusetts Institute of Technology
January 2005
[February 2005]

Signature of Author: .....
Department of Biology
January 24, 2005

Certified by: .....
Peter S. Kim
Professor, Department of Biology
President, Merck Research Laboratories
Co-Thesis Supervisor

.....
Pamela J. Bjorkman
Professor, Department of Biology
Investigator, HHMI
California Institute of Technology
Co-Thesis Supervisor

Accepted by: .....
Stephen P. Bell
Professor, Department of Biology
Chairman, Biology Graduate Committee

Acknowledgements

First, I would like to thank my advisors, Peter Kim and Pamela Bjorkman, for providing me the opportunity to work in their labs. I hope I have learned from their remarkable ability to design experiments, interpret data and communicate the story effectively. Many thanks go to my thesis committee including Frank Solomon, Harvey Lodish and Bob Sauer, for their time and invaluable input on all of my research projects, as well as their flexibility in scheduling (and often rescheduling) committee meetings. I would also like to acknowledge my undergraduate advisors, Michael Rossmann and Richard Kuhn, who first introduced me to research and have continued to support and encourage me throughout graduate school.

I had the great fortune of meeting wonderful friends during graduate school. I greatly value Briana Burton, Soni Lacefield Shimoda and Kristen Johnson for their continued friendship through the years and distance. I also thank Sarah Frew, Allison Bardin, John Lo, Monica Wolf and my classmates from Bio98 for fun memories.

I thank all members of the former Kim lab, especially Michael Kay, Debbie Eckert, Pete Carr and Itay Rousso who have been great labmates and wonderful friends. I appreciate everything I learned from them and I hope we all remain life long friends. I also thank Ben Chen and Feng Yang who taught me lot about tissue culture, and Pam Dvorak and Susan Tocio who have always provided excellent administrative help with a smile on their face.

Although I would have liked to have more time at MIT, I am grateful for my experiences at Caltech. Many thanks go out to all members of the Bjorkman lab, past and present. I have learned a great deal from each and every one. Anthony West deserves a huge thank you for patiently helping me with crystallography, teaching me how to use the Biacore and often helping with mysterious computer and/or printer problems. I owe a lot of gratitude to Peter Snow for his generosity in giving time, energy and expertise to help me express challenging proteins. I wish he could see the end result of my thesis. I also thank Inder Nangiana and Cynthia Jones for the proteins they expressed; Andy Herr for many helpful discussions; Darcy Ballister and Evelyn Cheung for setting up crystal trays; and Marta Murphy for lots of administrative help. Finally, thanks to Beth and Tom Tubman, Libby and Mike Sprague, and Mindy Davis for their help in lab and for the many fun outings.

Thank you to Zsuzsi Hamburger for being a great sister, friend and labmate, for all of the emergency trips, the pep talks and help with crystallography and making figures. Thank you to my parents for always seeing, and trying to point out, the positive side of every situation, for being my biggest fans and for their unconditional love and support. I also thank Nagyi for always supporting and encouraging me and reminding me of the importance of balance in my life.

Finally, and most importantly, I thank my husband, Rejon Sarrazine, for being especially understanding, supportive, patient and helpful, and for his sacrifices including moving to Boston and then to LA so I that I could pursue my Ph.D.

Structural Studies of the Human Polymeric Immunoglobulin Receptor

by

Agnes Eva Hamburger

Submitted to the Department of Biology on January 24, 2005
in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Biology

Abstract

The human polymeric immunoglobulin receptor, pIgR, is a glycosylated type I transmembrane protein expressed on the basolateral surface of secretory epithelial cells. pIgR plays a key role in mucosal immunity and, together with bound immunoglobulins (Igs), provides a first line of specific defense against pathogens and their toxins. pIgR binds dimeric IgA (dIgA) and pentameric IgM (pIgM) produced by local plasma cells and transports these polymeric Igs to the apical surface of the cell where the complexes are cleaved from the membrane and deposited into mucosal secretions. The Fc portion of dIgA initially interacts non-covalently with the N-terminal domain (D1) of pIgR, followed by a covalent interaction with D5. In order to gain insight into the molecular details of the initial interaction, we solved the 1.9 Å resolution crystal structure of D1 of pIgR. The structure reveals a folding topology similar to variable Ig domains with differences in the complementarity determining regions (CDRs). CDR1, the primary determinant in dimeric IgA binding, contains a single helical turn. CDR2, the main determinant in binding to pIgM is very short and contains a potentially critical glutamic acid involved in pIgM binding. CDR3 points away from the other CDRs, preventing dimerization of D1 analogous to the variable heavy and light chains in antibodies. Surface plasmon resonance studies showed that D1, regardless of its glycosylation state, binds dIgA with an equilibrium dissociation constant of 300 nM in the absence of other pIgR domains, but does not bind to monomeric IgA1-Fcα. The structure of D1 allows interpretation of previous mutagenesis studies and structure-based comparisons with other IgA and IgM receptors. To further characterize the interaction between intact pIgR and dIgA, we have also initiated structural studies of the other extracellular domains of pIgR, both alone and in complex with the Fc portion of dIgA. Finally, we have also undertaken structural studies of pIgR in complex with the choline-binding protein A, CbpA, a protein found on the surface of *Streptococcus pneumoniae*, a pathogen that uses pIgR as a receptor to invade the human mucosal epithelium.

Thesis Supervisors:
Peter S. Kim, Professor of Biology, Massachusetts Institute of Technology; President, Merck Research Laboratories
Pamela J. Bjorkman, Professor of Biology and Investigator, HHMI, California Institute of Technology

Table of Contents

CHAPTER 1

Introduction

Introduction

The majority of human pathogens and toxins enter susceptible hosts at mucosal surfaces. Since the adult human has approximately 400 m^2 of respiratory and gastrointestinal mucosae (Norderhaug et al., 1999), immunologic protection at these surfaces is essential. Innate defense and local adaptive immune molecules work together to protect these surfaces. Nonspecific antimicrobial agents such as bile, digestive enzymes, acid, and lysozyme, in combination with the mechanical motions of cilia in the respiratory tract and smooth muscle contractions in the intestinal, reproductive and respiratory tracts can clear some pathogens trapped in the mucosae (Lamm, 1997). However, a more specific neutralizing defense is provided by secretory immunoglobulins (SIgs), which are capable of binding and neutralizing specific antigens.

SIgs include polymeric IgA (mainly dimers containing joining (J) chain, but also trimers and tetramers (Vaerman et al., 1995)) and polymeric IgM (J chain-containing pentamers), which are secreted by local plasma cells in the lamina propria, just beneath the epithelium. J chain-lacking IgM hexamers are also secreted by these plasma cells, but only J chain-containing polymeric Igs (pIgs) bind the transmembrane polymeric Ig receptor (pIgR) (Vaerman et al., 1998a), also known as the transmembrane secretory component (SC), expressed on the basolateral surface of epithelial cells. The pIgR-pIg complexes are transported to the apical surface where they are released by the cleavage of pIgR from the membrane, forming SIgs (Figure 1). Actively transcytosed secretory IgA (SIgA), and to some extent secretory IgM (SIgM), provide protection by binding to pathogens and blocking their adherence to, and invasion of, epithelial cells, a process known as immune exclusion. pIgR is constitutively expressed and transcytosed to the apical surface, where uncomplexed pIgR is also cleaved to release free SC (Brandtzaeg, 1973). In addition to actively transporting pIgs across the

epithelium, pIgR also protects the antibodies from proteolytic cleavage in the harsh environment of the mucosae (Crottet and Corthesy, 1998), especially in the gastrointestinal tract.

Immunoglobulin structure

Antibodies are tetramers of two heavy and two light chains, arranged in an approximately Y shape. The light chain contains an N-terminal variable and C-terminal constant region, while the heavy chain contains one N-terminal variable and three or four C-terminal regions. The Fab (fragment antigen binding) is composed of a light chain and the two N-terminal Ig domains of the heavy chain, and is connected by a hinge region to the Fc (fragment crystalline, named for its ability to crystallize readily), which is a dimer of the two C-terminal domains of the heavy chains. The Fab fragments recognize and specifically bind antigens at the antibody-combining site, and the Fc fragments interact with Fc receptors that either elicit immune effector functions, transport the antibody to specific cellular locations, or stabilize the antibody to protect it from proteolytic degradation. The five classes of antibody isotypes (IgG, IgE, IgD, IgM and IgA) are determined by the features of their heavy chains, which differ in their sequences, the number and position of their disulfide bonds, the number and type of carbohydrates attached, the length of their hinge regions, and the number of C-terminal domains (Janeway et al., 1999).

IgG, the most abundant serum Ig, was the first antibody isotype for which high-resolution structural information became available. Crystal structures of IgG Fab fragments and Fab-antigen complexes revealed their four domain β-barrel arrangement and showed that in some cases the antibody combining site is a concave pocket that the antigen fits into, but in other cases the antibody protrudes into the antigen (reviewed in (Wilson et al., 1991)). Similar to the Fab structures, Fc fragments also have four β-barrel Ig domains with extensive contact surface

between the Cγ3 domains and no contact between the Cγ2 domains (Huber et al., 1976). A carbohydrate attached to the inner surface of the Cγ2 domain interacts with this domain through both polar and hydrophobic residues (Deisenhofer, 1981). Intact IgG structures have provided several snapshots of these flexible molecules with an overall Y or T shape but each with an asymmetric arrangement of the Fab and Fc fragments, consistent with antibodies' dynamic nature (Harris et al., 1992; Harris et al., 1998). Another Ig isotype, IgE, is present at low concentrations in serum and is produced mainly against parasites. The interaction of IgE with its receptor is responsible for inflammation and allergic reactions. IgE contains an Ig constant domain in place of the hinge region. The crystal structure of Fcε revealed the overall structure is very similar to Fcγ, including the position of the N-linked carbohydrate in between the Cε3 domains, which is an analogous location to the carbohydrate on Cγ2 in IgG (Wan et al., 2002; Wurzburg et al., 2000).

IgA has two subclasses, IgA1 and IgA2, which differ from each other in the number of N-linked oligosaccharide attachment sites and in the hinge region. IgA1 contains a heavily O-linked glycosylated 23-residue hinge region and two N-linked glycosylation sites, whereas IgA2 has a 13-residue deletion that removes the five O-linked glycosylation sites, and contains additional N-linked oligosaccharide attachment sites. Serum IgA is predominantly monomeric IgA1 while mucosal IgA is mainly dimeric with a relative increase in IgA2 (Kerr, 1990; Kett et al., 1986; Mestecky and McGhee, 1987). Early electron microscopy studies of IgA showed that the attachment of the Fab and Fc fragments resembled that of IgG (Svehag and Bloth, 1970). More recently, small-angle X-ray and neutron solution scattering and homology modeling experiments, however, suggest that the hinge regions of IgA1 (Boehm et al., 1999) and IgA2 (Furtado et al., 2004) cause an extended Fab and Fc conformation that is different from the

arrangement seen in IgG. This results in a predominantly T shape for IgA versus the overall Y shape for IgG. These studies also suggested that IgA2, with its shorter hinge region, is significantly more compact than IgA1, and that the C-terminal 18-residue extension (the tailpiece) in both IgA subtypes is probably folded up against the Cα3 domain (Boehm et al., 1999).

Although a high-resolution crystal structure of intact IgA is presently not available, crystal structures of the Fab and Fc fragments have been solved. Murine Fab fragment crystal structures confirmed the β-sandwich conformation of the individual domains, with an overall arrangement similar to IgG Fab structures (Satow et al., 1986; Suh et al., 1986). A recent crystal structure of Fcα (Herr et al., 2003), however, has revealed differences between the Fc fragments of IgA and previously determined structures of Fcγ and Fcε. It should be noted that the structure of Fcα was extracted from the co-crystal structure of Fcα in complex with the FcαRI receptor, but significant conformational changes between the FcαRI-bound Fcα structure and free Fcα are not expected (Herr et al., 2003). Although the overall architectures of Fcα, Fcγ and Fcε are very similar, the different features of Fcα include the presence of unusual disulfide bonds in the Cα2 domain and at the base of the hinge region, and external, rather than internal, N-linked oligosaccharides (Herr et al., 2003) (Figure 2).

Humoral immunity at mucosal surfaces

IgA, found in both serum and mucosal secretions, is the most abundant human Ig isotype and the principal mucosal antibody (Kerr, 1990). IgA in mucosal secretions has been described as the "first line of defense", while serum IgA is a "second line of defense" against antigens that have penetrated the epithelial barrier. dIgA is delivered to mucosal secretions by pIgR, a receptor that

actively transcytoses J chain-containing pIgs from the basolateral side of epithelial cells (Mostov, 1994). Once in secretions, SIgA binds pathogens and their toxins and prevents their attachment to, and penetration of, the host. SIgA can neutralize pathogens by directly blocking interactions between bacterial adhesins and their cellular receptors or by inhibiting the movement of the bacteria by crosslinking them or interacting with their flagella (reviewed in (Lamm, 1997)). Binding can occur to specific antigens by the IgA antigen-binding site (Armstrong and Dimmock, 1992; Lamm, 1997; Outlaw and Dimmock, 1990) or nonspecifically to bacterial lectins by carbohydrate moieties on IgA or pIgR (Wold et al., 1990). In addition to its barrier function in mucosal secretions, SIgA is also a major component in human breast milk and provides passive immunization to newborns (reviewed in (Brandtzaeg, 2003; Cleary, 2004; Van de Perre, 2003)).

IgA has long been considered non-inflammatory because it does not bind and activate compliment by the classical pathway (Russell et al., 1989). However, many studies have now shown that aggregated serum IgA triggers cellular functions such as phagocytosis, antibody-dependent cell-mediated cytotoxicity (ADCC), degranulation and respiratory burst similar to IgG after binding to its receptor, FcαRI (reviewed in (Monteiro and Van De Winkel, 2003)). Intracellular signaling by IgA, IgG and IgE receptors is transduced via the same intracellular protein, the γ chain (Monteiro and Van De Winkel, 2003). Monomeric and dimeric forms of IgA, but not SIgA, can elicit these cellular responses in the absence of an integrin co-receptor (van Egmond et al., 2000; Vidarsson et al., 2001). As described above, the main function of SIgA is immune exclusion at mucosal surfaces, where the body is in constant contact with antigens from both pathogens and commensal bacteria and ingested food. Therefore, it would actually be disadvantageous for SIgA to elicit inflammatory signals to foreign substances

because the body would be in a constant state of mucosal inflammation, which would eventually damage the protective barrier of the epithelial lining. However the ability of IgA to trigger these cellular functions after interacting with antigens that have already crossed the epithelium provides additional protection in the serum. The structural basis for the ability of monomeric and dimeric IgA but not SIgA to initiate phagocytosis is described later.

IgM, the first antibody produced in the humoral response to infection, is also present in both serum and mucosal secretions (Janeway et al., 1999). While SIgA is the primary Ig in secretions, SIgM is also present at lower concentrations and clears pathogens via similar mechanisms (Norderhaug et al., 1999). Similarly to IgA, serum IgM is able to activate complement very effectively, while mucosal IgM does not (Davis et al., 1988; Randall et al., 1990; Wiersma et al., 1998). In patients with IgA deficiency, IgM is thought to substitute for the function of IgA (Brandtzaeg et al., 1987).

A minority of the plasma cells in the lamina propria secrete IgG. While there is no known active transport of IgG to mucosal secretions in humans, damage to the epithelial layer can result in passively-transferred IgG molecules at mucosal surfaces with low proteolytic activity, such as the respiratory and reproductive tracts. Therefore, IgG provides some, albeit minimal, protection at mucosal surfaces (Lamm, 1997; Norderhaug et al., 1999).

Polymeric immunoglobulins

IgA and IgM, unlike IgG, IgD and IgE, are unique in that they can form higher ordered oligomers. While serum IgA is monomeric, mucosal IgA is predominantly dimeric, with trimers and tetramers present at lower levels (Vaerman et al., 1995). Similarly, serum IgM is hexameric, and mucosal IgM is pentameric (Davis et al., 1988; Niles et al., 1995). Because of their

multivalency, pIgs can effectively crosslink multiple pathogens to limit their mobility and facilitate their clearance (Lamm, 1997). pIgs, especially IgM, bind their antigens with relatively low affinity, an effect that is compensated for by the high avidity effects due to the presence of four identical binding sites in the case of dIgA and ten sites in pIgM.

Polymeric forms of both IgA and IgM contain a C-terminal 18-residue extension, called the tailpiece, which is required for J chain incorporation and polymerization. The tailpiece sequences in human IgA and IgM differ at seven amino acid positions (Yoo et al., 1999), but both contain a penultimate cysteine residue that is essential for efficient polymerization and a highly conserved N-linked glycosylation site that is required for efficient polymer assembly (Atkin et al., 1996; Wiersma et al., 1997). Besides the tailpiece, additional structural elements within the Cα3, and to a lesser extent Cα2, domains of IgA and the Cμ4 and Cμ3 domains of IgM are required for J chain incorporation and polymer formation (Yoo et al., 1999).

Based on mutagenesis experiments and electron microscopy, a model for dIgA structure has been proposed (Bastian et al., 1995; Feinstein et al., 1971; Garcia-Pardo et al., 1981; Krugmann et al., 1997) (Figure 3A). One J chain molecule covalently bridges two IgA molecules arranged in an end-to end configuration (Garcia-Pardo et al., 1981). J chain Cys15 forms a disulfide bond with tailpiece Cys471 in the first Fcα homodimer, and J chain Cys69 forms a second disulfide bond with Cys471 in the tailpiece from the second Fcα homodimer. The two remaining tailpieces, one from each homodimer, are linked directly to each other by a disulfide bond between their penultimate cysteine residues (Bastian et al., 1995; Krugmann et al., 1997).

Pentameric IgM contains J chain and, like pIgA, is actively transported to mucosal secretions by pIgR, but hexameric IgM lacks J chain, and therefore does not bind to or get

secreted by pIgR (Randall et al., 1990). The polymerization state of IgM depends on the amount of J chain, where in the absence of J chain IgM is secreted as hexamers and an increase in J chain production results in an increase in pentameric IgM (Niles et al., 1995). J chain is not found in small IgM assembly intermediates, suggesting that it does not get incorporated until late in the assembly (Brewer and Corley, 1997; Randall et al., 1990). Pentameric IgM appears to be the substrate for J chain incorporation. A model proposes that five IgM molecules are joined by J chain, excluding a sixth IgM molecule that would normally be added to form hexameric IgM in the absence of J chain (Brewer and Corley, 1997) (Figure 3B).

Joining (J) chain

J chain, a 15 kDa polypeptide, was first identified about 35 years ago as a component of SIgA (Halpern and Koshland, 1970) (Figure 3C) and SIgM (Mestecky et al., 1971). This polypeptide is expressed by immunocytes containing various Ig classes (Brandtzaeg, 1974), however, in the absence of IgA or IgM, J chain is degraded intracellularly (Mosmann et al., 1978). The number of J chain molecules in dIgA and pIgM is commonly accepted to be one (Zikan et al., 1986), however early immunochemical studies suggested dIgA may contain two and pIgM may contain three to four J chains (Brandtzaeg, 1975b).

Although the exact function of J chain is not clear, it appears to be essential for the specific transport of pIgs by pIgR, and while it facilitates or stabilizes polymer formation, it is not essential for Ig polymerization (Hendrickson et al., 1995) (Vaerman et al., 1998b) (Johansen et al., 2001). In J chain knockout mice, a higher ratio of monomeric to dimeric IgA was observed, but low levels of dIgA were still present (Hendrickson et al., 1995). Increased serum IgA and reduced biliary and fecal IgA levels were also reported because J chain-deficient IgA

are not actively transported by the hepatobiliary transport mechanism in rodents, and the lack of transfer of IgA from knockout mice was confirmed *in vitro* in Madin-Darby Canine Kidney (MDCK) cells (Hendrickson et al., 1995). Further analyses of these mice showed that SC was not associated with IgA, and although IgA levels were decreased in bile and feces, they were unaffected in mucosal and glandular secretions, suggesting the possibility of an alternative mechanism for the transport of J chain-deficient IgA from wild type IgA (Hendrickson et al., 1996). J chain knockout mice showed no change in serum IgM levels in one study (Hendrickson et al., 1995), but a reduction in serum IgM with impaired ability to activate complement was observed in another study (Erlandsson et al., 1998).

J chain has a hydrophobic secretion signal that directs its synthesis into the endoplasmic reticulum for processing and secretion. Mature human J chain contains 137 amino acids including eight cysteine residues (Max and Korsmeyer, 1985) of which six are involved in intramolecular disulfide bonds and two disulfide bonds to the tailpiece penultimate cysteine in polymeric Igs (Bastian et al., 1992; Bastian et al., 1995; Frutiger et al., 1992). J chain also has an N-linked oligosaccharide attachment site, which appears to be important for the efficient assembly of dIgA (Krugmann et al., 1997). J chain is highly conserved across different species and is also found in invertebrates and amphibians that lack adaptive immunity, suggesting it originated as a component of the ancient innate immune system (Takahashi et al., 1996). In addition to sequence conservation, J chain is also functionally conserved since it is interchangeable between species to form pIgA molecules capable of binding SC, although with lower affinity (Johansen et al., 2000).

The three dimensional structure of J chain remains elusive. It is normally found as an extracellular protein in complex with IgA or IgM and it is found as a free protein only inside Ig

secreting plasma cells. It is sequestered within pIgs so it is not readily available, and separation from pIgs requires reduction of the intermolecular disulfide bonds, which results in the reduction of the intramolecular disulfide bonds as well (Morrison and Koshland, 1972). While J chain does not resemble any known protein, several models have been proposed about its secondary structure. In the first model, a two domain structure was proposed in which the N-terminal domain was predicted to contain β-sheet structure and a mixture of α-helical, β-sheet and random coil structures in the C-terminal domain (Cann et al., 1982). In the second model, J chain was modeled as an eight stranded (A-H) β-sandwich based on its circular dichroism (CD) spectrum and amino acid profiles for β-sheet propensity and hydrophobicity (Zikan et al., 1985) despite having only a low degree of sequence similarity with other Ig domains. However, the assigned disulfide bonds in this model were incorrect (Frutiger et al., 1992). A third model based on conserved sequence features and the inter- and intra-molecular disulfide bond pattern of J chain was also proposed (Frutiger et al., 1992). In this model J chain contains a domain with two antiparallel β-sheets held together by two disulfide bonds, which is associated with a second domain with both α-helical and β-strand propensities (Frutiger et al., 1992). A similar two domain, although less specific, model with mostly β-strands, extended coils and two short α-helices based on disulfide bonds and predicted secondary structure was recently proposed (Johansen et al., 2001).

IgA and IgM receptors

There are two functional classes of Fc receptors (FcRs). One class of receptors is involved in transporting Igs to specific places and protecting them from proteolytic degradation, and include the pIgR and the neonatal Fc receptor (FcRn). The second class, consisting of the FcγRs, FcεRs

and FcαRI, is responsible for eliciting effector immune responses, such as phagocytosis, endocytosis, ADCC, B cells activation and activation of cytokines and inflammation. Fc receptors can also be grouped into three distinct structural groups (Wurzburg and Jardetzky, 2003) (Figure 4). The receptors in the first group are composed of tandem Ig domains and include the IgG receptors (FcγRI, FcγRII and FcγRIII) (Hulett and Hogarth, 1994; van de Winkel and Anderson, 1991), the high affinity IgE receptor (FcεRI) (Metzger, 1991), the IgA specific receptor (FcαRI) (Maliszewski et al., 1990), and two receptors that bind both IgA and IgM (pIgR (Mostov et al., 1984) and Fcα/μR (Shibuya et al., 2000)). The second group resembles MHC class I homologs, and its sole member is the IgG receptor, FcRn (Simister and Mostov, 1989). The third group belongs to the C-type lectin superfamily and includes the low-affinity IgE receptor (FcεRII) (Kikutani et al., 1986) and the recently identified chicken FcY receptor (FcRY) (West et al., 2004).

Polymeric Ig receptor

The most extensively studied IgA and IgM receptor is pIgR, also known as the secretory component, which was identified almost 40 years ago as a component of pIgA isolated from musocal secretions (Brandtzaeg, 1996). This glycoprotein was shown to occur in both a transmembrane bound and a secreted free form, expressed by secretory epithelial cells (Brandtzaeg, 1974; Tomasi et al., 1965). Since then, pIgR transport of J chain-containing pIgs has been demonstrated both *in vivo* and *in vitro*. Expression of the rabbit pIgR cDNA in polarized MDCK cells, which normally do not express pIgR, results in the basolateral expression of pIgR and the transport of the receptor to the apical surface in a ligand-independent manner (Mostov and Deitcher, 1986). Similar results were observed in MDCK cells transfected with the

human pIgR cDNA (Tamer et al., 1995). Although pIgR transcytosis is constitutive, binding by dIgA stimulates increased transcytosis (Song et al., 1994). Knocking out the mouse pIgR gene results in a drastic increase of serum IgA, a slight increase in serum IgG and IgE, but unchanged IgM levels (Shimada et al., 1999). Additionally, IgA levels were reduced but not negligible in bile, feces and intestinal secretions, further suggesting dIgA is mainly transported by pIgR although a small amount may be secreted via an alternative pathway (Shimada et al., 1999).

pIgR is a type I transmembrane glycoprotein consisting of a 620 residue extracellular region arranged as five tandem Ig-like domains, a 23 residue transmembrane anchor and a 103 residue cytoplasmic region. Human pIgA and pIgM bind to the extracellular region of SC with an affinity of approximately 10 nM (Natvig et al., 1997; Roe et al., 1999). Binding of human pIgR to dIgA takes place in two steps. In the first step, the pIgR N-terminal domain (D1), which contains regions analogous to the three antigen-binding complementarity determining regions (CDR) loops of Ig variable domains, binds the Cα3, and possibly Cα2, domains of dIgA (Frutiger et al., 1986; Geneste et al., 1986). In the second step, Cys467 of the human pIgR domain 5 (D5) forms a disulfide bridge with Cys311 in the Cα2 domain of the second IgA molecule in dIgA (Fallgreen-Gebauer et al., 1993) in a late transcytotic compartment (Chintalacharuvu et al., 1994), however this disulfide bond does not form in all species (Knight et al., 1975). Also, pIgM, in contrast to dIgA, does not become covalently bound to human pIgR (Brandtzaeg, 1975a). Further, transport of pIgs by pIgR is species dependent. For example, human pIgR binds and transports both pIgA and pIgM, but pIgR from some other species (e.g. rabbit, rodents and chicken) only bind pIgA (Socken and Underdown, 1978; Underdown et al., 1992; Wieland et al., 2004).

Mutagenesis and peptide binding experiments have been used to localize the pIgR D1 regions responsible for binding pIgA and pIgM. For dIgA binding, all three pIgR D1 CDR loops were shown to be essential, in addition to the EF loop, which is located on the opposite side of the domain from the CDRs (Coyne et al., 1994). In pIgM binding, pIgR D1 CDR2 appears to contribute most to the binding, however, for maximal binding the CDR1 and CDR3 loops are also required (Roe et al., 1999). Peptide binding studies also identified residues in dIgA that participate in binding to pIgR. The potential binding sites are in the Cα3 domain of dIgA and involve residues 402-410 in the DE loop and residues 430-443 in the FG loop (Hexham et al., 1999; White and Capra, 2002) (Figure 5A). Since J chain-containing pIgs are preferentially bound and transcytosed by pIgR (Vaerman et al., 1998a), J chain may also contact pIgR directly.

The intracellular trafficking of pIgR and SIgA has been studied extensively (Rojas and Apodaca, 2002). The 103 residue cytoplasmic tail of pIgR contains a number of cellular sorting signals. Deletion of the cytoplasmic tail results in a truncated pIgR that is able to bind ligand, but is transported from the Golgi apparatus directly to the apical surface where it can be cleaved to SC (Mostov et al., 1986). Newly synthesized wild type protein is targeted from the *trans*-Golgi network to the basolateral surface of epithelial cells via a 17 residue signal immediately C-terminal to the transmembrane region (Casanova et al., 1991). pIgR-pIg complexes are initially internalized by clathrin-coated pits and delivered to basolateral early endosomes (Mostov, 1994). After sorting in this compartment, pIgR-pIgs are targeted to common recycling endosomes. Next, the complex is either delivered to apical recycling and from there, or directly from the common recycling endosome (Mostov, 1994), the protein arrives at the apical surface where pIgR is cleaved from the cell by an unidentified protease releasing free SC or SIg. Multiple C-terminal truncations of released pIgR have been identified, indicating multiple cleavage sites

exist (Eiffert et al., 1984; Asano et al., 2004). Some pIgR escapes cleavage, however, and is internalized and delivered to apical early endosomes and recycled to either the apical surface or the basolateral surface by a process known as reverse transcytosis. Recently, a subcomplex of the mammalian retromer, a multimeric complex that mediates intracellular sorting of a vacuolar enzyme-transporting receptor (Vps10p) (Seaman et al., 1998), has been implicated in the proper sorting of pIgR (Verges et al., 2004).

Protection from pathogens has been proposed to occur at three different stages during pIgR mediated transport of dIgA (Monteiro and Van De Winkel, 2003). As discussed previously, SIgA in external secretions can bind and crosslink pathogens thereby slowing their mobility and inhibiting their ability to adhere to epithelial cells (Lamm, 1997). Additionally, SIgA can also intercept and neutralize pathogens, primarily viruses, intracellularly during transepithelial transport, assuming the pathogen and SIgA go through the same intracellular compartment (Mazanec et al., 1992). Finally, dIgA can bind bacteria and viruses that have already invaded target cells at the basolateral surface of the protective epithelial layer, and pIgR can shuttle the antibody-antigen complex to the apical surface, safely removing the pathogen (Kaetzel et al., 1991; Lamm et al., 1992).

FcαRI

Another well-characterized IgA receptor is FcαRI (also known as CD89), which is a membrane-bound glycoprotein with two extracellular Ig domains, a transmembrane anchor and a cytoplasmic tail lacking recognized signaling motifs (Maliszewski et al., 1990; Monteiro and Van De Winkel, 2003). Although FcαRI is a member of the Ig superfamily of Fc receptors, it shares higher sequence similarity with the killer inhibitory receptors (KIRs) and the leukocyte

Ig-like receptors (LIRs) (Wende et al., 1999). FcαRI is expressed on the surface of monocytes, eosinophils, neutrophils and macrophages, and binds both monomeric and dimeric forms of IgA (Monteiro and Van De Winkel, 2003). FcαRI mediates immune effector responses such as phagocytosis, ADCC, respiratory burst and cytokine release (Monteiro et al., 1990). Monomeric and polymeric IgA, but not SIgA, can bind FcαRI to trigger phagocytosis through the clustering of the receptor (van Egmond et al., 2000; Vidarsson et al., 2001).

The recent crystal structures of FcαRI (Herr et al., 2003; Ding et al., 2003) and the complex between FcαRI and IgA1-Fc (Herr et al., 2003) have shed some light on IgA-mediated immune responses. FcαRI contains two Ig domains oriented at approximately right angles to each other (Ding et al., 2003; Herr et al., 2003). Although the FcγRs (Maxwell et al., 1999; Sondermann et al., 1999; Sondermann et al., 2000) and FcεRI (Garman et al., 1998) also have overall bent structures, FcαRI domains are rotated to resemble the KIR and LIR folds (Chapman et al., 2000; Fan et al., 1997; Wurzburg and Jardetzky, 2003). Another difference between the Fcα-FcαRI structure from the Fcγ-FcγRs (Sondermann et al., 2000) and Fcε-FcεRI (Garman et al., 2000) structures is that the latter interact with a 1:1 stoichiometry at the hinge region, whereas two molecules of FcαRI bind to one Fcα homodimer at the Cα2-Cα3 interface (Herr et al., 2003). This binding surface is analogous to the Cγ2-Cγ3 interface of Fcγ, where a number of proteins, including FcRn, bind to IgG (Martin et al., 2001). The cocrystal structure also revealed the FcαRI binding regions on Fcα, which include residues in the FG loop of the Cα3 domain of Fcα (Herr et al., 2003) (Figure 5B). This loop was shown to be one of the binding sites for pIgR D1 (White and Capra, 2002) (Figure 5A) and is only ~10Å away from Fcα Cys311 (Herr et al., 2003), which forms a disulfide bond with pIgR D5. The overlapping binding site (Figure 5C), and the close proximity of the pIgR D5 disulfide bridge, has been suggested (Herr et al., 2003) to

explain why SIgA does not bind to or activate FcαRI in the absence of an integrin coreceptor (van Egmond et al., 2000; Vidarsson et al., 2001).

Fcα/μR

Fcα/μR, a relatively recently-identified receptor, can bind to monomeric and polymeric IgA as well as IgM (McDonald et al., 2002; Shibuya et al., 2000). It is expressed on the surface of B lymphocytes, macrophages and mesangial cells and has been identified as a candidate for a receptor responsible for the deposition of immune complexes in IgA nephropathy (McDonald et al., 2002), a condition that can result in kidney failure. Fcα/μR expressed on B cells can mediate the endocytosis of IgM-coated microbes and may therefore also play a role in antigen presentation (Shibuya et al., 2000). Additionally, the authors suggested that uptake of IgM through Fcα/μR plays a role in priming T helper lympyhocytes (Shibuya et al., 2000) based on the observations that mice lacking SIgM have a delayed development of specific IgG antibodies to T cell dependent antigens (Ehrenstein et al., 1998) and dissemination of pathogens in peripheral but not secondary lymphoid organs (Ochsenbein et al., 1999). Sequence analysis of Fcα/μR predicts an approximately 50 kDa extracellular region with an N-terminal Ig domain that shares 43% sequence identity with the human pIgR D1, a transmembrane anchor, and a cytoplasmic tail that contains a di-leucine motif in the mouse protein, which is potentially involved in receptor internalization (Shibuya et al., 2000).

Alternative IgA receptors

The asialoglycoprotein receptor (ASGPR) and transferrin receptor (TfR) have also been shown to interact with IgA. ASGPR, a C-type lectin, recognizes galactose and N-acetylgalactosamine

on desialylated glycoproteins and mediates the endocytosis of serum glycoproteins, including IgA (Stockert et al., 1982). The ASGPR recognition site appears to be the O-linked carbohydrates in the hinge region of IgA1 (Stockert et al., 1982). ASGPR is thought to provide a degradative pathway to maintain glycoprotein homeostasis in the blood (Stockert, 1995). TfR (also known as CD71) has also been shown to interact with IgA1, and suggested to be involved in tissue deposition of IgA (Moura et al., 2001). TfR was implicated in the pathogenesis of IgA nephropathy because its expression is upregulated in cultured mesangial cells and on glomerular mesangial cells in patients with the disease (Moura et al., 2001).

Streptococcus pneumoniae and pIgR

Treatment of human infection by gram-positive bacteria is becoming increasingly challenging mainly due to the emergence of antibiotic-resistant bacterial strains. *Streptococcus pneumoniae* (*S. pneumoniae* or pneumococcus), a bacterium that colonizes the upper respiratory tract is an example of such a pathogen (Tuomanen et al., 1995). A number of pneumococcal virulence factors have been identified including the polysaccharide capsule, cell-wall components and a variety of bacterial proteins that facilitate adhesion and invasion, or inhibit host defense (Jedrzejas, 2001). One of these virulence factors, the choline-binding protein, is secreted by the pneumococcus and is anchored to the cell surface by binding to the terminal phosphorylcholine moieties of teichoic or lipoteichoic acid structures on the bacterial cell wall (Jedrzejas, 2001).

The choline binding protein A, CbpA, is anchored to choline on the surface of *S. pneumoniae* via a highly conserved choline-binding motif, which consists of about ten tandem repeats of ~20 residues at the C-terminus (Jedrzejas, 2001). CbpA was shown to be important for bacterial adherence (Rosenow et al., 1997) and the host receptor for *S. pneumoniae* was

identified as the human pIgR (Zhang et al., 2000). An insertional knockout of *cbpA*, or antibodies against either CbpA or pIgR, abolished bacterial adherence and invasion, and two CbpA domains, R1 and R2, were shown to bind to the extracellular domains of human pIgR (Zhang et al., 2000). R1 and R2 share 89% sequence identity and are highly conserved in CbpA from many strains of the bacteria. Peptide mapping and mutagenesis results highlighted the importance of a conserved hexapeptide motif (YRNYPT) within the R domains (Elm et al., 2004b; Hammerschmidt et al., 2000). Domain deletion, mutagenesis and peptide binding studies have recently revealed that domains 3 and 4 of pIgR (D3-4) are necessary and sufficient to bind and uptake CbpA (Elm et al., 2004a; Lu et al., 2003). Interestingly, while SIgA and free SC in mucosal secretions inhibit pneumococcal invasion, pneumococci translocate across nasopharyngeal epithelial cells using the pIgR reverse transcytosis pathway (Zhang et al., 2000).

When I started this project, an extensive amount of work had been done to characterize IgA-based immunity, however there was no high-resolution structural information available for any of the proteins involved. Shortly after I began, the crystal structures of FcαRI (Herr et al., 2003; Ding et al., 2003) and the cocrystal structure of FcαRI-Fcα were published (Herr et al., 2003), but the structures of pIgR, J chain, dIgA (or dFcα), and CbpA remained unknown. For my thesis project I wanted to use structural biology to contribute to the understanding of the molecular mechanisms between these key IgA-based immunological proteins. The subsequent chapters describe the structural characterization of the human polymeric Ig receptor and two of its binding partners, dIgA and CbpA. In Chapter 2, I describe the 1.9 Å resolution crystal structure of the N-terminal, ligand-binding domain of pIgR (D1) and present complementary surface plasmon resonance binding experiments between pIgR D1 and dIgA. Based on the D1 structure, I

interpret previous mutagenesis results and make comparisons to other IgA receptors. This work was done in collaboration with Dr. Anthony P. West, Jr. who taught me how to use the biosensor instrument and the crystallographic software used to solve the structure. Chapter 3 describes the solution structure of an adhesion domain (R2) of CbpA, the *S. pneumoniae* surface protein responsible for pIgR binding. Additional biophysical experiments are also presented that were used to characterize the binding properties of CbpA domains and pIgR. This work was done in collaboration with several groups at St. Jude Children's Research Hospital. My contribution to the project was providing purified pIgR and performing the initial binding experiments to show that the recombinant protein was able to bind to CbpA. In Chapter 4, I describe ongoing work toward crystallizing the full-length extracellular region of pIgR, the Fc region of J chain-containing dIgA (dFcα), CbpA, and the pIgR binding regions of CbpA, R1 and R2, all of them alone and in complex with each other.

References

Armstrong, S. J., and Dimmock, N. J. (1992). Neutralization of influenza virus by low concentrations of hemagglutinin-specific polymeric immunoglobulin A inhibits viral fusion activity, but activation of the ribonucleoprotein is also inhibited. J Virol *66*, 3823-3832.

Asano, M., Takenouchi-Ohkubo, N., Matsumoto, N., Ogura, Y., Nomura, H., Suguro, H., and Moro, I. (2004). Multiple cleavage sites for polymeric immunoglobulin receptor. Immunology *112*, 583-589.

Atkin, J. D., Pleass, R. J., Owens, R. J., and Woof, J. M. (1996). Mutagenesis of the human IgA1 heavy chain tailpiece that prevents dimer assembly. J Immunol *157*, 156-159.

Bastian, A., Kratzin, H., Eckart, K., and Hilschmann, N. (1992). Intra- and interchain disulfide bridges of the human J chain in secretory immunoglobulin A. Biol Chem Hoppe Seyler *373*, 1255-1263.

Bastian, A., Kratzin, H., Fallgren-Gebauer, E., Eckart, K., and Hilschmann, N. (1995). Intra- and inter-chain disulfide bridges of J chain in human S-IgA. Adv Exp Med Biol *371A*, 581-583.

Boehm, M. K., Woof, J. M., Kerr, M. A., and Perkins, S. J. (1999). The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. J Mol Biol *286*, 1421-1447.

Brandtzaeg, P. (1973). Structure, synthesis and external transfer of mucosal immunoglobulins. Ann Immunol (Paris) *124*, 417-438.

Brandtzaeg, P. (1974). Presence of J chain in human immunocytes containing various immunoglobulin classes. Nature *252*, 418-420.

Brandtzaeg, P. (1975a). Human secretory immunoglobulin M. An immunochemical and immunohistochemical study. Immunology *29*, 559-570.

Brandtzaeg, P. (1975b). Immunochemical studies on free and bound J chain of human IgA and IgM. Scand J Immunol *4*, 439-450.

Brandtzaeg, P. (1996). History of oral tolerance and mucosal immunity. Ann N Y Acad Sci *778*, 1-27.

Brandtzaeg, P. (2003). Mucosal immunity: integration between mother and the breast-fed infant. Vaccine *21*, 3382-3388.

Brandtzaeg, P., Karlsson, G., Hansson, G., Petruson, B., Bjorkander, J., and Hanson, L. A. (1987). The clinical condition of IgA-deficient patients is related to the proportion of IgD- and IgM-producing cells in their nasal mucosa. Clin Exp Immunol *67*, 626-636.

Brewer, J. W., and Corley, R. B. (1997). Late events in assembly determine the polymeric structure and biological activity of secretory IgM. Mol Immunol *34*, 323-331.

Cann, G. M., Zaritsky, A., and Koshland, M. E. (1982). Primary structure of the immunoglobulin J chain from the mouse. Proc Natl Acad Sci U S A *79*, 6656-6660.

Casanova, J. E., Apodaca, G., and Mostov, K. E. (1991). An autonomous signal for basolateral sorting in the cytoplasmic domain of the polymeric immunoglobulin receptor. Cell *66*, 65-75.

Chapman, T. L., Heikema, A. P., West, A. P., Jr., and Bjorkman, P. J. (2000). Crystal structure and ligand binding properties of the D1D2 region of the inhibitory receptor LIR-1 (ILT2). Immunity *13*, 727-736.

Chintalacharuvu, K. R., Tavill, A. S., Louis, L. N., Vaerman, J. P., Lamm, M. E., and Kaetzel, C. S. (1994). Disulfide bond formation between dimeric immunoglobulin A and the polymeric immunoglobulin receptor during hepatic transcytosis. Hepatology *19*, 162-173.

Cleary, T. G. (2004). Human milk protective mechanisms. Adv Exp Med Biol *554*, 145-154.

Coyne, R. S., Siebrecht, M., Peitsch, M. C., and Casanova, J. E. (1994). Mutational analysis of polymeric immunoglobulin receptor/ligand interactions. Evidence for the involvement of multiple complementarity determining region (CDR)-like loops in receptor domain I. J Biol Chem *269*, 31620-31625.

Crottet, P., and Corthesy, B. (1998). Secretory component delays the conversion of secretory IgA into antigen-binding competent F(ab')2: a possible implication for mucosal defense. J Immunol *161*, 5445-5453.

Davis, A. C., Roux, K. H., and Shulman, M. J. (1988). On the structure of polymeric IgM. Eur J Immunol *18*, 1001-1008.

Deisenhofer, J. (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution. Biochemistry *20*, 2361-2370.

Ding, Y., Xu, G., Yang, M., Yao, M., Gao, G. F., Wang, L., Zhang, W., and Rao, Z. (2003). Crystal structure of the ectodomain of human FcalphaRI. J Biol Chem *278*, 27966-27970.

Ehrenstein, M. R., O'Keefe, T. L., Davies, S. L., and Neuberger, M. S. (1998). Targeted gene disruption reveals a role for natural secretory IgM in the maturation of the primary immune response. Proc Natl Acad Sci U S A *95*, 10089-10093.

Eiffert, H., Quentin, E., Decker, J., Hillemeir, S., Hufschmidt, M., Klingmuller, D., Weber, M. H., and Hilschmann, N. (1984). [The primary structure of human free secretory component and the arrangement of disulfide bonds]. Hoppe Seylers Z Physiol Chem *365*, 1489-1495.

Elm, C., Braathen, R., Bergmann, S., Frank, R., Vaerman, J. P., Kaetzel, C. S., Chhatwal, G. S., Johansen, F. E., and Hammerschmidt, S. (2004a). Ectodomains 3 and 4 of human polymeric Immunoglobulin receptor (hpIgR) mediate invasion of Streptococcus pneumoniae into the epithelium. J Biol Chem *279*, 6296-6304.

Elm, C., Rohde, M., Vaerman, J. P., Chhatwal, G. S., and Hammerschmidt, S. (2004b). Characterization of the interaction of the pneumococcal surface protein SpsA with the human polymeric immunoglobulin receptor (hpIgR). Indian J Med Res *119 Suppl*, 61-65.

Erlandsson, L., Andersson, K., Sigvardsson, M., Lycke, N., and Leanderson, T. (1998). Mice with an inactivated joining chain locus have perturbed IgM secretion. Eur J Immunol *28*, 2355-2365.

Fallgreen-Gebauer, E., Gebauer, W., Bastian, A., Kratzin, H. D., Eiffert, H., Zimmermann, B., Karas, M., and Hilschmann, N. (1993). The covalent linkage of secretory component to IgA. Structure of sIgA. Biol Chem Hoppe Seyler *374*, 1023-1028.

Fan, Q. R., Mosyak, L., Winter, C. C., Wagtmann, N., Long, E. O., and Wiley, D. C. (1997). Structure of the inhibitory receptor for human natural killer cells resembles haematopoietic receptors. Nature *389*, 96-100.

Feinstein, A., Munn, E. A., and Richardson, N. E. (1971). The three-dimensional conformation of M and A globulin molecules. Ann N Y Acad Sci *190*, 104-121.

Frutiger, S., Hughes, G. J., Hanly, W. C., Kingzette, M., and Jaton, J. C. (1986). The amino-terminal domain of rabbit secretory component is responsible for noncovalent binding to immunoglobulin A dimers. J Biol Chem *261*, 16673-16681.

Frutiger, S., Hughes, G. J., Paquet, N., Luthy, R., and Jaton, J. C. (1992). Disulfide bond assignment in human J chain and its covalent pairing with immunoglobulin M. Biochemistry *31*, 12643-12647.

Furtado, P. B., Whitty, P. W., Robertson, A., Eaton, J. T., Almogren, A., Kerr, M. A., Woof, J. M., and Perkins, S. J. (2004). Solution structure determination of monomeric human IgA2 by X-ray and neutron scattering, analytical ultracentrifugation and constrained modelling: a comparison with monomeric human IgA1. J Mol Biol *338*, 921-941.

Garcia-Pardo, A., Lamm, M. E., Plaut, A. G., and Frangione, B. (1981). J chain is covalently bound to both monomer subunits in human secretory IgA. J Biol Chem *256*, 11734-11738.

Garman, S. C., Kinet, J. P., and Jardetzky, T. S. (1998). Crystal structure of the human high-affinity IgE receptor. Cell *95*, 951-961.

Garman, S. C., Wurzburg, B. A., Tarchevskaya, S. S., Kinet, J. P., and Jardetzky, T. S. (2000). Structure of the Fc fragment of human IgE bound to its high-affinity receptor Fc epsilonRI alpha. Nature *406*, 259-266.

Geneste, C., Iscaki, S., Mangalo, R., and Pillot, J. (1986). Both Fc alpha domains of human IgA are involved in in vitro interaction between secretory component and dimeric IgA. Immunol Lett *13*, 221-226.

Halpern, M. S., and Koshland, M. E. (1970). Noval subunit in secretory IgA. Nature *228*, 1276-1278.

Hammerschmidt, S., Tillig, M. P., Wolff, S., Vaerman, J. P., and Chhatwal, G. S. (2000). Species-specific binding of human secretory component to SpsA protein of Streptococcus pneumoniae via a hexapeptide motif. Mol Microbiol *36*, 726-736.

Harris, L. J., Larson, S. B., Hasel, K. W., Day, J., Greenwood, A., and McPherson, A. (1992). The three-dimensional structure of an intact monoclonal antibody for canine lymphoma. Nature *360*, 369-372.

Harris, L. J., Skaletsky, E., and McPherson, A. (1998). Crystallographic structure of an intact IgG1 monoclonal antibody. J Mol Biol *275*, 861-872.

Hendrickson, B. A., Conner, D. A., Ladd, D. J., Kendall, D., Casanova, J. E., Corthesy, B., Max, E. E., Neutra, M. R., Seidman, C. E., and Seidman, J. G. (1995). Altered hepatic transport of immunoglobulin A in mice lacking the J chain. J Exp Med *182*, 1905-1911.

Hendrickson, B. A., Rindisbacher, L., Corthesy, B., Kendall, D., Waltz, D. A., Neutra, M. R., and Seidman, J. G. (1996). Lack of association of secretory component with IgA in J chain-deficient mice. J Immunol *157*, 750-754.

Herr, A. B., Ballister, E. R., and Bjorkman, P. J. (2003). Insights into IgA-mediated immune responses from the crystal structures of human FcalphaRI and its complex with IgA1-Fc. Nature *423*, 614-620.

Hexham, J. M., White, K. D., Carayannopoulos, L. N., Mandecki, W., Brisette, R., Yang, Y. S., and Capra, J. D. (1999). A human immunoglobulin (Ig)A calpha3 domain motif directs polymeric Ig receptor-mediated secretion. J Exp Med *189*, 747-752.

Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, M., and Palm, W. (1976). Crystallographic structure studies of an IgG molecule and an Fc fragment. Nature *264*, 415-420.

Hulett, M. D., and Hogarth, P. M. (1994). Molecular basis of Fc receptor function. Adv Immunol *57*, 1-127.

Janeway, C. A., Travers, P., Walport, M., and Capra, J. D. (1999). Immunobiology: the immune system in health and disease, 4th edn (New York, Garland Publishing).

Jedrzejas, M. J. (2001). Pneumococcal virulence factors: structure and function. Microbiol Mol Biol Rev *65*, 187-207 ; first page, table of contents.

Johansen, F. E., Braathen, R., and Brandtzaeg, P. (2000). Role of J chain in secretory immunoglobulin formation. Scand J Immunol *52*, 240-248.

Johansen, F. E., Braathen, R., and Brandtzaeg, P. (2001). The J chain is essential for polymeric Ig receptor-mediated epithelial transport of IgA. J Immunol *167*, 5185-5192.

Kaetzel, C. S., Robinson, J. K., Chintalacharuvu, K. R., Vaerman, J. P., and Lamm, M. E. (1991). The polymeric immunoglobulin receptor (secretory component) mediates transport of immune complexes across epithelial cells: a local defense function for IgA. Proc Natl Acad Sci U S A *88*, 8796-8800.

Kerr, M. A. (1990). The structure and function of human IgA. Biochem J *271*, 285-296.

Kett, K., Brandtzaeg, P., Radl, J., and Haaijman, J. J. (1986). Different subclass distribution of IgA-producing cells in human lymphoid organs and various secretory tissues. J Immunol *136*, 3631-3635.

Kikutani, H., Inui, S., Sato, R., Barsumian, E. L., Owaki, H., Yamasaki, K., Kaisho, T., Uchibayashi, N., Hardy, R. R., Hirano, T., and et al. (1986). Molecular structure of human lymphocyte receptor for immunoglobulin E. Cell *47*, 657-665.

Knight, K. L., Vetter, M. L., and Malek, T. R. (1975). Distribution of covalently bound and non-covalently bound secretory component on sbuclasses of rabbit secretory IgA. J Immunol *115*, 595-598.

Krugmann, S., Pleass, R. J., Atkin, J. D., and Woof, J. M. (1997). Structural requirements for assembly of dimeric IgA probed by site-directed mutagenesis of J chain and a cysteine residue of the alpha-chain CH2 domain. J Immunol *159*, 244-249.

Lamm, M. E. (1997). Interaction of antigens and antibodies at mucosal surfaces. Annu Rev Microbiol *51*, 311-340.

Lamm, M. E., Robinson, J. K., and Kaetzel, C. S. (1992). Transport of IgA immune complexes across epithelial membranes: new concepts in mucosal immunity. Adv Exp Med Biol *327*, 91-94.

Lu, L., Lamm, M. E., Li, H., Corthesy, B., and Zhang, J. R. (2003). The human polymeric immunoglobulin receptor binds to Streptococcus pneumoniae via domains 3 and 4. J Biol Chem *278*, 48178-48187.

Maliszewski, C. R., March, C. J., Schoenborn, M. A., Gimpel, S., and Shen, L. (1990). Expression cloning of a human Fc receptor for IgA. J Exp Med *172*, 1665-1672.

Martin, W. L., West, A. P., Jr., Gan, L., and Bjorkman, P. J. (2001). Crystal structure at 2.8 A of an FcRn/heterodimeric Fc complex: mechanism of pH-dependent binding. Mol Cell *7*, 867-877.

Max, E. E., and Korsmeyer, S. J. (1985). Human J chain gene. Structure and expression in B lymphoid cells. J Exp Med *161*, 832-849.

Maxwell, K. F., Powell, M. S., Hulett, M. D., Barton, P. A., McKenzie, I. F., Garrett, T. P., and Hogarth, P. M. (1999). Crystal structure of the human leukocyte Fc receptor, Fc gammaRIIa. Nat Struct Biol *6*, 437-442.

Mazanec, M. B., Kaetzel, C. S., Lamm, M. E., Fletcher, D., and Nedrud, J. G. (1992). Intracellular neutralization of virus by immunoglobulin A antibodies. Proc Natl Acad Sci U S A *89*, 6901-6905.

McDonald, K. J., Cameron, A. J., Allen, J. M., and Jardine, A. G. (2002). Expression of Fc alpha/mu receptor by human mesangial cells: a candidate receptor for immune complex deposition in IgA nephropathy. Biochem Biophys Res Commun *290*, 438-442.

Mestecky, J., and McGhee, J. R. (1987). Immunoglobulin A (IgA): molecular and cellular interactions involved in IgA biosynthesis and immune response. Adv Immunol *40*, 153-245.

Mestecky, J., Zikan, J., and Butler, W. T. (1971). Immunoglobulin M and secretory immunoglobulin A: presence of a common polypeptide chain different from light chains. Science *171*, 1163-1165.

Metzger, H. (1991). The high affinity receptor for IgE on mast cells. Clin Exp Allergy *21*, 269-279.

Monteiro, R. C., Kubagawa, H., and Cooper, M. D. (1990). Cellular distribution, regulation, and biochemical nature of an Fc alpha receptor in humans. J Exp Med *171*, 597-613.

Monteiro, R. C., and Van De Winkel, J. G. (2003). IgA Fc receptors. Annu Rev Immunol *21*, 177-204.

Morrison, S. L., and Koshland, M. E. (1972). Characterization of the J chain from polymeric immunoglobulins (IgA-IgM-immunological specificity-primary structure). Proc Natl Acad Sci U S A *69*, 124-128.

Mosmann, T. R., Gravel, Y., Williamson, A. R., and Baumal, R. (1978). Modification and fate of J chain in myeloma cells in the presence and absence of polymeric immunoglobulin secretion. Eur J Immunol *8*, 94-101.

Mostov, K. E. (1994). Transepithelial transport of immunoglobulins. Annu Rev Immunol *12*, 63-84.

Mostov, K. E., de Bruyn Kops, A., and Deitcher, D. L. (1986). Deletion of the cytoplasmic domain of the polymeric immunoglobulin receptor prevents basolateral localization and endocytosis. Cell *47*, 359-364.

Mostov, K. E., and Deitcher, D. L. (1986). Polymeric immunoglobulin receptor expressed in MDCK cells transcytoses IgA. Cell *46*, 613-621.

Mostov, K. E., Friedlander, M., and Blobel, G. (1984). The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. Nature *308*, 37-43.

Moura, I. C., Centelles, M. N., Arcos-Fajardo, M., Malheiros, D. M., Collawn, J. F., Cooper, M. D., and Monteiro, R. C. (2001). Identification of the transferrin receptor as a novel immunoglobulin (Ig)A1 receptor and its enhanced expression on mesangial cells in IgA nephropathy. J Exp Med *194*, 417-425.

Natvig, I. B., Johansen, F. E., Nordeng, T. W., Haraldsen, G., and Brandtzaeg, P. (1997). Mechanism for enhanced external transfer of dimeric IgA over pentameric IgM: studies of diffusion, binding to the human polymeric Ig receptor, and epithelial transcytosis. J Immunol *159*, 4330-4340.

Niles, M. J., Matsuuchi, L., and Koshland, M. E. (1995). Polymer IgM assembly and secretion in lymphoid and nonlymphoid cell lines: evidence that J chain is required for pentamer IgM synthesis. Proc Natl Acad Sci U S A *92*, 2884-2888.

Norderhaug, I. N., Johansen, F. E., Schjerven, H., and Brandtzaeg, P. (1999). Regulation of the formation and external transport of secretory immunoglobulins. Crit Rev Immunol *19*, 481-508.

Ochsenbein, A. F., Fehr, T., Lutz, C., Suter, M., Brombacher, F., Hengartner, H., and Zinkernagel, R. M. (1999). Control of early viral and bacterial distribution and disease by natural antibodies. Science *286*, 2156-2159.

Outlaw, M. C., and Dimmock, N. J. (1990). Mechanisms of neutralization of influenza virus on mouse tracheal epithelial cells by mouse monoclonal polymeric IgA and polyclonal IgM directed against the viral haemagglutinin. J Gen Virol *71 (Pt 1)*, 69-76.

Randall, T. D., King, L. B., and Corley, R. B. (1990). The biological effects of IgM hexamer formation. Eur J Immunol *20*, 1971-1979.

Roe, M., Norderhaug, I. N., Brandtzaeg, P., and Johansen, F. E. (1999). Fine specificity of ligand-binding domain 1 in the polymeric Ig receptor: importance of the CDR2-containing region for IgM interaction. J Immunol *162*, 6046-6052.

Rojas, R., and Apodaca, G. (2002). Immunoglobulin transport across polarized epithelial cells. Nat Rev Mol Cell Biol *3*, 944-955.

Rosenow, C., Ryan, P., Weiser, J. N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H. R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol *25*, 819-829.

Russell, M. W., Reinholdt, J., and Kilian, M. (1989). Anti-inflammatory activity of human IgA antibodies and their Fab alpha fragments: inhibition of IgG-mediated complement activation. Eur J Immunol *19*, 2243-2249.

Satow, Y., Cohen, G. H., Padlan, E. A., and Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 A. J Mol Biol *190*, 593-604.

Seaman, M. N., McCaffery, J. M., and Emr, S. D. (1998). A membrane coat complex essential for endosome-to-Golgi retrograde transport in yeast. J Cell Biol *142*, 665-681.

Shibuya, A., Sakamoto, N., Shimizu, Y., Shibuya, K., Osawa, M., Hiroyama, T., Eyre, H. J., Sutherland, G. R., Endo, Y., Fujita, T., *et al.* (2000). Fc alpha/mu receptor mediates endocytosis of IgM-coated microbes. Nat Immunol *1*, 441-446.

Shimada, S., Kawaguchi-Miyashita, M., Kushiro, A., Sato, T., Nanno, M., Sako, T., Matsuoka, Y., Sudo, K., Tagawa, Y., Iwakura, Y., and Ohwaki, M. (1999). Generation of polymeric immunoglobulin receptor-deficient mouse with marked reduction of secretory IgA. J Immunol *163*, 5367-5373.

Simister, N. E., and Mostov, K. E. (1989). An Fc receptor structurally related to MHC class I antigens. Nature *337*, 184-187.

Socken, D. J., and Underdown, B. J. (1978). Comparison of human, bovine and rabbit secretory component-immunoglobulin interactions. Immunochemistry *15*, 499-506.

Sondermann, P., Huber, R., and Jacob, U. (1999). Crystal structure of the soluble form of the human fcgamma-receptor IIb: a new member of the immunoglobulin superfamily at 1.7 A resolution. Embo J *18*, 1095-1103.

Sondermann, P., Huber, R., Oosthuizen, V., and Jacob, U. (2000). The 3.2-A crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. Nature *406*, 267-273.

Song, W., Bomsel, M., Casanova, J., Vaerman, J. P., and Mostov, K. (1994). Stimulation of transcytosis of the polymeric immunoglobulin receptor by dimeric IgA. Proc Natl Acad Sci U S A *91*, 163-166.

Stockert, R. J. (1995). The asialoglycoprotein receptor: relationships between structure, function, and expression. Physiol Rev *75*, 591-609.

Stockert, R. J., Kressner, M. S., Collins, J. C., Sternlieb, I., and Morell, A. G. (1982). IgA interaction with the asialoglycoprotein receptor. Proc Natl Acad Sci U S A *79*, 6229-6231.

Suh, S. W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S., and Davies, D. R. (1986). The galactan-binding immunoglobulin Fab J539: an X-ray diffraction study at 2.6-A resolution. Proteins *1*, 74-80.

Svehag, S. E., and Bloth, B. (1970). Ultrastructure of secretory and high-polymer serum immunoglobulin A of human and rabbit origin. Science *168*, 847-849.

Takahashi, T., Iwase, T., Takenouchi, N., Saito, M., Kobayashi, K., Moldoveanu, Z., Mestecky, J., and Moro, I. (1996). The joining (J) chain is present in invertebrates that do not express immunoglobulins. Proc Natl Acad Sci U S A *93*, 1886-1891.

Tamer, C. M., Lamm, M. E., Robinson, J. K., Piskurich, J. F., and Kaetzel, C. S. (1995). Comparative studies of transcytosis and assembly of secretory IgA in Madin-Darby canine kidney cells expressing human polymeric Ig receptor. J Immunol *155*, 707-714.

Tomasi, T. B., Jr., Tan, E. M., Solomon, A., and Prendergast, R. A. (1965). Characteristics of an Immune System Common to Certain External Secretions. J Exp Med *121*, 101-124.

Tuomanen, E. I., Austrian, R., and Masure, H. R. (1995). Pathogenesis of pneumococcal infection. N Engl J Med *332*, 1280-1284.

Underdown, B. J., Switzer, I., and Jackson, G. D. (1992). Rat secretory component binds poorly to rodent IgM. J Immunol *149*, 487-491.

Vaerman, J. P., Langendries, A., Giffroy, D., Brandtzaeg, P., and Kobayashi, K. (1998a). Lack of SC/pIgR-mediated epithelial transport of a human polymeric IgA devoid of J chain: in vitro and in vivo studies. Immunology *95*, 90-96.

Vaerman, J. P., Langendries, A., and Vander Maelen, C. (1995). Homogenous IgA monomers, dimers, trimers and tetramers from the same IgA myeloma serum. Immunol Invest *24*, 631-641.

Vaerman, J. P., Langendries, A. E., Giffroy, D. A., Kaetzel, C. S., Fiani, C. M., Moro, I., Brandtzaeg, P., and Kobayashi, K. (1998b). Antibody against the human J chain inhibits polymeric Ig receptor-mediated biliary and epithelial transport of human polymeric IgA. Eur J Immunol *28*, 171-182.

Van de Perre, P. (2003). Transfer of antibody via mother's milk. Vaccine *21*, 3374-3376.

van de Winkel, J. G., and Anderson, C. L. (1991). Biology of human immunoglobulin G Fc receptors. J Leukoc Biol *49*, 511-524.

van Egmond, M., van Garderen, E., van Spriel, A. B., Damen, C. A., van Amersfoort, E. S., van Zandbergen, G., van Hattum, J., Kuiper, J., and van de Winkel, J. G. (2000). FcalphaRI-positive liver Kupffer cells: reappraisal of the function of immunoglobulin A in immunity. Nat Med *6*, 680-685.

Verges, M., Luton, F., Gruber, C., Tiemann, F., Reinders, L. G., Huang, L., Burlingame, A. L., Haft, C. R., and Mostov, K. E. (2004). The mammalian retromer regulates transcytosis of the polymeric immunoglobulin receptor. Nat Cell Biol *6*, 763-769.

Vidarsson, G., van Der Pol, W. L., van Den Elsen, J. M., Vile, H., Jansen, M., Duijs, J., Morton, H. C., Boel, E., Daha, M. R., Corthesy, B., and van De Winkel, J. G. (2001). Activity of human IgG and IgA subclasses in immune defense against Neisseria meningitidis serogroup B. J Immunol *166*, 6250-6256.

Wan, T., Beavil, R. L., Fabiane, S. M., Beavil, A. J., Sohi, M. K., Keown, M., Young, R. J., Henry, A. J., Owens, R. J., Gould, H. J., and Sutton, B. J. (2002). The crystal structure of IgE Fc reveals an asymmetrically bent conformation. Nat Immunol *3*, 681-686.

Wende, H., Colonna, M., Ziegler, A., and Volz, A. (1999). Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. Mamm Genome *10*, 154-160.

West, A. P., Jr., Herr, A. B., and Bjorkman, P. J. (2004). The chicken yolk sac IgY receptor, a functional equivalent of the mammalian MHC-related Fc receptor, is a phospholipase A2 receptor homolog. Immunity *20*, 601-610.

White, K. D., and Capra, J. D. (2002). Targeting mucosal sites by polymeric immunoglobulin receptor-directed peptides. J Exp Med *196*, 551-555.

Wieland, W. H., Orzaez, D., Lammers, A., Parmentier, H. K., Verstegen, M. W., and Schots, A. (2004). A functional polymeric immunoglobulin receptor in chicken (Gallus gallus) indicates ancient role of secretory IgA in mucosal immunity. Biochem J *380*, 669-676.

Wiersma, E. J., Chen, F., Bazin, R., Collins, C., Painter, R. H., Lemieux, R., and Shulman, M. J. (1997). Analysis of IgM structures involved in J chain incorporation. J Immunol *158*, 1719-1726.

Wiersma, E. J., Collins, C., Fazel, S., and Shulman, M. J. (1998). Structural and functional analysis of J chain-deficient IgM. J Immunol *160*, 5979-5989.

Wilson, I. A., Rini, J. M., Fremont, D. H., Fieser, G. G., and Stura, E. A. (1991). X-ray crystallographic analysis of free and antigen-complexed Fab fragments to investigate structural basis of immune recognition. Methods Enzymol *203*, 153-176.

Wold, A. E., Mestecky, J., Tomana, M., Kobata, A., Ohbayashi, H., Endo, T., and Eden, C. S. (1990). Secretory immunoglobulin A carries oligosaccharide receptors for Escherichia coli type 1 fimbrial lectin. Infect Immun *58*, 3073-3077.

Wurzburg, B. A., Garman, S. C., and Jardetzky, T. S. (2000). Structure of the human IgE-Fc C epsilon 3-C epsilon 4 reveals conformational flexibility in the antibody effector domains. Immunity *13*, 375-385.

Wurzburg, B. A., and Jardetzky, T. S. (2003). The IgA receptor complex: a two-for-one deal. Nat Struct Biol *10*, 585-587.

Yoo, E. M., Coloma, M. J., Trinh, K. R., Nguyen, T. Q., Vuong, L. U., Morrison, S. L., and Chintalacharuvu, K. R. (1999). Structural requirements for polymeric immunoglobulin assembly and association with J chain. J Biol Chem *274*, 33771-33777.

Zhang, J. R., Mostov, K. E., Lamm, M. E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell *102*, 827-837.

Zikan, J., Mestecky, J., Kulhavy, R., and Bennett, J. C. (1986). The stoichiometry of J chain in human secretory dimeric IgA. Mol Immunol *23*, 541-544.

Zikan, J., Novotny, J., Trapane, T. L., Koshland, M. E., Urry, D. W., Bennett, J. C., and Mestecky, J. (1985). Secondary structure of the immunoglobulin J chain. Proc Natl Acad Sci U S A *82*, 5905-5909.

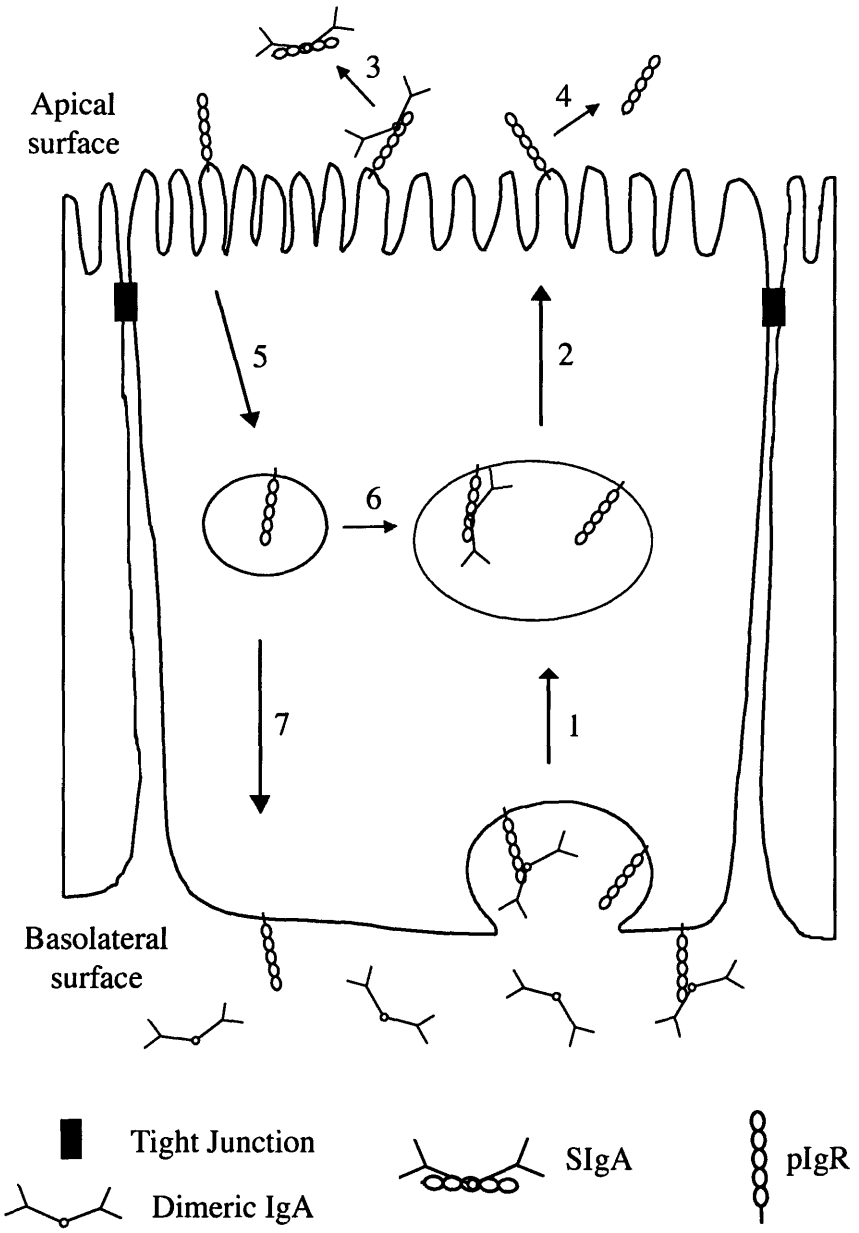


Figure 1. Model for pIgR transcytosis. Newly synthesized pIgR is expressed on the basolateral surface of epithelial cells where it binds to dIgA secreted by local plasma cells in the lamina propria. pIgR-dIgA complexes are initially internalized by clathrin-coated pits (1) and are delivered to the apical surface (2) where pIgR is cleaved from the cell releasing SIgA (3) or free SC (4). pIgR that escapes cleaveage is internalized (5), and recycled to either the apical surface (6) or the basolateral surface by a process known as reverse transcytosis (7).

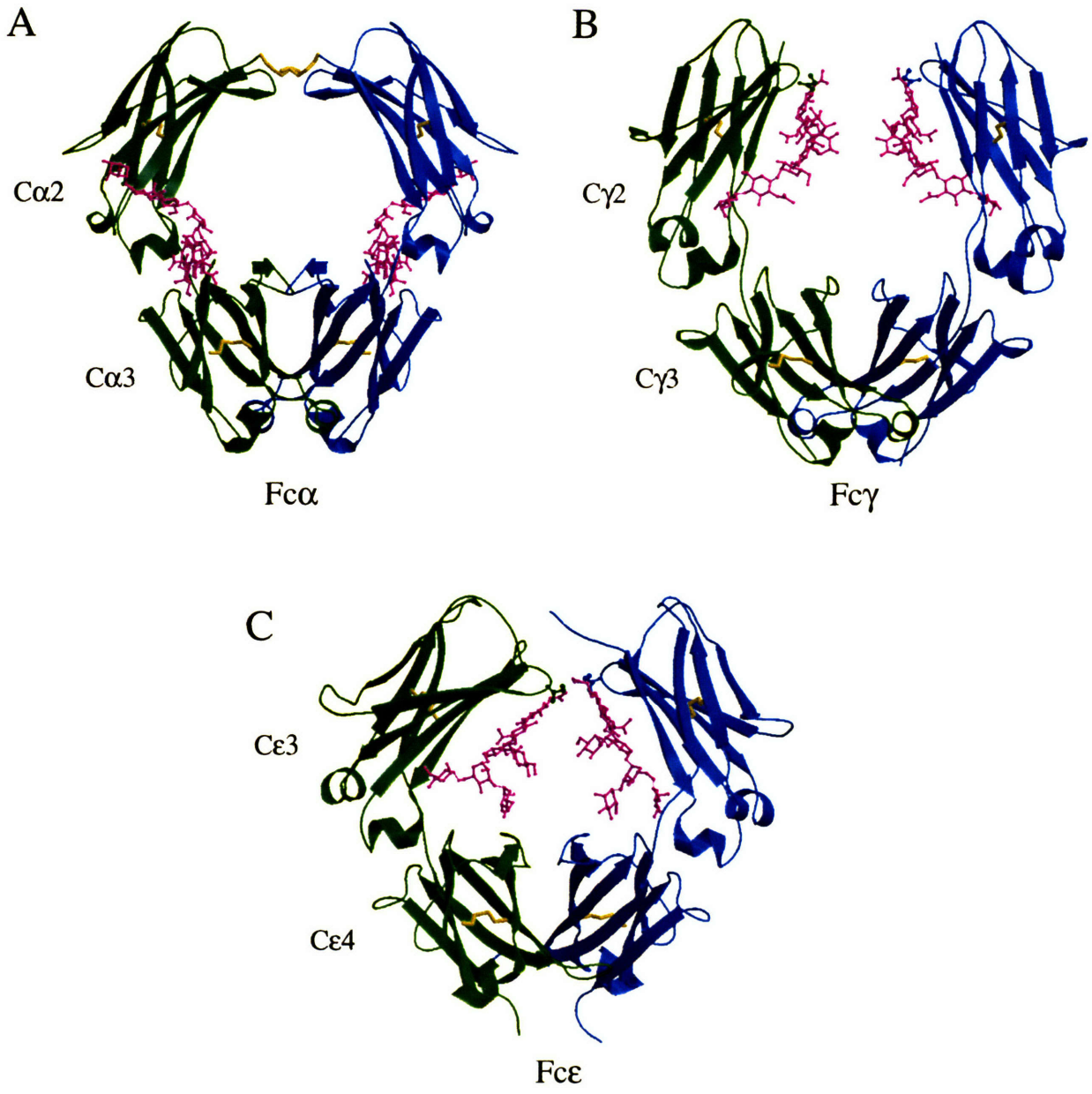


Figure 2. Crystal structures of Fc fragments. Ribbon diagrams showing front views of (A) Fcα (PDB code 1OW0), (B) Fcγ (PDB code 1DN2), and (C) Fcε (PDB code 1O0V). One monomer in each Fc homodimer is shown in green, the other in blue. N-linked carbohydrates are shown in magenta and disulfide bonds are yellow in ball-and-stick representation.

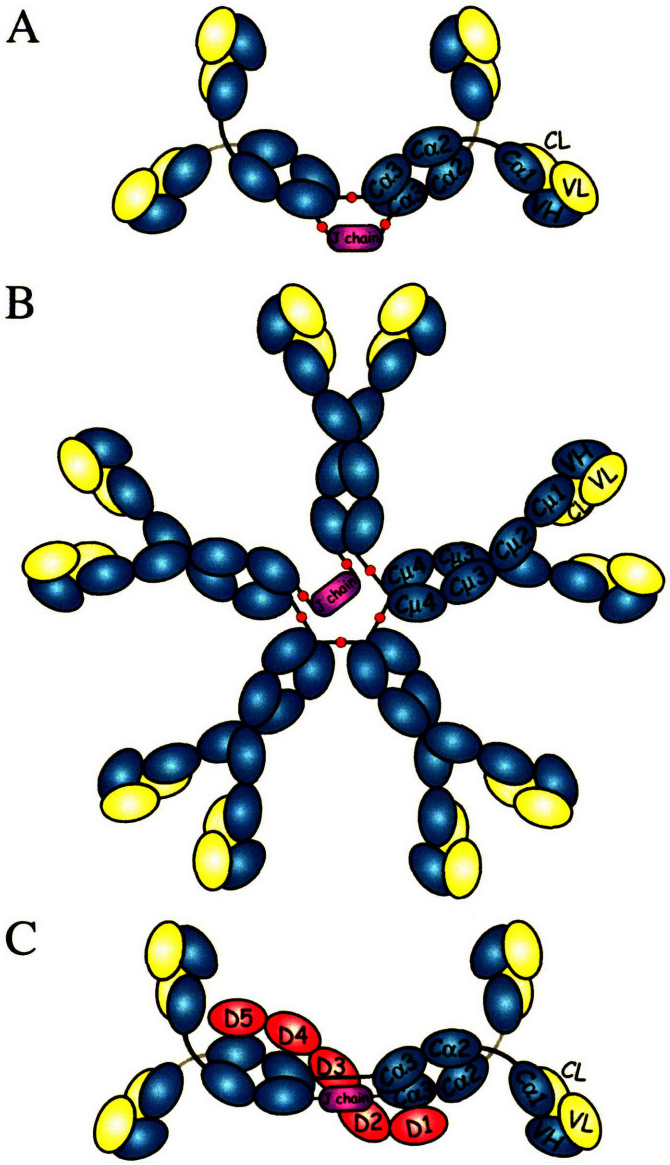


Figure 3. Putative models for the structures of dIgA, pIgM and SIgA. Monomeric Igs are composed of two light chains (yellow) and two heavy chains (blue). One molecule of J chain (pink) bridges two IgA molecules in an end-to-end arrangement to form dIgA (A), or five IgM molecules in a cyclic configuration to form pIgM (B). One molecule of pIgR D1-5 (red) binds noncovalently to Cα3 (and possibly Cα2) of one IgA via D1 and covalently to Cα2 in the second IgA molecule in dIgA via D5, to form SIgA (C). Red spheres in (A) and (B) indicate tailpiece-tailpiece and tailpiece-J chain disulfide bonds. Figure adapted from (Rojas and Apodaca, 2002).

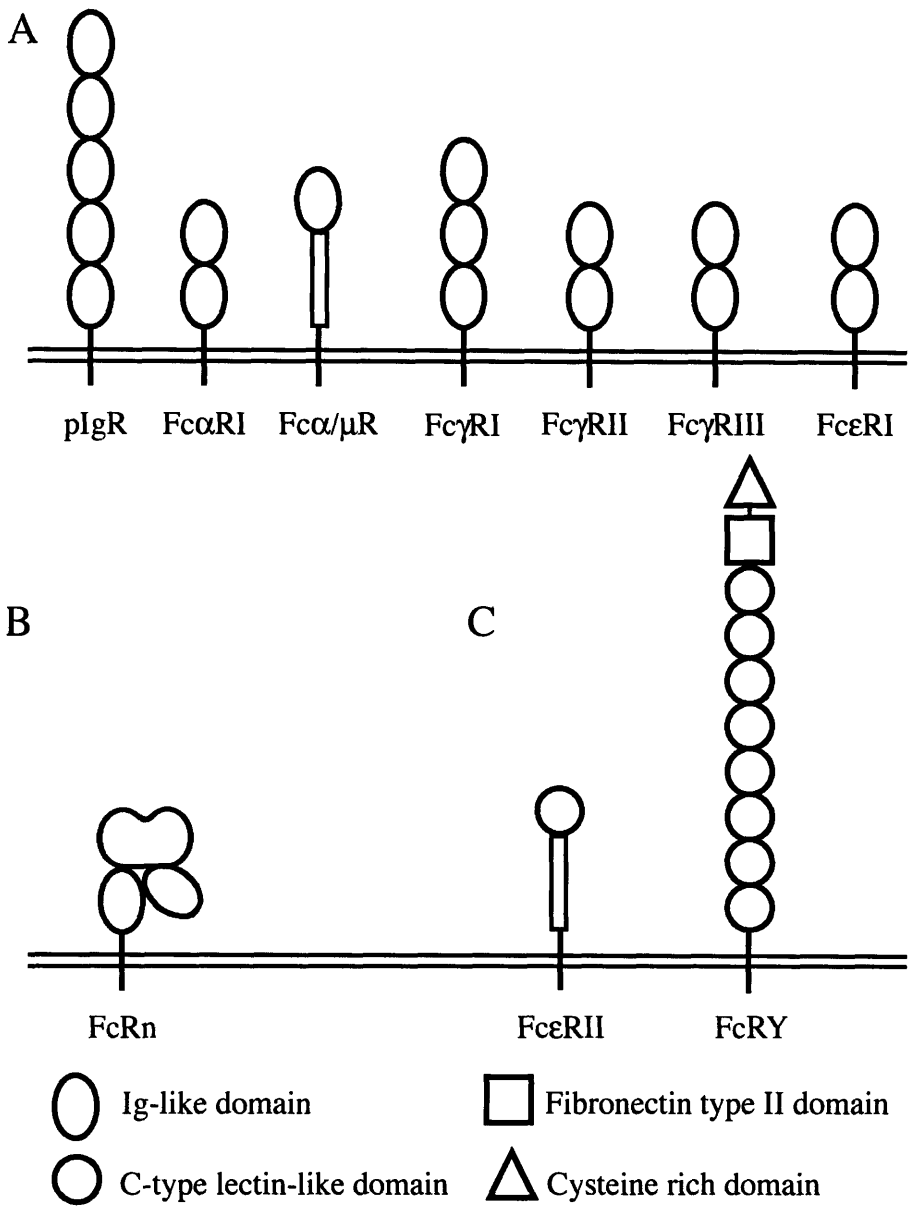


Figure 4. Schematic representation of the three structural groups of Fc receptors. The first group is composed of tandem Ig domains, and include the FcγRs (FcγRI, FcγRII and FcγRIII), the high affinity IgE receptor (FcεRI), the IgA specific receptor (FcαRI), and two receptors that bind both IgA and IgM (pIgR and Fcα/μR) (A). FcRn resembles MHC class I homologs (B) and the third group belongs to the C-type lectin superfamily and includes the low-affinity IgE receptor (FcεRII) (FcεRII) and the chicken FcY receptor (FcRY).

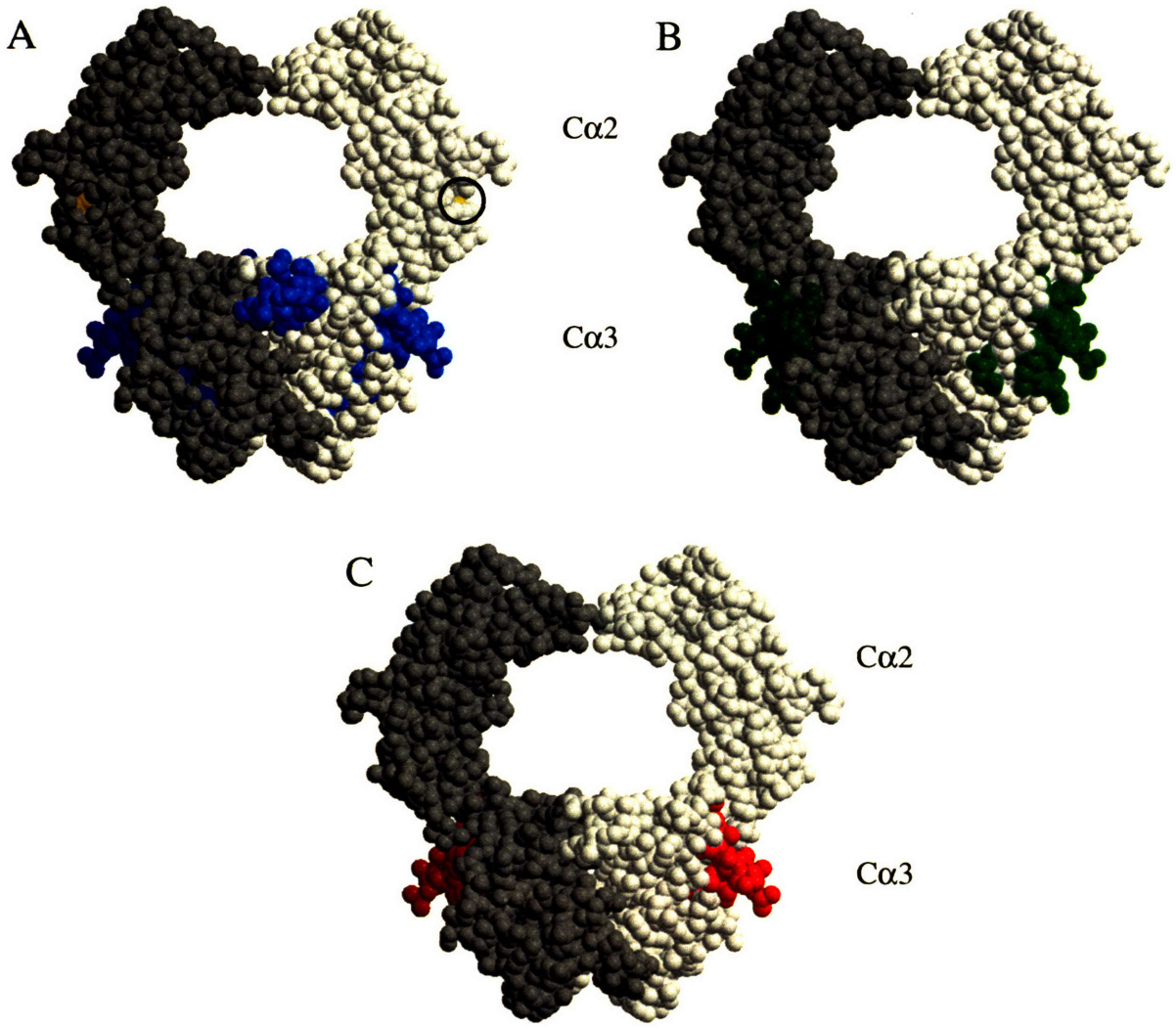


Figure 5. Receptor binding sites on Fcα. The two monomers in the Fcα homodimer structure are colored grey and light grey. The residues in the Cα3 domain implicated in binding to pIgR D1 are shown in blue (A) and the residues involved in binding to FcαRI are shown in green. Overlapping residues in the two binding sites are shown in red (C). Circles in (A) indicate Cys311 (yellow) in Cα2, the residue that forms a disulfide bond with pIgR D5.

CHAPTER 2

Crystal Structure of a Polymeric Immunoglobulin-Binding Fragment of the Human Polymeric Immunoglobulin Receptor

This chapter describes the 1.9 Å resolution crystal structure of pIgR D1 and characterization of the interaction between D1 and dimeric IgA. My contribution to this work included construction of the pIgR D1 and Fcα-tp expression vectors, establishing a stable mammalian cell line expressing Fcα-tp, protein purification, performing the binding experiments, crystallization of D1, data collection and solving the structure of D1. Dr. Anthony P. West Jr. taught me how to use the biosensor instrument and the crystallographic software used to solve the structure.

Crystal Structure of a Polymeric Immunoglobulin Binding Fragment of the Human Polymeric Immunoglobulin Receptor

Agnes E. Hamburger,[1,3] Anthony P. West, Jr.,[1] and Pamela J. Bjorkman[1,2,*]

[1]Division of Biology 114-96

[2]Howard Hughes Medical Institute

California Institute of Technology

Pasadena, California 91125

[3]Division of Biology

Massachusetts Institute of Technology

Cambridge, Massachusetts 02142

*To whom correspondence should be addressed:

Tel.: 626-395-8350; Fax: 626-792-3683; E-mail: bjorkman@caltech.edu

Running Title: Structure of a pIg Binding Fragment of pIgR

Summary

The polymeric immunoglobulin receptor (pIgR) is a type I transmembrane protein that delivers dimeric IgA (dIgA) and pentameric IgM to mucosal secretions. Here, we report the 1.9 Å resolution X-ray crystal structure of the N-terminal domain of human pIgR, which binds dIgA in the absence of other pIgR domains with an equilibrium dissociation constant of 300 nM. The structure of pIgR domain 1 reveals a folding topology similar to immunoglobulin variable domains, but with differences in the counterparts of the complementarity determining regions (CDRs), including a helical turn in CDR1 and a CDR3 loop that points away from the other CDRs. The unusual CDR3 loop position prevents dimerization analogous to the pairing of antibody variable heavy and variable light domains. The pIgR domain 1 structure allows interpretation of previous mutagenesis results and structure-based comparisons between pIgR and other IgA receptors.

Introduction

Polymeric immunoglobulins (pIgs) at mucosal surfaces provide the first line of defense against pathogens and toxins. Polymeric IgA (mainly dimeric IgA; dIgA) is the predominant Ig found in secretions, with pentameric IgM (pIgM) present at lower levels (Norderhaug et al., 1999b). Unlike other antibody isotypes, IgA and IgM can form polymers via an 18-residue extension at their C termini called the tailpiece. The joining (J) chain, a 15 kDa polypeptide, promotes the oligomerization of IgA and IgM to form pIgs by cross-linking the homodimeric Fc regions of two or five antibody molecules, respectively (Halpern and Koshland, 1970; Mestecky et al., 1971). In dIgA, J-chain forms two intermolecular disulfide bonds, the first with one of two tailpieces in the first Fc homodimer and the second with a tailpiece from the second Fc

homodimer. The two IgA heavy chains that are not covalently attached to J-chain form a disulfide bond to each other via Cys471 in their tailpieces (Bastian et al., 1992).

The polymeric immunoglobulin receptor (pIgR) transports pIgs across mucosal epithelia into mucosal secretions. pIgR is expressed on the basolateral surface of epithelial cells, where it binds selectively to J-chain-containing pIgs secreted by local plasma cells (Brandtzaeg and Prydz, 1984; Radl et al., 1971; Vaerman et al., 1998a). Human pIgR binds and transports both dIgA and pIgM, whereas pIgRs from other species (e.g., rabbit, rodents and chicken) only bind dIgA (Socken and Underdown, 1978; Underdown et al., 1992; Wieland et al., 2004). Biochemical and mutagenesis studies revealed that pIgR binding to dIgA takes place in two steps. First, the N-terminal pIgR domain (D1) makes a noncovalent interaction with the Cα3, and possibly Cα2, domains of one of the Fc regions of dIgA (Frutiger et al., 1986; Geneste et al., 1986; Hexham et al., 1999). In the second step, Cys467 in the C-terminal domain of the extracellular portion of human pIgR (D5) makes a disulfide bond with Cys311 in the Cα2 domain of the second IgA molecule (Fallgreen-Gebauer et al., 1993). pIgR/pIg complexes are transcytosed to the apical surface of the epithelial monolayer, where the pIgR/pIg complex is released from the membrane by the proteolytic cleavage of pIgR, forming secretory Ig (SIg). The cleaved ectodomain of human pIgR, also known as secretory component (SC), is covalently attached via a disulfide bond to dIgA or noncovalently to pIgM (Brandtzaeg, 1975; Mostov et al., 1980). Free SC that is not complexed to a pIg is also released into secretions (Brandtzaeg, 1973).

pIgR is a glycosylated type I transmembrane protein, consisting of a 620 residue extracellular region, a 23 residue transmembrane region and a 103 residue cytoplasmic tail. The extracellular region contains five domains (domains 1-5; D1-D5) that share sequence similarity

with Ig variable (V) regions (Mostov et al., 1984). Thus, the pIgR D1-D5 domains are predicted to be β sandwich structures each containing two β sheets (composed of strands A, B, E, and D and strands C″, C′, C, F, G and A′ in the case of D1 –D4, and strands A, B, E, D and strands A′, C, F, and G in the case of D5). D1-D4 each contain regions homologous to the three antigen binding complementarity-determining region (CDR) loops of Ig variable domains.

Mutagenesis and peptide binding studies have identified probable binding sites on pIgR and dIgA. pIgR D1 is necessary and sufficient for binding to dIgA (Frutiger et al., 1986) and all three of its CDR loops have been implicated in the interaction with the antibody (Coyne et al., 1994). Potential pIgR binding sites on dIgA involve the dIgA Cα3 domain and include residues 402-410 in the loop connecting strands D and E and residues 430-443 in the FG loop (Hexham et al., 1999; White and Capra, 2002). Since pIgR preferentially binds and transcytoses dIgA molecules containing J-chain, pIgR D1 may also directly interact with J-chain (Brandtzaeg and Prydz, 1984; Vaerman et al., 1998a; Vaerman et al., 1998b). In order to contribute to the molecular description of the interaction between pIgR and pIgs, we determined the 1.9 Å crystal structure of a pIg binding fragment of the human pIgR consisting of the N-terminal or D1 domain. The structure reveals an Ig variable-like domain that allows interpretation of previous mutagenesis studies and comparison with other IgA receptors.

Results and Discussion

Biosensor Binding Experiments

The ligand binding domain of human pIgR, an Fc receptor that is specific for dimeric and polymeric forms of IgA and IgM, was expressed in *E. coli* and refolded from inclusion bodies. A glycosylated form of the protein was also expressed in baculovirus-infected insect cells. D1

was tested for binding to dimeric and monomeric versions of the IgA Fc region. For the dimeric version of IgA, we used dIgA/J-chain complexes purified from human serum (Vaerman et al., 1995), and for the monomeric version, we used recombinant IgA Fc regions expressed in CHO cells. One Fc region was expressed without the tailpiece (Fcα) as described (Herr et al., 2003b), and the other Fc region included the tailpiece (Fcα–tp). Binding was evaluated using a surface plasmon resonance-based binding assay. Both the glycosylated and nonglycosylated forms of D1 bind to dIgA with an equilibrium dissociation constant (K_D) of ~300 nM (Figure 1A), consistent with other binding studies (Bakos et al., 1994). Both versions of D1 show greatly reduced or negligible binding to the monomeric Fcα and Fcα-tp proteins at concentrations up to 5 μM (Figure 1B). By contrast, another IgA Fc receptor, FcαRI, binds dIgA, Fcα-tp and Fcα when injected at a concentration of 5 μM (data not shown) (Herr et al., 2003b).

Structure of pIgR D1

The 1.9 Å crystal structure of the human pIgR D1 was solved by multiple isomorphous replacement with anomalous scattering (MIRAS). As predicted by sequence analysis, the folding topology of pIgR D1 resembles the topologies of Ig variable heavy (V_H) and variable light (V_L) domains (Mostov et al., 1984). Like Ig variable domains, pIgR D1 contains ten β strands assembled into two antiparallel β sheets with strands A, B, E and D on one face and C″, C′, C, F, G and A′ on the other (Figure 2A-2C). Five residues that are characteristic of Ig regions (Williams and Barclay, 1988) are found in the expected positions in pIgR D1: Cys22 and Cys92, which form a disulfide bond linking the two β sheets, Trp37, the "invariant" tryptophan (Figure 2D) that packs into the hydrophobic core, and Arg63 and Asp86, which form a salt bridge (Figure 3). In addition, pIgR D1 contains a second disulfide bond between Cys38 and

Cys46 that link the C and C′ strands, which is also found in NKp44, a natural killer cell cytotoxicity activating receptor (Cantoni et al., 2003).

Antibody V_H and V_L domains each contain three hypervariable complementarity-determining region (CDR) loops (CDR1, CDR2, and CDR3), which together form the antigen combining site in a V_H-V_L heterodimer (Alzari et al., 1988). Although the pIgR D1 structure closely resembles an isolated Ig variable domain, the positions of the three CDR loops in D1 differ substantially from their antibody counterparts. Comparative analyses of the CDR regions in antibody structures have allowed classification of CDR loops into a set of canonical structures (Chothia and Lesk, 1987). Although the sequence of the D1 CDR1 loop shares similarity with sequences found in type 1 of the six canonical CDR1 structures in Ig Vκ domains (Bakos, 1993), the CDR1 structure in pIgR D1 differs considerably. Unlike any of the canonical structures for CDR1 loops in V_L or V_H domains (Al-Lazikani et al., 1997), the pIgR D1 CDR1 includes a single helical turn composed of residues that are highly conserved among sequences of pIgR from different species (Figure 3). By contrast, CDR1 loops in V_H and V_L domains are composed of extended structures joined by short links or hairpin turns (Chothia et al., 1989). CDR2 and CDR3 in pIgR D1 do not appear to resemble any of the canonical sequence or structure patterns. The pIgR CDR2 loop is very short, with only two residues in the loop region between the C′ and C″ strands. By contrast, CDR2 loops in Vκ domains consist of three residues that form a classic γ-turn (Al-Lazikani et al., 1997). Unlike its position in antibody V_H and V_L domains, the CDR3 loop of pIgR D1 is tilted toward the C″C′CFGA′ sheet, away from the other CDRs (Figure 4A). A conserved hydrophobic residue, Tyr36, is buried at the interface of the D1 CDR3 and the C″C′CFGA′ sheet (Figure 4B). The position of the CDR3 loop is stabilized by hydrogen bonds between Asn97 within the loop and two residues on the C″C′CFGA′ sheet: Arg34 in the C strand

and Thr48 in the C′ strand. The observed conformation of the D1 CDR3 loop is likely to represent its conformation in solution because the loop is not involved in crystal contacts. The different orientation of the pIgR CDR3 loop as compared to the CDR3 loops of Ig V domains (Figure 4A) is consistent with why D1 does not form dimers analogous to antibody V_H-V_L heterodimers (Figure 4C) or V_L-V_L homodimers. When pIgR D1 is superimposed on an Ig V_H-V_L combining site (Zdanov et al., 1994) (Figure 4C) to create a D1 dimer, the CDR3 regions clash with the CC′ loop from the partner molecule (Figure 4D). Although two different dimers of pIgR D1 are observed in the crystal packing, neither is analogous to Ig V_H-V_L dimers. The D1 dimers in the crystals are unlikely to be biologically relevant because there are two different dimeric arrangements and the D1 protein migrates in the position expected for a monomer on a gel filtration column (data not shown).

Although the protein used for our structural studies is not glycosylated because it was expressed in bacteria, the pIgR D1 structure can be used to locate residues to which carbohydrate would normally be attached. Human pIgR D1 contains two N-linked carbohydrates attached to Asn65 and Asn72 (Hughes et al., 1999). These residues are located on the D strand and the DE loop, respectively, and their sidechains are accessible to solvent (Figure 2B). Studies comparing the binding of bacterially expressed nonglycosylated D1 with glycosylated D1 expressed in insect cells revealed nearly identical binding affinities for dIgA (Figure 1A), consistent with earlier studies showing that pIgR glycosylation is not necessary for specific binding to dIgA (Bakos et al., 1991). Carbohydrates may be required, however, for efficient transport or release of the pIgR ectodomain during transcytosis of dIgA, since nonglycosylated pIgR is released from the apical surface at low levels (Matsumoto et al., 2003). In addition, carbohydrates on SC

contribute to stabilization of SIgs by protecting it from proteolytic degradation (Crottet and Corthesy, 1998).

Structure-Based Interpretations of Binding Studies

Previous mutational and peptide-mapping experiments to define the regions of pIgR that form the binding sites for dIgA and pIgM can now be interpreted using the pIgR D1 structure. CDR1 of pIgR D1 was proposed to make essential contacts with dIgA because it is the most highly conserved region of D1 among different species of pIgR, and a synthetic peptide corresponding to residues 15-37 of SC, a region that contains the D1 CDR1 loop, is capable of binding pIgs (Bakos et al., 1991). Interestingly, this peptide binds Igs indiscriminately, interacting with pIgs as well as with monomeric IgA and IgG (Bakos et al., 1991). The promiscuous binding of this peptide suggests that other parts of D1, in addition to the CDR1 loop, likely contribute to pIg binding specificity. The D1 structure reveals a solvent-exposed helix within the CDR1 loop containing residues with the potential to make direct contacts with pIgs (Figure 5).

Point mutations in the CDR1 region of the rabbit pIgR D1 highlighted the importance of three charged residues (two arginines and a lysine, which correspond to Arg31, Arg34 and Lys35 in the sequence of human pIgR D1). Substitution of these residues with alanine abolished pIgR binding to dIgA (Coyne et al., 1994). The human pIgR D1 structure shows that Arg31 (rabbit Arg37) is solvent exposed and could interact directly with dIgA. Although Arg34 and Lys35 were predicted to be in the CDR1 loop (Coyne et al., 1994), an analysis of main chain hydrogen bonds in the D1 structure shows that they are both in the C strand. Arg34 is the first residue in the strand, as its main chain carbonyl group forms a hydrogen bond with the main chain amino group of Gly95 in the F strand. Human D1 Arg34 (rabbit Arg40) points in the opposite direction from Arg31 and is occluded by the CDR3 loop, forming a hydrogen bond to CDR3 loop residue

Asn97 (Figure 4B). Arg34 is thus unlikely to interact directly with a ligand, but may affect ligand binding indirectly by stabilizing the CDR3 loop. This interpretation is consistent with the results of mutagenesis experiments involving rabbit pIgR in which replacing the rabbit counterpart of Arg34 with a glutamic acid, a residue that can still form a hydrogen bond with Asn97 in CDR3, had no affect on ligand binding, whereas substitution to an alanine, which cannot form the hydrogen bond, abolished binding (Coyne et al., 1994). Lys35 (rabbit Lys41) is buried between the two β sheets and is also unlikely to represent a direct point of contact with the antibody. However, it stabilizes the position of the CDR1 loop by forming hydrogen bonds with Arg31 and Thr33 in the conserved helical turn in the CDR1 loop (Figure 5). Single point mutations in surrounding residues corresponding to human pIgR residues Thr27, Ser28, Asn30, His32 and Thr33 resulted in decreased dIgA binding (Coyne et al., 1994). Residues Ser28 through His32 comprise the helical turn in CDR1 and, together with the adjacent Thr27 and Thr33, are surface exposed in positions that could interact with dIgA directly. The rabbit counterpart of human D1 Val29 was not substituted because it was predicted to be buried (Bakos et al., 1993; Coyne et al., 1994), however the crystal structure reveals that it is a surface-exposed residue located on the helix within the CDR1 loop (Figure 5). Although its role in binding dIgA is difficult to predict because it points in the opposite direction from another critical residue, Arg31, solvent-exposed hydrophobic residues are often located at binding interfaces (Dall'Acqua et al., 1996; Kelley and O'Connell, 1993; Lebron and Bjorkman, 1999; Tsumoto et al., 1995; Vaughn et al., 1997; Wells and de Vos, 1996).

In addition to CDR1, mutagenesis and binding experiments using rabbit pIgR have implicated the other CDRs in binding dIgA. Replacing the CDR2 or CDR3 regions in D1 with their corresponding regions from D2 results in complete loss of dIgA binding (Coyne et al.,

1994). The human pIgR D1 structure reveals that CDR3 is positioned away from the other CDRs (Figure 2B), unlike the CDR3 loops of Ig V domains, which form a contiguous binding surface with CDRs 1 and 2 (Alzari et al., 1988). It may therefore be difficult for a ligand contacting the CDR1 and CDR2 regions of pIgR D1 to simultaneously contact CDR3. Thus it is possible that the loss of affinity upon replacing the D1 CDR3 with another sequence may be due to steric interference with dIgA binding rather than alteration of a direct binding surface. The results of other loop replacements are harder to interpret using the pIgR D1 structure: replacing the rabbit D1 EF loop with its counterpart in D2 led to the loss of dIgA binding, whereas replacing the C"D loop had little effect (Coyne et al., 1994). Both loops are on the opposite side of the D1 domain from the CDRs (Figure 5) and are not expected to contact the ligand.

Although human pIgR binds both dIgA and pIgM, rabbit pIgR only binds dIgA (Roe et al., 1999; Socken and Underdown, 1978). By taking advantage of these binding differences, studies using site-directed mutants and human/rabbit chimeric pIgR molecules have identified regions required for binding pIgM (Roe et al., 1999). All three CDR loops were shown to contribute to binding pIgM, but to different extents. For maximal binding of rabbit pIgR to pIgM, replacing all three CDR loops with their human counterparts was required, but CDR2 appears to be the most critical: replacing rabbit CDR2 with the human equivalent transferred pIgM binding, and replacing human CDR2 with the rabbit sequence resulted in substantial loss of binding (Roe et al., 1999). Sequence alignment shows large differences in the human and rabbit CDR2 loops, including a deletion of a charged residue, Glu54, in the rabbit CDR2 (Figure 3). Thus, deletion of the surface-exposed Glu54 may eliminate a point of electrostatic interaction between pIgR and pIgM. Interestingly, only bovine and human pIgR D1 regions contain the

glutamic acid in their CDR2 loops (Figure 3), and these are the only species of pIgR that show high affinity pIgM binding (Socken and Underdown, 1978).

The Human pIgR D1 Structure as a Model for Similar Domains

Amino acid sequence alignment of pIgR D1 sequences from eight different species reveals conservation of residues identified as critical for the human pIgR D1 structure (Figure 3), thus the human pIgR D1 structure can be used as a first-order model for the D1 domains of pIgR proteins from other species. In addition to the five residues characteristic of Ig V domains (asterisks in Figure 3), the cysteines that participate in the second disulfide bond and the three CDR loops (which are 67-100% conserved in CDR1, 50-100% in CDR2 and 40-80% in CDR3), residues that participate in anchoring the CDR1 and CDR3 loops are also highly conserved. All of the residues comprising the helical turn in CDR1 are either conserved or conservatively substituted in all eight species, and Lys35, a buried residue that stabilizes the helix, is also conserved (Figure 3). Tyr36, the residue buried under the CDR3 loop (Figure 4B), is conserved in 7 of 8 species of pIgR, and conservatively substituted by a phenylalanine in the eighth species (rabbit) (Figure 3). Three other residues that interact to stabilize the CDR3 loop position are either conserved or conservatively substituted in D1 sequences: CDR3 residue Asn97 (a polar residue in all sequences), Arg34 (conserved), and Thr48 (conserved) (Figure 3). Conservation of these critical residues suggests that the helical turn in CDR1 and the unusual CDR3 location observed in the human pIgR D1 structure are preserved in other pIgR D1 domains. The most notable difference in the D1 sequence alignments is seen in the CC' loop in all of the sequences.

The pIgR D1 structure can also be used to make predictions about the structures of the remaining domains in the human pIgR extracellular region. The characteristic disulfide bond seen in members of the Ig superfamily (Williams and Barclay, 1988) is retained in all five

domains. Also, pIgR domains 3, 4 and 5 each contain cysteines in positions to make the second disulfide bond that bridges the C and C′ β strands of D1 (Figure 2C). In addition, D5 contains a third disulfide bond that rearranges as a result of the covalent association with dIgA, whereby Cys467 of pIgR is linked to Cys311 in the IgA heavy chain (Fallgreen-Gebauer et al., 1993). The characteristic salt-bridge in Ig-like domains, which involves Arg63 and Asp86 in D1, is conserved in D2-D4, and the "invariant" tryptophan is present in D4 and D5 (Figure 3). By contrast with these conserved sequence features, which suggest an overall similar folding topology to that observed for the human D1 structure, the three CDR loops in human pIgR D1 do not share significant sequence identity with their counterparts in D2-D5 (Figure 3), suggesting local differences in loop structures. Such differences are not unexpected since the CDR loops are implicated in binding to dIgA, and only D1 can bind to dIgA in isolation (Frutiger et al., 1986). Overall, the sequences of D2-D5 are more variable than D1 across species (Piskurich et al., 1995). This observation is not surprising given the differences across species that D2-D5 make to dIgA binding. In the case of human pIgR, D2 and D3 enhance pIgR's affinity for dIgA (Norderhaug et al., 1999a). The presence of D2-D3 in bovine pIgR also increases ligand binding affinity (Beale, 1988), but D2-D5 in rabbit pIgR do not contribute significantly to dIgA binding (Frutiger et al., 1986). Thus, the full length ectodomain (D1-D5), an alternatively spliced version (lacking D2-D3), and D1 of rabbit pIgR bind to pIgs with similar affinities (Deitcher and Mostov, 1986; Frutiger et al., 1986). Murine pIgR D2-D3 are not necessary for high affinity noncovalent binding to dIgA (Crottet and Corthesy, 1999), and chicken pIgR contains only four extracellular Ig-like domains, with no ortholog to D2 of mammalian pIgRs (Wieland et al., 2004).

Sequence alignment suggests that human pIgR D1 also shares structural similarities with the corresponding region of another IgA and IgM receptor, the Fcα/μ receptor (Fcα/μR) (Figure 3) (Shibuya et al., 2000). Fcα/μR, which can bind to monomeric and polymeric IgA and IgM, but not IgG, is expressed by the majority of B lymphocytes and macrophages (McDonald et al., 2002; Shibuya et al., 2000). Human pIgR D1 and the Ig-like domain of Fcα/μR share 43% sequence identity. The five characteristic residues in Ig V domains (Figure 3) and the second disulfide bond linking strands C and C′ are conserved between pIgR D1 and Fcα/μR, suggesting similar tertiary structures. In addition, Fcα/μR shares 67% sequence identity with pIgR D1 in one of the identified ligand binding sites, the CDR1 loop, including complete conservation of all of the residues that comprise the helical turn within this loop in pIgR D1. The unusual CDR3 position is potentially another feature shared by the two proteins since the buried D1 residues Tyr36, Arg34 and Thr48 are conserved in Fcα/μR, and D1 Asn97 is replaced by a glutamic acid. A hydrogen bond between Fcα/μR Glu97 and Thr48 (analogous to the Asn97 to Thr48 hydrogen bond in pIgR D1) and a salt bridge between Fcα/μR Glu97 and Arg34 (replacing the hydrogen bond in pIgR D1 between Asn97 and Arg34) (Figure 4B) would preserve the downward orientation of the CDR3 loop. The overall conservation of structurally important residues within the CDR loops of Fcα/μR and pIgR D1 suggests a similar mode of interaction with the common ligands of these receptors.

Another IgA receptor, FcαRI, shares an overlapping binding site on the IgA Fc region with pIgR that includes the FG loop of Fcα (Herr et al., 2003a; Hexham et al., 1999; White and Capra, 2002). The overlap of the binding site has been suggested (Herr et al., 2003a) to explain why secretory IgA (dIgA plus the pIgR ectodomain) cannot bind to or activate FcαRI in the absence of an integrin co-receptor (van Egmond et al., 2000; Vidarsson et al., 2001). Although

pIgR and FcαRI appear to recognize at least some of the same portion of Fcα, their structures do not share detailed common features beyond the fact that both molecules are Ig superfamily members. The crystal structure of an FcαRI/Fcα complex shows that the site on FcαRI at the interface with Fcα involves residues in the FcαRI BC loop, the D strand, the DE loop, and the FG loop (Herr et al., 2003a). The BC and FG loops in FcαRI are topologically equivalent to CDR1 and CDR3, respectively, but are not technically CDRs because FcαRI does not resemble an Ig variable domain. Further, the two Ig-like domains of FcαRI lack C″ strands and therefore a loop equivalent to the CDR2 region. Thus, although pIgR and FcαRI bind to the same or to an overlapping site on Fcα, they do so with different recognition modes and folding topologies. Further structural studies of pIgR in complex with dimeric Fcα will be required to compare the recognition properties of pIgR, Fcα/μR, and FcαRI, and to fully understand the mechanism by which pIgR is specific for the binding and transport of polymeric Igs.

Experimental Procedures

Protein Expression and Purification

The cDNA encoding the full-length human pIgR was kindly provided by Roland Strong (Fred Hutchinson Cancer Research Center, Seattle, WA). For bacterial expression, pIgR D1 (encoding residues Lys1-Val109) was subcloned into the pET20b expression vector (Novagen) in frame with the C-terminal 6x-His tag using the following primers: 5'-G GAA TCC CAT ATG AAG AGT CCC ATA TTT GGT CC-3' and 5'-GG AAT TCA CTC GAG GAC CTC CAG GCT GAC-3'. The protein was expressed in *E. coli* BL21(DE3) cells (Novagen) by induction at OD_{600}=0.5-0.6 with isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 0.4 mM for 5 hours at 37°C. pIgR D1 inclusion bodies were solubilized in 8 M guanidine-HCl and 10 mM

DTT. The protein was refolded by the rapid dilution method in refolding buffer (100 mM Tris [pH 8.0], 400 mM L-arginine, 2 mM EDTA, 0.5 mM oxidized glutathione and 5 mM reduced glutathione) (Garboczi et al., 1992). The refolded protein was concentrated in a stirred-cell pressurized concentrator (Amicon) and loaded onto a Superdex 75 26/60 column (Amersham Biosciences) for size exclusion chromatography.

For insect cell expression to produce glycosylated protein, pIgR D1 containing a C-terminal 6x-His tag was subcloned into the baculovirus transfer vector pAcGP67A (BD Biosciences) in frame with the gp67 secretion signal. Recombinant baculovirus was generated by cotransfection of the transfer vector with linearized viral DNA (Baculogold; BD Biosciences). pIgR D1 was harvested from the supernatant of baculovirus infected High 5 insect cells, which was concentrated and buffer exchanged into TBS (20 mM Tris [pH 8.0], 150 mM NaCl) and purified by Ni-NTA affinity followed by size exclusion chromatography on a Superdex 200 16/60 column (Amersham Biosciences).

Human pIgA isolated from the sera of a patient with a pIgA-producing myeloma was kindly provided by Jean-Pierre Vaerman (Catholic University of Louvain, Brussels, Belgium) (Song et al., 1995; Vaerman et al., 1995). pIgA was further purified by size exclusion chromatography on a Superdex 200 HR 10/30 column (Amersham Biosciences) to separate dIgA from higher order polymers. The peak corresponding to dIgA (two IgA molecules linked by J-chain) was used for the binding studies. Soluble FcαRI and monomeric Fcα (Fcα homodimers lacking a tailpiece that are not linked by J-chain) was purified from the supernatants of stably transfected Chinese hamster ovary (CHO) cells as described previously (Herr et al., 2003b). Monomeric Fcα with the 18 residue tailpiece (Fcα-tp) and an N-terminal 6x-His tag was generated and purified as described for Fcα (Herr et al., 2003b).

Protein concentrations were determined spectrophotometrically at 280 nm using extinction coefficients of 13,370 $M^{-1}cm^{-1}$ for pIgR D1 and 64,940 $M^{-1}cm^{-1}$ for monomeric Fcα and Fcα-tp, calculated from their amino acid sequences using the ProtParam tool on the ExPASy Proteomics Server (Gill and von Hippel, 1989).

Binding Studies

Surface plasmon resonance (SPR) biosensor assays were carried out using a BIAcore 2000 instrument (Pharmacia Biosensor, Uppsala, Sweden). In this system, binding between a molecule coupled to a biosensor chip (the "ligand") and a second molecule injected over the chip (the "analyte") results in changes in the SPR signal that are read out in real time as resonance units (RUs) (Malmqvist, 1993). dIgA was covalently coupled to a reagent-grade CM5 sensor chip (Biacore) at three different densities (412, 763 and 1426 RUs) using the primary amine coupling method described in the BIAcore manual. The first flow cell was mock coupled with buffer only for background subtraction. A two-fold dilution concentration series (from 10.24 μM to 20 nM glycosylated D1 or from 6.4 μM to 12.5 nM nonglycosylated D1) of 180 μL of pIgR D1 (the analyte) was injected over the chip at 5 μL/min in 50 mM HEPES (pH 8.0), 150 mM NaCl and 0.005% (v/v) P20, and the binding reactions were allowed to closely approach or to reach equilibrium. Two-minute injections of 1.5 M $MgCl_2$ were used to regenerate the surface of the chip between injections. The sensorgrams were processed and analyzed with the Scrubber software package (BioLogic Software, Campbell, Australia). Equilibrium dissociation constants (K_Ds) were derived by nonlinear regression analysis of plots of R_{eq} (the equilibrium binding response) versus the log of the concentration of analyte, and the resulting binding data were fit to a single-site binding model. Data collection and binding analyses were performed identically for the glycosylated and nonglycosylated forms of pIgR D1. To compare the binding of pIgR D1 to

monomeric and dimeric versions of the Fc region of IgA, Fcα, Fcα-tp, and dIgA were coupled to a CM5 biosensor chip at densities of 2457 RUs, 1411 RUs and 3153 RUs, respectively, as described above. 100 μL of 5 μM glycosylated D1, nonglycosylated D1, or FcαRI were injected over the chip at 50 μL/min in 50 mM HEPES (pH 8.0), 150 mM NaCl, 0.005% (v/v) P20.

Crystal Growth and Data Collection

Crystals were grown by vapor diffusion in 1:1 hanging drops containing bacterially-expressed pIgR D1 (~10 mg/mL in 20 mM HEPES (pH 7.0), 150 mM NaCl), 20% (w/v) polyethylene glycol 8000 (PEG 8000), 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate pH 6.5 and reproduced by streak seeding. Before data collection, crystals were cryopreserved in 23% (w/v) PEG 8000, 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate (pH 6.5) with 15% (v/v) glycerol. Heavy atom derivatives were prepared by soaking crystals in the following solutions: 1 mM di-μ-iodobis (ethylenediamine) diplatinum (II) nitrate (PIP), 40 mM trimethyl lead acetate, or 100 mM gadolinium (III) chloride. Data from a native crystal and the PIP derivative were collected at –170°C using an R-AXIS IV mounted on a Rigaku RU-200 rotating anode generator. This native data set (Native I) was used for initial phase determination and model building. A higher resolution native data set (Native II), which was used for refinement, and the lead and gadolinium derivative data sets were collected on beamline 8.2.2 at the Advanced Light Source (ALS, Berkeley, CA) at –170°C (Table 1).

Structure Determination and Refinement

Data were processed and scaled with DENZO and SCALEPACK (Otwinowski and Minor, 1997). The data could be indexed in either primitive monoclinic ($P2_1$) or C-centered orthorhombic ($C222_1$) space groups with unit-cell parameters a = 42.1 Å, b = 156.4 Å, c = 53.9 Å; β = 113.0° or a = 42.1 Å, b = 99.2 Å and c = 156.4 Å, respectively. The overall scaling

statistics for $C222_1$ were significantly worse than for $P2_1$ (R_{sym} = 14.3% (47.7%) versus R_{sym} = 8.5% (34.3%)) and the h0l plane did not show mm symmetry, indicating the crystals were monoclinic. After examination of systematic absences in the 0k0 reflections, the space group was assigned as $P2_1$. The asymmetric unit contains six molecules with a solvent content of ~40% (V_M=2.1 $Å^3$/Da) as calculated from the Matthews coefficient (Matthews, 1968).

For merging the native and derivative data sets, two options for indexing had to be considered. Two identical but nonequivalent unit cells can be defined in monoclinic space groups when cosβ=-a/(2c) (Rudolph et al., 2004), a condition met by the D1 unit cell parameters. Each derivative data set was compared to two versions of the Native I data set: as originally indexed, and reindexed using the transformation (h, k, l) to (h, -k, -h-l). The correct indexing of each derivative as compared to the native I data was determined by comparing R_{merge} values, which were systematically higher when the opposite indexing had been chosen for the two data sets being compared (Tucker et al., 1996).

In addition to indexing ambiguities, the D1 diffraction data exhibited evidence of pseudomerohedral twinning, a rare type of twinning that can occur in monoclinic and orthorhombic crystal systems when the unit cell parameters meet certain conditions (Rudolph et al., 2004). One such condition, as discussed above, is when cosβ=-a/(2c), in which case twinned crystals in the monoclinic system can mimic an orthorhombic space group. Since the D1 crystal parameters meet this condition, all data sets were checked for twinning. The cumulative distribution of intensities (I/<I>), as calculated with TRUNCATE (Collaborative Computational Project No. 4., 1994) showed a sigmoidal shape, suggesting twinning in all of the crystals. However, the second moment of intensities ($<|I|^2>/(<|I|>)^2$) for some of the data sets was 2.0, which is the expected value for untwinned data. For the data sets with the seemingly

contradictory twin test results, further examination of a related value ($<|F|>^2/<F^2>$) confirmed twinning (Brunger et al., 1998). The estimated twin fractions calculated with CNS (Brunger et al., 1998) ranged from 20.7% to 37.3% (Table 1). Similar cases of pseudomerohedreal twinning in monoclinic crystal forms have been described by others (De La Fortelle and Bricogne, 1997; Larsen et al., 2002; Rudolph et al., 2004).

Heavy atom positions and initial phases were derived with the programs SOLVE (Terwilliger and Berendzen, 1999) and SHARP (De La Fortelle and Bricogne, 1997; Terwilliger and Berendzen, 1999) using the Native I, trimethyl lead acetate, gadolinium chloride and PIP derivative data sets with no corrections for twinning. An initial electron density map was calculated to 2.7 Å and solvent flattened with Solomon (Collaborative Computational Project No. 4., 1994). The first four molecules in the asymmetric unit were located in the initial experimental map using the program MOLREP (Vagin and Teplyakov, 1997) and a "β strands-only" model of NKp44 (PDB code 1HKF) (Cantoni et al., 2003) (nonconserved side chains truncated to alanine and residues 5-9, 14-15, 25-31, 40-45, 53-55, 69-71, 80-85, 95-101 and 112 omitted). The remaining two molecules were located using the real space search program ESSENS (Kleywegt and Jones, 1997). Maps were calculated by solvent flattening and histogram matching with DM in CCP4 (Collaborative Computational Project No. 4., 1994). NKp44 (Cantoni et al., 2003) served as a starting point for model building with the program O (Jones and Kjeldgaard, 1997).

Refinement was performed using the CNS suite of programs (Brunger et al., 1998). The test set of reflections for calculating R_{free} was generated with the thin shell method in DATAMAN (Kleywegt and Jones, 1996) to minimize the bias from the 6-fold noncrystallographic symmetry (NCS) and twin-related reflections. Initial refinement was

conducted using the 2.5 Å Native I data with NCS constraints, grouped temperature (B) factors, bulk solvent and anisotropic temperature-factor corrections. In subsequent refinement using the 1.9 Å Native II data set, the NCS constraints were relaxed to NCS restraints (300 kcal/mol $Å^2$) and individual B factors were calculated, taking into account the twin fraction of 37.3% and maintaining the same set of test reflections for calculating R_{free}. NCS restraints were limited to the main chain atoms of the β strands, excluding regions that differed in the six molecules (mostly at crystal contacts and in the loops). R_{cryst} and R_{free} improved by 5.1% and 3.4%, respectively, after inclusion of the corrections for twinning.

The final model (R_{cryst} = 18.3%, R_{free} = 24.4%) contains six D1 domains (residues 2-109) arranged as three dimers, five or six residues of the C-terminal 6x-His tag in two of the six D1 domains (molecules B and A, respectively), 158 water molecules, and two Mg^{2+} ions. Each of the two Mg^{2+} ions are coordinated by three histidine residues in the 6x-His tags from two molecules (His112 and His115 from molecule A with His112 from molecule F coordinate one ion and His112 and His115 from molecule B with His112 from molecule E from the adjacent asymmetric unit coordinate the other). Residue 2 in molecules B, C and D, the CC′ loop (residues 41-45) in molecules A, B, C and D and the 6x-His tag for molecules C, D, E and F were disordered and are not included in the refined model and 2.8% of the residues were modeled as alanine. The electron density is weak for residues 97-101 (with B factors averaging 49.6 $Å^2$, compared with an average of 32.3 $Å^2$ for the rest of the model). Ramachandran plot statistics were calculated using PROCHECK (Laskowski et al., 1993). Automated structural comparisons of pIgR D1 were done using the DALI server (Holm and Sander, 1993). Least-squares alignments of pIgR D1 and its closest structural homolog, the V_L and V_H domains of a mouse scFv (PDB code 1MFA), were done using O (Jones and Kjeldgaard, 1997). Sequence

alignments were performed with T-Coffee (Notredame et al., 2000). Figures were generated with Molscript (Kraulis, 1991), Bobscript (Esnouf, 1997), and Raster 3D (Esnouf, 1997; Kraulis, 1991; Merrit and Murphy, 1994).

Acknowledgements

We thank the staff at the Advanced Light Source (ALS) beamline 8.2.2 for technical support and Drs. H. Feinberg, Z.A. Hamburger and E.R. Sprague for crystallographic discussions, and Dr. Peter Snow of the Caltech Protein Expression Facility for insect cell expression of D1. We also thank Dr. A.B. Herr and members of the Bjorkman laboratory for critical reading of the manuscript. A.E.H. was supported by the Whitehead Institute for Biomedical Research and acknowledges Dr. Peter S. Kim for earlier support. A.P.W. was supported by a Career Development Award from the Burroughs-Wellcome Fund.

Accession Numbers

Atomic coordinates have been deposited in the Protein Data Bank with accession code 1XED.

References

Al-Lazikani, B., Lesk, A. M., and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. J Mol Biol *273*, 927-948.

Alzari, P. M., Lascombe, M. B., and Poljak, R. J. (1988). Three-dimensional structure of antibodies. Annu Rev Immunol *6*, 555-580.

Bakos, M. A., Kurosky, A., Czerwinski, E. W., and Goldblum, R. M. (1993). A conserved binding site on the receptor for polymeric Ig is homologous to CDR1 of Ig V kappa domains. J Immunol *151*, 1346-1352.

Bakos, M. A., Kurosky, A., and Goldblum, R. M. (1991). Characterization of a critical binding site for human polymeric Ig on secretory component. J Immunol *147*, 3419-3426.

Bakos, M. A., Widen, S. G., and Goldblum, R. M. (1994). Expression and purification of biologically active domain I of the human polymeric immunoglobulin receptor. Mol Immunol *31*, 165-168.

Bastian, A., Kratzin, H., Eckart, K., and Hilschmann, N. (1992). Intra- and interchain disulfide bridges of the human J chain in secretory immunoglobulin A. Biol Chem Hoppe Seyler *373*, 1255-1263.

Beale, D. (1988). Cyanogen bromide cleavage of bovine secretory component and its tryptic fragments. Int J Biochem *20*, 873-879.

Brandtzaeg, P. (1973). Structure, synthesis and external transfer of mucosal immunoglobulins. Ann Immunol (Paris) *124*, 417-438.

Brandtzaeg, P. (1975). Human secretory immunoglobulin M. An immunochemical and immunohistochemical study. Immunology *29*, 559-570.

Brandtzaeg, P., and Prydz, H. (1984). Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins. Nature *311*, 71-73.

Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta crystallogr D Biol Crystallogr *54*, 905-921.

Cantoni, C., Ponassi, M., Biassoni, R., Conte, R., Spallarossa, A., Moretta, A., Moretta, L., Bolognesi, M., and Bordo, D. (2003). The three-dimensional structure of the human NK cell receptor NKp44, a triggering partner in natural cytotoxicity. Structure (Camb) *11*, 725-734.

Chothia, C., and Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol *196*, 901-917.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., and et al. (1989). Conformations of immunoglobulin hypervariable regions. Nature *342*, 877-883.

Collaborative Computational Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. acta Crystallogr D Biol Crystallogr *50*, 760-763.

Coyne, R. S., Siebrecht, M., Peitsch, M. C., and Casanova, J. E. (1994). Mutational analysis of polymeric immunoglobulin receptor/ligand interactions. Evidence for the involvement of multiple complementarity determining region (CDR)-like loops in receptor domain I. J Biol Chem *269*, 31620-31625.

Crottet, P., and Corthesy, B. (1998). Secretory component delays the conversion of secretory IgA into antigen-binding competent F(ab')2: a possible implication for mucosal defense. J Immunol *161*, 5445-5453.

Crottet, P., and Corthesy, B. (1999). Mapping the interaction between murine IgA and murine secretory component carrying epitope substitutions reveals a role of domains II and III in covalent binding to IgA. J Biol Chem *274*, 31456-31462.

Dall'Acqua, W., Goldman, E. R., Eisenstein, E., and Mariuzza, R. A. (1996). A mutational analysis of the binding of two different proteins to the same antibody. Biochemistry *35*, 9667-9676.

De La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorpous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol *276*, 472-494.

Deitcher, D. L., and Mostov, K. E. (1986). Alternate splicing of rabbit polymeric immunoglobulin receptor. Mol Cell Biol *6*, 2712-2715.

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J Mol Graph Model *15*, 132-134, 112-133.

Fallgreen-Gebauer, E., Gebauer, W., Bastian, A., Kratzin, H. D., Eiffert, H., Zimmermann, B., Karas, M., and Hilschmann, N. (1993). The covalent linkage of secretory component to IgA. Structure of sIgA. Biol Chem Hoppe Seyler *374*, 1023-1028.

Frutiger, S., Hughes, G. J., Hanly, W. C., Kingzette, M., and Jaton, J. C. (1986). The amino-terminal domain of rabbit secretory component is responsible for noncovalent binding to immunoglobulin A dimers. J Biol Chem *261*, 16673-16681.

Garboczi, D. N., Hung, D. T., and Wiley, D. C. (1992). HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. Proc Natl Acad Sci U S A *89*, 3429-3433.

Geneste, C., Iscaki, S., Mangalo, R., and Pillot, J. (1986). Both Fc alpha domains of human IgA are involved in in vitro interaction between secretory component and dimeric IgA. Immunol Lett *13*, 221-226.

Gill, S. C., and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem *182*, 319-326.

Halpern, M. S., and Koshland, M. E. (1970). Noval subunit in secretory IgA. Nature *228*, 1276-1278.

Herr, A. B., Ballister, E. R., and Bjorkman, P. J. (2003a). Insights into IgA-mediated immune responses from the crystal structures of human FcalphaRI and its complex with IgA1-Fc. Nature *423*, 614-620.

Herr, A. B., White, C. L., Milburn, C., Wu, C., and Bjorkman, P. J. (2003b). Bivalent binding of IgA1 to FcalphaRI suggests a mechanism for cytokine activation of IgA phagocytosis. J Mol Biol *327*, 645-657.

Hexham, J. M., White, K. D., Carayannopoulos, L. N., Mandecki, W., Brisette, R., Yang, Y. S., and Capra, J. D. (1999). A human immunoglobulin (Ig)A calpha3 domain motif directs polymeric Ig receptor-mediated secretion. J Exp Med *189*, 747-752.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J Mol Biol *233*, 123-138.

Hughes, G. J., Reason, A. J., Savoy, L., Jaton, J., and Frutiger-Hughes, S. (1999). Carbohydrate moieties in human secretory component. Biochim Biophys Acta *1434*, 86-93.

Jones, T. A., and Kjeldgaard, M. (1997). Electron density map interpretation. methods enzymol *277*, 173-208.

Kelley, R. F., and O'Connell, M. P. (1993). Thermodynamic analysis of an antibody functional epitope. Biochemistry *32*, 6828-6835.

Kleywegt, G. J., and Jones, T. A. (1996). xdlMAPMAN and xdlDATAMAN - Programs for Reformatting, Analysis and Manipulation of Biomacromolecular Electron-Density Maps and Reflection Data Sets. Acta crystallogr D Biol Crystallogr *52*, 826-828.

Kleywegt, G. J., and Jones, T. A. (1997). Template Convolution to Enhance or Detect Structural Features in Macromolecular Electron-Density Maps. Acta Crystallogr D Biol Crystallogr *53*, 179-185.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr *24*, 946-950.

Larsen, N. A., Heine, A., de Prada, P., Redwan el, R., Yeates, T. O., Landry, D. W., and Wilson, I. A. (2002). Structure determination of a cocaine hydrolytic antibody from a pseudomerohedrally twinned crystal. Acta Crystallogr D Biol Crystallogr *58*, 2055-2059.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr *26*, 283-291.

Lebron, J. A., and Bjorkman, P. J. (1999). The transferrin receptor binding site on HFE, the class I MHC-related protein mutated in hereditary hemochromatosis. J Mol Biol *289*, 1109-1118.
Malmqvist, M. (1993). Biospecific interaction analysis using biosensor technology. Nature *361*, 186-187.

Matsumoto, N., Asano, M., Ogura, Y., Takenouchi-Ohkubo, N., Chihaya, H., Chung-Hsing, W., Ishikawa, K., Zhu, L., and Moro, I. (2003). Release of non-glycosylated polymeric immunoglobulin receptor protein. Scand J Immunol *58*, 471-476.

Matthews, B. W. (1968). Solvent content of protein crystals. J Mol Biol *33*, 491-497.

McDonald, K. J., Cameron, A. J., Allen, J. M., and Jardine, A. G. (2002). Expression of Fc alpha/mu receptor by human mesangial cells: a candidate receptor for immune complex deposition in IgA nephropathy. Biochem Biophys Res Commun *290*, 438-442.

Merrit, E. A., and Murphy, M. E. P. (1994). Raster3d, a program for photorealistic molecular graphics. Acta crystallogr D Biol Crystallogr *50*, 869-873.

Mestecky, J., Zikan, J., and Butler, W. T. (1971). Immunoglobulin M and secretory immunoglobulin A: presence of a common polypeptide chain different from light chains. Science *171*, 1163-1165.

Mostov, K. E., Friedlander, M., and Blobel, G. (1984). The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. Nature *308*, 37-43.

Mostov, K. E., Kraehenbuhl, J. P., and Blobel, G. (1980). Receptor-mediated transcellular transport of immunoglobulin: synthesis of secretory component as multiple and larger transmembrane forms. Proc Natl Acad Sci U S A *77*, 7257-7261.

Norderhaug, I. N., Johansen, F. E., Krajci, P., and Brandtzaeg, P. (1999a). Domain deletions in the human polymeric Ig receptor disclose differences between its dimeric IgA and pentameric IgM interaction. Eur J Immunol *29*, 3401-3409.

Norderhaug, I. N., Johansen, F. E., Schjerven, H., and Brandtzaeg, P. (1999b). Regulation of the formation and external transport of secretory immunoglobulins. Crit Rev Immunol *19*, 481-508.

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

Otwinowski, Z., and Minor, W. (1997). Processing of X-Ray Diffraction Data Collected in Oscillation Mode. Methods enzymol *276*, 307-326.

Piskurich, J. F., Blanchard, M. H., Youngman, K. R., France, J. A., and Kaetzel, C. S. (1995). Molecular cloning of the mouse polymeric Ig receptor. Functional regions of the molecule are conserved among five mammalian species. J Immunol *154*, 1735-1747.

Radl, J., Klein, F., van den Berg, P., de Bruyn, A. M., and Hijmans, W. (1971). Binding of secretory piece to polymeric IgA and IgM paraproteins in vitro. Immunology *20*, 843-852.

Roe, M., Norderhaug, I. N., Brandtzaeg, P., and Johansen, F. E. (1999). Fine specificity of ligand-binding domain 1 in the polymeric Ig receptor: importance of the CDR2-containing region for IgM interaction. J Immunol *162*, 6046-6052.

Rudolph, M. G., Wingren, C., Crowley, M. P., Chien, Y. H., and Wilson, I. A. (2004). Combined pseudo-merohedral twinning, non-crystallographic symmetry and pseudo-translation in a monoclinic crystal form of the gammadelta T-cell ligand T10. Acta Crystallogr D Biol Crystallogr *60*, 656-664.

Shibuya, A., Sakamoto, N., Shimizu, Y., Shibuya, K., Osawa, M., Hiroyama, T., Eyre, H. J., Sutherland, G. R., Endo, Y., Fujita, T., *et al.* (2000). Fc alpha/mu receptor mediates endocytosis of IgM-coated microbes. Nat Immunol *1*, 441-446.

Socken, D. J., and Underdown, B. J. (1978). Comparison of human, bovine and rabbit secretory component-immunoglobulin interactions. Immunochemistry *15*, 499-506.

Song, W., Vaerman, J. P., and Mostov, K. E. (1995). Dimeric and tetrameric IgA are transcytosed equally by the polymeric Ig receptor. J Immunol *155*, 715-721.

Terwilliger, T. C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr *55 (Pt 4)*, 849-861.

Tsumoto, K., Ogasahara, K., Ueda, Y., Watanabe, K., Yutani, K., and Kumagai, I. (1995). Role of Tyr residues in the contact region of anti-lysozyme monoclonal antibody HyHEL10 for antigen binding. J Biol Chem *270*, 18551-18557.

Tucker, A. D., Rowsell, S., Melton, R. G., and Pauptit, R. A. (1996). A new crystal form of carboxypeptidase G2 from Pseudomonas sp. strain RS-16 which is more amenable to structure determination. Acta Crystallogr D Biol Crystallogr *52*, 890-892.

Underdown, B. J., Switzer, I., and Jackson, G. D. (1992). Rat secretory component binds poorly to rodent IgM. J Immunol *149*, 487-491.

Vaerman, J. P., Langendries, A., Giffroy, D., Brandtzaeg, P., and Kobayashi, K. (1998a). Lack of SC/pIgR-mediated epithelial transport of a human polymeric IgA devoid of J chain: in vitro and in vivo studies. Immunology *95*, 90-96.

Vaerman, J. P., Langendries, A., and Vander Maelen, C. (1995). Homogenous IgA monomers, dimers, trimers and tetramers from the same IgA myeloma serum. Immunol Invest *24*, 631-641.

Vaerman, J. P., Langendries, A. E., Giffroy, D. A., Kaetzel, C. S., Fiani, C. M., Moro, I., Brandtzaeg, P., and Kobayashi, K. (1998b). Antibody against the human J chain inhibits polymeric Ig receptor-mediated biliary and epithelial transport of human polymeric IgA. Eur J Immunol *28*, 171-182.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. J Appl Crystallogr *30*, 1022-1025.

van Egmond, M., van Garderen, E., van Spriel, A. B., Damen, C. A., van Amersfoort, E. S., van Zandbergen, G., van Hattum, J., Kuiper, J., and van de Winkel, J. G. (2000). FcalphaRI-positive liver Kupffer cells: reappraisal of the function of immunoglobulin A in immunity. Nat Med *6*, 680-685.

Vaughn, D. E., Milburn, C. M., Penny, D. M., Martin, W. L., Johnson, J. L., and Bjorkman, P. J. (1997). Identification of critical IgG binding epitopes on the neonatal Fc receptor. J Mol Biol *274*, 597-607.

Vidarsson, G., van Der Pol, W. L., van Den Elsen, J. M., Vile, H., Jansen, M., Duijs, J., Morton, H. C., Boel, E., Daha, M. R., Corthesy, B., and van De Winkel, J. G. (2001). Activity of human IgG and IgA subclasses in immune defense against Neisseria meningitidis serogroup B. J Immunol *166*, 6250-6256.

Wells, J. A., and de Vos, A. M. (1996). Hematopoietic receptor complexes. Annu Rev Biochem *65*, 609-634.

White, K. D., and Capra, J. D. (2002). Targeting mucosal sites by polymeric immunoglobulin receptor-directed peptides. J Exp Med *196*, 551-555.

Wieland, W. H., Orzaez, D., Lammers, A., Parmentier, H. K., Verstegen, M. W., and Schots, A. (2004). A functional polymeric immunoglobulin receptor in chicken (Gallus gallus) indicates ancient role of secretory IgA in mucosal immunity. Biochem J *380*, 669-676.

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily--domains for cell surface recognition. Annu Rev Immunol *6*, 381-405.

Zdanov, A., Li, Y., Bundle, D. R., Deng, S. J., MacKenzie, C. R., Narang, S. A., Young, N. M., and Cygler, M. (1994). Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-A resolution. Proc Natl Acad Sci U S A *91*, 6423-6427.

Table 1. Data Collection, Phasing, and Refinement Statistics for pIgR D1

Data Set	Wavelength (Å)	Resolution (Å)	No. Unique Reflections/Total Reflections	Completeness (%)[a]	Twin Fraction (%)	R_{merge} (%)[b]	I/σ	Phasing Power[c]
Native I	1.54	2.5	21,524/	97.4	26.5	8.5	19.6	
		(2.59 - 2.50)	76,172	(80.7)		(34.3)	(2.5)	
Native II	1.078	1.9	49,755/	99.6	37.3	7.8	17.7	
		(1.93-1.90)	185,798	(99.1)		(46.2)	(2.6)	
PIP	1.54	3.2	10,464/	98.5	20.7	9.7	11.3	0.4
		(3.31 -3.20)	28,678	(94.3)		(20.8)	(5.3)	
Gd	1.7108	2.8	27,967/	89.6	30.3	10.4	8.6	1.0
		(2.85 -2.80)	47,501	(58.7)		(41.3)	(1.5)	
Pb	0.9509	2.0	79,649/	94.1	37.0	6.5	12.2	0.6
		(2.03 -2.00)	153,910	(92.7)		(23.8)	(3.5)	

Refinement Statistics ($P2_1$)		Number of Nonhydrogen Atoms	
Resolution (Å)	30-1.9	Protein	4993
Number of reflections in working set	47,117 (93.4%)	Water	158
Number of reflections in test set	2,476 (4.9%)	Mg^{2+}	2
R_{cryst} (%)[d]	18.3		
R_{free} (%)[e]	24.4		

Rmsd from ideality		Ramachandran plot quality (%)	
		Nonglycine residues in	
		most favored	87.3
		additionally allowed	12.3
Bond lengths (Å)	0.007	generously allowed	0.4
Bond angles (deg)	1.30	disallowed	0.0

Values in parentheses indicate data in the highest resolution shell.

[a] Completeness is defined as the number of independent reflections/total theoretical number.

[b] R_{merge}= 100 x $\Sigma(|I-\langle I\rangle|)/\Sigma I$, where I is the integrated intensity of a given reflection.

[c] Rms f_h/E (phasing power), where f_h is the heavy atom structure factor amplitude and E is the lack of closure error. The phasing power statistics were derived using twinned data and may therefore not be accurate.

[d] $R_{cryst}(F)=\Sigma_h||F_{obs}(h)|-|F_{calc}(h)||/\Sigma_h|F_{obs}(h)|$, where $|F_{obs}(h)|$ and $|F_{calc}(h)|$ are the observed and calculated structure factor amplitudes for the hkl reflection.

[e] R_{free} is calculated from reflections in a test set not included in the atomic refinement.

Table 1. Data Collection, phasing and refinement statistics for pIgR D1.

Figure 1. Biosensor Analyses of pIgR D1 Binding to dIgA

(A) Equilibrium binding data for biosensor experiments in which nonglycosylated (expressed in bacteria) and glycosylated (expressed in insect cells) versions of pIgR D1 were injected over dIgA immobilized at three different densities (412, 763 and 1426 RUs). The plot shows the equilibrium binding response (R_{eq}) versus the log of concentration of the indicated proteins. Best-fit binding curves based on a 1:1 binding model are superimposed on the binding data.

(B) Sensorgram from binding experiments in which 5 μM glycosylated or nonglycosylated forms of D1 were injected over monomeric (with and without the tailpiece) and dimeric versions of the IgA Fc region. The injected protein is indicated in front of an arrow pointing to the immobilized protein.

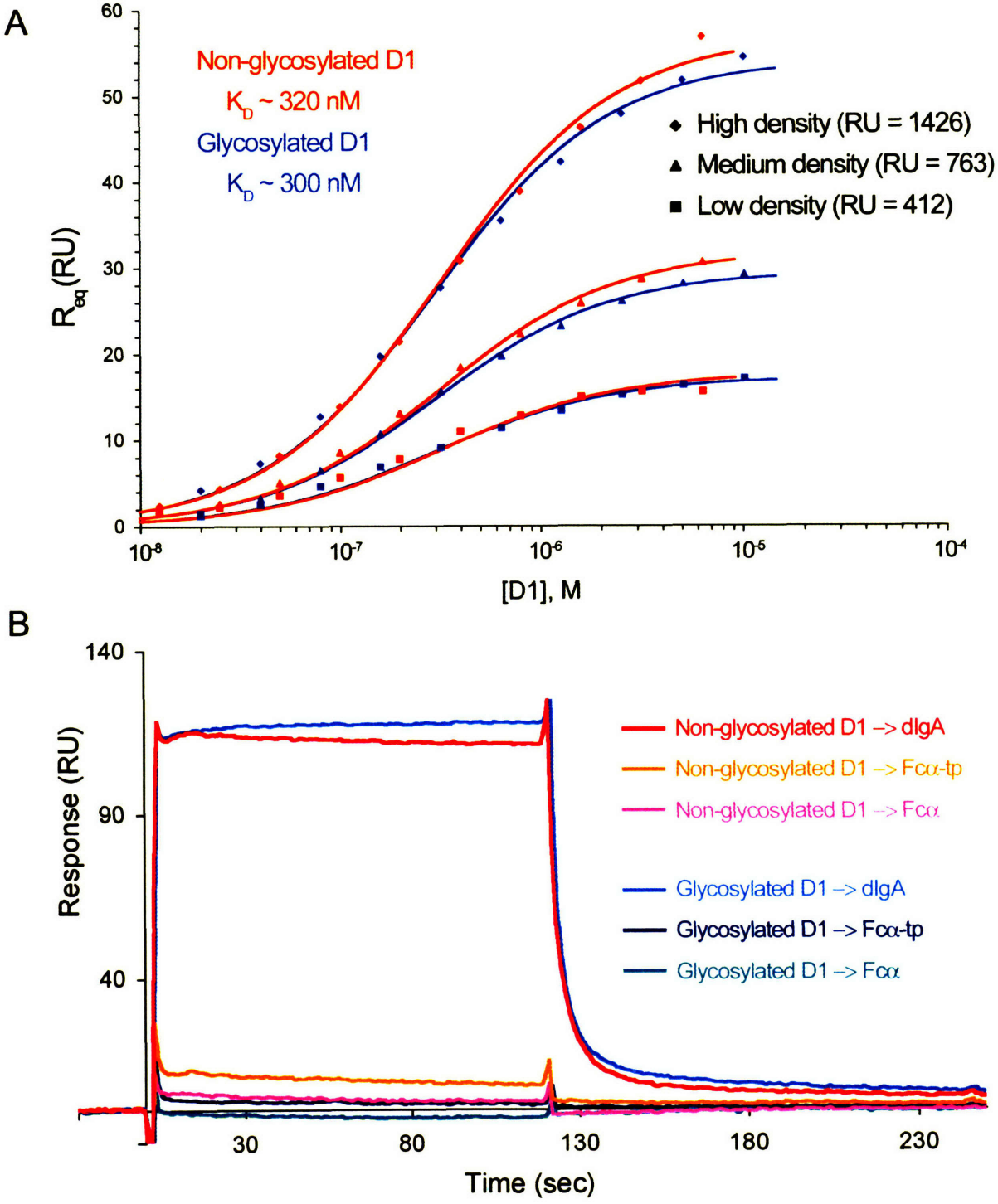


Figure 1. Biosensor Analyses of pIgR D1 Binding to dIgA

Figure 2. The Structure of Human pIgR D1

(A) Cα trace (molecule E). The 6xHis tag from molecule A (dashed line) was superimposed onto molecule E.

(B) Ribbon diagram showing side and front views of pIgR D1. β strands A, B, E, D are shown in blue, β strands C″, C′, C, F, G, A′ are in green and the three CDR loops (including the α helix within CDR1) are red. Cysteines involved in disulfide bonds are shown in yellow in ball-and-stick representation and the locations of potential N-linked glycosylation sites are indicated by pink spheres.

(C) Topology diagram of pIgR D1. The color schemes for the β strands, α helix and CDR loops are the same as in (A).

(D) Stereoview of the pIgR D1 model in the region of the "invariant" tryptophan, Trp37, superimposed on a 2.7 Å experimental electron density map (calculated with MIRAS phases) contoured at 1.0 σ.

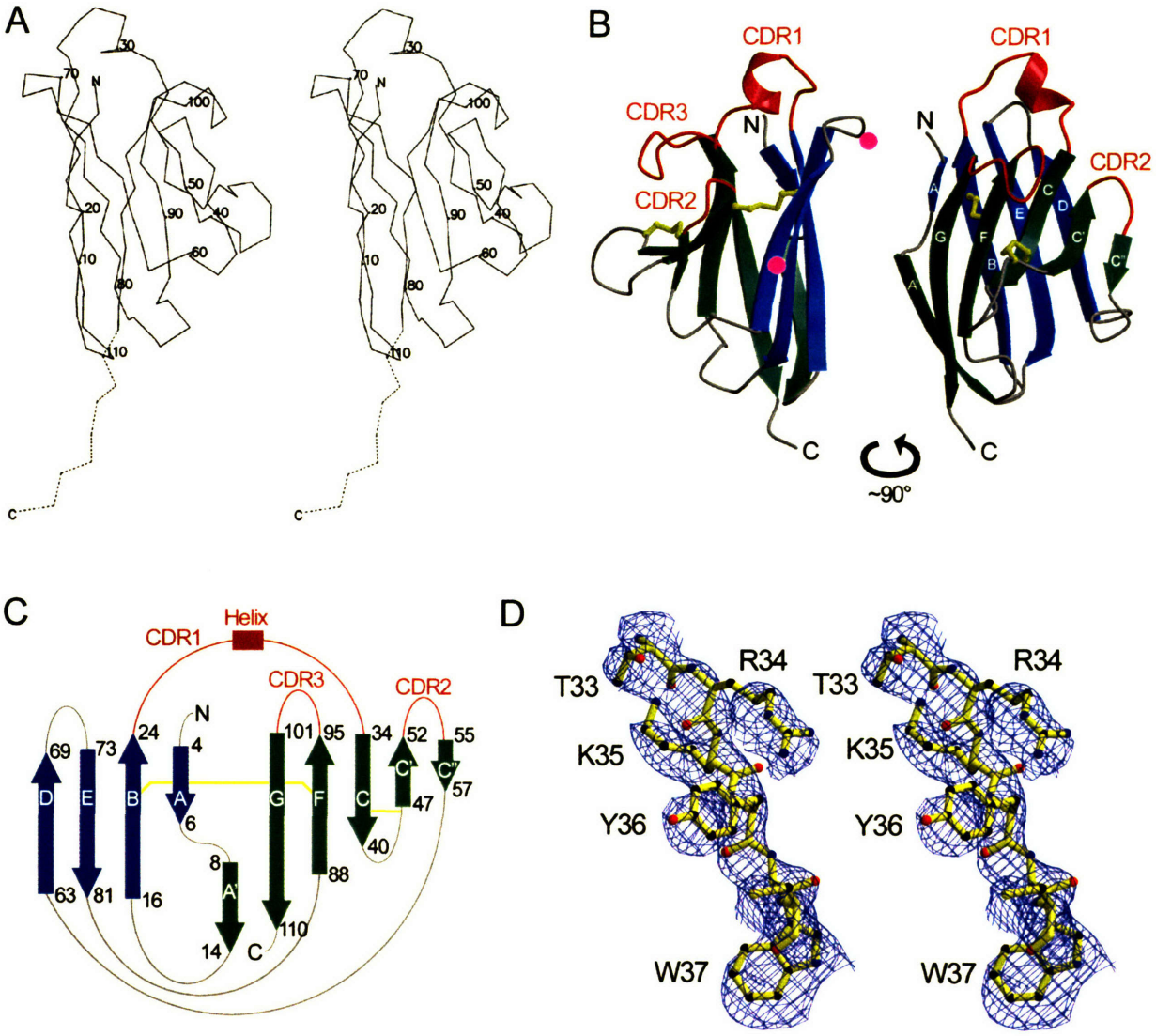


Figure 2. The Structure of Human pIgR D1

Figure 3. Amino Acid Sequence Alignment of Human pIgR D1 and Related Proteins

Top panel: pIgR D1 from eight different species and the Ig-like domain from the human Fcα/μR were aligned. Cysteines involved in disulfide bonds are shown in red, conserved residues are highlighted in yellow and conservatively substituted residues are highlighted in blue. Asterisks indicate the positions of five characteristic residues in Ig V domains. Sequences within the three CDR loops are shown in bold and underlined. Crystallographically determined secondary structure elements are shown above the sequences. Accession codes for the pIgR sequences are: P01833 (human), P01832 (rabbit), O70570 (mouse), P15083 (rat), P81265 (bovine), AAK69593 (*Macropus eugenii* (tammar wallaby)), AAD41688 (*Trichosurus vulpecula* (silver-grey brushtail possum)), and AAP69598 (*Gallus gallus* (chicken)); and AAL51154 (human Fcα/μR).

Bottom panel: The D1 domain of pIgR is aligned with the other domains within the human pIgR extracellular region.

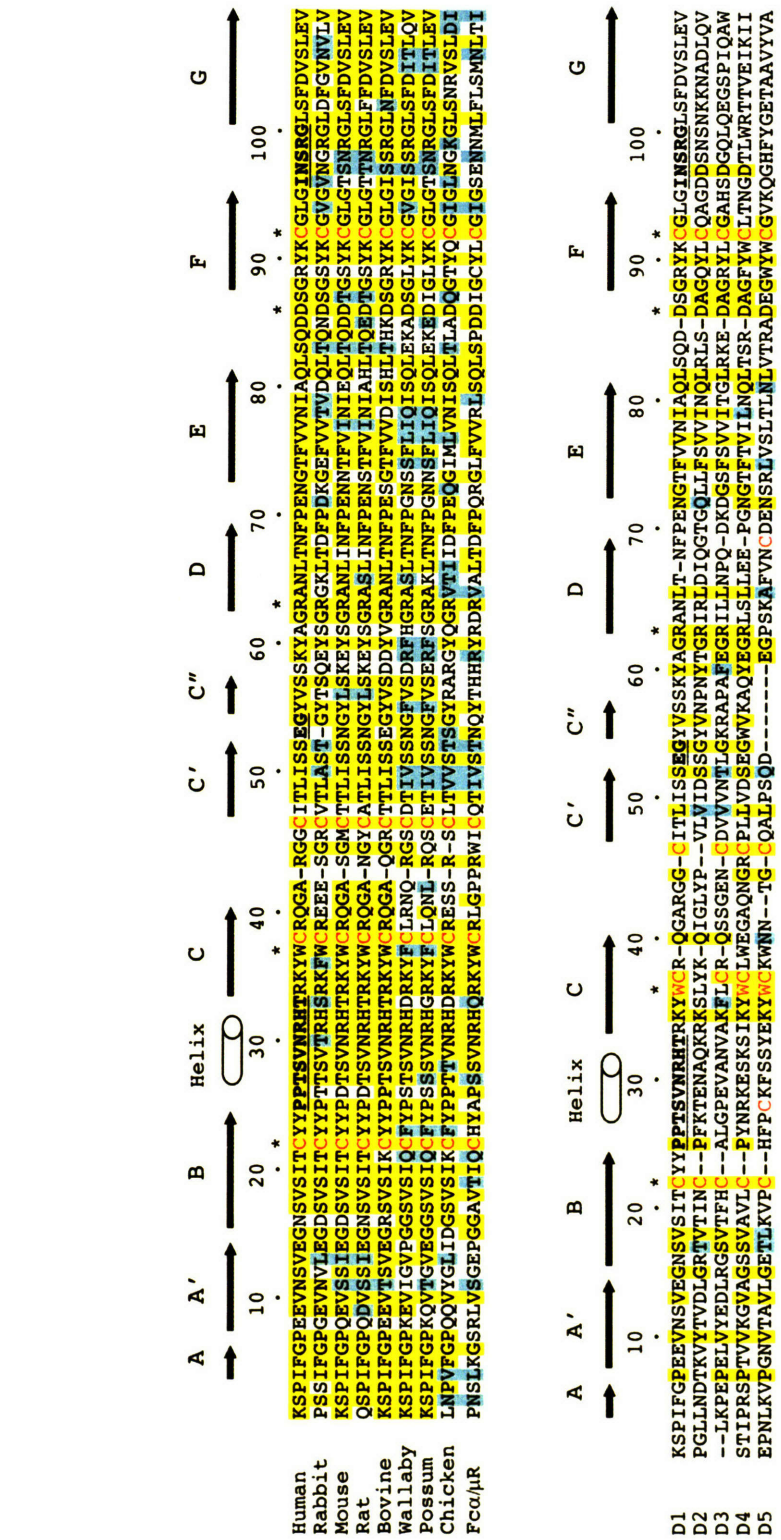


Figure 3. Amino Acid Sequence Alignment of Human pIgR D1 and Related Proteins

Figure 4. Structural Consequences of the Position of the pIgR D1 CDR3 Loop

(A) Stereoview of the superposition of pIgR D1 (red) with an Ig V_L domain (blue) derived from a mouse single-chain Fv (PDB code 1MFA).

(B) Stereoview of the region in the vicinity of the pIgR D1 CDR3 loop showing Tyr36 buried at the interface between CDR3 and the C″C′CFGA′ sheet. A ball-and-stick representation of Tyr36 is shown in red. Atoms in residues that stabilize this position of CDR3 are highlighted with an atom-based color code (carbon: black; oxygen: red; nitrogen: blue), and hydrogen bonds are indicated as black dashed lines.

(C-D) Comparison of an Ig V_H-V_L heterodimer (PDB code 1MFA) (panel C) with a computer model of a pIgR D1 homodimer (panel D) created by superimposing D1 on the V_H and V_L domains of a V_H-V_L heterodimer (panel C). The rmsd values are 1.39 Å for the V_H-D1 superposition (calculated for 101 Cα atoms) and 1.38 Å for the V_L-D1 superposition (calculated for 78 Cα atoms). CDR3 loops are highlighted in pink.

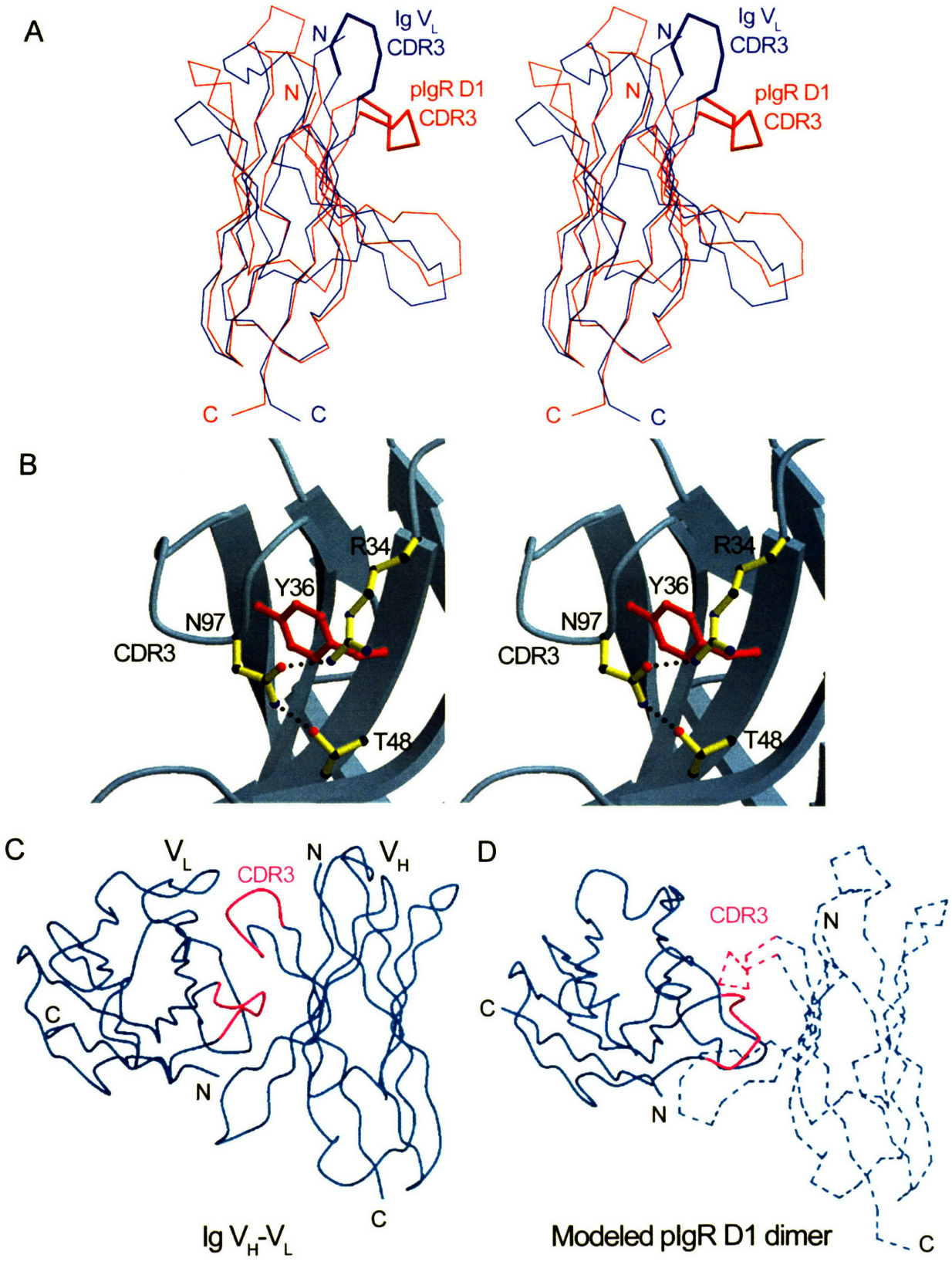


Figure 4. Structural Consequences of the Position of the pIgR D1 CDR3 Loop

Figure 5. Mutagenesis Data Mapped onto Structure

The positions of substitutions that abolished (red) or decreased (blue) pIgA binding to rabbit pIgR (Coyne et al., 1994) are mapped onto the human pIgR D1 structure. A close-up of the CDR1 region is shown in the upper right, and the sequence of CDR1 in human and rabbit pIgR D1 is shown in the lower right. Val29 (black), which is solvent exposed in the D1 structure, was assumed to be buried and was therefore not substituted (Bakos et al., 1993; Coyne et al., 1994).

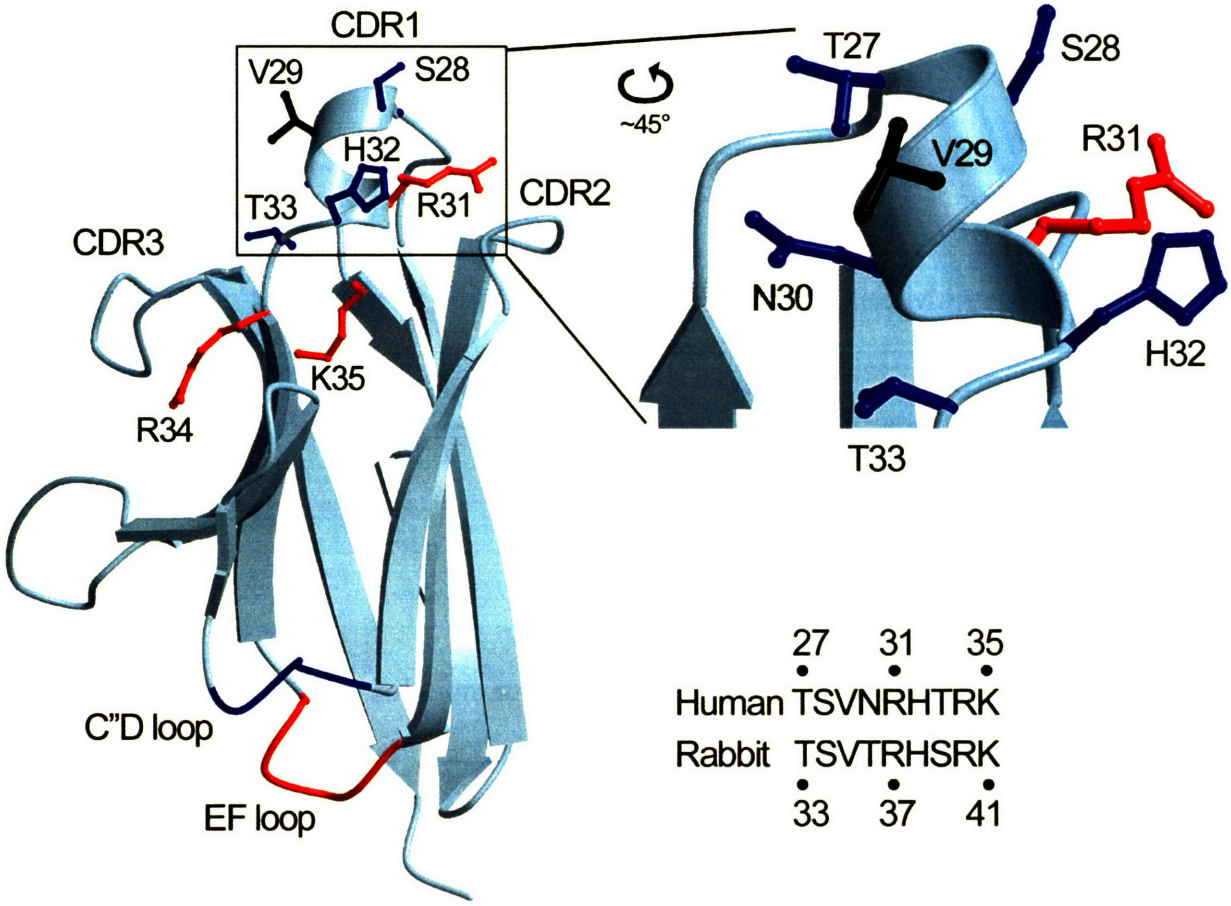


Figure 5. Mutagenesis Data Mapped onto Structure

CHAPTER 3

Solution Structure of Choline Binding Protein A, the Major Adhesin of *Streptococcus pneumoniae*

This chapter describes the solution structure of an adhesion domain (R2) of CbpA, the *S. pneumoniae* surface protein responsible for pIgR binding, and biophysical experiments characterizing the binding properties of CbpA domains and pIgR. This work was done in collaboration with several groups at St. Jude Children's Research Hospital and my contribution included providing purified pIgR and performing the initial binding experiments to show that the recombinantly expressed pIgR binds CbpA, R1 and R2.

Solution Structure of Choline Binding Protein A, the Major Adhesin of *Streptococcus pneumoniae*

[1,3] Rensheng Luo, [2,3] Beth Mann, [4, 5] William S. Lewis, [6] Arthur Rowe, [2,7] Richard Heath, [1,11] Michael L. Stewart, [8] Agnes E. Hamburger, [1] Siva Sivakolundu, [1] Eilyn R. Lacy, [8,9] Pamela J. Bjorkman, [2,10,11] Elaine Tuomanen, and [1,11] Richard W. Kriwacki.

Departments of [1] Structural Biology and [2] Infectious Diseases, [4] Hartwell Center for Bioinformatics and Biotechnology, and [7] Division of Protein Sciences, Department of Infectious Diseases, St. Jude Children's Research Hospital, 332 N. Lauderdale St., Memphis, Tennessee 38105, USA; [6] National Centre for Macromolecular Hydrodynamics, University of Nottingham, School of Biosciences, Sutton, Bonington, Leicestershire, LE12 5RD, UK; [8] Division of Biology, [9] Howard Hughes Medical Institute California Institute of Technology, Pasadena, California 91125, USA; Departments of [10] Pediatrics, and [11] Molecular Sciences, University of Tennessee Health Sciences Center, Memphis, Tennessee, 38163, USA

[3] These authors contributed equally to this work.

[5] Current address: Plant Sciences Institute, Iowa State University, 0077 Roy J. Carver Co-Laboratory, Ames, IA 50011.

Correspondence should be addressed to E.T. (elaine.tuomanen@stjude.org; 901-495-3486) or R.W.K. (richard.kriwacki@stjude.org; 901-495-3290)

Abstract

Streptococcus pneumoniae (pneumococcus) remains a significant health threat world-wide, especially to the young and old. While some of the biomolecules involved in pneumococcal pathogenesis are known and understood in mechanistic terms, little is known about the molecular details of bacterium/host interactions. We report here the solution structure of the "repeated" adhesion domains (domains R1 and R2) of the principal pneumococcal adhesin, choline binding protein A (CbpA). Further, we provide insights into the mechanism by which CbpA binds its human receptor, polymeric immunoglobulin receptor (pIgR). The R domains, comprised of 12 imperfect copies of the leucine zipper heptad motif, adopt a unique 3-α-helix, raft-like structure. Each pair of α-helices is anti-parallel and conserved residues in the loop between Helixes 1 and 2 exhibit a novel "tyrosine fork" structure that is involved in binding pIgR. This and other structural features that we show are conserved in most pneumococcal strains appear to generally play an important role in bacterial adhesion to pIgR. Interestingly, pneumococcus is the only bacterium known to adhere to and invade human cells by binding to pIgR.

Running title

Solution structure of the pneumococcal adhesin, CbpA

Introduction

Streptococcus pneumoniae (pneumococcus) remains the most common invasive bacterial agent leading to hospitalization in all age groups (Schuchat et al., 2001), with the majority of these cases affecting either children or the elderly. Penicillin remains the primary mode of treatment but the emergence of antibiotic resistance has intensified the search for new therapeutic approaches (Whitney et al., 2000). While disease prevention through vaccination is partially effective, this approach is less effective in the youngest and oldest patients (Butler et al., 1993). Hence, studies of disease mechanisms may provide insights into new anti-bacterial therapies. The pneumococcus invades human nasopharyngeal epithelial (NE) cells and enters the blood stream through a process termed reverse transcytosis mediated by polymeric immunoglobulin receptor (pIgR) (Zhang et al., 2000). The normal function of pIgR is to transport secretory IgA (sIgA) from the basolateral to the apical surface of NE cells (Mostov and Kaetzel, 1999). A protein on the bacterial surface, choline binding protein A (CbpA), binds specifically to an extracellular domain of pIgR and hijacks the endocytosis machinery to translocate pneumococci across NE cells into the blood stream. While the participation of CbpA in pneumococcal adhesion and invasion is well established (Rosenow et al., 1997), the molecular details of these processes are not understood.

CbpA (also referred to as PspC, SpsA, and PbcA) is one of 15 proteins identified in the genome of the TIGR4 strain (Tettelin et al., 2001) that exhibit multiple C-terminal repeats of a ~19 amino acid motif that binds choline moieties present on the bacterial cell wall (Gosink and Tuomanen, 2000). The mechanism of bacterial surface attachment by CbpA can be understood on the basis of structural studies of another choline binding protein (Cbp), LytA. [Note: We use the generic term CbpA to refer to the CbpA protein from the TIGR4 strain.] The C-terminal

domain of LytA contains seven repeats of the choline binding motif and adopts a unique β-solenoid structure, with choline groups binding between β-hairpin "steps" of the staircase-like structure (Fernandez-Tornero et al., 2001). Similarly, CbpA has eight repeats of this motif in its C-terminus (Fig. 1A) and is predicted to bind surface-exposed choline groups in similar manner (Fernandez-Tornero et al., 2001). The N-terminal domains of the 15 TIGR4 Cbps differ widely, with functions ranging from bacterial autolysis (LytA) and glucosaminidase activity (LytB) to adhesion to human NE cells (CbpA) (Gosink et al., 2000; Tettelin et al., 2001). The sequence of the CbpA N-terminus (residues 39-514 (Tettelin et al., 2001)) exhibits numerous repeats of the leucine zipper (LZ) motif (Landschulz et al., 1988) that cluster within five domains termed A, B, R1, R2 and C (Fig. 1A). Domains A, B and C are 21 to 25 amino acids in length and are predicted to form coiled-coil dimers (Fig. 1B) (Lupas et al., 1991). The ~110 amino acid-long, "repeated" domains, R1 and R2 (78% identical) (Zhang et al., 2000) are also predicted to form self-associated, coiled-coil structures. We refer to these as "R" domains. The segments that connect these domains are predicted to lack secondary structure. In general, LZ motifs mediate intra- or inter-molecular α-helix/α-helix interactions, and are well understood as the basis for parallel self-association into α-helical dimers, trimers, and tetramers (Lupas, 1996). LZ motifs also mediate anti-parallel α-helix/α-helix interactions (Oakley and Hollenbeck, 2001). To date, however, the R domains of CbpA have not been structurally characterized.

CbpA R domains bind the extracellular, immunoglobulin-like (Ig-like) domains of pIgR termed secretory component (SC) (Hammerschmidt et al., 1997; Zhang et al., 2000). SC covalently binds to IgA dimers and, after transcytosis from the basolateral to apical surface of NE cells, the SC/IgA_2 complex is proteolytically cleaved from pIgR to release sIgA (Mostov, 1994). A conserved hexapeptide motif within the R domains is required for binding to sIgA

(Hammerschmidt et al., 2000). Here we present the solution structure of domain R2 of CbpA containing this motif determined using NMR spectroscopy. Further, we used this structure to model that for domain R1. Surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) were used to gain insight into the mechanism by which CbpA binds components of pIgR. Our results provide the first molecular insights into the structurally novel mechanism by which pneumococcus binds pIgR and subsequently invades human cells.

Results

Conserved features of CbpA sequences from many pneumococcal strains

The R domains of CbpA from the TIGR4 strain of *Streptococcus pneumoniae* are highly conserved in CbpA sequences from other pneumococcal strains (Iannelli et al., 2002). We compared CbpA sequences (excluding the choline binding motifs) from 47 pneumococcal strains and quantified R domain conservation (Fig. 2A and B). Thirty-nine sequences exhibit R1 and R2 domains that are ≥50% identical to the TIGR4 domains, six exhibit one R2-like (≥50% identical) domain, and one exhibits one R1-like (93% identical) domain. Residues at 22 positions are identical in these R1 and R2 domains, 87 in total, including five residues within the hexapeptide motif discussed earlier. Thus, our structural and biophysical results for the R domains of the TIGR4 strain are relevant to CbpA from virtually all known pneumococcal strains.

Solution Structure of CbpA, Domains R1 and R2

The secondary structure of N-terminal CbpA domains was elucidated using circular dichroism (CD) and NMR spectroscopy. Domains R1 and R2 (78 % identical) each contain 12 imperfect copies of the LZ motif and span residues 175-285 and 327-442, respectively. Eight of the 24 differences between R1 and R2 correspond to substitution of Glu by Lys, or *vice versa* (Fig. 1C).

The CbpA R domain LZ motifs are similar to those found in coiled-coil proteins (Fig. 1C) (Lupas et al., 1991); however, the amino acid at position **d** of the heptad motif, which is usually Leu in the classical LZ motif, is most frequently Ala in the R domains.

CbpA-R1 and -R2 possess extensive α-helical structure on the basis of CD spectra (Fig. 3B) and we used NMR spectroscopy to determine the CbpA-R2 structure in solution. Secondary $^{13}C_{\alpha}$ chemical shifts for CbpA-R2 (Fig. 4A), which are indicative of secondary structure, clearly revealed three α-helices: Helix 1 (residues 330-357), Helix 2 (residues 366-390) and Helix 3 (residues 396-425). These adopt an unusual three-α-helix, raft-like structure through anti-parallel α-helix/α-helix interactions (Fig. 4B). The α-helices, while anti-parallel, are not exactly co-axial due to slight coiling (Fig. 4C). The crossing angle for Helices 1 and 2 is -178° and that for Helices 2 and 3 is -176°. In addition, we generated a homology model of CbpA-R1 based on the structure of CbpA-R2. The structures of CbpA-R2 (Fig. 4B) and -R1 indicate that many hydrophobic residues are buried at the two α-helix/α-helix interfaces and that the protein surfaces are dominated by polar and charged residues (Fig. 4D). CbpA-R2 exhibits a bi-lobed electrostatic potential, with one face of the structure highly electronegative (red contours in Fig. 4D, left) and the other electropositive (blue contours). The differences in charged residues between CbpA-R2 and –R1 cause increased electronegative character on one face of CbpA-R1 and reduced electropositive character on the other (Fig. 4D, right).

Peptides with the sequence RNYPT bind to sIgA and SC, and mutations in this motif in SpsA, a CbpA variant with one R domain, abolish binding to sIgA (Hammerschmidt et al., 2000). The YPT motif, which occurs once in each of the two TIGR4 CbpA R domains (Fig. 1C), is found in the loop between Helix 1 and Helix 2 in the CbpA-R2 structure (Fig. 4B and E). Interestingly, these and other conserved residues in the loop protrude into solvent from an

electrostatically neutral region of CbpA-R2, with two partially stacked Tyr residues forming a "tyrosine fork" structure (Fig. 4D, left, and E). Most residues in the Helix 1/Helix 2 loop are conserved in CbpA-R1 and thus the homology model of domain R1 exhibits a similar tyrosine fork feature. While the differences in charged residues mentioned above affects the electrostatic properties of domain R1, the potentials in the vicinity of the YPT motif in this domain are very similar to those of the R2 domain (Fig. 4G, right), suggesting that the negative charge of this region, and the protruding YPT motif, are involved in binding to pIgR. Many of the residues in this region have significant hydrophobic character, including Tyr 358 and 363, Pro 359, and Thr 360 and 362 (Fig. 4E) (Creighton, 1993; Nozaki and Tanford, 1971), suggesting that hydrophobic interactions are involved in receptor binding.

Proteins that contain multiple LZ motifs often self-associate. However, the results of equilibrium analytical ultracentrifugation (AUC) experiments showed that CbpA-R1 and -R2 are predominantly monomers in solution (Suppl. Fig. 1C & D). Analysis of sedimentation data for both domains did, however, provide evidence for very weak self-association at high protein concentration (Suppl. Table 1). Further, we showed using equilibrium AUC that CbpA-R1 and Cbp-R2 do not interact to form hetero-oligomers (data not shown). Importantly, these results show that the principal role of the CbpA LZ motifs is to mediate the folding, not oligomerization, of the R domains.

Structural properties of CbpA-N (residues 39-174)

The role of the N-terminal domain of CbpA (containing LZ domains A and B, Fig. 1A) in adhesion and invasion is not known. To provide preliminary insights into function, we characterized the structure and self-association properties of CbpA-N (residues 39-174). CD results show that CbpA-N adopts α-helical secondary structure (Fig. 3B) and that it has a stable

fold (Fig. 3C). As observed for CbpA-R1 and -R2, AUC analysis showed that CbpA-N is monodisperse; however, a small degree of oligomerization was observed at 0 and 50 mM NaCl (Suppl. Table 1).

The full CbpA N-terminus (residues 39-442) is comprised of three α-helical domains

We used CD to compare the structure of a 404 amino acid fragment of CbpA (CbpA-NR12; Fig. 3A) with that of the three domains within this fragment (CbpA-N, -R1 and -R2). CD spectra for CbpA-N, -R1, -R2 and -NR12 (Fig. 3B) showed that all constructs are highly α-helical. Further, we monitored ellipticity at 222 nm for each construct as the temperature was raised from 5 °C to 95 °C to characterize thermal stability. CbpA-R1 and -R2 exhibited thermal unfolding transitions at 44 °C and 47 °C, respectively (Fig. 3C). CbpA-N also exhibits a single thermal transition but at a much higher temperature, 71 °C (Fig. 3C). Based on these data we conclude that CbpA-N, -R1 and -R2 are individually folded, stable α-helical domains. The thermal melting curve for CbpA-NR12 is a composite of those for the individual domains, exhibiting two thermal transitions, one at 45 °C and a second at 71 °C. These data suggest strongly that the structure of the individual CbpA domains is preserved in the multi-domain construct, CbpA-NR12.

We also compared the structure of CbpA-R2 with that of the R2 domain within CbpA-NR12 by analyzing 2D ^{1}H-^{15}N TROSY spectra. The spectrum of CbpA-R2 (Fig. 5A) exhibited the appropriate number of resonances, most of which appeared outside the crowded central region. The resonance pattern of CbpA-R2 was well reproduced in the spectrum of CbpA-NR12. The similarity between the two spectra is illustrated by the blue overlays in Fig. 5A and 5B; the blue layer identifies resonances that appeared at identical positions in the two spectra. In Fig. 5A, the spectrum of CbpA-R2 is the red colored layer; however, most resonances appear blue because they also appeared in Fig. 5B. In Fig. 5B, approximately 40 isolated resonances

(excluding obvious side chain resonances) are colored blue, indicating that they also appeared in Fig. 5A. CbpA-R2 is comprised of 119 amino acids and the fact that resonances for one third of these appeared at identical chemical shift values in the two spectra suggests that the structure of CbpA-R2 is the same in the two constructs. These data provide further support for the conclusion that the N-terminus of CbpA (residues 39-442) is comprised of three independent, α-helical domains.

Role of CbpA "R" domains in pIgR binding

We studied the binding of domains of CbpA to components of pIgR using several approaches to understand the underlying interaction mechanism. First, we used ELISA to measure the extent to which CbpA constructs bind to sIgA. In agreement with previous studies (Hammerschmidt et al., 2000; Zhang et al., 2000), our results show that multi-domain fragments of CbpA, including CbpA-R12 and -NR12, bind to immobilized sIgA (Suppl. Fig. 2). However, this assay detected only a weak interaction between sIgA and CbpA-N, -R1 and -R2. These results suggested that multiple CbpA domains cooperate in binding sIgA.

To clarify the role of domains R1 and R2 in pIgR binding, we performed binding experiments using surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC). We immobilized either sIgA or SC expressed in Sf9 insect cells (SC-D15) on SPR biosensor surfaces and monitored CbpA binding. The experiments were performed at several different concentrations of each CbpA construct (Suppl. Fig. 2B) and a global fitting procedure was used to determine association (k_a) and dissociation (k_d) rate constants (Fig. 6A, Suppl. Table 2). Further, these rate constants were used to calculate the apparent Gibbs free energy of binding (ΔG_{SPR}) for each reaction ($\Delta G_{SPR} = -RT \ln(k_d/k_a)$; Fig. 6A). SPR data for CbpA fragments binding sIgA and SC-D15 showed similar trends and, therefore, only those for binding to sIgA

are illustrated in Fig. 6A and Suppl. Fig. 2B. Kinetic constants for all reactions (sIgA and SC-D15) are given in Suppl. Table 2. Because CbpA-NR12 and -R12 each contain two "R" domains, each with one YPT motif, we determined whether the individual R domains bound to components of pIgR. CbpA-R1 and -R2 bound to the sIgA surface more rapidly than did those containing two R domains (CbpA-R12 and -NR12) (Fig. 6A and Suppl. Table 2), as would be expected if association was controlled by diffusion. These constructs dissociated from the sIgA surface with rates that depended on the number of R domains; CbpA-R1 and -R2 dissociated more rapidly than did CbpA-R12 and -NR12. CbpA-R1 and -R2 exhibited similar ΔG_{SPR} values (-13.4 kcal/mol and -12.8 kcal/mol, respectively), and the value for CbpA-NR12 was only slightly more negative (-14.0 kcal/mol). In this surface-based assay, the presence of two R domains in CbpA-NR12 appears to enhance binding to sIgA only slightly.

While the SPR data provided important insights into CbpA:sIgA (and SC) interactions, due to the surface-based format of this assay, ΔG_{SPR} values may not reflect binding behavior in solution. Further, the analysis of binding stoichiometry from SPR data can be problematic. To overcome these limitations, we used ITC to monitor binding of the CbpA constructs to SC-D15. Solutions of CbpA-R1, CbpA-R2, CbpA-R12, or CbpA-NR12 were injected into a solution of SC-D15 and the evolved heat was measured (Fig. 6B and Table 1). These data were fit using standard equations to give values of the enthalpy (ΔH_{ITC}), Gibbs free energy (ΔG_{ITC}), and entropy ($-T\Delta S_{ITC}$) of binding. In addition, the CbpA:SC-D15 mole ratio (N) for each construct was determined (Table 1). Binding to SC-D15 was studied because only this protein could be obtained with sufficient purity and in sufficient quantities. The absolute values of ΔG_{ITC} are between 3 to 4 kcal/mol smaller than those determined from SPR-derived rate constants; however, only a small difference was observed between the values for the different CbpA

constructs, which parallels the results from SPR (Table 1 and Fig. 6A). Surface immobilization of sIgA and SC-D15 in the SPR experiments may promote association relative to the situation in solution, and give rise to anomalously large values of ΔG_{SPR}. Despite this caveat, ΔG_{SPR} values can be compared to give an accurate measure of relative binding affinities for different CbpA constructs. The ITC results show that *two* moles of CbpA-R1 or CbpA-R2, or *one* mole of CbpA-R12 or CbpA-NR12, bind to *one* mole of SC-D15. This latter result, along with the similarity of the ΔG_{ITC} values and kinetic constants for CbpA-R1 and -R2 binding SC-D15, strongly suggests that there are two thermodynamically indistinguishable binding sites in SC-D15 for CbpA R domains. ΔG_{ITC} values for CbpA-R12 and -NR12, each with two R domains, are only marginally larger in absolute magnitude than those for CbpA-R1 and -R2 which, at first, may seem inconsistent with the presence of two binding sites in SC-D15. However, the values of the entropy parameter ($-T\Delta S$) for the former constructs are twice those for the latter (Table 1), which largely nullifies the enthalpic benefits (in CbpA-R12 and -NR12) associated with having two R domains. The large entropic penalty of binding for constructs with two R domains may arise from restriction upon binding of a highly flexible polypeptide linker between domains R1 and R2.

Mutation of residues within the YPT motif of CbpA interferes with sIgA binding (Hammerschmidt et al., 2000). To further confirm our binding model, we prepared three CbpA-NR12 constructs in which the Tyr residue of the YPT motif was mutated to Gly in the R1 domain (CbpA-NR12-Y205G), the R2 domain (CbpA-NR12-Y358G) or both domains (CbpA-NR12-Y205G/Y358G) and used SPR to determine k_a, k_d and ΔG_{SPR} values (Fig. 6A). Similar mutant CbpA constructs in which the Pro residue of the YPT motif was mutated to Gly were also prepared and analyzed. We reasoned that Gly would readily mimic the backbone conformation

of Tyr and Pro residues and, therefore, other than eliminating a side chain, this substitution would not disrupt the structure of the R domains, as was shown by CD studies of CbpA-NR12-Y205G/Y358G (data not shown). The results for the GPT and YGT mutants binding to sIgA and SC-D15 were similar (Suppl. Table 2), and only those for the GPT mutants binding sIgA are discussed here. The k_a values for the mutant constructs binding to immobilized sIgA were similar to that of CbpA-NR12. However, the k_d values were different and depended on the number of intact YPT motifs. These rates are on average 5-fold larger (faster) than that observed for CbpA-NR12. Mutation of both YPT motifs increased the dissociation rate further to a value that is approximately 500-fold larger than that for CbpA-NR12. The value of ΔG_{SPR} for CbpA-NR12 binding to sIgA is -13.2 kcal mol^{-1} while those for the single GPT mutants are -12.2 kcal mol^{-1} and -12.3 kcal mol^{-1} (for CbpA-NR12-Y205G and -Y358G, respectively). The ΔG_{SPR} value for the double GPT mutant, CbpA-NR12-Y205G/Y358G, is -9.8 kcal mol^{-1}. These results show that one R domain, either R1 or R2, is sufficient for CbpA to bind sIgA (and SC) with high affinity. Affinity is diminished by about 1 kcal mol^{-1} if one residue in the YPT motif in either R1 or R2 in CbpA-NR12 is mutated and by about 3.5 kcal mol^{-1} if one residue in each YPT motif in CbpA-NR12 is mutated (Fig. 6A). CbpA-N failed to bind immobilized sIgA or SC-D15 in SPR experiments at concentrations up to 1.5 μM (data not shown), showing that the R domains within CbpA-NR12 are the principal determinants of interactions with sIgA (and SC). Further, the similarity of binding data for CbpA-R1 and -R2, and the two single GPT mutants of CbpA-NR12, indicates that the R1 and R2 domains interact similarly with the two R domain binding sites within sIgA and SC. It has been shown recently that CbpA binds to Ig-like domains 3 and 4 of secretory component (Elm et al., 2004a; Lu et al., 2003). It is possible that the two binding sites we have identified are found within these two domains.

Role of CbpA "R" domains in adhesion to nasopharyngeal epithelia

To understand whether there were parallels between our *in vitro* findings on the effects of mutations within the YPT motifs on CbpA binding to pIgR and bacterial adhesion to cellular receptors, we performed adhesion assays with pneumococci and human NE cells (Detroit cells). We expressed wild-type and mutant CbpA proteins in a CbpA null pneumococcal strain (ΔCbpA$^-$) and measured adhesion of these bacteria to Detroit cells (Fig. 6C). While the reproducibility of results from this biological assay is inferior to that associated with our SPR results, the adhesion results clearly show the same trends as do results from SPR. For example, mutation of the Tyr residue in one of the two YPT motifs does not have a statistically significant effect on the number of pneumococci that adhere to epithelial cells. However, mutation of Tyr within both YPT motifs reduces adhesion to the level associated with pneumococci that lack CbpA. These results show that, while CbpA is not the sole determinant of adhesion to human target cells, incorporation of mutations in two out of 693 amino acids in the CbpA sequence that significantly decrease affinity for sIgA and SC *in vitro* (Fig. 6A) have similar effects on adhesion to human cells.

Discussion

Novel structural features of CbpA domains

The mode of attachment of pneumococci to the surface of human epithelial cells is uniquely characterized by the absence of pili or fibrillar bacterial protrusions. Rather, the pneumococcus utilizes a novel mechanism to adhere to and invade human cells; this mechanism is mediated in part by CbpA which is secreted and recaptured onto the bacterial surface. Amongst pathogenic bacteria, the use of choline moieties on the bacterial surface as binding sites for functional

proteins is uniquely essential to pneumococci (Rosenow et al., 1997; Tomasz, 1967). We show here that domains R1 and R2 of CbpA adopt a strikingly simple structure comprised of three α-helices that bundle together through anti-parallel interactions into a flat, raft-like structure. While we have not experimentally determined the structure of CbpA-R1, the similarity of its primary sequence, CD spectra and thermodynamic and kinetic properties to those of the R2 domain strongly suggest that our homology model is accurate. The R domains are unusual in their helical topology and in the highly polar and charged nature of their molecular surfaces. A prominent feature with mixed polar and hydrophobic character—the tyrosine fork—near the loop between Helices 1 and 2 contains several conserved residues that are likely to play a key role in binding to Ig-like domains of pIgR. Most other conserved residues (Figs. 1C and 4D) are hydrophobic and are involved in α-helix/α-helix interactions. These may be conserved to preserve the helical topology of the R domains. Other conserved residues are charged and/or polar and are generally surface exposed; these residues may also play roles in interactions with pIgR.

The arrangement of the three α-helices of CbpA-R2 is different from that of other LZ proteins in that the anti-parallel α-helices are nearly co-axial, with crossing angles near -180° (-178° for Helices 1 and 2, -176° for Helices 2 and 3). We identified four proteins with one pair of anti-parallel α-helices comprised of LZ sequences that exhibit structural similarity to CbpA-R2 (Pdb file 1ROP, Dali Z-score, 5.4; 1GRJ, 4.9; 1CXZ, 4.8; and 1AQT, 4.2) (Holm and Sander, 1993; Oakley and Hollenbeck, 2001). The α-helix crossing angles in these proteins are -161°, -153°, -160°, and -159°, respectively. These crossing angles arise due to the bulky nature of the hydrophobic side chains of Ile, Leu and Val at positions **a** and **d** in the heptad repeats, which forces the helices to twist in order to pack "knobs-into-holes" (Crick, 1953). The LZ motifs in CbpA-R1 and -R2 differ from the classical motif in that alanine is often found at position **d** (Fig.

1C). The presence of a limited number of small alanine side chains at the α-helix/α-helix interface may allow nearly perfect anti-parallel packing of α-helices into the flat, raft-like structure of CbpA-R2.

Phylogenetic analysis reveals conserved structural features of CbpA

Our CbpA mutagenesis results showed that preservation of the Tyr and Pro residues within at least one YPT motif is essential for high-affinity binding to sIgA. Our results are consistent with the findings of others. In particular, Elm et al. (Elm et al., 2004b) showed that a fragment of SpsA (residues 37-283) with a single R2-like domain (SpsA is also termed PspC 2.1, which constitutes Group 4 in Fig. 2) was a competitive inhibitor of pneumococcal adherence to human cells that express pIgR. Importantly, this group also showed that a related CbpA fragment in which Tyr 201 (within the YPT motif) was mutated to Asp was not an inhibitor of adherence. We also observed that the isolated domains, CbpA-R1 and -R2, bind sIgA and SC with similar kinetic and thermodynamic constants. This suggests that the conserved features of the R domains are associated with receptor binding. Further, we reasoned that the residues in these domains that are most highly conserved amongst different CbpA sequences are involved in receptor binding. Our phylogenetic analysis showed that 22 residues that are conserved in all sequences (Figs. 1C and 4D) are found in the C-terminal portion of Helix 1, in the loop connecting Helices 1 and 2, and in the N-terminal portion of Helix 2. These residues, many of which are hydrophobic, are evenly distributed on the two faces and within the two α-helix/α-helix interfaces of the CbpA-R2 structure. The majority are found in Helix 2. For the R2 domain of the TIGR4 strain, an abundance of acidic residues in the midst of the conserved residues gives rise to a strong, negative electrostatic potential on one face of CbpA. However, the electronegative character near the loop containing the YPT motif is neutralized, in part due to the influence of two conserved

basic residues, Arg 356 and Lys 364 (Fig. 4E). This causes the side chains of the two Tyr residues (Tyr 358 and 363) of the tyrosine fork to protrude from one face of the raft-like structure into a region of low electrostatic potential (< ±1kt). In addition, several other conserved residues (Pro 359, Thr 360, and Thr 362) protrude prominently from the loop on the opposite face of the raft-like structure. Together, our results from mutagenesis and binding experiments and this structure-based analysis, strongly suggest that these conserved residues play important roles in binding to SC. While these residues sit in a node in the electrostatic potential map, it is possible that the SC surface has electropositive character that serves, in the initial stages of binding, to attract the electronegative surface of CbpA-R2 and possibly other CbpA R domains.

Materials and Methods

Construction of CbpA expression plasmids and Expression and purification of CbpA proteins. Expression plasmids for CbpA proteins were prepared using standard methods, as discussed in detail under Suppl. Materials. CbpA proteins for structural and biophysical studies were expressed in *E. coli* and purified using standard procedures, as given under Suppl. Materials.

Expression of SC-D15 in Sf9 cells and purification. A gene encoding pIgR D1-D5 (residues 1-589 of the mature protein with a C-terminal 6x His-tag) was subcloned into the baculovirus transfer vector pAcGP67b (Pharmingen) in frame with the gp67 hydrophobic secretion signal. Recombinant virus was generated by cotransfection of the transfer vector with linearized viral DNA (Baculogold; Pharmingen). SC-D15 was harvested from the supernatants of baculovirus-infected High 5 Cells (Invitrogen), which were concentrated and buffer exchanged into 20 mM

Tris pH 8.0, 150 mM NaCl. SC-D15 was purified on Ni-NTA resin (Qiagen) followed by gel filtration chromatography (Superdex 200, Amersham-Pharmacia).

Adhesion experiments. Pneumococci were grown in C+Y media (Lacks and Hotchkiss, 1960) to early stationary phase. All strains were grown in the presence of chloramphenicol (5 $\mu g\ ml^{-1}$), the ΔCbpA- mutant strain with erythromycin (1 $\mu g\ ml^{-1}$), and those transformed with the pNE1/CbpA vectors, with spectinomycin (500 $\mu g\ ml^{-1}$). Cells were pelleted and resuspended in fluorescein isothiocyanate (FITC; 1 mg/ml; Sigma) (Rosenow et al., 1997; Gosink et al., 2000). After incubation at 25 °C for 30 minutes, the bacteria were washed with PBS containing Ca^{2+} and Mg^{2+} (Mediatech), and diluted to 1 x 10^7 cfu/ml. Confluent monolayers of Detroit 562 epithelial cells (ATCC) were established in 96 well Terasaki trays and activated prior to infection with human TNFα (10 ng/well) for 2 hours. The cells were infected with 10^5 cfu of labeled bacteria for 30 minutes at 37 °C. The cells were washed four times with PBS containing Ca^{2+} and Mg^{2+} and fixed with 2.5% (vol/vol) glutaraldehyde. The number of adherent bacteria was determined by counting using fluorescence microscopy. We used the following procedure to determine that pneumococci transformed with pNE1/CbpA vectors expressed equal amounts of wild-type and mutant CbpA on the bacterial surface. The total bacterial CbpA and surface-bound CbpA (extracted from the bacterial surface with 10% choline) were quantified by Western analysis using a polyclonal anti-CbpA antibody generated against the CbpA N-terminus (residues 39-443) in rabbits. The antibody was purified using affinity chromatography before use.

NMR spectroscopy. Isotope labeled samples for NMR studies were prepared by culturing BL21(DE3) cells in isotope-labeled MOPS-based minimal media (Neidhardt et al., 1974). For

^{15}N-labeling, ^{15}N-ammonium chloride was used; for ^{13}C/^{15}N-labeling, ^{13}C-glucose was also used; and for ^{2}H/^{13}C/^{15}N-labeling, $^{2}H_2O$ was also used. 1 mM NMR samples were prepared in 10 mM potassium phosphate, 50 mM NaCl, pH 6.5, 5 % (v/v) $^{2}H_2O$ and 0.02 % (w/v) sodium azide. Gradient-, sensitivity-enhanced 2D ^{1}H-^{15}N TROSY spectra (Rance et al., 1999; Weigelt, 1998; Zhu et al., 1999) for CbpA-R2 and -NR12 were recorded using identical acquisition parameters: ^{1}H: spectral width, 14,368 Hz; data size, 1k complex points; ^{15}N: spectral width, 2,433 Hz; data size, 128 complex points. A Cosine window function and zero-filling were applied prior to Fourier transformation in each dimension. Backbone resonance assignments for ^{2}H/^{13}C/^{15}N-labeled CbpA-R2 were determined through the analysis of multiple 3D and 4D spectra, including 3D constant time- (CT-) HNCA, CT-HN(CO)CA (Yamazaki et al., 1994b), CT-HN(CA)CB, and CT-HN(COCA)CB (Yamazaki et al., 1994a), and 4D HNCOCA and HNCACO (Mulder et al., 2000). In addition, a 4D ^{1}H-^{15}N HSQC-NOESY-^{1}H-^{15}N HSQC spectrum was analyzed to confirm backbone assignments and to obtain local backbone helical distance restraints. Side-chain assignments were made through the analysis of 3D C(CO)NH-TOCSY, H(CCO)NH-TOCSY, HBHA(CBCACO)NH, HCCH-COSY and HCCH-TOSCY spectra. Additional 3D and 4D NOESY spectra, including 3D NOESY-^{1}H-^{15}N HSQC, 3D NOESY-^{1}H-^{13}C HSQC, 4D ^{1}H-^{15}N HSQC-NOESY-^{1}H-^{13}C HSQC and 4D ^{1}H-^{13}C HSQC-NOESY-^{1}H-^{13}C HSQC, were used to obtain distance restraints for structure calculations. Spectra were recorded at 25 °C using several spectrometers: a Varian Inova 600, a Varian INOVA 900, and a Bruker AVANCE 800. Spectra were processed using NMRPipe software (Delaglio et al., 1995) and analyzed using Felix software (Accelerys, Inc.). The ^{1}H dimensions of spectra were referenced to external TSP and the ^{13}C and ^{15}N dimensions were referenced indirectly using the appropriate ratios of

gyromagnetic ratios (Cavanagh et al., 1996). Secondary $^{13}C_\alpha$ chemical shift values ($\Delta\delta^{13}C_\alpha$) were calculated as described (Schwarzinger et al., 2001).

Structure Calculation, Refinement and Analysis. CbpA-R2 structures were calculated using interproton distance restraints estimated from 3D and 4D NOESY cross peak volumes. Restraint lower bounds were set to 1.8 Å and upper bounds to 2.5, 3.5, and 6.0 Å for large, medium, and small volumes, respectively. The program TALOS (Cornilescu et al., 1999) was used to estimate backbone dihedral angles psi (ψ) and phi (ϕ) based on $^1H_\alpha$, ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$, and $^{13}C'$ chemical shift values within the α-helical segments of CbpA-R2. In addition, hydrogen-bond (H-bond) restraints were used within α-helices. Amide protons involved in H-bonds were identified on the basis of slow exchange with ^2H_2O. Restraints are summarized in Table 2. Structures were calculated using torsion angle dynamics (TAD) (Stein et al., 1997) within CNS (Brunger et al., 1998). The TAD protocol was performed in stages: (1) 75 ps high-temperature TAD (50,000 K) followed by (2) cooling to 1,000 K over the course of 75 ps and ramping of the van der Waals scaling term from 0.1 to 1.0. (3) The molecules were further cooled to 300 K over the course of 20 ps using conventional Cartesian dynamics followed finally by (4) 10,000 steps of conjugate gradient energy minimization. The NOE (dihedral) restraint energy term was 150 kcal mol^{-1} (100 kcal mol^{-1}) for stages 1-3 and 100 kcal mol^{-1} (300 kcal mol^{-1}) for the last. Two hundred structures were calculated, and the 40 lowest-energy structures were further refined by using the SANDER module of AMBER 8.0 (Case et al., 2004). Solvent was represented by the Generalized-Born (GB) model (Xia et al., 2002). Structures were first energy minimized for 1 ps without restraints followed by 40 ps of simulated annealing from 400 K to 0 K with all restraints. The distance and angle restraint force constants were 20 kcal mol^{-1} Å^{-2} and 2 kcal mol^{-1} Å^{-2}, respectively. Statistics for the 20 lowest-energy structures are given in Table 2. The program Define_S was used to

determine the α-helix crossing angles for CbpA-R2 and other LZ-containing proteins (Richards and Kundrot, 1988). Further, we compared the structure of CbpA-R2 with those of other proteins using the Dali server (http://www.ebi.ac.uk/dali/Interactive.html).

Homology modeling of CbpA-R1. We used the program MOE (Chemical_Computing_Group) to calculate a homology model of CbpA-R1 using the experimental structure of CbpA-R2 as the template. Briefly, MOE first aligned the sequence of the target (R1) with that of the template (R2). Next, the side chains of non-identical residues were built and side chain conformations were adjusted by reference to a rotomer library followed by energy minimization using the AMBER '94 forcefield.

SPR measurements. SPR experiments were performed at 25 °C using a BIACORE 3000 instrument (Biacore, Inc.). sIgA and SC-D15 were covalently attached to different carboxymethyl-dextran coated gold surfaces (CM-5 Chip, Biacore) using the manufacturer's procedures. Binding was measured by flowing CbpA proteins in 10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P20, pH 7.4 (HBS-EP buffer, Biacore) at a flow rate of 20 μL/min through the reference and sIgA-containing flow cells in sequence. A blank was also run consisting of only buffer. Following the injection, dissociation of CbpA proteins was measured by flowing only buffer through the cells. The chip surfaces were regenerated by injecting 20 μL of 0.1% SDS through the cells. Data reported is the difference in SPR signal between the flow cell containing sIgA and the reference cell. Contributions to the signal from buffer were removed by subtraction of the blank (buffer only) injection from the reference-subtracted signal (Morton and Myszka, 1998). Triplicate injections were made and the average taken. Data were analyzed by simultaneously fitting association and dissociation phases at all concentrations using the BiaEvaluation software (Biacore, Inc.). The model used for fitting was the simplest that was able

to account for the observed binding. The concentrations of CbpA constructs used in these experiments are given in the legend to Suppl. Fig. 3.

ITC experiments. ITC experiments were performed at 25 °C using a VP-ITC (Microcal) calorimeter. A preliminary 2 μL injection of either CbpA-R1, CbpA-R2, CbpA-R12, or CbpA-NR12 (all at 40 μM) was followed by multiple injections of 5 μL at 300 sec intervals into a solution of 4 μM SC-D15. All proteins were dissolved in 20 mM sodium phosphate, pH 6.5, 100 mM NaCl. Thermodynamic parameters were obtained using Origin software (OriginLab) using a single binding site model after correcting heat values for the effect of dilution. The results of titrations with the CbpA constructs and SC-D15 indicated that the effective concentration of SC-D15 was 1.42 μM, indicating that only 36% of SC molecules were competent for binding CbpA. This conclusion was confirmed using gel filtration chromatography, as follows. In the absence of CbpA, SC-D15 migrated as a single species. Addition of an excess of CbpA-R1 caused only ~40% of the parent SC peak intensity to shift to a more rapidly migrating species (corresponding to the CbpA-R1:SC-D15 complex). Formation of this complex was complete upon addition of approximately 0.8 molar equivalents of CbpA-R1. We used this effective concentration of SC-D15 (1.42 μM) when analyzing raw heat data to obtain CbpA:SC-D15 mole ratios (Table 1).

Acknowledgments.

The authors acknowledge Weixing Zhang for assistance with NMR experiments, Daniel Stokes and David Cardwell for assistance with protein production, Abby Parrill for generating the homology model of CbpA-R1, Cheon-Gil Park and Ross Hilliard for help with preparation of SC-D15 and ITC experiments, Charles Ross for computer support, and Charles Galea for helpful comments on the manuscript. Ranjith Muhandiram and Lewis Kay are acknowledged for graciously providing Varian NMR pulse programs. The authors thank Dr. Eriks Kupce of Varian for recording the 900 MHz 2D TROSY spectrum of CbpA-R2, and Peter M. Snow and the Caltech Protein Expression Facility for expression of SC-D15. A.E.H. was supported by the Whitehead Institute for Biomedical Research. The authors acknowledge support from the American Lebanese Syrian Associated Charities (ALSAC), the National Cancer Institute (CA82491 to RWK), National Center for Research Resources (RR014675 for a Biacore 3000 instrument) and a Cancer Center (CORE) Support Grant (CA 21765). The authors dedicate this manuscript to the memory of Peter M. Snow who died tragically in a bicycling accident in Maine, USA, on Wednesday, August 4th.

References

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P.G.-K., R. W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905-921.

Butler, J.C., Breiman, R.F., Campbell, J.F., Lipman, H.B., Broome, C.V. and Facklam, R.R. (1993) Pneumococcal polysaccharide vaccine efficacy. An evaluation of current recommendations. *JAMA*, **270**, 1826-1831.

Case, D.A., Darden, T.A., Cheatham, I., T. E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Wang, B., Pearlman, D.A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J.W., Ross, W.S. and Kollman, P.A. (2004) AMBER 8. University of California, San Francisco.

Cavanagh, J., Fairbrother, W.J., Palmer III, A.G. and Skelton, N.J. (1996) *Protein NMR Spectroscopy*. Academic Press, New York.

Chemical_Computing_Group. MOE version 2004.02. Montreal, Quebec, Canada.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, **13**, 289-302.

Creighton, T.E. (1993) *Proteins: Structures and Molecular Properties*. W. H. Freeman & Co., New York, NY.

Crick, F.H.C. (1953) The packing of aplha-helices: simple coiled-coils. *Acta Crystallogr.*, **6**, 689-697.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMR Pipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

Elm, C., Braathen, R., Bergmann, S., Frank, R., Vaerman, J.P., Kaetzel, C.S., Chhatwal, G.S., Johansen, F.E. and Hammerschmidt, S. (2004a) Ectodomains 3 and 4 of Human Polymeric Immunoglobulin Receptor (hpIgR) Mediate Invasion of Streptococcus pneumoniae into the Epithelium. *J Biol Chem*, **279**, 6296-6304. Epub 2003 Dec 6293.

Elm, C., Rohde, M., Vaerman, J.P., Chhatwal, G.S. and Hammerschmidt, S. (2004b) Characterization of the interaction of the pneumococcal surface protein SpsA with the human polymeric immunoglobulin receptor (hpIgR). *Indian J Med Res*, **119**, 61-65.

Fernandez-Tornero, C., Lopez, R., Garcia, E., Gimenez-Gallego, G. and Romero, A. (2001) A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nat. Struct. Biol.*, **8**, 1020-1024.

Gosink, K. and Tuomanen, E. (2000) Streptococcus pneumoniae: Invasion and Inflammation. In Fischetti, V.A. (ed.), *Gram-positive Pathogens*. American Society for Microbiology, Washington, D. C.

Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I. and Masure, H.R. (2000) Role of novel choline binding proteins in virulence of Streptococcus pneumoniae. *Infect. Immun.*, **68**, 5690-5695.

Hammerschmidt, S., Talay, S.R., Brandtzaeg, P. and Chhatwal, G.S. (1997) SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol. Microbiol.*, **25**, 1113-1124.

Hammerschmidt, S., Tillig, M.P., Wolff, S., Vaerman, J.P. and Chhatwal, G.S. (2000) Species-specific binding of human secretory component to SpsA protein of Streptococcus pneumoniae via a hexapeptide motif. *Mol. Microbiol.*, **36**, 726-736.

Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233**, 123-138.

Iannelli, F., Oggioni, M.R. and Pozzi, G. (2002) Allelic variation in the highly polymorphic locus pspC of Streptococcus pneumoniae. *Gene*, **284**, 63-71.

Lacks, S. and Hotchkiss, R.D. (1960) A study of the genetic material determining an enzyme in Pneumococcus. *Biochim Biophys Acta*, **39**, 508-518.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **240**, 1759-1764.

Lu, L., Lamm, M.E., Li, H., Corthesy, B. and Zhang, J.R. (2003) The human polymeric immunoglobulin receptor binds to Streptococcus pneumoniae via domains 3 and 4. *J. Biol. Chem.*, **278**, 48178-48187. Epub 42003 Sep 48117.

Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.*, **21**, 375-382.

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

Morton, T.A. and Myszka, D.G. (1998) Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Methods Enzymol.*, **295**, 268-294.

Mostov, K.E. (1994) Transepithelial transport of immunoglobulins. *Ann. Rev. Immunol.*, **12**, 63-84.

Mostov, K.E. and Kaetzel, C.S. (1999) Immunoglobulin transport and the polymeric immunoglobulin receptor. In *Mucosal Immunology*. Academic Press, New York.

Mulder, F.A., Ayed, A., Yang, D., Arrowsmith, C.H. and Kay, L.E. (2000) Assignment of 1H(N), 15N, 13C(alpha), 13CO and 13C(beta) resonances in a 67 kDa p53 dimer using 4D-TROSY NMR spectroscopy. *J Biomol NMR*, **18**, 173-176.

Neidhardt, F.C., Bloch, P.L. and Smith, D.F. (1974) Culture medium for enterobacteria. *J. Bact.*, **119**, 736-747.

Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281-296.

Nozaki, Y. and Tanford, C. (1971) The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J Biol Chem*, **246**, 2211-2217.

Oakley, M.G. and Hollenbeck, J.J. (2001) The design of antiparallel coiled coils. *Curr. Opin. Struct. Biol.*, **11**, 450-457.

Rance, M., Loria, J.P. and Palmer, A.G.r. (1999) Sensitivity improvement of transverse relaxation-optimized spectroscopy. *J Magn Reson*, **136**, 92-101.

Richards, F.M. and Kundrot, C.E. (1988) Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins*, **3**, 71-84.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A. and Masure, H.R. (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. *Mol. Microbiol*, **25**, 819-829.

Schuchat, A., Hilger, T., Zell, E., Farley, M.M., Reingold, A., Harrison, L., Lefkowitz, L., Danila, R., Stefonek, K., Barrett, N., Morse, D. and Pinner, R. (2001) Active bacterial core surveillance of the emerging infections program network. *Emerg Infect Dis*, **7**, 92-99.

Schwarzinger, S., Kroon, G.J.A., Foss, T.R., Chung, J., Wright, P.E. and Dyson, H.J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Amer. Chem. Soc.*, **123**, 2970-2978.

Stein, E.G., Rice, L.M. and Brunger, A.T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J Magn Reson*, **124**, 154-164.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M. (2001) Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science*, **293**, 498-506.

Tomasz, A. (1967) Choline in the cell wall of a bacterium: novel type of polymer-linked choline in Pneumococcus. *Science*, **157**, 694-697.

Weigelt, J. (1998) Single Scan, Sensitivity- and Gradient-Enhanced TROSY for Multidimensional NMR Experiments. *J. Amer. Chem. Soc.*, **120**, 10778-10779.

Whitney, C.G., Farley, M.M., Hadler, J., Harrison, L.H., Lexau, C., Reingold, A., Lefkowitz, L., Cieslak, P.R., Cetron, M., Zell, E.R., Jorgensen, J.H. and Schuchat, A. (2000) Increasing prevalence of multidrug-resistant Streptococcus pneumoniae in the United States. *N Engl J Med*, **343**, 1917-1924.

Xia, B., Tsui, V., Case, D.A., Dyson, H.J. and Wright, P.E. (2002) Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water. *J Biomol NMR*, **22**, 317-331.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994a) A suite of triple resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity. *J. Amer. Chem. Soc.*, **116**, 11655-11666.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D.L., Dahlquist, F.W., Arrowsmith, C.H. and Kay, L.E. (1994b) An HNCA pulse scheme for the backbone assignment of 15N, 13C, 2H-labeled proteins: Application to a 37-kDa Trp repressor-DNA complex. *J. Am. Chem. Soc.*, **116**, 6464-6465.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M. and Tuomanen, E. (2000) The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell*, **102**, 827-837.

Zhu, G., Kong, X. and Sze, K. (1999) Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *J Biomol. NMR*, **13**, 77-81.

Table 1. Thermodynamic parameters obtained from analysis of ITC data for CbpA fragments binding to SC-D15 in solution.

Construct	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	–TΔS (kcal mol^{-1})	$N_{normalized}$* (moles CbpA/moles SC)
CbpA-R1	-9.82 ± 0.01	-26.60 ± 0.05	+16.78 ± 0.10	1.86 ± 0.01
CbpA-R2	-10.23 ± 0.02	-21.19 ± 0.07	+10.96 ± 0.10	2.19 ± 0.01
CbpA-R12	-10.39 ± 0.02	-46.30 ± 0.14	+35.91 ± 0.10	0.87 ± 0.01
CbpA-NR12	-10.40 ± 0.01	-38.88 ± 0.07	+28.48 ± 0.10	1.07 ± 0.01

* $N_{normalized}$ values were determined using 1.42 μM as the effective SC-D15 protein concentration (see Experimental Procedures for details). ΔG, ΔH and –TΔS values are in units of kcal per mole of CbpA construct.

Table 2. Statistics of 20 lowest-energy structures of CbpA-R2 based on solution NMR data.

Total number of NOEs	**2,292**
Intra-residue	922
Inter-residue	1,370
Sequential	622
Medium-range (i, i up to ± 4)	599
Long-range (i, i ± 5 or larger)	149
Total number of dihedral restraints	**166**
ψ	83
ϕ	83
RMSD from mean structure (Å)	
All residues within structural region	
Backbone heavy atoms (328-425)	1.1 ± 0.2
All heavy atoms	1.8 ± 0.4
Helixes 1-3 (residues 330-357, 365-390, 395-425)	
Backbone heavy atoms	0.6 ± 0.1
All heavy atoms	1.5 ± 0.3
Distance restraint violations (average number per structure)	
Restraints violated by >0.50 Å	1.5
Maximum distance restraint violation (Å)	1.7
Torsion angle restraint violations (average number per structure)	
ψ restraints violated by >5°	0
ϕ restraints violated by >5°	0
Ramachandran ψ, ϕ statistics	
Angles in most favored region	86.7%
Angles in allowed region	10.2%
Angles in generously allowed region	2.4%
Angles in disallowed region	0.7%

Figure 1. Domain structure of CbpA (from the TIGR4 strain of *Streptococcus pneumoniae*). (**A**) Domains labeled A, B, R1, R2 and C exhibit multiple repeats of the LZ motif, and the domain labeled CBD (choline binding domain) contains eight repeats of the choline binding motif. (**B**) Leucine zipper probability for CbpA determined using COILS (Lupas et al., 1991). (**C**) Sequences of domains R1 and R2. The letters **a-g** and horizontal lines indicate the locations of the 12 LZ heptad motifs. Within the LZ motifs, hydrophobic residues are colored brown (L, V, I, and A); acidic residues, red (D and E); and basic residues, blue (K and R). Sites in the R1 and R2 sequences where an acidic residue is swapped for a basic residue, or *vice versa*, are indicated by red and blue shading, respectively. Residues of the conserved RNYPT motif are colored green and the three α-helices of CbpA-R2 are indicated by black rectangles. Residues that are conserved in 87 R domain sequences from 47 pneumococcal strains are also given.

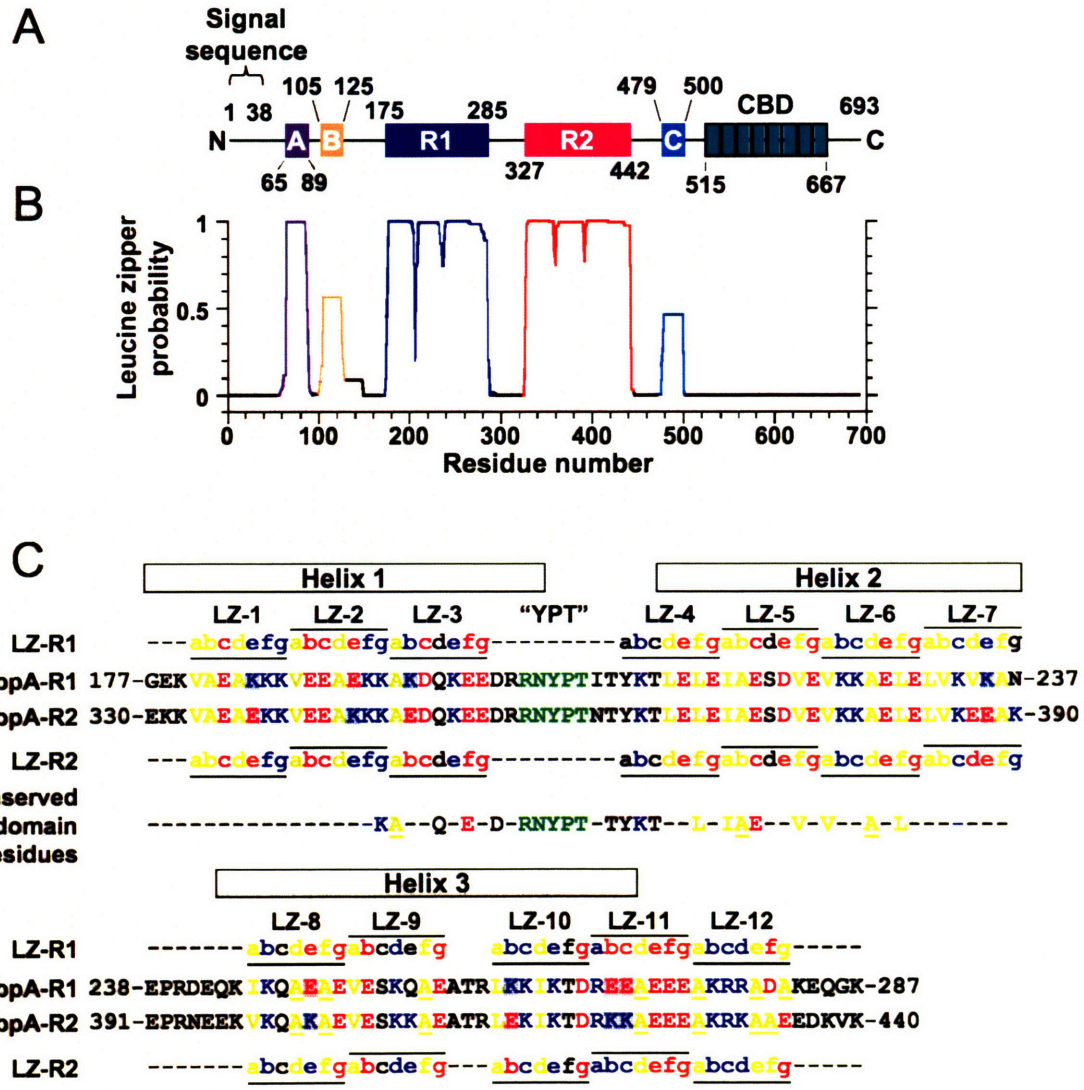


Figure 1. Domain structure of CbpA (from the TIGR4 strain of *Streptococcus pneumoniae*).

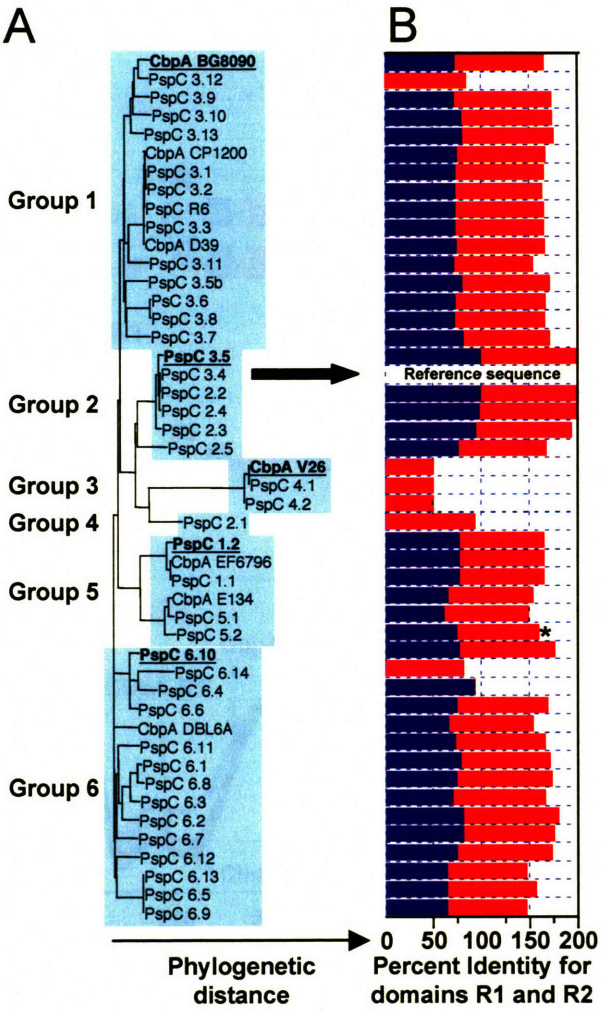


Figure 2. R domain structure is conserved in CbpA sequences from most pneumococcal strains. (**A**) Phylogenetic tree showing the relationships between 47 CbpA sequences, which cluster into six groups. The TIGR4 sequence is also termed PspC 3.4 (marked by arrow). (**B**) The histogram illustrates the percentage of amino acid identity (relative to TIGR4 CbpA) for the R1 and/or R2 domains (R1, blue bars; R2, red bars). Details of this analysis are found in Supplementary data.

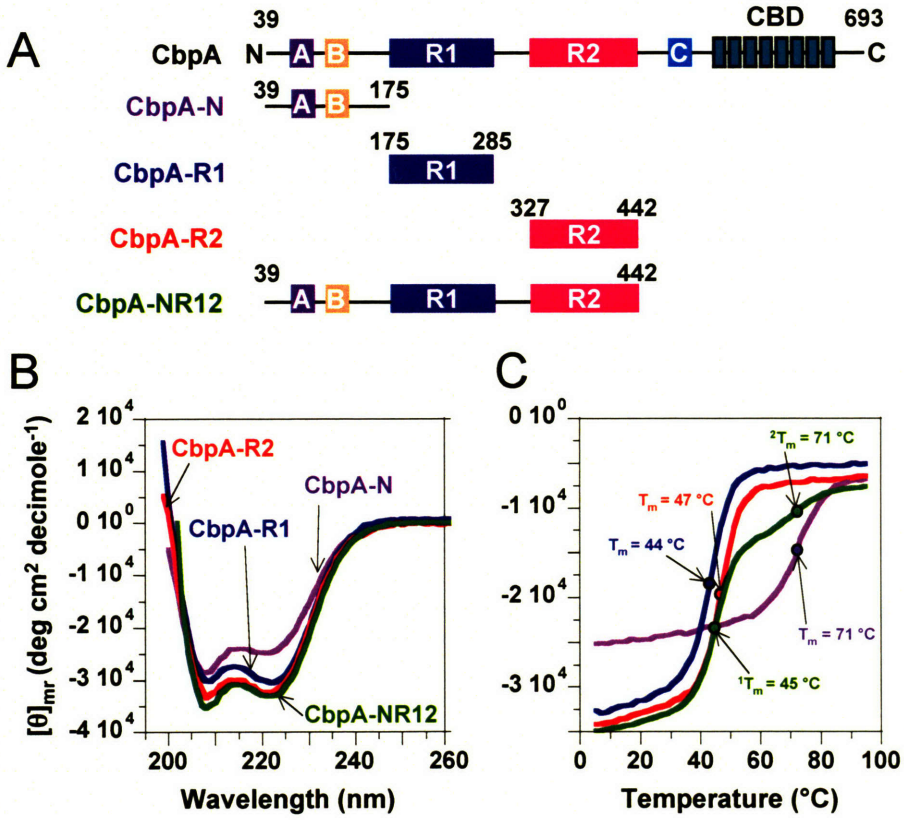


Figure 3. Secondary structure of CbpA domains. (**A**) CbpA constructs used in this study. (**B**) CD spectra for CbpA-N (violet trace), -R1 (blue trace), -R2 (red trace), and -NR12 (green trace). (**C**) Thermal denaturation traces obtained by measuring CD ellipticity at 222 nm at different temperatures. The coloring scheme is as in (B).

Figure 4. Solution structure of CbpA R domains. (**A**) Secondary $^{13}C_{\alpha}$ chemical shift values for CbpA-R2 showing the three α-helices. Resonances for residues marked by red asterisks are unassigned. (**B**) Superposed ensemble of 20 lowest-energy structures of CbpA-R2 obtained from solution NMR data; the backbone and select hydrophobic residues (in brown color) at the two α-helix/α-helix interfaces are illustrated. Helix 1 is colored red; Helix 2 blue; and Helix 3 green. The location of the YPT motif is noted. (**C**) End-on view of the three α-helices of domain R2, colored as in (B). (**D**) Contour maps of electrostatic potentials (± 1kt; red, negative potential; blue, positive potential) for CbpA-R2 (left) and homology model of CbpA-R1 (right) generated using GRASP (Nicholls et al., 1991). The α-helices are colored as in (B). Tyr 358 & 363 (magenta; labeled "tyrosine fork"), Pro 359 (yellow), and Thr 360 & 362 (orange) are also illustrated. The C_{α} atoms of other conserved residues are illustrated as colored spheres (Lys 346, Arg 356 and Lys 364 (blue); Glu 352, Asp 354 and Glu 372 (red); Gln 350 & Thr 365 (gray); Ala 347, Ile 368, Ile 370, Ala 371, Val 375, Val 377, Ala 80 and Leu 382 (yellow)). (**E**) Close-up view of conserved residues in loop between Helix 1 and Helix 2 of CbpA-R2 that protrude into a region of neutral electrostatic potential. Key residues are illustrated as in (D).

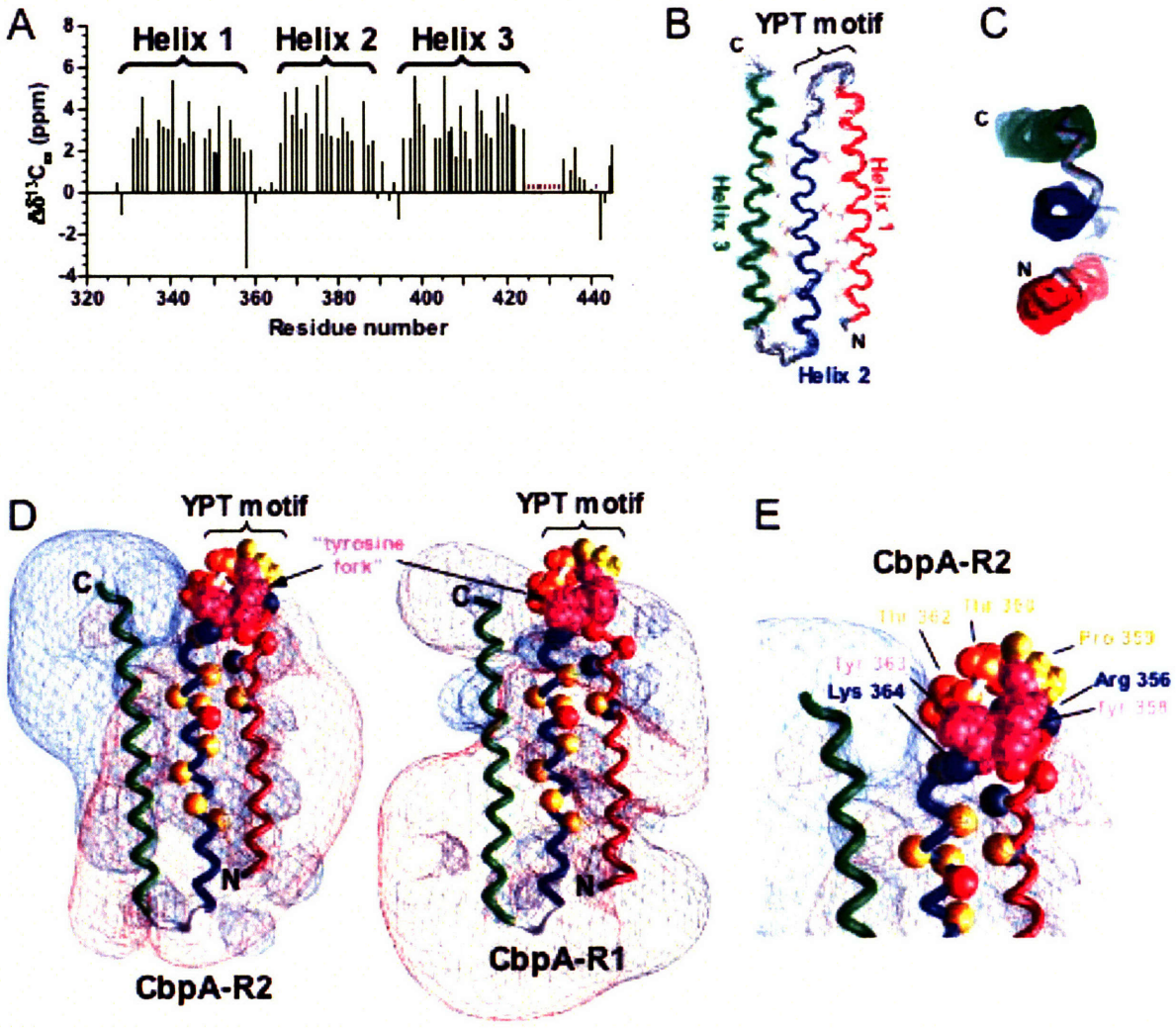


Figure 4. Solution structure of CbpA R domains.

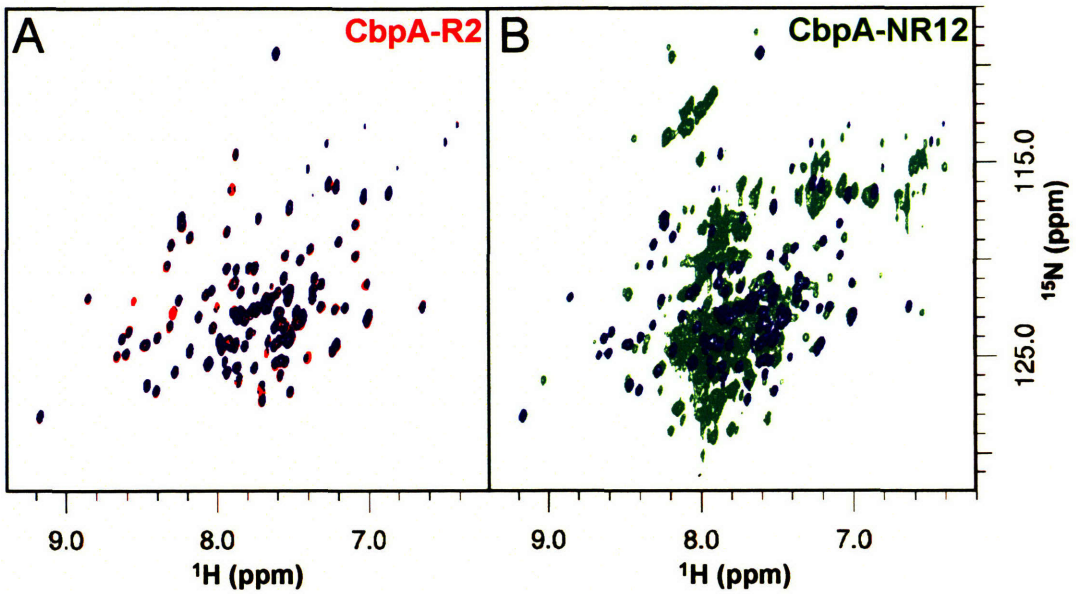


Figure 5. 800 MHz 2D ^{1}H-^{15}N TROSY spectra of ^{2}H/^{13}C/^{15}N-labeled CbpA-R2 (**A**) and -NR12 (**B**). The individual spectra are illustrated in red (A) and green (B) ink. The blue overlays in panels A and B identify resonances that appear at identical positions in the two spectra.

Figure 6. Insights into the molecular mechanism of CbpA/pIgR interactions. (**A**) Histogram illustrating the association (k_a) and dissociation (k_d) rate constants derived from SPR data for different CbpA constructs interacting with immobilized sIgA. The rate constants were obtained by fitting equations for a 1:1 binding model to raw data like those illustrated in Supplementary Figure 2B. (**B**) Raw ITC data (top) and binding curves (bottom) for CbpA-R1 (blue), -R2 (red), -R12 (black) and -NR12 (green) binding to SC-D15. In the bottom panels, the colored circles show raw data points and the black lines show the fit of equations for a single binding site model to the raw data. Only every second data point is illustrated although all points were included in the analysis. The CbpA fragment:SC-D15 mole ratios at the reaction endpoints are noted in each bottom panel. (**C**) Adhesion of pneumococci to NE (Detroit) cells. ΔCbpA$^-$ pneumococci were transformed with pNE1 plasmids that encoded full-length CbpA or CbpA with one or two mutations within the YPT motif. ΔCbpA$^-$ (control) corresponds to results for untransformed ΔCbpA$^-$ bacteria.

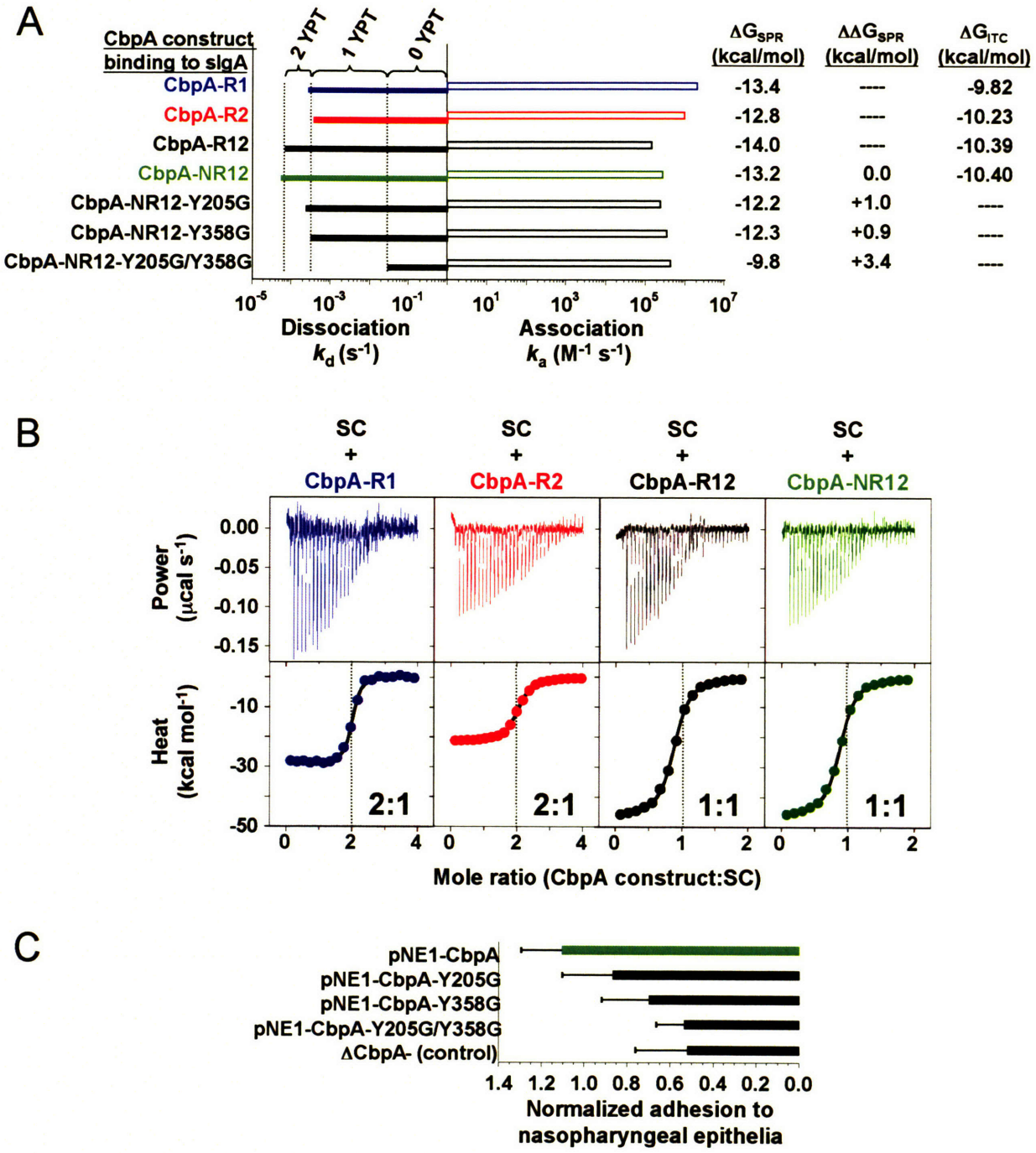


Figure 6. Insights into the molecular mechanism of CbpA/pIgR interactions.

Supplementary Material

Materials and Methods

Construction of CbpA expression plasmid. The following segments of CbpA genomic DNA from the TIGR4 strain of *Streptococcus pneumoniae* (accession number: nucleotide, AAK76241; protein, AAK76241) were sub-cloned into plasmids for protein expression in *E. coli*: residues 39-174 (CbpA-N), 175-289 (CbpA-R1), 329-443 (CbpA-R2), 175-443 (CbpA-R12), 39-321 (CbpA-NR1) and 39-442 (CbpA-NR12). CbpA-N, CbpA-R1, CbpA-R2, and CbpA-R12 were subcloned into the bacterial expression vector pET28a (Novagen). CbpA-NR1 and CbpA-NR12 were subcloned into pQE-30 (Qiagen). Site-directed mutagenesis was used to introduce mutations into the expression plasmid for CbpA-NR12 (pQE-CbpA-NR12), as follows. In each construct, the following residues were mutated to Gly, Tyr 205 (pQE-CbpA-NR12-Y205G), Tyr 358 (pQE-CbpA-NR12-Y358G), Tyr 205 and Tyr 358 (pQE-CbpA-NR12-Y205G/Y358G), Pro 206 (pQE-CbpA-NR12-P206G), Pro 359 (pQE-CbpA-NR12-P359G), and Pro 206 and Pro 359 (pQE-CbpA-NR12-P206G/P359G). Further, full-length genomic DNA for CbpA was subcloned into pNE1 (Bartilson et al., 2001) (pNE1-CbpA) to express CbpA in ΔCbpA$^-$ pneumococci. Site-directed mutagenesis was used to prepare pNE1 plasmids for expression of full-length CbpA with the same mutations as described above; these plasmids were pNE1-CbpA-Y205G; pNE1-CbpA-Y358G; pNE1-CbpA-Y205G/Y358G; pNE1-CbpA-P206G; pNE1-CbpA-P359G; and pNE1-CbpA-P206G/P359G.

Expression and purification of CbpA proteins. The pET and pQE CbpA plasmids were used to express CbpA fragments as His-tagged polypeptides in *E. coli* BL21(DE3) (Studier et al., 1990) using standard procedures. Proteins were purified using Ni^{2+}-affinity chromatography

(Amersham-Pharmacia resins), followed in some cases (CbpA-R1, -R2, and -R12) by thrombin (Calbiochem) cleavage to remove the His-tag. CbpA-N, -NR1 and -NR12 exhibited secondary site cleavage by thrombin within the CbpA sequence and, therefore, were prepared with intact His tags. Proteins were further purified by gel filtration chromatography (Superdex 200, Amersham Biosciences) in 20 mM sodium phosphate, pH 6.5, 20 mM NaCl, and 0.02% (w/v) sodium azide. CbpA proteins contained exogenous N-terminal residues, as follows: CbpA-N, MGSSHHHHHHSSGLVPRGSHM; CbpA-NR1 and -NR12, MRGSHHHHHHGSM; and CbpA-R1, -R2 and -R12, GSHM. Protein concentrations were determined using absorbance at 280 nm (Gill and von Hippel, 1989) and extinction coefficients determined using the ProteinParameters tool at the ExPASy web site (http://us.expasy.org/tools/protparam.html).

CD experiments. Circular dichroism (CD) spectra (Fig. 3B and C) were recorded at 25 °C using an AVIV 62A DS spectropolarimeter in 1.0 cm path length quartz cells. CbpA samples were prepared by dialysis against 1 mM sodium phosphate, pH 6.5, 50 mM NaCl. In thermal denaturation experiments, the ellipticity at 222 nm was recorded after 1 min. equilibration at temperatures from 5 °C to 95 °C in 2 °C steps.

CbpA-sIgA binding experiments using ELISA. To analyze CbpA-sIgA binding, each well of a 96 well plate was coated with 0.5 mg purified recombinant CbpA protein. Wells were blocked for 1 hour with rabbit serum (diluted 1:50 with 10 mM sodium phosphate, pH 7.2, 150 mM NaCl (PBS)), and 0.5 mg sIgA protein (ICN) was applied for 1 hour. After washing with PBS, an antibody to secretory component (Sigma) was applied after a 1:2000 dilution in PBS for 1 hour. Biotinylated anti-mouse IgG (Vector Labs) and ABC detection reagent were then applied

according to the manufacturer's protocol. Turbo TMB (Pierce) was used as a chromogenic substrate. Reactions were stopped with 1 M sulfuric acid, and the absorbance at 450 nm of each well was determined with a plate reader. All reactions were performed in triplicate at room temperature.

AUC experiments. Sedimentation equilibrium experiments were performed with CbpA-N, CbpA-R1, CbpA-R2 and CbpA-A at 20 °C in a Beckman XL-A analytical ultracentrifuge, at operational speeds in the range 27,000 – 40,000 rpm. The polypeptide concentrations for CbpA-N, CbpA-R1, and CbpA-R2 ranged from 0.01 mM to 4 mM. Polypeptides were dissolved in 20 mM sodium phosphate buffer, pH 6.5, in the presence of 0 mM, 50 mM or 200 mM NaCl. Depending on the protein concentration either 12 mm or 3 mm optical path length cells were used. Data were manipulated using the Beckman XL-A software.

The software INVEQ (Rowe, in preparation) was employed to analyze sedimentation equilibrium data. INVEQ fits the data set to the following equation:

$$r = \{(\ln(c_r/c_i) + 0.5*(\sigma_w/(1+2BMc_r))*r_i2) / (0.5*(\sigma_w/(1+2BMc_r)))\}0.5 \qquad \text{(Equation 1)}$$

where r is any radial position at which the solute concentration c has the value c_r, and r_i and c_i are the values of these parameters at a defined reference position. The latter radial position is usually taken as being the data point closest to the meniscus. The parameter σ is the reduced molecular weight of the solute, defined as

$$\sigma = M(1 - \bar{v}\rho)\omega^2/2RT \qquad \text{(Equation 2)}$$

where M is the molecular weight of the solute, $\bar{v}$ (ml/g) its partial specific volume, ρ is the density of the solvent, ω (radians/sec) the angular velocity of the rotor, R is the gas constant and T the temperature (°K).

For any monomer-dimer equilibrium system it is simple from a knowledge of the equilibrium association constant (K_a) and the solute molar concentration to define the mole fraction α of monomers which have dimerised, from which the weight averaged σ value ((σ_w, in mass rather than molar units) can be derived. Representing the thermodynamic non-ideality by a single second virial coefficient term BM (formally the B value is a $B_{1,1}$ term) assumed to be the same for both monomer and dimer, the apparent σ value ($\sigma_{apparent}$) to be used in the equilibrium equation is given by

$$\sigma_{apparent} = (\sigma_w/(1+2BMc_r)). \qquad \text{(Equation 3)}$$

It should be noted that equation 1 is simply a usual form of the equation for sedimentation equilibrium *inverted* to give r = f(c) rather than the normal c = f(r) format. Although apparently trivial, this is important for stability in curve fitting. The more usual (c = f(r)) format becomes recursive when terms covering self-association and/or non-ideality are introduced. The INVEQ format avoids this problem, and by providing a more rigorous way of fitting for K_a than is employed in direct fitting methods (Rowe, in preparation) it becomes possible to float both K_a and the non-ideality term (BM) in the fitting algorithm. Using this approach, it was possible to estimate weak (K_d up to 100 mM) interaction coefficients despite the inevitable presence of a non-ideality term of similar numerical magnitude.

All data were analyzed using locally written programs in the software Pro Fit™ (Quantumsoft). In order to obtain a fit when lower solute concentrations were being studied we either fixed BM = 5 ml/g, a typical value for this system, and/or we floated the baseline offset E. The latter quantity can justifiably be set equal to zero for higher c values, where absorption optics are in use, but small errors can cause problems with more dilute systems.

CbpA sequence analysis. Methods. The analysis of CbpA (Fig. 2) was performed using Vector NTI 9.0.0 software (Informax). The accession numbers for CbpA sequences are as given in Tables 1 and 2 of Iannelli, *et al.* (Iannelli et al., 2002). The sequences of the C-terminal choline binding domains were deleted before analysis so that relationships within the N-terminal segments could be more clearly observed. The phylogenetic tree illustrated in Fig. 2A was generated using the Align feature in Vector NTI, which uses the Neighbor Joining (NJ) algorithm of Saitou and Nei (Saitou and Nei, 1987). For Fig. 2B, all CbpA sequences were individually aligned with the sequences of domains R1 and R2 from the TIGR4 strain. If similarity to one or both domains was identified, the percentage identity was determined. Also, whether the sequence RNYPT was identified within the R domains was noted. The sequences of the R1 and/or R2 domains were aligned; the consensus sequence illustrated in Fig. 1C shows residues that are identical in all 87 R domain sequences. The sequence of the R2 domain of PspC 5.2 used in this analysis consisted of residues 308-371 which correspond to Helices 1 and 2 of the CbpA TIGR4 R2 domain.

Results. The N-terminal segments of the 47 CbpA sequences can be divided into six phylogenetically related groups. The CbpA sequences in Groups 1, 2, 5 and 6, including the TIGR4 sequence (also referred to as PspC 3.4) (Iannelli et al., 2002), are highly related and all

but four (PspC 3.12, PspC 5.2, PspC 6.4, and PspC 6.14) exhibit two R domains that are very similar to domains R1 and R2 of the TIGR4 sequence (Fig. 2B). The sequence of PspC 3.12 (Group 1) exhibits only one R domain, which is most similar to the TIGR4 R2 domain. PspC 5.2 (Group 5) exhibits a typical R1 domain but lacks a complete R2 domain. However, a segment of this sequence is 85% identical to the Helix 1 and Helix 2 segment of domain R2 of CbpA-TIGR4, including the conserved residues illustrated in Fig. 1C. PspC 6.14 (Group 6) exhibits a single R2-like domain while PspC 6.4 (Group 6) exhibits a single R1-like domain. A lone sequence constitutes group 4, PspC 2.1, which contains a single R2-like domain. R domain sequence similarity is illustrated in Fig. 2B. Group 3 is comprised of three CbpA and four CbpA-like sequences that each exhibits a single R2-like domain, with identity to that of the TIGR4 sequence ranging from 50% to 78%. All of the aforementioned sequences contain the conserved consensus sequence (Fig. 1C), including the YPT motif and other conserved residues (Fig. 1C and 4D).

References for Supplementary Material

Bartilson, M., Marra, A., Christine, J., Asundi, J.S., Schneider, W.P. and Hromockyj, A.E. (2001) Differential fluorescence induction reveals Streptococcus pneumoniae loci regulated by competence stimulatory peptide. *Mol Microbiol*, **39**, 126-135.

Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, **182**, 319-326.

Iannelli, F., Oggioni, M.R. and Pozzi, G. (2002) Allelic variation in the highly polymorphic locus pspC of Streptococcus pneumoniae. *Gene*, **284**, 63-71.

Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.*, **94**, 12366-12371.

Rowe, A.J. (in preparation) Weak interactions: optimal algorithms for their study in the AUC. In Scott, D.J., Harding, S.E. and Rowe, A.J. (eds.), *Modern Analytical Ultracentrifugation: Techniques & Methods*. Royal Society of Chemistry, London.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4**, 406-425.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60-89.

Supplementary Table 1. Values of the non-ideality term (2BM) and equilibrium dissociation constant (K_D) derived from analysis of sedimentation equilibrium centrifugation data for CbpA domains.

	Self-association parameters		
Construct	**2BM (ml g^{-1})**	**K_D (mM)**	**Polypeptide concentration (mM)**
CbpA-R1, 0 mM NaCl	4.2	22.2	4.5
CbpA-R1, 50 mM NaCl	15.0	2.4	1.9
CbpA-R1, 200 mM NaCl	8.0	4.0	3.4
CbpA-R2, 0 mM NaCl	13.4	75.8	2.4
CbpA-R2, 50 mM NaCl	9.5	106.0	2.7
CbpA-R2, 200 mM NaCl	15.0	67.3	1.9
CbpA-N, 0 mM NaCl	10.0*	10.0	0.4
CbpA-N, 50 mM NaCl	10.0*	10.0	0.4
CbpA-N, 200 mM NaCl	10.0*	>100.0	0.4

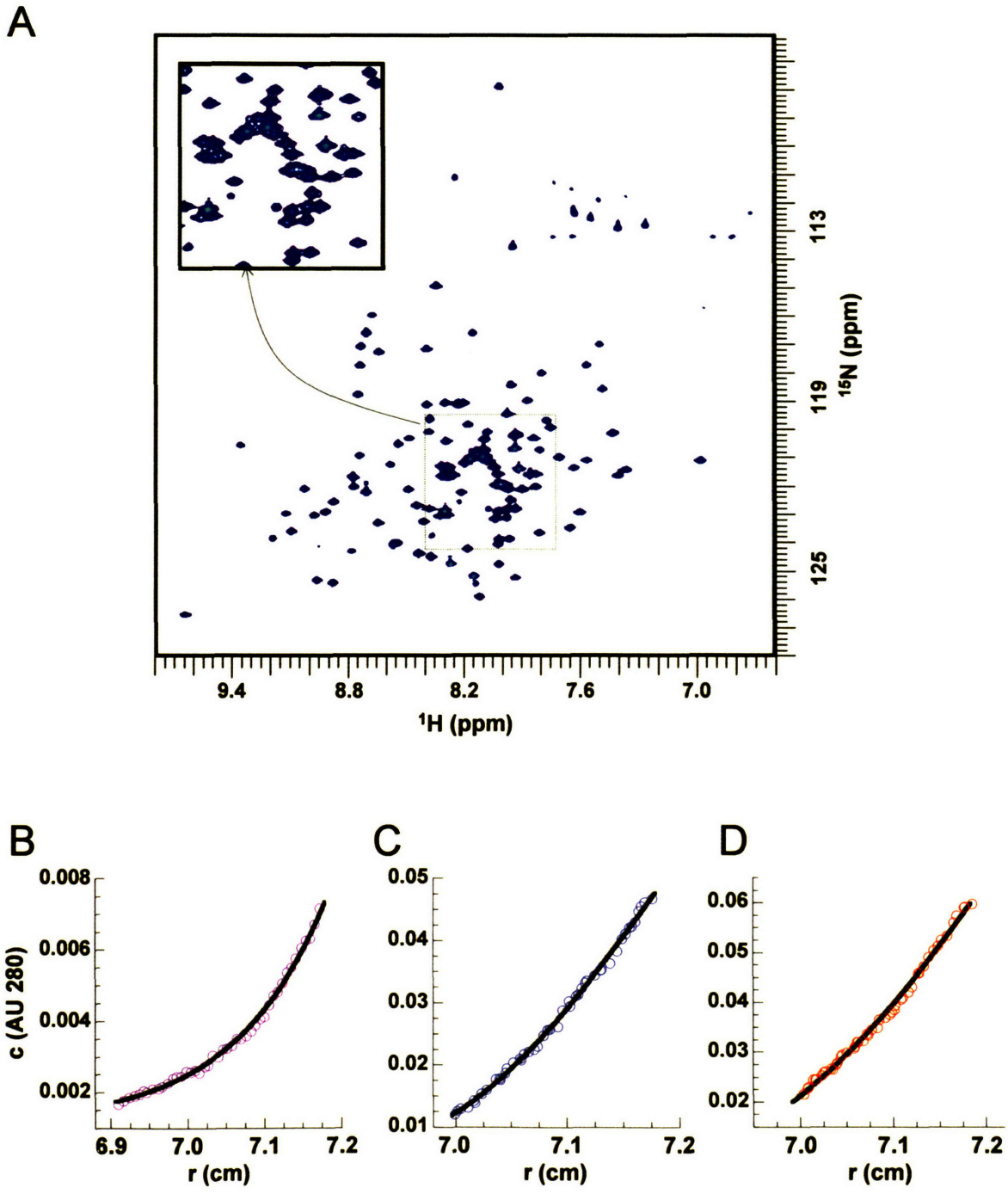
* The non-ideality term (2BM) was set to these values to analyze AUC data under these conditions.

Supplementary Table 2. Association (k_a) and dissociation (k_d) rate constants obtained from analysis of surface plasmon resonance data for CbpA fragments binding to immobilized sIgA or SC-D15. Results from triplicate measurement are given.

	Binding to sIgA		Binding to SC	
Construct	k_a (s^{-1} M^{-1})	k_d (s^{-1})	k_a (s^{-1} M^{-1})	k_d (s^{-1})
CbpA-R1	$1.93 \pm 0.10 \times 10^{6}$	$2.71 \pm 0.13 \times 10^{-4}$	$2.11 \pm 0.21 \times 10^{5}$	$5.65 \pm 0.53 \times 10^{-4}$
CbpA-R2	$9.13 \pm 0.59 \times 10^{5}$	$3.71 \pm 0.13 \times 10^{-4}$	$4.38 \pm 0.84 \times 10^{4}$	$4.99 \pm 0.14 \times 10^{-4}$
CbpA-R12	$1.38 \pm 0.03 \times 10^{6}$	$7.06 \pm 0.47 \times 10^{-5}$	$4.04 \pm 0.34 \times 10^{5}$	$2.50 \pm 0.38 \times 10^{-4}$
CbpA-NR12	$2.55 \pm 0.07 \times 10^{5}$	$5.57 \pm 1.17 \times 10^{-5}$	$5.01 \pm 0.20 \times 10^{4}$	$4.83 \pm 0.30 \times 10^{-5}$
CbpA-NR12-Y205G	$2.22 \pm 0.63 \times 10^{5}$	$2.41 \pm 0.0.94 \times 10^{-4}$	$9.69 \pm 3.59 \times 10^{4}$	$8.49 \pm 3.68 \times 10^{-5}$
CbpA-NR12-P206G	$1.27 \pm 0.01 \times 10^{6}$	$2.72 \pm 0.12 \times 10^{-4}$	$1.93 \pm 0.71 \times 10^{5}$	$4.06 \pm 0.57 \times 10^{-5}$
CbpA-NR12-Y358G	$3.29 \pm 0.02 \times 10^{5}$	$3.14 \pm 0.26 \times 10^{-4}$	$1.27 \pm 0.03 \times 10^{5}$	$2.51 \pm 0.21 \times 10^{-4}$
CbpA-NR12-P359G	$6.64 \pm 0.09 \times 10^{5}$	$3.30 \pm 0.13 \times 10^{-4}$	$6.66 \pm 3.62 \times 10^{4}$	$1.66 \pm 0.62 \times 10^{-4}$
CbpA-NR12-Y205G/Y358G	$4.17 \pm 1.26 \times 10^{5}$	$2.90 \pm 0.64 \times 10^{-2}$	$1.74 \pm 0.86 \times 10^{5}$	$1.83 \pm 0.43 \times 10^{-2}$
CbpA-NR12-P206G/P359G	$1.03 \pm 0.09 \times 10^{5}$	$1.91 \pm 0.27 \times 10^{-2}$	ND*	ND

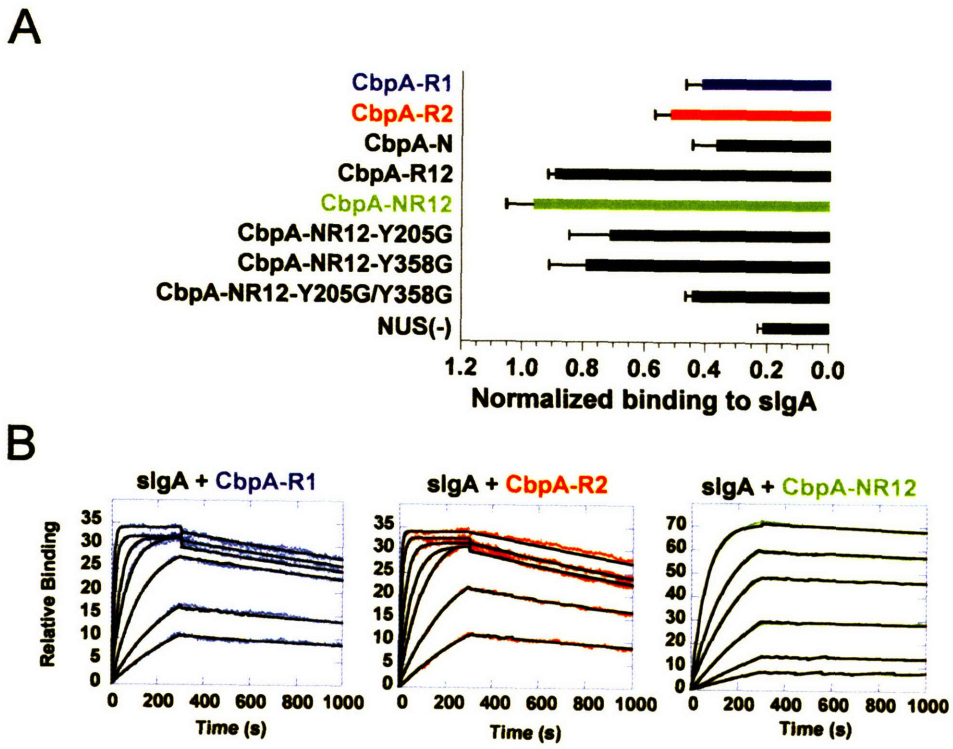
* ND, not determined.

Supplementary Figure 1. Characterization of CbpA domains by NMR (**A**) and AUC (**B-D**). (**A**) 900 MHz 2D ^{1}H-^{15}N TROSY spectrum of $^{2}H/^{13}C/^{15}N$-labeled CbpA-R2. Resonances of many residues overlap due to the repeating nature of the amino acid sequence (Fig. 1C). We overcame resonance overlap by using TROSY (Pervushin et al., 1997) to narrow ^{1}H and ^{15}N resonances and by using high magnetic field strengths. The crowded, central region is expanded in the upper left panel. Representative sedimentation equilibrium data obtained for (**B**) CbpA-N (0.4 mM), (**C**) CbpA-R1 (1.9 mM), and (**D**) CbpA-R2 (2.7 mM) dissolved in 20 mM sodium phosphate, pH 6.5, 50 mM NaCl. Experimental data points are shown as colored circles and the fit of Equation 1 to the raw data as a solid black line.



Supplementary Figure 1. Characterization of CbpA domains by NMR (**A**) and AUC (**B-D**).

Supplementary Figure 2. Molecular mechanism of CbpA/pIgR interactions. (**A**) Binding of wild-type and mutant CbpA fragments to sIgA based on results from ELISA. Results for CbpA-R1 (blue), CbpA-R2 (red) and CbpA-NR12 (green) are emphasized. NUS(-) corresponds to background binding in the absence of a CbpA construct. (**B**) Raw SPR data for CbpA-R1 (blue), CbpA-R2 (red) and CbpA-NR12 (green) binding to immobilized sIgA. The black lines show the fit of equations for a 1:1 binding model to the experimental data (colored points). The concentrations of CbpA constructs in the solutions that flowed over the sIgA or SC-D15 surfaces were as follows (based on amino acid analysis): CbpA-R1, 1462, 731, 292, 146, 73.1, 29.2, 14.6 nM; CbpA-R2, 1450, 725, 290, 145, 72.5, 29, 14.5 nM; CbpA-N, 1584, 792, 396, 198, 99, 49.5, 24.8 nM; CbpA-R12, 4.0, 2.0, 1.0, 0.4, 0.2, 0.1, 0.04 nM; CbpA-NR1, 665, 333, 133, 66.5, 33.3, 13.3, 6.7 nM; and CbpA-NR12, 207, 82.9, 41.4, 20.7, 8.3, and 4.1 nM. The concentrations of mutant CbpA-NR12 constructs used were as follows: 500, 250, 125, 62.5, 31.8, 12.5 and 6.3 nM.



Supplementary Figure 2. Molecular mechanism of CbpA/pIgR interactions.

CHAPTER 4

Expression and Crystallization Trials of pIgR, Fcα/μR, dFcα and CbpA

Introduction

As described in the previous chapters, pIgR plays a key role in mucosal immunity by transporting pIgs to mucosal surfaces. Although the recent crystal structures of the pIg-binding domain (D1) of pIgR (Hamburger et al., 2004) (Chapter 2) and the Fc fragment of IgA (Fcα) (Herr et al., 2003) have shed some light on IgA-based immunity, many questions remain to be answered. For example, the extracellular region of pIgR contains five tandem Ig-like domains (Mostov et al., 1984) and the detailed structural features of D2-5, and the overall arrangement of the five domains, remain unknown. The high-resolution structure of dIgA also has not been determined. Previous work has shown that dIgA is arranged in an end-to-end configuration with one J chain molecule covalently bridging two IgA molecules (Garcia-Pardo et al., 1981) via a C-terminal 18-residue extension in IgA, called the tailpiece (tp). The Fcα structure revealed the arrangement of the Cα2 and Cα3 domains (Herr et al., 2003), but did not contain tp. Further, nothing is known about the high-resolution structure of J chain, or about the precise arrangement of the two IgA (or Fcα-tp) molecules with J chain in dIgA (or dFcα). Ultimately, the cocrystal structure of the pIgR-dFcα complex will be necessary to fully understand the molecular details of the mechanism by which pIgR binds dIgA.

Another point of interest is to determine the molecular basis for the specificity of pIgR for pIgA and pIgM. FcαRI binds both monomeric and dimeric forms of IgA (Monteiro and Van De Winkel, 2003), and Fcα/μR can bind to monomeric IgA in addition to pIgA and pIgM (McDonald et al., 2002; Shibuya et al., 2000). Further structural studies of pIgR and Fcα/μR are needed to be able to compare the recognition properties of these receptors with those of FcαRI (Ding et al., 2003; Herr et al., 2003). Although detailed structural information about the human Fcα/μR is currently unavailable, sequence analysis predicts an approximately 50 kDa

extracellular region with an N-terminal Ig-like domain that shares 43% sequence identity with the human pIgR D1 (Shibuya et al., 2000).

Although pIgR provides immunologic protection at mucosal surfaces, it is also used by *Streptococcus pneumoniae* (pneumococcus) to invade human cells (Zhang et al., 2000). CbpA, the major pneumococcus adhesin (Rosenow et al., 1997), contains two regions, R1 and R2, which bind to D3-4 of human pIgR (Zhang et al., 2000)((Elm et al., 2004; Lu et al., 2003). The recent solution structure of R2 revealed a unique three-α-helical structure, with the three helices in an antiparallel orientation and a "tyrosine fork" involved in pIgR binding between the first two helices (Luo et al., 2004) (Chapter 3). However, to determine the molecular details of the interaction between pIgR and CbpA, the structures of full-length CbpA, D3-4 of pIgR, and the complex between the two proteins need to be elucidated.

In order to address the topics mentioned above, I set out to obtain high-resolution crystal structures of these proteins, alone and in complex with each other. To obtain diffraction quality crystals of pIgR, dFcα, Fcα/μR and CbpA, milligram quantities of these proteins are necessary. With large and flexible proteins, such as these, a "divide and conquer" approach is sometimes required either to get high expression levels, or to obtain diffracting crystals. Therefore, I made a series of expression constructs (Table 1) based on secondary structure predictions, sequence alignments with proteins of known structures, and available biochemical and mutagenesis data.

Another consideration for maximizing protein expression levels is what expression system to use (Table 1). The extracellular region of pIgR contains seven N-linked glycosylation sites (Hughes et al., 1999), which seem to be required for the efficient transport of pIgR during transcytosis or for release from the apical surface (Matsumoto et al., 2003). However, these carbohydrates are not necessary for specific binding to dIgA (Bakos et al., 1991). Therefore, I

tried expressing pIgR fragments in both bacteria and baculovirus-infected insect cells. Since the extracellular portion of Fcα/μR also has several N-linked oligosaccharide attachments sites (Shibuya et al., 2000), baculovirus-infected insect cells were utilized for protein expression. An *E. coli* expression system was used to express CbpA.

The most challenging of these proteins to express was dFcα because of the complexity of this protein. First, two molecules of Fcα-tp form a homodimer and then two Fcα-tp homodimers and one molecule of J chain together from dFcα. The human IgA tp contains a cysteine residue and an N-linked glycosylation site that are required for efficient polymer assembly (Atkin et al., 1996). Mature human J chain contains 137 amino acids including eight cysteine residues (Max and Korsmeyer, 1985) of which six are involved in intramolecular disulfide bonds and two intermolecular disulfide bonds to the tp penultimate cysteine in polymeric Igs (Bastian et al., 1992; Bastian et al., 1995; Frutiger et al., 1992). J chain also has an N-linked oligosaccharide attachment site, which appears to be important for the efficient assembly of dIgA (Krugmann et al., 1997). J chain Cys15 forms a disulfide bond with tailpiece Cys471 in the first Fcα homodimer, and J chain Cys69 forms a second disulfide bond with Cys471 in the tailpiece from the second Fcα homodimer. The other two tailpieces, one from each homodimer, are linked directly to each other by a disulfide bond (Bastian et al., 1995; Krugmann et al., 1997). Due to the presence of essential N-linked oligosaccharide attachment sites and the extensive disulfide bonding between J chain and Fcα-tp, both baculovirus-infected insect cells and stably transfected mammalian cells were used to coexpress the two proteins.

After optimizing the expression and purification conditions, I set up extensive crystallization screens of the proteins alone and in complex with each other (Table 2). I have

obtained small, disk-shaped crystals of the full-length pIgR ectodomain, although the current crystals do not diffract. Additional crystallization and optimization trials are ongoing.

Materials and Methods

Expression and purification of pIgR

The cDNA encoding the full-length human pIgR was kindly provided by Roland K. Strong at the Fred Hutchinson Cancer Research Center in Seattle, WA. Three constructs for bacterial expression were generated: D1, MBP-D1 and D3-4 (Table 1). Expression, refolding and purification of D3-4 (residues N214-V448) were performed as described for D1 (residues K1-V109) (Hamburger et al., 2004) (Chapter 2). Briefly, the protein was expressed in *E. coli* BL21(DE3) cells (Novagen) in inclusion bodies, which were solubilized in 8 M guanidine-HCl and 10 mM DTT. The protein was refolded by rapid dilution in refolding buffer (100 mM Tris pH 8.0, 400 mM L-arginine, 2 mM EDTA, 0.5 mM oxidized glutathione, and 5 mM reduced glutathione) (Garboczi et al., 1992). The refolded protein was purified by gel filtration chromatography. D1 was also subcloned into the pMAL-p2x expression vector (New England Biolabs) for periplasmic expression as a fusion with the maltose binding protein (MBP). *E. coli* BL21(DE3) cells (Novagen) were induced with 0.3 mM isopropyl-β-D-thiogalactoside (IPTG) at an $OD_{600} = 0.5$ for 3-5 hours. The cells were lysed by osmotic shock and the protein was purified on an amylose resin column. D1 was cleaved from MBP with Factor Xa protease (New England Biolabs) according to the manufacturer's recommendations.

Seven additional pIgR constructs were generated by PCR for protein expression using baculovirus-infected insect cells (Table 1). Six of the genes (D1, D1-3, D145, D1-5, SC577 and SC607) were first subcloned into the pET20b vector in frame with the C-terminal 6xHis tag, and then re-amplified by PCR to acquire the His tag, and subcloned into the pAcGP67A baculovirus

expression vector (BD Biosciences) in frame with the gp67 hydrophobic secretion signal. D3-4 was amplified by PCR and subcloned directly into a modified version of pAcGP67A, containing an N-terminal 6xHis tag and Factor Xa cleavage site. Expression and purification of the proteins were carried out as previously described for SC607, which was referred to as SC-D15 (Luo et al., 2004) (Chapter 3). Briefly, recombinant baculovirus was generated by cotransfection of a transfer vector with linearized viral DNA. Recombinant proteins were isolated from the supernatants of infected High five cells, which were concentrated and buffer exchanged into TBS (20 mM Tris pH 8.0, 150 mM NaCl) and purified by NiNTA affinity and size exclusion chromatography. Protein purity was analyzed by SDS-PAGE.

Expression and purification of Fcα/μR

The cDNA encoding the full-length human Fcα/μR was kindly provided by Kenneth J. McDonald at the Renal Unit, Western Infirmary in Glasgow, UK. Four Fcα/μR constructs were generated for protein expression using baculovirus infected insect cells (Table 1). The full-length ectodomain (ecto; residues L1-R434) and the N-terminal Ig domain (residues L1-A165) were subcloned into the pFastBacI vector (Life Technologies) with its native secretion signal sequence and a C-terminal 6xHis tag and Factor Xa cleavage site. Ecto and the Ig domains were also subcloned into the pAcGP67A expression vector in frame with the gp67 hydrophobic signal sequence and an N-terminal 6xHis tag and Factor Xa cleavage site. Protein expression and purification was carried out as described above.

Expression and purification of dFcα (Fcα-tp and J-chain) and dIgA

The cDNA encoding the human J chain was kindly provided by K. R. Chintalacharuvu at the University of California, Los Angeles. For coexpression in baculovirus infected insect cells, Fcα-tp and J chain were subcloned into the pFastBacDual expression vector (Life Technologies)

(Table 1). J chain, including its natural hydrophobic signal sequence (M1-D159, the numbering reflects inclusion of the secretion signal), was subcloned into the pFastBacDual multiple cloning site (mcs) 1. Fcα-tp (S240-Y473, encoding the IgA1 Cα2 and Cα3 domains and the 18 residue tailpiece) was first subcloned into a modified version of pAcGP67A (containing an N-terminal 6xHis tag and Factor Xa cleavage site) in frame with the gp67 hydrophobic secretion signal. The gene was then re-amplified by PCR to acquire the signal sequence, 6xHis tag and protease cleavage site and subcloned into pFastBacDual mcs2. To increase the amount of J chain expressed, and therefore to increase the levels of secreted dimeric Fcα, J chain (Q23-D159) was also subcloned into the pAcGP67B expression vector in frame with the gp67 secretion signal for co-infection with the pFastBacDual(Fcα-tp/J chain) construct. Protein expression and purification was performed as described above.

For expression in mammalian cells, Fcα-tp (S240-Y473) containing an N-terminal 6x-His tag and Factor Xa cleavage site was subcloned into the in pBJ5-GS-MCS3 expression vector in frame with the rat IgG2a hydrophobic secretion signal (Martin and Bjorkman, 1999). J chain (residues M1-D159) was subcloned into two mammalian expression vectors, pBJ5-GS-MCS and pcDNA3.1(-) (Invitrogen). Chinese hamster ovary (CHO) cells were cotransfected with pBJ5-GS-MCS3(Fcα-tp) and pBJ5-GS-MCS(J chain) and stable colonies were selected and amplified using methionine sulfoximine (MSX) as described previously (Martin and Bjorkman, 1999). To increase the amount of secreted dFcα, the stably transfected CHO cells were retransfected with J chain in the pcDNA3.1(-) vector. Selection and amplification of stable cell lines using 100 μM MSX and 750 μg/mL Geneticin (Gibco) were carried out. A CHO cell line expressing Fcα-tp alone was also generated as described above, using 100 μM MSX selection.

Human dIgA was isolated from the sera of a patient with a pIgA-producing myeloma (Vaerman et al., 1995) and purified as described previously (Hamburger et al., 2004) (Chapter 2).

Expression and purification of CbpA, R1 and R2

The *S. pneumoniae* TIGR4 strain CbpA gene was kindly provided by Elaine Tuomanen at St. Jude Children's Research Hospital in Memphis, TN. CbpA (E39-A444), R1 (T174-A280) and R2 (L327-A433) were amplified by PCR and subcloned into pET19b in frame with an N-terminal 10xHis tag and Enterokinase cleavage site (Table 1). The proteins were expressed in *E. coli* BL21(DE3) cells (Novagen) by induction with 0.4 mM IPTG at an OD_{600} = 0.5 for 5 hours at 37 °C. The cells were lysed by sonication and the soluble proteins were purified by NiNTA affinity followed by size exclusion chromatography on a Superdex 75 16/20 column (Amersham Pharmacia).

Binding studies

Surface plasmon resonance (SPR) binding assays were performed using a BIAcore 2000 instrument (Pharmacia Biosensor, Uppsala, Sweden). Using this system, binding between a molecule coupled to a biosensor chip and a second molecule injected over the chip results in changes in the SPR signal that are monitored in real time as resonance units (RUs) (Malmqvist, 1993). To test whether the pIgR constructs were folded properly, dIgA was covalently coupled at three different densities (412, 763, and 1426 RUs) to a reagent-grade CM5 sensor chip (Biacore) using primary amine coupling, and the various pIgR proteins (the analytes) were each injected over the chip. Similarly, to test whether CbpA, R1 and R2 were folded properly, each of the three proteins was coupled to a CM5 sensor chip and pIgR SC607 was injected over the chip. To test the activity of recombinantly-expressed dFcα from CHO cells, pIgR SC607 was coupled

to the biosensor chip at three different densities (2615, 5599, and 6567 RUs) and dFcα was injected over the chip. Fcα-tp was also coupled to a chip and FcαRI was injected over the chip. In each case one flow cell was mock coupled with buffer for background subtraction. To obtain equilibrium dissociation constants (K_Ds) for D1 and D1-3, the sensorgrams were processed and analyzed with the Scrubber software package (BioLogic Software, Australia). K_Ds were derived by nonlinear regression analysis of plots of R_{eq} (the equilibrium binding response) versus the log of the concentration of the injected analyte, and the binding data were fit to a single-site binding model.

Protein crystallization

Purified proteins were concentrated to ~5-110 mg/mL in an Amicon Ultra centrifugal filter device with 10 kDa molecular weight cut-off for crystallization trials. Protein concentrations were determined spectrophotometrically at 280 nm using extinction coefficients calculated from the amino acid sequence of each protein using the ProtParam tool on the ExPAsy Proteomics Server (Gill and von Hippel, 1989). Crystallization trials with grid screens and commercially available sparse matrix screens were set up using both the hanging drop and sitting drop vapor diffusion methods (Table 2).

Results and Discusssion

Expression, purification and characterization of pIgR

To produce sufficient amounts of pIgR for crystallization, three bacterial constructs (D1, MBP-D1 and D3-4) were generated and tested for expression levels (Table 1). MBP-D1 was expressed in the periplasm of *E.coli*, and purified on amylose resin. Cleavage of the fusion protein with Factor Xa protease resulted in the release of insoluble D1 and not pursued further. Expression, refolding, purification and characterization of D1 was described in Chapter 2

(Hamburger et al., 2004), and the same methods were used to generate D3-4. Refolded D3-4 migrates in the position expected for the folded monomer on a gel filtration column (data not shown).

Seven baculovirus expression constructs (D1, D1-3, D145, D3-4, D1-5, SC577, and SC607) were also produced to test expression levels in baculovirus infected insect cells (Table 1). D1-5 expressed at less than 1 mg/L of insect cell supernatant and was not further pursued. Slightly longer C-terminal truncations of the full-length extracellular region of pIgR, SC577 and SC607, and the other constructs (D1, D1-3, D145 and D3-4) all expressed and were secreted between 6 and 35 mg/L (Table 1). The recombinant proteins were harvested from baculovirus-infected insect cell supernatants and purified by NiNTA and gel filtration chromatography to over 95% purity (Figure 1). Each protein migrates in the position expected for the monomer on a gel filtration column (data not shown).

To determine if the pIgR proteins were folded properly, a surface plasmon resonance (SPR) binding assay using human dIgA was utilized. dIgA was isolated from the serum of a patient with a pIgA-producing myeloma (Vaerman et al., 1995) and dIgA was purified as described (Hamburger et al., 2004) (Chapter 2). dIgA was coupled to a biosensor chip and the various pIgR proteins were injected. Both D1 proteins (refolded nonglycosylated, and insect cell-expressed glycosylated) bind dIgA with an equilibrium dissocation constant (K_D) of ~300 nM (Hamburger et al., 2004) (Chapter 2). In the same binding assay, D1-3 bound dIgA with a K_D of ~1 μM (data not shown). Since the N-terminal D1 domain of pIgR is necessary for binding to dIgA (Frutiger et al., 1986), D3-4 was not expected to bind dIgA and therefore not tested. However, in collaboration with Richard Kriwacki's laboratory at St. Jude Children's Research Hospital, the ability of D3-4 to bind to CbpA was confirmed (data not shown). The

other constructs (D145, SC577 and SC607) all contain Cys467 in D5, which forms a disulfide bridge with Cys311 in Cα2 (Fallgreen-Gebauer et al., 1993). Due to the covalent nature of this interaction we were unable to obtain off-rates or equilibrium dissociation constants for these proteins. We were, however, able to show qualitatively that the binding interaction took place and approximately 33% of the injected proteins remained covalently bound to the biosensor chip (data not shown). The low pH and high salt conditions used to regenerate the chip after injection of D1 and D1-3 did not result in obtaining the original baseline after injection of proteins containing D5 because of the covalent interaction with dIgA.

Expression and purification of Fcα/μR

To try to obtain large amounts of Fcα/μR sufficient for crystallization trials, baculovirus-infected insect cells were used to express four different constructs (Table 1). Two of the constructs contained the putative full-length ectodomain of the protein (residues L1-R434 of the mature protein): one with the natural hydrophobic secretion signal and a C-terminal 6xHis tag, while the other contained the baculovirus gp67 secretion signal and an N-terminal 6xHis tag. The other two constructs contained the N-terminal region of the ectodomain, including the Ig-like domain (residues L1-A165 of the mature protein). Similar to the full-length constructs, one of them contained the natural secretion signal with a C-terminal 6xHis tag, while the other had the gp67 secretion signal and an N-terminal 6xHis tag. Analyses of the baculovirus infected insect cell supernatants by SDS-PGE revealed that expression levels were very low (less than 1 mg/L), regardless of the secretion signal used or the position of the affinity tag. Nonetheless, protein was purified from the supernatants of baculovirus infected insect cells by NiNTA affinity and gel filtration chromatography. The gel filtration profile indicated that the majority of the protein was aggregated and eluted in the void volume. Therefore, due to the low expression levels and the

apparent misfolding of the protein, further characterization and crystallization trials were not attempted.

Expression, purification and characterization of Fcα-tp and J-chain

In order to produce dFcα capable of binding to pIgR, expression levels in both baculovirus infected insect cells and mammalian CHO cells were tested. For expression in insect cells, J chain and Fcα-tp (encoding IgA1 Cα2 and Cα3 domains and the 18 residue tailpiece) were coexpressed using a dicistronic expression vector, pFastBacDual (Life Technologies). The protein was purified by NiNTA and gel filtration chromatography, which revealed the presence of both monomeric Fcα as well as polymeric Fcα (pFcα) in the preparation (Figure 2A). Others have observed that polymerization of another polymeric immunoglobulin, pIgM, depends on the amount of J chain present, where an increase in J chain production results in an increase in J chain-containing IgM secretion (Niles et al., 1995). Therefore, in order to try to increase the ratio of secreted dFcα to monomeric Fcα, the J chain gene was subcloned into another expression vector, pAcGP67B, for coinfection with the dicistronic vector. However, coinfection with the two recombinant baculoviruses resulted in the drastic decrease in expression levels of both proteins (P.M. Snow and A.E.H., unpublished observations). Later protein preparations, using a new viral stock of recombinant baculovirus containing the dicistronic vector showed an increase in the amount of pFcα and dFcα relative to the amount of monomeric Fcα (Figure 2B).

For expression in mammalian cells, CHO cells were cotransfected with Fcα-tp and J chain in the expression vectors pBJ5-GS-MCS3 and pBJ5-GS-MCS, respectively, and a stable cell line expressing both proteins was generated using MSX selection. As was the case for the insect cells, both monomeric Fcα and pFcα were purified from the supernatants of the CHO cells (Figure 2C). The pFcα peak was broad with two shoulders. The peak that eluted first was

assumed to be higher ordered oligomers, and the peak that eluted second was assumed to be dFcα. To increase the amount of secreted dFcα, the stably-transfected CHO cells were retransfected with the J chain gene in another vector, pcDNA3.1(-) (Invitrogen), which uses a different selection marker, G418 or Geneticin. Selection and amplification of a stable cell line was repeated using both MSX and Geneticin (Gibco). Although the amount of secreted pFcα increased proportionally over the amount of monomeric Fcα in the new cell line, both complexes were nonetheless present (Figure 2D). Biosensor analyses of the purified proteins revealed that dFcα bound to pIgR SC607 coupled to the chip, but no binding was observed when monomeric Fcα was injected (Figure 3).

Expression, purification and characterization of CbpA, R1 and R2

The pIgR-binding protein on the surface of *Streptococcus pneumoniae*, CbpA, and two smaller pIgR-binding CbpA fragments, R1 and R2, were expressed in *E. coli* and purified by NiNTA and gel filtration chromatography. An SPR binding assay with SC577 coupled to the biosensor chip and the CbpA proteins injected confirmed that the recombinantly expressed CbpA, R1 and R2 were folded properly and retained their ability to bind to pIgR (data not shown).

Protein crystallization trials

Proteins estimated to be over 95% pure by SDS-PAGE were concentrated to 5-110 mg/mL. Protein concentrations were determined spectrophotometrically at 280 nm using extinction coefficients calculated from the amino acid sequence of each protein. Crystallization trials with pIgR constructs (refolded D1 and D3-4 as well as insect cell-expressed D1, D1-3, D3-4, D145, SC577 and SC607); dIgA fragments (CHO cell-expressed Fcα-tp, dFcα and commercially available IgA (Sigma)), and CbpA constructs (bacterially-expressed CbpA, R1 and R2) were set up using both grid screens and commercially available sparse matrix screens (Table 2).

Although an extensive number of conditions were tested, diffracting crystals have not yet been obtained. SC577 has yielded small crystals but attempts to improve these disk-shaped crystals (Figure 4A) have not been successful. SDS-PAGE analyses of the highly reproducible disks suggest they are protein crystals (Figure 4B), however X-ray diffraction was not observed.

In addition to setting up crystallization trials with the individual proteins, we also tried to cocrystallize several protein complexes (Table 2). Soluble pIgR and dIgA are thought to interact with a 1:1 stoichiometry (Morrison et al., 1997; Rindisbacher et al., 1995). Therefore, D1 or SC577 were each mixed with dFcα in a 1:1 molar ratio prior to crystallization. ITC results presented in Chapter 3 show that the stoichiometry of the CbpA:pIgR interaction is 1:1, while the R1 and/or R2:pIgR ratio is 2:1 (Luo et al., 2004). Consequently, D3-4 or SC607 were each mixed with CbpA in a 1:1 molar ratio, and both refolded and insect cell-expressed D3-4 were each mixed with R1 and R2 with a 1:1:1 molar ratio prior to cocrystallization.

Summary

In the absence of high-resolution structural information, previous electron microscopy studies, mutagenesis results and binding data have been used to propose models for the relative arrangements of pIgR and dIgA (including J chain) in SIgA (reviewed in (Norderhaug et al., 1999b; Rojas and Apodaca, 2002)). Given the complexity of the interactions between, and the flexibility within, the molecules in SIgA, many potential structural arrangements are possible. By combining data from available literature and results from the studies presented in Chapters 2 and 4, two of the more simple but structurally accurate models are described below (Figure 5).

In monomeric IgA, tp appears to fold up against the Cα3 domain (Boehm et al., 1999), however its position in dIgA is unknown. Since dIgA is arranged in an end-to-end orientation with disulfide bonds between the tp penultimate cysteine residues either directly to each other or

to J chain, it is possible that at least the tps that are covalently bound to each other extend away from the Cα3 domains, toward the partner molecule (Figure 5). Alternatively, if the tps remain folded against the Cα3 domains, the two IgA molecules would be overlapping, with J chain between two of the Cα3 domains.

Inhibition studies with Fab fragments specific for IgA have shown that the Cα3, and possibly Cα2, domains of dIgA are involved in noncovalent binding to pIgR (Geneste et al., 1986). Peptide binding studies have further implicated two loops in the Cα3 domain in the binding interaction (Hexham et al., 1999; White and Capra, 2002) (Chapter 1, Figure 5A). Additionally, the fact that J chain is a required component in pIgs for binding to pIgR suggests that J chain may also form part of the binding interface with D1. Therefore it is likely that D1 of pIgR interacts through noncovalent interactions with one of the two dIgA Cα3 domains whose tp is covalently bound to J chain (Figure 5).

The SPR based binding experiments described in Chapters 2 and 4 showed that D1 binds to dIgA with an equilibrium dissociation constant of 300 nM (Hamburger et al., 2004) (Chapter 2), D1-3 binds with an equilibrium dissociation constant of 1 μM, and D145 retains its ability to bind to dIgA through both noncovalent and covalent interactions (Chapter 4). Together, these results suggest that D2-3 do not contribute to the binding affinity of human pIgR for dIgA. Similarly, D2-3 in rabbit (Deitcher and Mostov, 1986; Frutiger et al., 1986) and murine (Crottet and Corthesy, 1999) pIgR are not necessary for high affinity noncovalent interaction with dIgA, and chicken pIgR has no ortholog to the mammalian D2 (Wieland et al., 2004). Therefore, in these species pIgR D2-3 are not expected to interact with dIgA (Figure 5A). However, contrary to our results, others have observed that D2-3 are required for human pIgR expressed on the surface of cultured cells to bind to and transcytose dIgA (Norderhaug et al., 1999a). It is difficult

to reconcile the difference between the two findings. Although the surface expressed D145 retained the ability to bind to pIgM, it was not recognized by an antibody specific for D5 (Norderhaug et al., 1999a), indicating the protein might have been misfolded in a way that artificially disrupted binding to dIgA, but not pIgM. However, bovine pIgR also utilizes D2-3, although to a lesser extent than D1, for binding dIgA (Beale, 1988), suggesting that in certain cases pIgR D2-3 might make additional contacts with dIgA, either through Fcα, J chain, or both (Figure 5B).

In human SIgA, pIgR D5 is covalently linked to the Cα2 domain of the second IgA molecule in dIgA (Fallgreen-Gebauer et al., 1993). Since the Cα2 and Cα3 domains are arranged symmetrically in monomeric Fcα (Herr et al., 2003) human pIgR D5 may bind to either monomer in the second Fcα homodimer with similar preference (Figure 5A). However, because of J chain, dIgA is asymmetric so in cases where D2-3 enhance binding between the two molecules, those domains may interact with dIgA such that it would orient D5 to be sterically closer to one of the two Fcα homodimers (Figure 5B).

The models presented here take into account currently available data and represent two likely arrangements of dIgA and pIgR. However, to fully understand the detailed intermolecular interactions in SIgA, a cocrystal structure of pIgR in complex with dFcα (or dIgA) is necessary.

Acknowledgements

I would like to acknowledge Dr. Peter Snow, Inderjit Nangiana and Cynthia Jones of the Caltech Protein Expression Facility for recombinant baculovirus-infected insect cells for protein production; Darcy Ballister, Jason Martineau, Evelyn Cheung and Kevin McKenzie for assistance with crystallization trials; Evelyn Cheung for establishing a CHO cell line expressing

monomeric Fcα-tp; Beth Huey Tubman for help with harvesting CHO cell supernatants and members of the Bjorkman laboratory for helpful discussions.

References

Atkin, J. D., Pleass, R. J., Owens, R. J., and Woof, J. M. (1996). Mutagenesis of the human IgA1 heavy chain tailpiece that prevents dimer assembly. J Immunol *157*, 156-159.

Bakos, M. A., Kurosky, A., and Goldblum, R. M. (1991). Characterization of a critical binding site for human polymeric Ig on secretory component. J Immunol *147*, 3419-3426.

Bastian, A., Kratzin, H., Eckart, K., and Hilschmann, N. (1992). Intra- and interchain disulfide bridges of the human J chain in secretory immunoglobulin A. Biol Chem Hoppe Seyler *373*, 1255-1263.

Bastian, A., Kratzin, H., Fallgren-Gebauer, E., Eckart, K., and Hilschmann, N. (1995). Intra- and inter-chain disulfide bridges of J chain in human S-IgA. Adv Exp Med Biol *371A*, 581-583.

Beale, D. (1988). Cyanogen bromide cleavage of bovine secretory component and its tryptic fragments. Int J Biochem *20*, 873-879.

Boehm, M. K., Woof, J. M., Kerr, M. A., and Perkins, S. J. (1999). The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. J Mol Biol *286*, 1421-1447.

Crottet, P., and Corthesy, B. (1999). Mapping the interaction between murine IgA and murine secretory component carrying epitope substitutions reveals a role of domains II and III in covalent binding to IgA. J Biol Chem *274*, 31456-31462.

Deitcher, D. L., and Mostov, K. E. (1986). Alternate splicing of rabbit polymeric immunoglobulin receptor. Mol Cell Biol *6*, 2712-2715.

Ding, Y., Xu, G., Yang, M., Yao, M., Gao, G. F., Wang, L., Zhang, W., and Rao, Z. (2003). Crystal structure of the ectodomain of human FcalphaRI. J Biol Chem *278*, 27966-27970.

Elm, C., Braathen, R., Bergmann, S., Frank, R., Vaerman, J. P., Kaetzel, C. S., Chhatwal, G. S., Johansen, F. E., and Hammerschmidt, S. (2004). Ectodomains 3 and 4 of human polymeric Immunoglobulin receptor (hpIgR) mediate invasion of Streptococcus pneumoniae into the epithelium. J Biol Chem *279*, 6296-6304.

Fallgreen-Gebauer, E., Gebauer, W., Bastian, A., Kratzin, H. D., Eiffert, H., Zimmermann, B., Karas, M., and Hilschmann, N. (1993). The covalent linkage of secretory component to IgA. Structure of sIgA. Biol Chem Hoppe Seyler *374*, 1023-1028.

Frutiger, S., Hughes, G. J., Hanly, W. C., Kingzette, M., and Jaton, J. C. (1986). The amino-terminal domain of rabbit secretory component is responsible for noncovalent binding to immunoglobulin A dimers. J Biol Chem *261*, 16673-16681.

Frutiger, S., Hughes, G. J., Paquet, N., Luthy, R., and Jaton, J. C. (1992). Disulfide bond assignment in human J chain and its covalent pairing with immunoglobulin M. Biochemistry *31*, 12643-12647.

Garboczi, D. N., Hung, D. T., and Wiley, D. C. (1992). HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. Proc Natl Acad Sci U S A *89*, 3429-3433.

Garcia-Pardo, A., Lamm, M. E., Plaut, A. G., and Frangione, B. (1981). J chain is covalently bound to both monomer subunits in human secretory IgA. J Biol Chem *256*, 11734-11738.

Geneste, C., Iscaki, S., Mangalo, R., and Pillot, J. (1986). Both Fc alpha domains of human IgA are involved in in vitro interaction between secretory component and dimeric IgA. Immunol Lett *13*, 221-226.

Gill, S. C., and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem *182*, 319-326.

Hamburger, A. E., West, A. P., Jr., and Bjorkman, P. J. (2004). Crystal structure of a polymeric immunoglobulin binding fragment of the human polymeric immunoglobulin receptor. Structure (Camb) *12*, 1925-1935.

Herr, A. B., Ballister, E. R., and Bjorkman, P. J. (2003). Insights into IgA-mediated immune responses from the crystal structures of human FcalphaRI and its complex with IgA1-Fc. Nature *423*, 614-620.

Hexham, J. M., White, K. D., Carayannopoulos, L. N., Mandecki, W., Brisette, R., Yang, Y. S., and Capra, J. D. (1999). A human immunoglobulin (Ig)A calpha3 domain motif directs polymeric Ig receptor-mediated secretion. J Exp Med *189*, 747-752.

Hughes, G. J., Reason, A. J., Savoy, L., Jaton, J., and Frutiger-Hughes, S. (1999). Carbohydrate moieties in human secretory component. Biochim Biophys Acta *1434*, 86-93.

Krugmann, S., Pleass, R. J., Atkin, J. D., and Woof, J. M. (1997). Structural requirements for assembly of dimeric IgA probed by site-directed mutagenesis of J chain and a cysteine residue of the alpha-chain CH2 domain. J Immunol *159*, 244-249.

Lu, L., Lamm, M. E., Li, H., Corthesy, B., and Zhang, J. R. (2003). The human polymeric immunoglobulin receptor binds to Streptococcus pneumoniae via domains 3 and 4. J Biol Chem *278*, 48178-48187.

Luo, R., Mann, B., Lewis, W. S., Rowe, A., Heath, R., Stewart, M. L., Hamburger, A. E., Sivakolundu, S., Lacy, E. R., Bjorkman, P. J., *et al.* (2004). Solution structure of choline binding protein A, the major adhesin of Streptococcus pneumoniae. Embo J.

Malmqvist, M. (1993). Biospecific interaction analysis using biosensor technology. Nature *361*, 186-187.

Martin, W. L., and Bjorkman, P. J. (1999). Characterization of the 2:1 complex between the class I MHC-related Fc receptor and its Fc ligand in solution. Biochemistry *38*, 12639-12647.

Matsumoto, N., Asano, M., Ogura, Y., Takenouchi-Ohkubo, N., Chihaya, H., Chung-Hsing, W., Ishikawa, K., Zhu, L., and Moro, I. (2003). Release of non-glycosylated polymeric immunoglobulin receptor protein. Scand J Immunol *58*, 471-476.

Max, E. E., and Korsmeyer, S. J. (1985). Human J chain gene. Structure and expression in B lymphoid cells. J Exp Med *161*, 832-849.

McDonald, K. J., Cameron, A. J., Allen, J. M., and Jardine, A. G. (2002). Expression of Fc alpha/mu receptor by human mesangial cells: a candidate receptor for immune complex deposition in IgA nephropathy. Biochem Biophys Res Commun *290*, 438-442.

Monteiro, R. C., and Van De Winkel, J. G. (2003). IgA Fc receptors. Annu Rev Immunol *21*, 177-204.

Morrison, M., Harris, K. S., and Roth, M. B. (1997). smg mutants affect the expression of alternatively spliced SR protein mRNAs in Caenorhabditis elegans. Proc Natl Acad Sci U S A *94*, 9782-9785.

Mostov, K. E., Friedlander, M., and Blobel, G. (1984). The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. Nature *308*, 37-43.

Niles, M. J., Matsuuchi, L., and Koshland, M. E. (1995). Polymer IgM assembly and secretion in lymphoid and nonlymphoid cell lines: evidence that J chain is required for pentamer IgM synthesis. Proc Natl Acad Sci U S A *92*, 2884-2888.

Norderhaug, I. N., Johansen, F. E., Krajci, P., and Brandtzaeg, P. (1999a). Domain deletions in the human polymeric Ig receptor disclose differences between its dimeric IgA and pentameric IgM interaction. Eur J Immunol *29*, 3401-3409.

Norderhaug, I. N., Johansen, F. E., Schjerven, H., and Brandtzaeg, P. (1999b). Regulation of the formation and external transport of secretory immunoglobulins. Crit Rev Immunol *19*, 481-508.

Rindisbacher, L., Cottet, S., Wittek, R., Kraehenbuhl, J. P., and Corthesy, B. (1995). Production of human secretory component with dimeric IgA binding capacity using viral expression systems. J Biol Chem *270*, 14220-14228.

Rojas, R., and Apodaca, G. (2002). Immunoglobulin transport across polarized epithelial cells. Nat Rev Mol Cell Biol *3*, 944-955.

Rosenow, C., Ryan, P., Weiser, J. N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H. R. (1997). Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. Mol Microbiol *25*, 819-829.

Shibuya, A., Sakamoto, N., Shimizu, Y., Shibuya, K., Osawa, M., Hiroyama, T., Eyre, H. J., Sutherland, G. R., Endo, Y., Fujita, T., *et al.* (2000). Fc alpha/mu receptor mediates endocytosis of IgM-coated microbes. Nat Immunol *1*, 441-446.

Vaerman, J. P., Langendries, A., and Vander Maelen, C. (1995). Homogenous IgA monomers, dimers, trimers and tetramers from the same IgA myeloma serum. Immunol Invest *24*, 631-641.

White, K. D., and Capra, J. D. (2002). Targeting mucosal sites by polymeric immunoglobulin receptor-directed peptides. J Exp Med *196*, 551-555.

Wieland, W. H., Orzaez, D., Lammers, A., Parmentier, H. K., Verstegen, M. W., and Schots, A. (2004). A functional polymeric immunoglobulin receptor in chicken (Gallus gallus) indicates ancient role of secretory IgA in mucosal immunity. Biochem J *380*, 669-676.

Zhang, J. R., Mostov, K. E., Lamm, M. E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. Cell *102*, 827-837.

Table 1. pIgR, Fcα/μR, dFcα and CbpA protein expression constructs. The table shows the name of the constructs; the system used for expression (bacteria, baculovirus infected insect cells or mammalian CHO cells); the expression vectors used, the name of the signal sequence used to direct the protein for secretion; the type and location (C- or N-terminal) of the affinity tag; the protease cleave site present to remove the affinity tag; the residues in the mature proteins; the approximate molecular weights of the proteins not taking into account potential glycosylation; the number of N-linked carbohydrate sites (listed as zero for all constructs expressed in bacteria); and the expression level of each protein in mg/L.

	Vector	Signal Sequence	Affinity Tag	Cleavage Site	Residues (mature protein)	MW (kDa)	# Glycos	Expression (~mg/L)
BACTERIA								
pIgR D1	pET20b	no	C-6xHis	no	K1-V109	13	0	refold; 10 mg/L
pIgR D3-4	pET20b	no	C-6xHis	no	N214-V448	27	0	refold; 25 mg/L
pIgR MBP-D1	pMAL-p2x	malE	N-MBP	Factor Xa	K1-V109	42-13	0	periplasm
CbpA full	pET19b	no	N-10xHis	Enterokinase	E39-A444	49	0	> 100 mg/L
CbpA R1	pET19b	no	N-10xHis	Enterokinase	T174-A280	16	0	> 65 mg/L
CbpA R2	pET19b	no	N-10xHis	Enterokinase	L327-A433	16	0	> 35 mg/L
BACULOVIRUS								
pIgR D1	pAcGP67A	gp67	C-6xHis	no	K1-V109	14	2	> 25 mg/L
pIgR D1-5	pAcGP67A	gp67	C-6xHis	no	K1-E546	62	7	< 1 mg/L
pIgR D1-3	pAcGP67A	gp67	C-6xHis	no	K1-E331	38	4	> 35 mg/L
pIgR SC577	pAcGP67A	gp67	C-6xHis	no	K1-K559	63	7	> 8 mg/L
pIgR SC607	pAcGP67A	gp67	C-6xHis	no	K1-E589	67	7	> 13 mg/L
pIgR D145	pAcGP67A	gp67	C-6xHis	no	K1-V109/S332-E589	42	5	> 15 mg/L
pIgR D3-4	pAcGP67A	gp67	N-6xHis	Factor Xa	N214-V448	29	1	> 6 mg/L
Fcα-tp/J chain	pFastBacDual	gp67/native	N-6xHis/no	Factor Xa/no	S240-Y473/Q23-D159	28/16	2/1	> 6 mg/L
Fcα-tp	pAcGP67A	gp67	N-6xHis	Factor Xa	S240-Y473	28	2	
J chain	pAcgp67b	gp67	no	no	Q23-D159	16	1	
Fcα/μR Ecto	pFastBacI	native	C-6xHis	FactorXa	L1-R434	48	2	< 1 mg/L
Fcα/μR Ecto	pAcGP67A	gp67	N-6xHis	FactorXa	L1-R434	49	2	< 1 mg/L
Fcα/μR Ig domain	pFastBacI	native	C-6xHis	FactorXa	L1-A165	20	1	< 1 mg/L
Fcα/μR Ig domain	pAcGP67A	gp67	N-6xHis	FactorXa	L1-A165	21	1	< 1 mg/L
CHO CELLS								
Fcα-tp	pBJ5-GS-MCS3	rIgG2a	N-6xHis	FactorXa	S240-Y473	28	2	> 6 mg/L
J chain	pBJ5-GS-MCS	native	no	no	Q23-D159	16	1	
J chain	pcDNA3.1(-)	native	no	no	Q23-D159	16	1	

Table 1. pIgR, Fcα/μR, dFcα and CbpA protein expression constructs

Table 2. Crystallization trials for pIgR, dFcα, CbpA and complexes. The table shows the different crystallization condition tested for each protein and protein complex. The names of the commercially available sparse matrix screens are listed, and the numbers in parentheses indicate the number of conditions in that kit. Crystal Screen I, Crystal Screen II, Peg/Ion, Salt Rx and the Index screens were purchased from Hampton Research; Cryo I, Cryo II, Wizard I, and Wizard II were purchased from Emerald BioStructures; and the Cation, Anion, and the PHClear Suite screens were purchased from Nextal.

	Crystal Screen I (50)	Crystal Screen II (48)	Peg/ Ion (48)	Wizard I (48)	Wizard II (48)	Cryo I (48)	Cryo II (48)	Salt Rx (96)	Index (96)	PH clear (96)	Cation (96)	Anion (96)	Various (additives, grids, etc.)
Refolded D1	✓	✓			✓	✓	✓						✓✓✓
D1	✓	✓	✓	✓	✓	✓	✓	✓					✓✓✓
D1-D3	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
D145	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	
Refolded D34	✓	✓		✓	✓				✓	✓	✓	✓	
D34			1-24					✓		✓	✓	✓	
SC577	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓✓✓
SC607	✓	✓							✓				
Fcα-tp	✓	✓	✓	✓	✓			✓	✓				✓
dFcα	✓	✓		✓	✓								
Sigma IgA	✓	✓		✓	✓								
CbpA	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
R1	✓	✓		✓	✓				✓	✓	✓	✓	✓
R2	✓	✓		✓	✓				✓	✓	✓	✓	✓
D1/dFcα	✓	✓		✓	✓								
SC577/dFcα	✓	✓		✓	✓								
SC607/CbpA	✓	✓		✓	✓				✓				
D34/CbpA	✓	✓		✓	✓								
Ref D34/R1/R2	✓	✓		✓	✓				✓	✓	✓	✓	
D34/R1/R2			1-24					✓		✓	✓	✓	

Table 2. Crystallization trials for pIgR, dFcα, CbpA and complexes

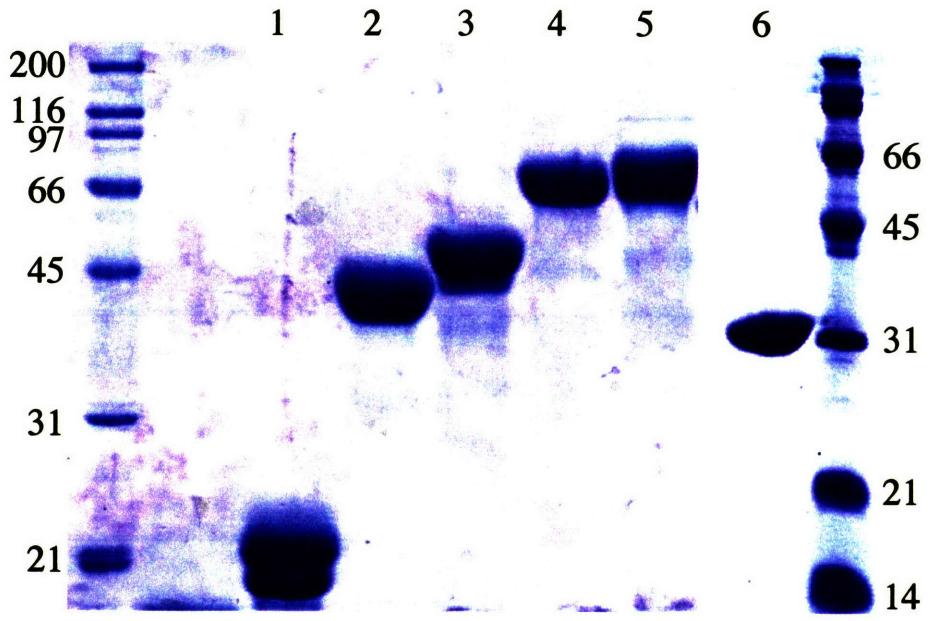


Figure 1. SDS-PAGE analysis of purified pIgR proteins. pIgR D1 (lane 1), D1-3 (lane 2), D145 (lane 3), SC577 (lane 4), SC607 (lane 5), and D3-4 (lane 6) were purified from the supernatants of baculovirus infected insect cells. Protein preparations were analyzed by SDS-PAGE prior to crystallization trials. The purity of each protein was estimated to be over 95%. D1 (lane 1) runs as a doublet due to partial glycosylation at the two N-linked carbohydrate sites.

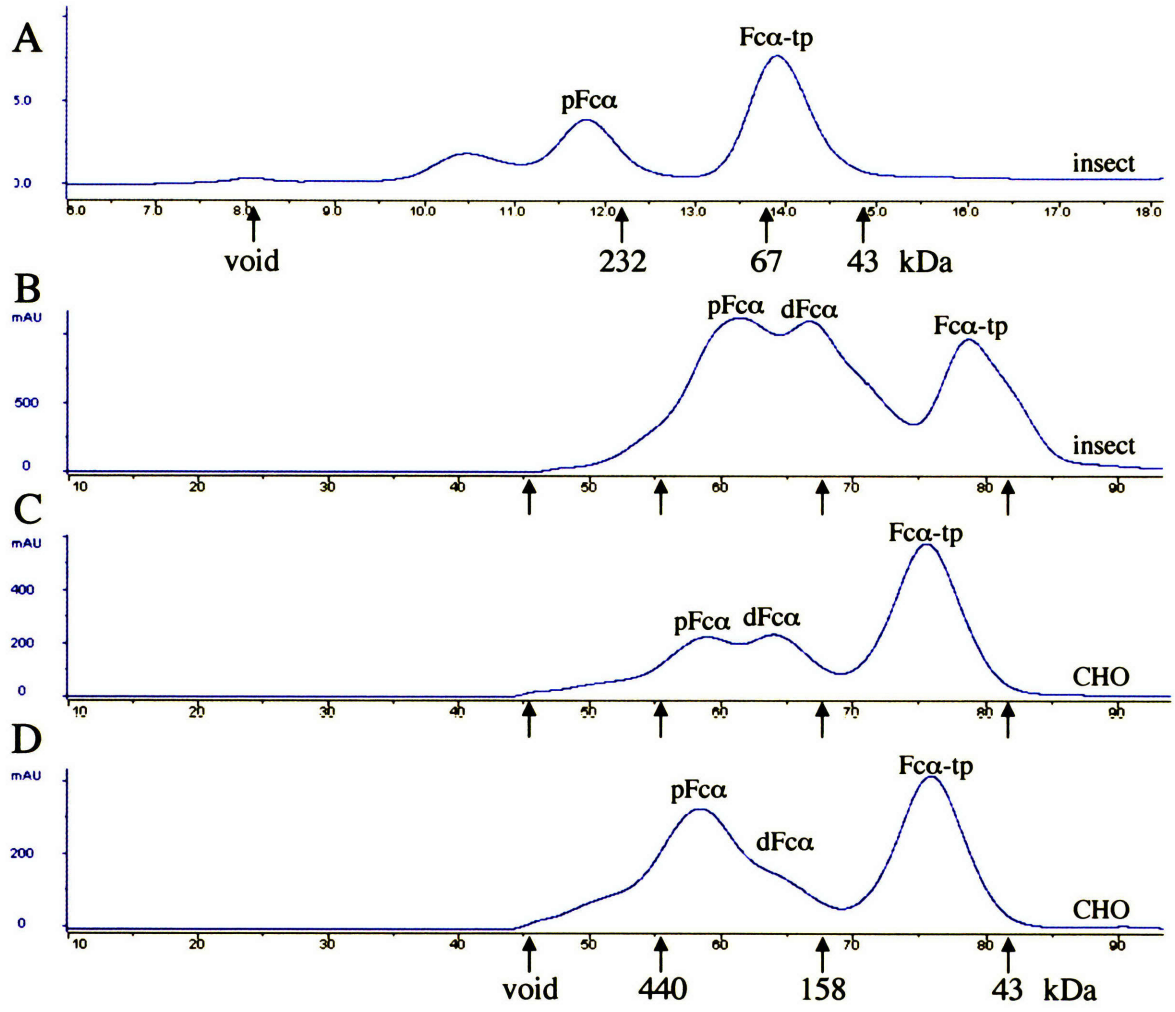


Figure 2. Size exclusion chromatography profiles of dFcα protein preparations. (A) Small scale coexpression of Fcα-tp and J chain in baculovirus infected insect cells results in the secretion of mainly monomeric Fcα. (B) Large scale coexpression of Fcα-tp and J chain using a different viral stock results in mainly dFcα and pFcα. (C) pFcα purified from the supernatants of CHO cells cotransfected with Fcα-tp and J chain, both under selection with MSX. (D) pFcα purified from the supernatants of CHO cells retransfected with J chain under Geneticin selection. Positions of molecular weight standards are indicated with arrows.

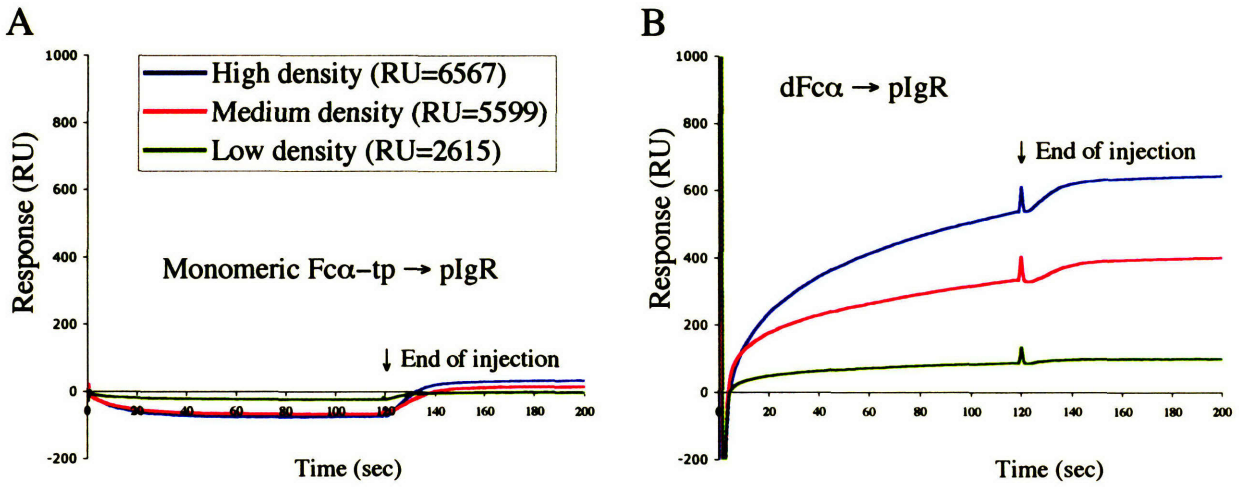


Figure 3. Biosensor analyses of monomeric and dimeric Fcα binding to pIgR. Sensorgrams from binding experiments in which monomeric Fcα-tp (A) or dFcα (B) were injected over pIgR SC607 immobilized on a biosensor chip at three different densities. The end of the injection is indicated with an arrow.

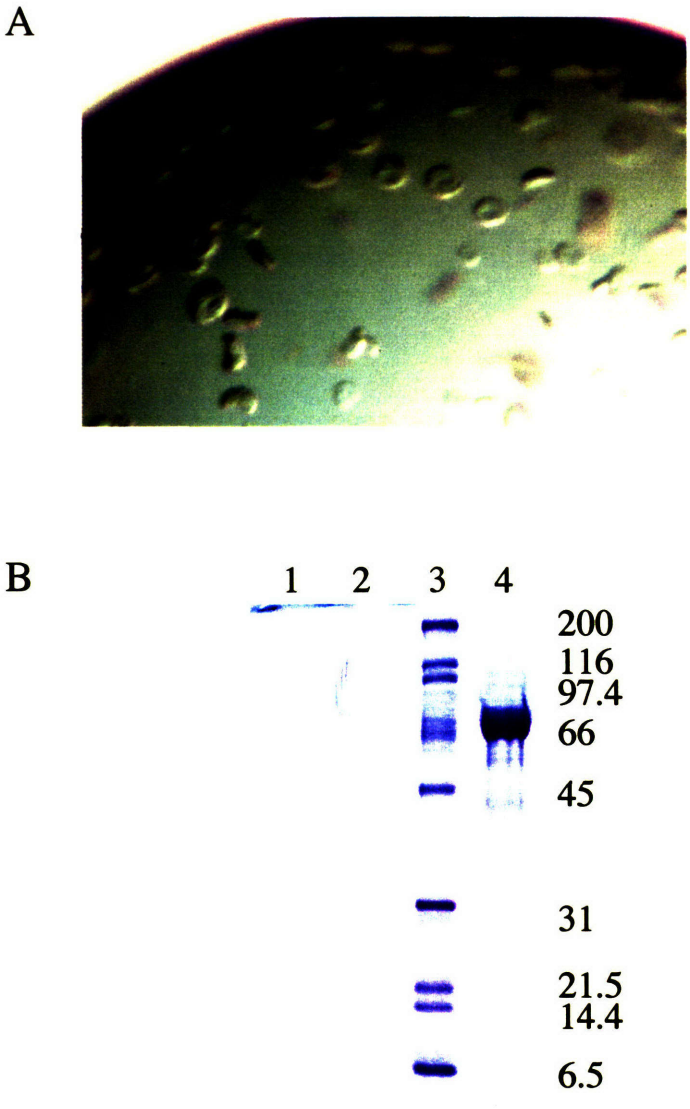
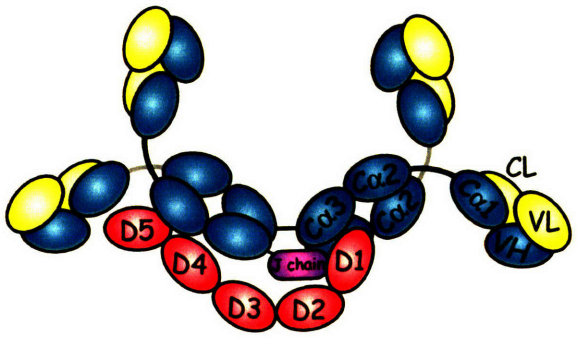


Figure 4. SDS-PAGE analysis of pIgR SC577 crystals. (A) Disk-shaped crystals (~120 μm in diameter) of SC577 were grown by hanging drop vapor diffusion and photographed in the drop. (B) Three to four crystals were removed from the drop, washed three times in the crystallization solution and analyzed by SDS-PAGE: last wash (lane 1), SC577 crystals (lane 2), molecular weight standards (lane 3) and SC577 protein used for the crystallization trials (lane 4).

A



B

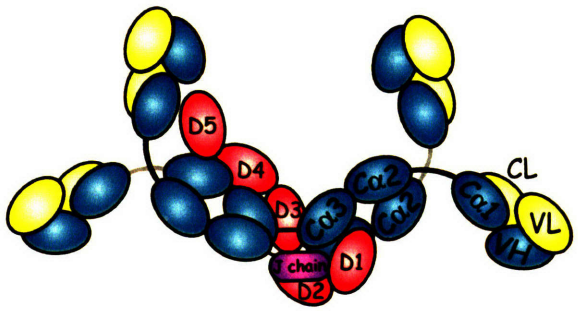


Figure 5. Putative models for the structure of SIgA. pIgR (red) D1 binds to the Cα3 domain, and possibly Cα2 domain and J chain, in the first IgA molecule in dIgA. (A) If pIgR D2-3 do not interact with, or contribute to, dIgA binding, D5 might interact with the Cα2 domain in the second IgA molecule that is sterically more accessible. (B) If pIgR D2-3 interact with dIgA, pIgR D5 might be oriented to interact with the Cα2 domain in the second IgA molecule that is on the opposite side of dIgA.

APPENDIX I

Expression and Characterization of a Native Retroviral Envelope Glycoprotein

Appendix I describes my efforts to try to express a soluble, trimeric form of the Moloney Murine Leukemia Virus envelope glycoprotein for structural studies.

Introduction

Enveloped viruses such as influenza and the human immunodeficiency virus (HIV) pose a great risk to human health. Although some important differences exist in the way they invade target cells, it is likely that different families of enveloped viruses use similar mechanisms. In each case the entry process requires fusion of the viral and cellular membranes, a process that is mediated by the interaction between a key viral surface glycoprotein, env, and a cellular receptor.

Retroviral envelope glycoproteins are synthesized as fusion-incompetent single chain precursors that are post-translationally cleaved by a cellular protease to yield the surface (SU) and the transmembrane (TM) subunits. The SU subunit is responsible for receptor recognition and binding. The ectodomain of TM contains a hydrophobic region, called the fusion peptide, which inserts into the membrane of the target cell. This initiates fusion between the viral and cellular membranes, allowing the virus to deposit its genetic information into the cytoplasm of the target cell. TM also contains a transmembrane region that anchors env to the viral membrane and a cytoplasmic region that can vary in size from just a few residues to well over a hundred. Env is thought to exist as a trimer of the SU/TM heterodimer on the surface of virions [1-3].

To date, the best-characterized surface protein involved in viral membrane fusion is hemagglutinin (HA) of the orthomyxovirus, influenza. Similar to retroviral env glycoproteins, HA is synthesized as a precursor protein, HA0, which is subsequently proteolyzed into two disulfide-bonded subunits, HA1 and HA2 [4, 5]. HA1 is responsible for binding to the cellular receptor, sialic acid, and is functionally equivalent to the retroviral SU subunit. HA2 is membrane bound and is responsible for mediating fusion, similar to the retroviral TM subunit. On the surface of virus particles, HA is a trimer of HA1/HA2 heterodimers [6].

The structure of the native HA revealed that HA2 forms a mostly α-helical stalk with the HA1 trimer at the membrane distal end contributing to the stalk, but primarily forming three globular domains that contain the receptor-binding domain. The length of the molecule is approximately 135 Å and the fusion peptide is buried in a pocket close to the base of HA [7]. These observations suggested that the protein must undergo major conformational changes for the fusion peptide to reach the target membrane. During infection, HA1 binds sialic acid at the plasma membrane, but unlike retroviruses, fusion does not occur at this point. Instead, the virus particle is taken up into the target cell by endocytosis. As the pH drops within the endosome, HA undergoes major, irreversible conformational changes, as revealed by the low-pH converted form [8]. HA1 is thought to dissociate from the complex, the amino-terminus of HA2 is displaced 100 Å and at least 36 residues are recruited to the viral membrane distal end of the triple stranded coiled coil seen in the native structure. This liberates and propels the fusion peptide into the host membrane by the "spring-loaded" mechanism [9], resulting in membrane fusion and the release of the viral core into the cytoplasm.

Based on striking structural and functional similarities to influenza HA, a membrane fusion model for HIV, and presumably other retroviruses, has been proposed [10]. The envelope protein is initially in its native, or pre-fusogenic, state on the surface of virions. Once SU (gp120 in the case of HIV) binds the cellular receptor (CD4 and chemokine coreceptor) conformational changes occur in SU that alter the SU/TM (gp120/gp41) interactions. This leads to conformational changes in the TM subunit resulting in the insertion of the fusion peptide into the target cell membrane. The TM domain then resolves into a fusion-active trimer-of-hairpins structure, in which the N-terminal region forms a trimeric coiled coil and the C-terminal region packs into the grooves of the coiled coil in an anti-parallel orientation [11]. This structural

rearrangement results in the apposition of the cellular and viral membranes and leads to membrane fusion. The fusion process can be inhibited in a dominant negative fashion by the addition of synthetic C and N peptides, corresponding to the C and N terminal regions of the prehairpin (Figure 1).

There are many high resolution structures of the fusion-active state of the membrane anchored ectodomain subunits, including the retroviral TMs: HIV-1 gp41, simian immunodeficiency virus (SIV) gp41, Moloney Murine Leukemia Virus (Mo-MLV) Mo-55, and human T cell leukemia virus (HTLV-1) gp21 [11-15]; the orthomyxovirus influenza HA [8, 16]; and the filovirus Ebola GP2 [17, 18]. Despite considerable sequence diversity among these viruses, they all share a number of common features. They all contain an N-terminal hydrophobic fusion peptide (not seen in the crystal structures), an adjacent coiled coil forming region, a disulfide bonded loop, and a second heptad repeat containing region at the C-terminus.

In contrast to the many fusogenic TM structures, there is only a limited amount of structural information about retroviral SU subunits. A monomeric HIV-1 gp120 structure at 2.5 Å resolution was solved, although the variable loops, V1, V2 and V3 were deleted and the construct was truncated at both the N- and C-termini. The protein is in complex with a two domain fragment of human CD4 and an Fab from a neutralizing antibody that blocks coreceptor binding, thus it may represent a non-native conformation [19]. A model of trimeric gp120 has been proposed, although the structure has not yet been solved [20]. The monomeric receptor-binding domain (RBD) of Friend MLV (Fr-MLV), which comprises approximately half of the SU subunit, has also been solved to 2.0 Å resolution [21]. A model for the trimeric packing of the Fr-RBD was also proposed [22]. However, currently there are no structures of trimeric retroviral SU/TM complexes. Therefore, the location of the fusion peptide, the contacts between

the SU and TM subunits and the interactions between the three SU/TM complexes within the env are unknown in the native state.

The availability of a high-resolution structure of a native envelope protein would provide a better understanding of retroviral binding and entry. This information would help to validate and complete the current model for viral membrane fusion. It may also have implications for structure-based drug discovery, the development of gene therapy vectors with defined cellular receptor specificity, and for the identification of new targets for viral entry inhibitors.

Ultimately we want to understand how human pathogenic retroviruses, such as HIV, invade cells. However, we expect that all retroviruses utilize very similar, if not identical, mechanisms to gain entry into target cells. Therefore, for technical reasons we chose to study one of the simple retroviruses, Mo-MLV, as a model system. Mo-MLV is expected to be a better candidate than HIV for structural studies for a number of reasons. For example, Mo-MLV is smaller in size, less heavily glycosylated, and the SU/TM subunit association is more stable (due to the disulfide bond) than in HIV.

In the case of Mo-MLV, the precursor gp85 is cleaved into the SU (gp70) and TM (p15E) subunits that remain covalently associated via a labile disulfide bond [23]. During or shortly after viral budding, p15E is further processed by the viral protease (PR), which removes the C-terminal 16 residues of the cytoplasmic tail, the R peptide [24, 25] (Figure 2). This cleavage is responsible for increasing the fusogenic ability of env [26, 27].

Mo-MLV is the prototype of the mammalian C-type retrovirus group. The murine viruses within this group are further classified according to their host range and superinfectivity resistance, a phenomenon in which prior infection with a retrovirus blocks further infection of the same cell by another retrovirus that uses the same receptor. Replication of Mo-MLV is

restricted to cells of mouse or rat origin, so it is classified as an ecotropic virus. The cellular receptor for ecotropic MLVs is the 14-pass transmembrane murine cationic amino acid transporter, mCAT-1 [28-30]. As with most other retroviruses, receptor binding and membrane fusion occur at the plasma membrane at neutral pH.

In order to better understand the mechanism that retroviruses utilize to invade target cells, we used biophysical and biochemical methods to characterize the Mo-MLV envelope glycoprotein in its native state. We used several different recombinant expression systems to try to obtain milligram quantities of purified protein necessary for crystallization trials. We also characterized the protein using analytical ultracentrifugation, circular dichroism and chemical cross-linking.

Materials and Methods

Construction of Mo-Env expression vectors

pCL-Eco [31] was kindly provided by Harvey F. Lodish (Whitehead Institute for Biomedical Research, MIT, Cambridge, MA). pCL-Eco, which contains the Mo-MLV env gene from strain J02255 of the virus, was used as the PCR template to generate expression constructs. See Table 1 for construct details.

Expression and purification of Mo-Env in bacteria

Constructs designed for bacterial cell expression (Table 1) were transformed into *E.coli* BL21(DE3)pLysS or AD494 cells (Novagen). Protein expression was induced with IPTG and analyzed by SDS-PAGE. Δ1-233, the protein with the highest expression level, was refolded from inclusion bodies and purified by Ni-NTA affinity followed by size exclusion chromatography on a Superdex 200 26/60 gel filtration column (Amersham Biosciences).

Expression of Mo-Env in Drosophila cells

Stable *Drosophila* cell lines expressing the Mo-MLV env were generated according to the manufacturer's DES protocol (Invitrogen). Briefly, cells were cotransfected with one of the two Mo-MLV env DNA constructs (Table 1) and the selection vector, pCoHYGO, using the calcium phosphate transfection reagents included in the kit. Stable cells were selected with 300 μg/mL hygromycin-B. Supernatant was continuously harvested from cells transfected with the pAc5.1/V5-HisA(env) construct, which utilizes the Ac5 promoter for constitutive expression of the gene of interest. Recombinant protein expression from the cells transfected with the pMT/BiP/V5-HisA(env) construct was induced with the addition of 500 μM copper sulfate. Supernatants were analyzed for protein expression by SDS-PAGE and western blotting.

Expression and purification of Mo-Env in baculovirus infected insect cells

Two different systems were used for protein expression in baculovirus infected insect cells. All of the constructs in the pFastBac1 vector (Table 1) were expressed in Sf9 cells using the Bac-to-Bac Baculovirus Expression System according to the manufacturer's protocol (GibcoBRL, Life Technologies). Briefly, recombinant bacmid DNA was isolated from DH10Bac cells transformed with the recombinant pFastBac1 vector, and used to transfect Sf9 cells. Recombinant baculovirus was harvested from the supernatant several days after the transfection and used to amplify the virus and to infect Sf9 cells for recombinant protein expression. The four constructs in the pAcGP67B and pAcSecG2T transfer vectors were cotransfected with linearized viral DNA (Baculogold, BD Biosciences) to generate recombinant baculovirus and the proteins were expressed in High Five insect cells. Secreted env proteins were harvested from the supernatants of infected cells, which were dialyzed against or buffer exchanged into Tris buffered saline (TBS; 20 mM Tris pH 8.0, 150 mM NaCl) and purified by Ni-NTA affinity (Qiagen) followed by size exclusion chromatography. Supernatants were also passed over an

immunoaffinity column prepared by coupling a purified monoclonal α-SU antibody from the supernatant of a rat hybridoma (83A25) [32] to cyanogen bromide activated sepharose 4B beads (Pharmacia).

Expression of Mo-Env in yeast

To test yeast as an expression host, Mo-MLV env constructs were made for expression in *Pichia pastoris* and *Saccharomyces cerevisiae* (Table 1). *Pichia* expression tests were performed according to the manufacturer's instructions (EasySelect *Pichia* Expression Kit, Invitrogen). Briefly, the Mo-MLV env genes were subcloned into the pPICZα expression vector in frame with the α-factor hydrophobic secretion signal. *Pichia pastoris* strains X-33 and GS115(*his4*) were transfected with the recombinant plasmid and protein expression was induced by the addition of methanol. Protein secretion was monitored by incubation of the media with Ni-NTA beads and analyzed by SDS-PAGE. The constructs for expression in *S. cerevisiae* (Table 1) were subcloned into the p423Gal1 vector and cells were transformed with the plasmids but expression was not tested.

Expression of Mo-Env proteins in mammalian cells

Stable mammalian cells were also generated to express the Mo-MLV env protein (Table 1). 293T cells were transiently transfected with full-length env constructs using the ProFection mammalian transfection system according to the manufacturer's protocol (Promega). Protein expression was analyzed 48 hours post-transfection by SDS-PAGE and western blotting. For the expression of secreted proein, stable CHO-K1 and CHO lecR- 3.2.8.1 cells [33] were generated as described [34]. Briefly, CHO cells were transfected with approximately 16 μg of DNA and stable colonies were selected with 100 μM methionine sulphoximine (MSX). Supernatants containing secreted protein were harvested and analyzed by SDS-PAGE and western blotting.

Chemical cross-linking

Chemical cross-linking experiments were carried out using a modified version of a published protocol [35]. Briefly, 6.7 μg of ΔC was incubated with 0.08, 0.8 or 4.0 mM concentrations of ethylene glycolbis(succinimidylsuccinate) (EGS; Pierce) in 10 μL reactions at room temperature for 30 minutes. 0.6 μL of Tris-HCl pH 7.4 was added to quench the reaction and incubated on ice for an additional 20 minutes. The cross-linked products were analyzed by SDS-PAGE. The same experiment was also performed with a constant concentration of 0.2 mM EGS and increasing amounts of ΔC (3.3, 6.7 and 10 μg). The cross-linked samples were analyzed by SDS-PAGE and western blotting using a monoclonal α-PentaHis antibody (Qiagen).

Circular dichroism analyses and thermal denaturation

CD experiments were performed with an AVIV 62A DS spectrometer. Absorbance readings from 190 to 250 nm were recorded on a 4.5 μM solution of ΔC in Hepes buffered saline (HBS; 20 mM Hepes pH 7.0, 150 mM NaCl) at 25 °C. The thermal denaturation of ΔC (4.5 μM in HBS) was monitored at 220 nm from 1 to 99 °C in 2 °C increments.

Sedimentation velocity analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation experiments were performed with a Beckman Coulter Optima XL-1 analytical ultracentrifuge. Two solutions of ΔRBD (~11.1 and ~31 μM in TBS) were centrifuged at 36,000 rpm at 20 °C, and two solutions of ΔC (2.3 μM in HBS and 2.9 μM in citrate buffered saline (CBS; 20 mM sodium citrate pH 5.6, 150 mM NaCl)) were centrifuged at 30,000 rpm at 20 °C. Absorbance readings for ΔRBD and ΔC were measured at 280 and 230 nm, respectively. The data were fit using the *c(s)* and *c(M)* analyses of SEDFIT, which calculates the differential sedimentation coefficient distribution or molar mass distribution, respectively, taking into account sample diffusion [36].

Crystallization attempts

RBD, ΔRBD and ΔC were concentrated to approximately 28, 20 and 7 mg/mL, respectively, and crystallization trials were set up using both the hanging drop and sitting drop vapor diffusion methods. Commercially available sparse matrix crystallization solutions (Hampton and Emerald BioStructures) as well as homemade grid screens were used.

Viral infectivity and inhibition assay

Mo-36 [37] (the Mo-MLV N peptide; D481-F515) and Mo-MLV C peptide (T536-E565) were synthesized on a Perkin Elmer Model 431 peptide synthesizer. Both peptides had an acetylated N terminus and C terminal amide. The peptides were purified by reverse-phase high-performance liquid chromatography on a Vydac C18 column using a water-acetonitrile gradient in the presence of 0.1% trifluoroacetic acid and lyophilized. α-Mo-92 [37] polyclonal antibodies were purified from rabbit serum on a protein A column (Pierce).

293T cells were cotransfected with pCL-Eco [31] and pMX-IRES-GFP [38] to generate a replication defective virus carrying the gene for GFP, Mo-GFP. 48 hours post-transfection, cell supernatants containing Mo-GFP virions were harvested. To titer the virus, NIH/3T3 mouse cells were then infected with increasing amounts of Mo-GFP in the presence of 4 μg/mL polybrene. Based in the viral titer, inhibition experiments were carried out with 4 μL of Mo-GFP in the presence of increasing concentrations of Mo-MLV C-peptide (0, 0.01, 0.1, 1, 10 and 50 μM) or α-Mo-92 antibodies (0, 1.6, 16, 32, 80 and 160 nM).

Results and Discussion

Structural studies of retroviral envelope proteins have been hindered in part because the glycoprotein complexes are difficult to produce in high quantities, many are heavily glycosylated and the instability of the SU/TM complexes often leads to dissociation of the two subunits. In

order to obtain a soluble, trimeric form of the native Mo-MLV env, we made a series of constructs for recombinant protein expression using a number of different expression systems including bacteria, stably transfected *Drosophila* cells, baculovirus infected insect cells, yeast and stable mammalian cells (Table 1).

Protein expression and purification in baculovirus infected insect cells

To allow for proper folding and processing to occur, baculovirus infected insect cells were used to express the Mo-MLV env protein in its native state. The original construct contained the native Mo-MLV hydrophobic signal sequence to direct the protein into the endoplasmic reticulum, the complete gp70 (SU) subunit and the ectodomain of p15E (TM), terminating just prior to the transmembrane anchoring domain resulting in secretion of the protein. Similar constructs have been used to successfully express a stable, soluble, trimeric ectodomain of the SIV gp160 in insect cells using a baculovirus expression system [35], and HIV gp140 [39] in mammalian cells.

Analyses of the supernatants of insect cells expressing the Mo-MLV env protein revealed that cleavage between the SU/TM subunits was only approximately 50% efficient using the baculovirus expression system (data not shown). Others have also noted that proteolytic processing is sometimes inefficient in insect cells when the level of protein synthesis is high. For example, HA cleavage is slow and inefficient when produced in insect cells using a baculovirus expression system compared to when expressed in a vertebrate cell line. Similar observations have been reported for other viral proteins such as the HIV-1 gp160 ([40] and references therein). The heterogeneity of the protein due to partial cleavage is expected to cause difficulties with crystallization. In addition, after cleavage of the precursor, the Mo-MLV SU and TM subunits remain linked through a labile disulfide bond. The majority of viral SU and

TM proteins derived from mammalian cells migrate independently by SDS-PAGE even in the absence of a reducing agent. If virions are pretreated with N-ethylmaleimide (NEM) they migrate as a covalent complex [23]. It has been proposed that this phenomenon is caused by an endogenous reducing agent in the Mo-MLV env, a free thiol from an unpaired cysteine, that can cause disulfide rearrangement and lead to the loss of the SU/TM association. The NEM would block the free thiol and prevent the disulfide rearrangement [22]. Regardless, in the absence of NEM long-term storage conditions could lead to the loss of the covalent SU/TM association and result in shedding, which would be disadvantageous in attempts to crystallize the protein.

Therefore, to stabilize the SU/TM complex and to avoid heterogeneity due to incomplete cleavage or shedding, we eliminated the SU/TM cleavage site (Table 1). The cleavage recognition site, RHKR, was mutated to RHAA using site directed mutagenesis. Expression of these mutants (ΔC) resulted in no detectable cleavage between the two subunits (data not shown). We expect this mutation to still represent the native state of the ectodomain for two reasons. First, the 2.8 Å resolution uncleaved influenza HA0 structure as compared to the cleaved HA structure showed mainly minor changes in the overall structure, and all of them were centered around the cleavage site [41]. Second, uncleaved Mo-MLV env is still incorporated into virions, although they are non-infectious [42].

To eliminate another source of potential heterogeneity among the secreted proteins, we also made a number of mutants to minimize the number of the N-linked glycosylation sites. The mature SU subunit of Mo-MLV contains seven N-linked oligosaccharides [43]. Mutagenesis studies have revealed that when each glycosylation site is mutated individually, five of the seven mutants produce viable and infectious virions. Mutations at positions 2 and 3 (Δ2 and Δ3, respectively) result in a detectable phenotype. Δ2 is temperature sensitive while Δ3 is lethal to

the virus. A triple mutant, Δ147, also results in viable virus [44]. Friend Murine Leukemia Virus (Fr-MLV), a related ecotropic MLV, contains eight N-linked glycosylation sites. The Fr-MLV Δ4 mutation is analogous to Mo-MLV Δ3, and is the only mutation to yield non-infectious virions [45]. These studies indicate that the oligosaccharide at position 3 in Mo-MLV and the corresponding one at position 4 in Fr-MLV may be sufficient for function.

To determine the minimum number of glycosylation sites required to retain full activity of the Mo-MLV env, we constructed Δ14567, a mutant lacking all five glycosylation sites which produced no phenotype when individually mutated [44]. The gene for this construct was made by site-specific oligonucleotide-directed mutagenesis (Quickchange, Stratagene) in the pcDNA1.1/Amp mammalian expression vector (Invitrogen). Wild type env and a Δ147 mutant were also generated to use as controls (Table 1). To generate a high-titer, helper-free retrovirus by transient transfection, the packaging cell line Anjou [46] was cotransfected with the pcDNA1.1/Amp plasmid carrying the env gene and pMX-IRES-GFP, a vector that contains the retroviral packaging signal and the gene for the green fluorescent protein (GFP) [38]. The resulting replication deficient virus is capable of undergoing a single round of infection and expresses GFP. NIH/3T3 mouse cells were infected and assayed for infection by fluorescence activated cell sorting (FACS) by monitoring GFP expression. However, we were unable to detect virions in the Anjou cell supernatants by western blotting and therefore could not test for infectivity in NIH/3T3 cells. Moreover, the expression level of secreted Δ14567 was even lower than for Δ147 in baculovirus infected insect cells, potentially indicating that the oligosaccharides at positions 5 and 6 aid in protein folding and may be necessary. Therefore, overexpression of Δ14567 was not pursued further.

The amount of secreted protein in the supernatants of baculovirus infected insect cells was very low with the original constructs (<0.5 mg/L), all of which contained the native Mo-MLV env hydrophobic signal sequence. To increase expression and secretion levels, the Mo-MLV secretion signal was replaced with the signal sequences of two secreted baculovirus proteins, the ecdysteriod UDPglycosyltransferase (EGT) and the envelope glycoprotein gp67 for baculovirus expression. This method increased expression of HIV-1 gp120 6- to 20-fold over using its native signal sequence [47]. To further optimize the chances of successfully overexpressing the Mo-MLV env, we also made a number of constructs with different combinations of truncation points, signal sequences, affinity tags, membrane anchors, intact glycosylation sites and the presence or absence of the SU/TM cleavage site (Table 1). The four best expressing proteins from baculovirus infected insect cells, RBD, ΔRBD, ΔC and Δ147C, were overexpressed in large-scale infections. These 6xHis tagged proteins were purified by Ni-NTA affinity and size exclusion chromatography to about 90% purity as determined by SDS-PAGE analysis. Secreted insect ferritin, a minor contaminant, was consistently present in the protein preparations of the lower expressing constructs (≤ 1 mg/L; ΔRBD and Δ147C) when isolated from the supernatant of baculovirus infected insect cells. Attempts to further purify these proteins with an immunoaffinity column prepared with a monoclonal α-SU antibody isolated from the supernatant of a rat hybridoma (83A25) failed because the protein did not bind to the column.

Protein expression in *Drosophila* S2 insect cells

The Mo-MLV ectodomain was also expressed in stably transfected cell lines derived from *Drosophila melanogaster*, Schneider 2 (S2) cells (DES, Invitrogen). Two stable cell lines were generated, each of them expressing the Δ147C glycosylation and cleavage mutant form of env

with enhanced signal sequences (Table 1). One cell line was transfected with the pAc5.1/V5-HisA(env) construct, which drives the constitutive expression of recombinant protein from the *Drosophila* actin 5C (Ac5) promoter and the env protein contained the EGT signal sequence. The second cell line was transfected with the pMT/BiP/V5-HisA(env) construct, which contains the *Drosophila* BiP signal sequence and an inducible metallothionein (MT) promoter. Both secreted proteins contained a C-terminal 6xHis tag, which was used for purification using Ni-NTA resin (Qiagen) and detection by western blotting. Although both versions of the protein were detected, the expression and secretion levels were very low (<0.5 mg/L), which was not sufficient for purification of the quantity of protein necessary for biochemical characterization and crystallization.

Protein expression and purification in bacterial cells

Mo-MLV env constructs (Table 1) were also expressed in *E.coli* BL21(DE3)pLysS and AD494(DE3)pLysS cells (Novagen). AD494(DE3)pLysS is a thioredoxin thioreductase-deficient strain with a mutation in the *trxB* gene, which allows disulfide bond formation and proper folding for some proteins in the *E.coli* cytoplasm (Novagen). However, bacterial expression of Mo-MLV env resulted in the production of insoluble protein localized to inclusion bodies using both cell lines. The highest expressing construct, Δ1-233, was purified from inclusion bodies and refolded. The refolded protein was completely aggregated and eluted from a Superdex 200 26/60 gel filtration column in the void volume (data not shown).

Protein expression in yeast

Several Mo-MLV env constructs were also made to test expression in two yeast strains, *Pichia pastoris* and *Saccharomyces cerevisiae* (Table 1). Δ147C and ΔRBD were transformed into *Pichia pastoris* X-33 and GS115(*his4*) cells and protein expression was induced with the

addition of methanol. Protein expression was not observed during analyses of the media by Ni-NTA affinity and SDS-PAGE. After failure to detect protein expression with *Pichia pastoris*, expression tests in *S. cerevisiae* were not pursued.

Protein expression in mammalian cells

Mammalian cells were also employed to overexpress the Mo-MLV env protein. Stably transfected wild type chinese hamster ovary cells (CHO) K1 and CHO lecR- 3.2.8.1 cells [33], a variant cell line which N-glycosylates with only simple oligosaccharides to decrease the heterogeneity in glycosylation sometimes seen in wild type CHO cells, were generated (Table 1). As with the other expression systems, very low expression levels were observed for all of the constructs, regardless of the signal sequence used.

Biochemical and biophysical characterization

To determine if the Mo-MLV env ectodomain secreted by baculovirus infected insect cells is a soluble, trimeric species we used gel filtration chromatography, chemical cross-linking and sedimentation velocity experiments. ΔRBD, ΔC, and Δ147C, but not RBD, eluted from size exclusion chromatography columns as multiple, overlapping peaks indicating the presence of multiple species (Figure 3A). The peak fractions in the region expected to contain the trimeric form of the env protein were pooled and used for further characterization. The oligomeric state of the Mo-MLV env was investigated by chemical cross-linking (Figure 3B). Various concentrations of ΔC were mixed with increasing concentrations of the cross-linking reagent, EGS, and analyzed by SDS-PAGE. This method was used to show that the SIV gp160 protein is trimeric [35]. The results with ΔC, however, suggested there was a mixture of different oligomeric states and aggregated protein present as a smear of high molecular weight bands were observed rather than distinct bands corresponding to ordered oligomers.

Sedimentation velocity analytical ultracentrifugation experiments were also performed to investigate the oligomeric state of the Mo-MLV env protein. Two solutions of ΔRBD were centrifuged at 36,000 rpm while sedimentation was monitored at 280 nm, whereas two solutions of ΔC were centrifuged at 30,000 rpm and absorbance was measured at 230 nm. The data for both experiments were fit using SEDFIT [36] to calculate the sedimentation coefficient, *c(s)*, and molar mass, *c(M)*, distributions of each sample. Analyses of the sedimentation coefficient distributions revealed the presence of a mixture of many different oligomeric species, primarily aggregates for both ΔRBD and ΔC (Figures 4A-B), suggesting that the proteins are misfolded and/or nonspecifically aggregated.

The stability of the protein was examined by thermal denaturation, monitoring circular dichroism (CD). CD analyses were preformed on a 4.5 μM solution of ΔC in HBS to detect secondary structural elements. The absorbance scan from 190 to 250 nm showed the presence of a mixture of secondary structural elements or possibly random coil (Figure 5A). A thermal denaturation experiment failed to detect a melting transition between 1 and 99 °C (Figure 5B). This further supports the hypothesis that the protein is in a partially folded or aggregated state rather than in a stably folded, trimeric conformation.

Initial crystallization attempts of ΔRBD resulted in the crystallization of a minor (~10%) contaminant, secreted insect ferritin. Further crystallization trials with ΔC yielded microcrystals that were too small to characterize. Optimization of the crystallization solutions and seeding attempts did not improve the crystals. RBD did not yield any crystals.

Viral infectivity and inhibition assay

To determine if the mechanism used by Mo-MLV to invade target cells is similar to other retroviruses, we established a viral infectivity and inhibition assay (Figure 6A). The HIV TM

(gp41) contains two regions, the N and C peptides, which are transiently exposed after SU (gp120) interacts with the cellular receptor and coreceptor. The N peptides form a trimeric coiled coil and the C peptides pack into the hydrophobic grooves on the N core in an antiparallel orientation to resolve into the fusion-active trimer of hairpin structure [11]. The gp41 transient intermediate has been identified as an attractive target for inhibition [48]. The fusion process can be inhibited in a dominant negative fashion by the addition of synthetic C or N peptides (reviewed in [49]). To inhibit the infection of mouse NIH/3T3 cells by Mo-GFP, Mo-36 [37], the functional equivalent of the N peptide and Mo-MLV C peptide were synthesized and purified by reverse-phase HPLC. Antibodies raised against the Mo-MLV N/C core, Mo-92 [37], were also purified to test these antibodies' ability to inhibit viral infection (Figure 6B). When Mo-MLV C peptide (up to 50 μM concentration) or α-Mo-92 antibodies (up to 160 nM concentration) were added to NIH/3T3 cells in the presence of 4 μL Mo-GFP, no inhibition of infection was observed (Figure 6C-D).

Acknowledgements

I would like to thank Peter S. Kim, in whose lab this project originated and former members of the former Kim lab, especially Michael S. Kay, Benjamin K. Chen, Feng Yang, and Peter A. Carr for helpful discussions and invaluable insights and Michael Burgess for peptide synthesis. I would also like to acknowledge Pamela J. Bjorkman for supporting this project at Caltech and Peter M. Snow at the Caltech Protein Expression Center for advice on and assistance with protein expression and purification. I also thank Andrew B. Herr for assistance with analytical ultracentrifugation experiments and Kyle S. Lassila for assistance with circular dichroism data collection.

References

1. Weiss, C.D., J.A. Levy, and J.M. White, *Oligomeric organization of gp120 on infectious human immunodeficiency virus type 1 particles.* J Virol, 1990. **64**(11): p. 5674-7.
2. Einfeld, D. and E. Hunter, *Oligomeric structure of a prototype retrovirus glycoprotein.* Proc Natl Acad Sci U S A, 1988. **85**(22): p. 8688-92.
3. Kamps, C.A., Y.C. Lin, and P.K. Wong, *Oligomerization and transport of the envelope protein of Moloney murine leukemia virus-TB and of ts1, a neurovirulent temperature-sensitive mutant of MoMuLV-TB.* Virology, 1991. **184**(2): p. 687-94.
4. Lazarowitz, S.G., R.W. Compans, and P.W. Choppin, *Influenza virus structural and nonstructural proteins in infected cells and their plasma membranes.* Virology, 1971. **46**(3): p. 830-43.
5. Skehel, J.J. and M.D. Waterfield, *Studies on the primary structure of the influenza virus hemagglutinin.* Proc Natl Acad Sci U S A, 1975. **72**(1): p. 93-7.
6. Wiley, D.C., J.J. Skehel, and M. Waterfield, *Evidence from studies with a cross-linking reagent that the haemagglutinin of influenza virus is a trimer.* Virology, 1977. **79**(2): p. 446-8.
7. Weis, W., et al., *Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid.* Nature, 1988. **333**(6172): p. 426-31.
8. Bullough, P.A., et al., *Structure of influenza haemagglutinin at the pH of membrane fusion.* Nature, 1994. **371**(6492): p. 37-43.
9. Carr, C.M. and P.S. Kim, *A spring-loaded mechanism for the conformational change of influenza hemagglutinin.* Cell, 1993. **73**(4): p. 823-32.
10. Chan, D.C. and P.S. Kim, *HIV entry and its inhibition.* Cell, 1998. **93**(5): p. 681-4.
11. Chan, D.C., et al., *Core structure of gp41 from the HIV envelope glycoprotein.* Cell, 1997. **89**(2): p. 263-73.
12. Weissenhorn, W., et al., *Atomic structure of the ectodomain from HIV-1 gp41.* Nature, 1997. **387**(6631): p. 426-30.
13. Malashkevich, V.N., et al., *Crystal structure of the simian immunodeficiency virus (SIV) gp41 core: conserved helical interactions underlie the broad inhibitory activity of gp41 peptides.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9134-9.
14. Fass, D., S.C. Harrison, and P.S. Kim, *Retrovirus envelope domain at 1.7 angstrom resolution.* Nat Struct Biol, 1996. **3**(5): p. 465-9.
15. Kobe, B., et al., *Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins.* Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4319-24.
16. Chen, J., J.J. Skehel, and D.C. Wiley, *N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 8967-72.
17. Weissenhorn, W., et al., *Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain.* Mol Cell, 1998. **2**(5): p. 605-16.
18. Malashkevich, V.N., et al., *Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9- A resolution.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2662-7.
19. Kwong, P.D., et al., *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.* Nature, 1998. **393**(6686): p. 648-59.

20. Kwong, P.D., et al., *Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus.* J Virol, 2000. **74**(4): p. 1961-72.
21. Fass, D., et al., *Structure of a murine leukemia virus receptor-binding glycoprotein at 2.0 angstrom resolution.* Science, 1997. **277**(5332): p. 1662-6.
22. Fass, D., *Thesis*, in *Biology*. 1997, Massachusetts Institute of Technology: Cambridge.
23. Pinter, A., J. Lieman-Hurwitz, and E. Fleissner, *The nature of the association between the murine leukemia virus envelope proteins.* Virology, 1978. **91**(2): p. 345-51.
24. Green, N., et al., *Sequence-specific antibodies show that maturation of Moloney leukemia virus envelope polyprotein involves removal of a COOH-terminal peptide.* Proc Natl Acad Sci U S A, 1981. **78**(10): p. 6023-7.
25. Schultz, A. and A. Rein, *Maturation of murine leukemia virus env proteins in the absence of other viral proteins.* Virology, 1985. **145**(2): p. 335-9.
26. Januszeski, M.M., et al., *Functional analysis of the cytoplasmic tail of Moloney murine leukemia virus envelope protein.* J Virol, 1997. **71**(5): p. 3613-9.
27. Rein, A., et al., *Function of the cytoplasmic domain of a retroviral transmembrane protein: p15E-p2E cleavage activates the membrane fusion capability of the murine leukemia virus Env protein.* J Virol, 1994. **68**(3): p. 1773-81.
28. Albritton, L.M., et al., *A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection.* Cell, 1989. **57**(4): p. 659-66.
29. Kim, J.W., et al., *Transport of cationic amino acids by the mouse ecotropic retrovirus receptor.* Nature, 1991. **352**(6337): p. 725-8.
30. Wang, H., et al., *Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter.* Nature, 1991. **352**(6337): p. 729-31.
31. Naviaux, R.K., et al., *The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses.* J Virol, 1996. **70**(8): p. 5701-5.
32. Evans, L.H., et al., *A neutralizable epitope common to the envelope glycoproteins of ecotropic, polytropic, xenotropic, and amphotropic murine leukemia viruses.* J Virol, 1990. **64**(12): p. 6176-83.
33. Stanley, P., *Selection of specific wheat germ agglutinin-resistant (WgaR) phenotypes from Chinese hamster ovary cell populations containing numerous lecR genotypes.* Mol Cell Biol, 1981. **1**(8): p. 687-96.
34. Martin, W.L. and P.J. Bjorkman, *Characterization of the 2:1 complex between the class I MHC-related Fc receptor and its Fc ligand in solution.* Biochemistry, 1999. **38**(39): p. 12639-47.
35. Chen, B., et al., *Expression, purification, and characterization of gp160e, the soluble, trimeric, ectodomain of the SIV envelope glycoprotein, gp-160.* J Biol Chem, 2000.
36. Schuck, P., *Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling.* Biophys J, 2000. **78**(3): p. 1606-19.
37. Fass, D. and P.S. Kim, *Dissection of a retrovirus envelope protein reveals structural similarity to influenza hemagglutinin.* Curr Biol, 1995. **5**(12): p. 1377-83.
38. Liu, X., et al., *Generation of mammalian cells stably expressing multiple genes at predetermined levels.* Anal Biochem, 2000. **280**(1): p. 20-8.
39. Zhang, C.W., et al., *Expression, purification, and characterization of recombinant HIV gp140. The gp41 ectodomain of HIV or simian immunodeficiency virus is sufficient to*

maintain the retroviral envelope glycoprotein as a trimer. J Biol Chem, 2001. **276**(43): p. 39577-85.

40. O'Reilly, D.R., L.K. Miller, and V.A. Luckow, *Baculovirus Expression Vectors.* A Laboratory Manual. 1994, New York: Oxford University Press, Inc.
41. Chen, J., et al., *Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation.* Cell, 1998. **95**(3): p. 409-17.
42. Zavorotinskaya, T. and L.M. Albritton, *Failure to cleave murine leukemia virus envelope protein does not preclude its incorporation in virions and productive virus-receptor interaction.* J Virol, 1999. **73**(7): p. 5621-9.
43. Rosner, M.R., L.S. Grinna, and P.W. Robbins, *Differences in glycosylation patterns of closely related murine leukemia viruses.* Proc Natl Acad Sci U S A, 1980. **77**(1): p. 67-71.
44. Felkner, R.H. and M.J. Roth, *Mutational analysis of the N-linked glycosylation sites of the SU envelope protein of Moloney murine leukemia virus.* J Virol, 1992. **66**(7): p. 4258-64.
45. Kayman, S.C., et al., *Mutational analysis of N-linked glycosylation sites of Friend murine leukemia virus envelope protein.* J Virol, 1991. **65**(10): p. 5323-32.
46. Pear, W.S., et al., *Production of high-titer helper-free retroviruses by transient transfection.* Proc Natl Acad Sci U S A, 1993. **90**(18): p. 8392-6.
47. Murphy, C.I., et al., *Enhanced expression, secretion, and large-scale purification of recombinant HIV-1 gp120 in insect cell using the baculovirus egt and p67 signal peptides.* Protein Expr Purif, 1993. **4**(5): p. 349-57.
48. Chan, D.C., C.T. Chutkowski, and P.S. Kim, *Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target.* Proc Natl Acad Sci U S A, 1998. **95**(26): p. 15613-7.
49. Eckert, D.M. and P.S. Kim, *Mechanisms of viral membrane fusion and its inhibition.* Annu Rev Biochem, 2001. **70**: p. 777-810.

Table 1. Mo-MLV envelope expression constructs

Expression of the Mo-MLV env was tested in five different expression systems: bacteria, stably transfected Drosophila cells, baculovirus infected insect cells, yeast and stably transfected mammalian cells. The table shows the expression host, expression vector, presence or absence on the SU/TM cleavage site recognition sequence, origin of the hydrophobic secretion signal, presence of an affinity tag, presence of a cleavage site for the removal of the affinity tag or another modification, the residues in the mature protein, the approximate molecular weight of the mature protein not taking into account the potential glycosylation, the number of potential N-linked glycosylation sites and the expression level in mg/L.

	Vector	SU/TM Cleavage Site	Signal Sequence	Affinity Tag	Cleavage site/ modification	Residues (mature protein)	MW (~kDa)	# N Glycos	Expression (~mg/L)
BACTERIA									
Wild type (ECTO)	pET20b	no	no	+/- C-6xHis	no	A1-D541	60	0	< 0.5
ΔRBD (Δ19-221)	pET20b	no	no	C-6xHis	no	A1-T18/I222-S571	41	0	> 1
ΔRBD (Δ1-232)	pET20b	no	no	C-6xHis	no	V233-S571	38	0	> 1
ΔRBD (Δ1-232)	pET17b	no	no	no	no	V233-T576	37	0	n.d.
RBD	pMAL-p2X	no	malE	N-MBP	Factor Xa	S2-P234	26+43	0	< 0.5
SU	pMAL-p2X	no	malE	N-MBP	Factor Xa	S2-R436	48+43	0	< 0.5
ΔRBD (Δ19-221)	pMAL-p2X	no	malE	N-MBP	Factor Xa	S2-T18/I222-D541	36+43	0	< 0.5
ECTO (574)	pMAL-p2X	no	malE	N-MBP	Factor Xa	S2-D541	59+43	0	< 0.5
ECTO (609)	pMAL-p2X	no	malE	N-MBP	Factor Xa	S2-T576	63+43	0	< 0.5
RBD	pMAL-c2X	no	no	N-MBP	Factor Xa	S2-P234	26+43	0	n.d.
SU	pMAL-c2X	no	no	N-MBP	Factor Xa	S2-R436	48+43	0	n.d.
DROSOPHILA S2									
EGTΔ147C	pAc5.1/V5-HisA	no	EGT	C-V5-6xHis	no	A1-S571	66	4	< 0.5
BiPΔ147C	pMT/BiP/V5-HisA	no	BiP	C-V5-6xHis	no	A1-S571	67	4	< 0.5
INSECT CELL									
Full length	pFastBac1	yes	native	no	no	A1-P632	70	7	< 0.5
Secreted	pFastBac1	yes	native	no	no	A1-T576	63	7	< 0.5
SU	pFastBac1	no	native	no	no	A1-R436	48	7	< 0.5
Full length-Δ147	pFastBac1	yes	native	no	no	A1-S571	70	4	n.d.
Full length-Δ14567	pFastBac1	yes	native	no	no	A1-S571	70	2	n.d.
Wild type (ECTO)	pFastBac1	yes	native	C-6xHis	+/-Factor Xa	A1-S571	64	7	< 0.5
ΔC	pFastBac1	no	native	C-6xHis	Factor Xa	A1-S571	64	7	< 0.5
Δ1	pFastBac1	yes	native	C-6xHis	Factor Xa	A1-S571	64	6	n.d.
Δ4	pFastBac1	yes	native	C-6xHis	Factor Xa	A1-S571	64	6	n.d.
Δ47	pFastBac1	yes	native	C-6xHis	Factor Xa	A1-S571	64	5	n.d.
Δ7C	pFastBac1	no	native	C-6xHis	Factor Xa	A1-S571	64	6	n.d.
Δ147	pFastBac1	yes	native	C-6xHis	Factor Xa	A1-S571	64	4	< 0.5
Δ147C	pFastBac1	no	native	C-6xHis	Factor Xa	A1-S571	64	4	< 0.5
Δ1457	pFastBac1	yes	native	C-6xHis	Factor Xa	A1-S571	64	3	n.d.
Δ14567	pFastBac1	yes	native	C-6xHis	Factor Xa	A1-S571	64	2	< 0.5
EGTΔ147C	pFastBac1	no	EGT	C-6xHis	Factor Xa	A1-S571	64	4	< 0.5
EGTΔ147CΔRBD	pFastBac1	no	EGT	C-6xHis	Factor Xa	A1-T18/I222-S571	42	3	< 0.5

Table 1. Mo-MLV envelope expression constructs n.d. not determined

INSECT CELLS									
ΔC-GPI	pFastBac1	no	native	no	GPI	A1-S571	66	7	< 0.5
Δ147C-GPI	pFastBac1	no	native	no	GPI	A1-S571	66	4	< 0.5
RBD	pFastBac1	no	native	+/- C-6xHis	+/-Factor Xa	A1-P234	28	2	3
Δ147C	pAcGP67B	no	gp67	C-6xHis	Factor Xa	A1-S571	65	4	0.5
ΔRBD	pAcGP67B	no	gp67	C-6xHis	Factor Xa	A1-T18/I223-S571	43	3	1
ΔC	pAcGP67B	no	gp67	C-6xHis	Factor Xa	S2-S571	64	7	2
ΔC	pAcSecG2T	no	gp67	N-GST C-6xHis	Thrombin Factor Xa	S2-S571	64+27	7	< 0.5
YEAST									
Pichia pastoris									
Δ147C	pPICZα	no	α-factor	C-6xHis	Factor Xa	A1-S571	64	4	< 0.5
ΔRBD	pPICZα	no	α-factor	C-6xHis	Factor Xa	A1-T18/I222-S571	42	3	< 0.5
S. cerevisiae									
Δ147C	p423Gal1	no	native	C-6xHis	Factor Xa	A1-S571	64	4	n.d.
ΔRBD	p423Gal1	no	native	C-6xHis	Factor Xa	A1-T18/I222-S571	42	3	n.d.
Δ147C-GPI	p423Gal1	no	native	no	GPI	A1-S571	66	4	n.d.
MAMMALIAN									
293T									
Full length-WT	pcDNA1.1/Amp	yes	native	no	no	A1-S571	70	7	n.d.
Full length-Δ147	pcDNA1.1/Amp	yes	native	no	no	A1-S571	70	4	n.d.
Full length-Δ14567	pcDNA1.1/Amp	yes	native	no	no	A1-S571	70	2	n.d.
CHO LecR- 3.2.8.1									
Wild type (ECTO)	pBJ5-GS-MCS	yes	native	C-6xHis	Factor Xa	A1-S571	64	7	< 0.5
ΔC	pBJ5-GS-MCS	no	native	C-6xHis	Factor Xa	A1-S571	64	7	< 0.5
CHO-K1									
Δ147C	pBJ5-GS-MCS3	no	rIgG2a	C-6xHis	Factor Xa	A1-S571	65	4	< 0.5
ΔRBD	pBJ5-GS-MCS3	no	rIgG2a	C-6xHis	Factor Xa	A1-T18/I222-S571	42	5	< 0.5
ΔC	pBJ5-GS-MCS3	no	rIgG2a	C-6xHis	Factor Xa	A1-S571	65	7	n.d.
Δ147C	pBJ5-GS-MCS	no	native	C-6xHis	Factor Xa	A1-S571	64	4	< 0.5
ΔRBD	pBJ5-GS-MCS	no	native	C-6xHis	Factor Xa	A1-T18/I222-S571	42	5	< 0.5
ΔC-GPI	pBJ5-GS-MCS	no	native	no	GPI	A1-S571	66	7	n.d.

Table 1 continued. Mo-MLV envelope expression constructs n.d. not determined

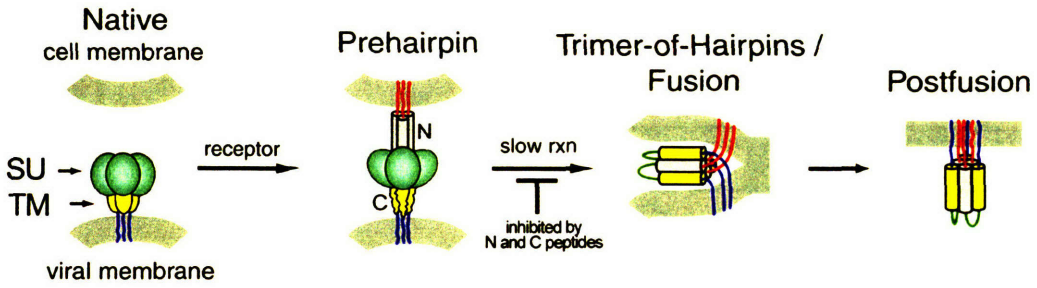


Figure 1. Model for retroviral membrane fusion and its inhibition

Prior to exposure to the cellular receptor, the viral envelope protein is in its native state. Binding of the viral SU subunit to the host receptor triggers major conformational changes in the TM subunit causing the N-terminal fusion peptide to be exposed and inserted into the target membrane, forming a transient prehairpin intermediate. The TM subunit N peptides, which form a trimeric coiled coil, also get exposed and the C peptides pack into the hydrophobic grooves on the N core in an antiparallel orientation. The prehairpin N and C peptide regions are vulnerable to inhibition by the addition of exogenous N and C peptides in a dominant negative manner. In the absence of inhibitors, the intermediate resolves into the fusion-active trimer of hairpins structure, brining the two membranes together and allowing fusion to occur. Figure adapted from [10].

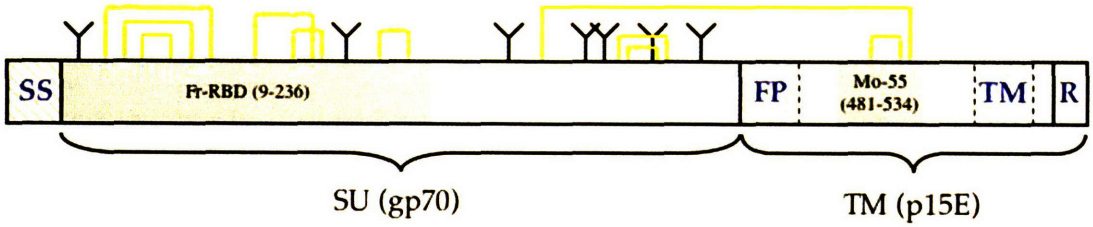


Figure 2. Schematic representation of the Mo-MLV full-length envelope protein

Mo-MLV is synthesized as a precursor protein, which undergoes post-translational modifications and processing: The signal sequence (SS) is cleaved off after insertion into the endoplasmic reticulum; seven N-linked oligosaccharides are added to the SU subunit (indicated by Y); a cleavage between the SU and TM subunits exposes the fusion peptide (FP); the transmembrane region (TM) anchors the protein to the lipid bilayer; and the R-peptide is removed during or shortly after budding. Disulfide bonds are shown with yellow lines, hydrophobic regions are shown with diagonal lines, and the shaded areas indicate regions of the MLVs for which X-ray crystal structures are available.

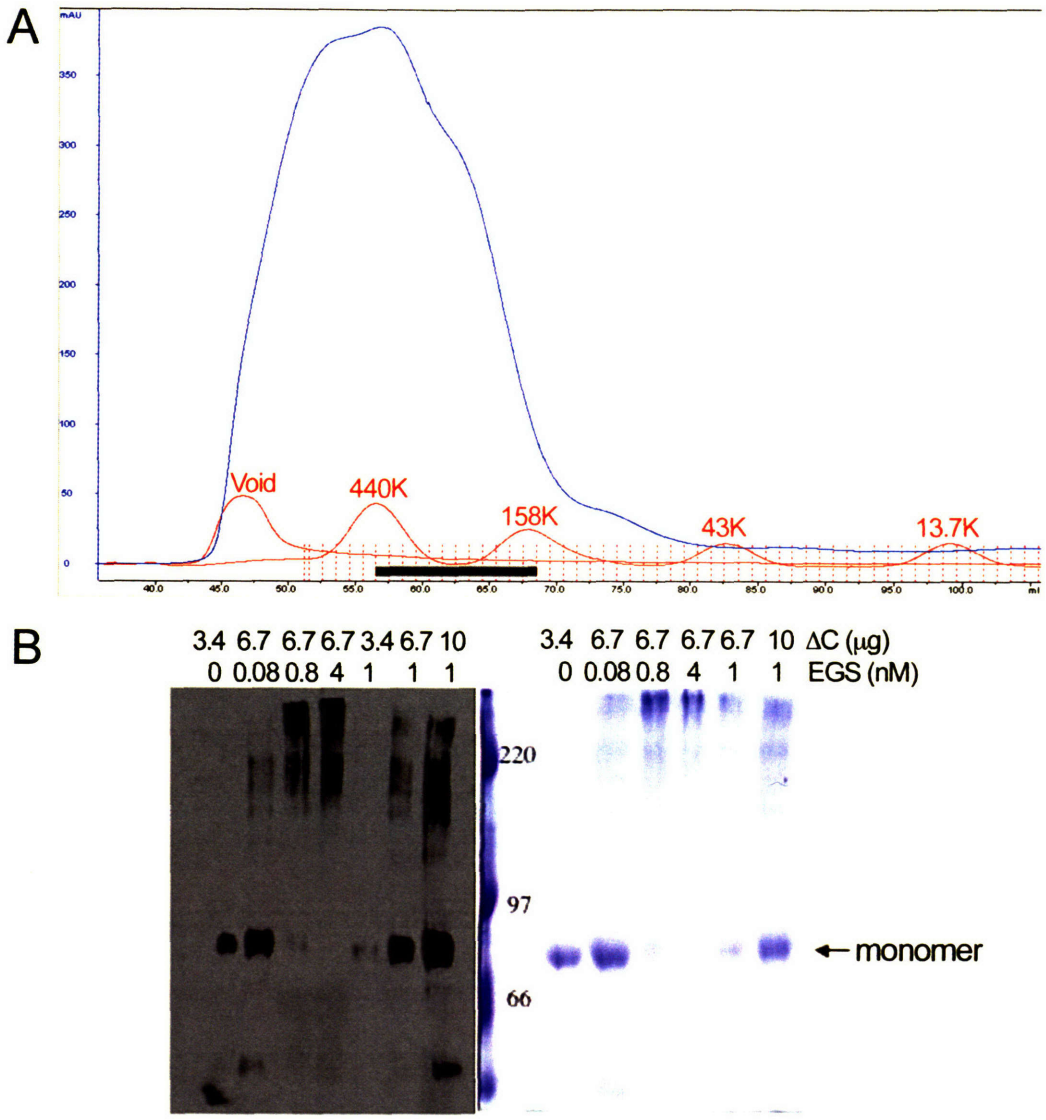
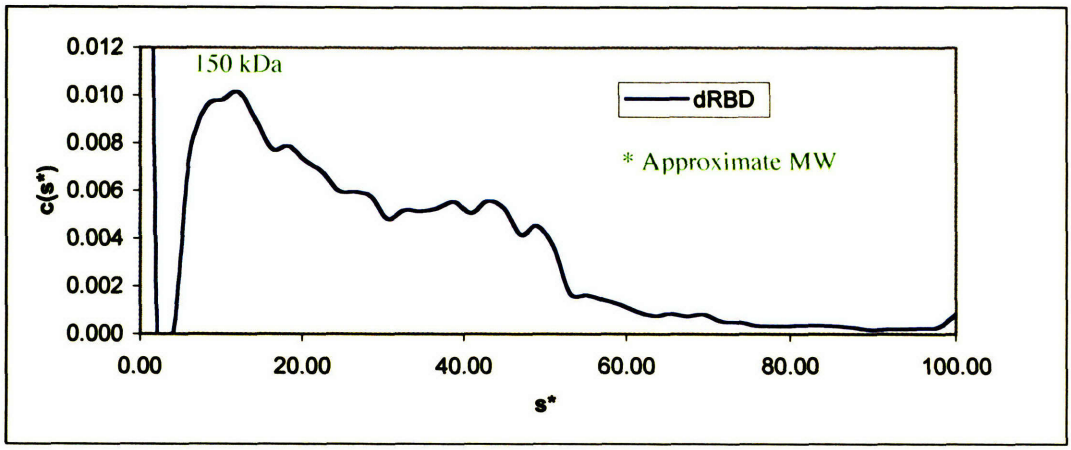


Figure 3. Gel-filtration and chemical cross-linking of Mo-MLV ΔC

(A) Gel filtration profile of ΔC (blue line) on a Superdex 200 16/60 column. Fractions that were pooled and used for characterization are marked with a black line. Molecular weight standards are shown in red.

(B) Chemical cross-linking of ΔC with various amounts of EGS as indicated by numbers above the lanes. Western blot probed with PentaHis antibody (Qiagen) is shown on the left and coomassie stained gel is shown on the right.

A



B

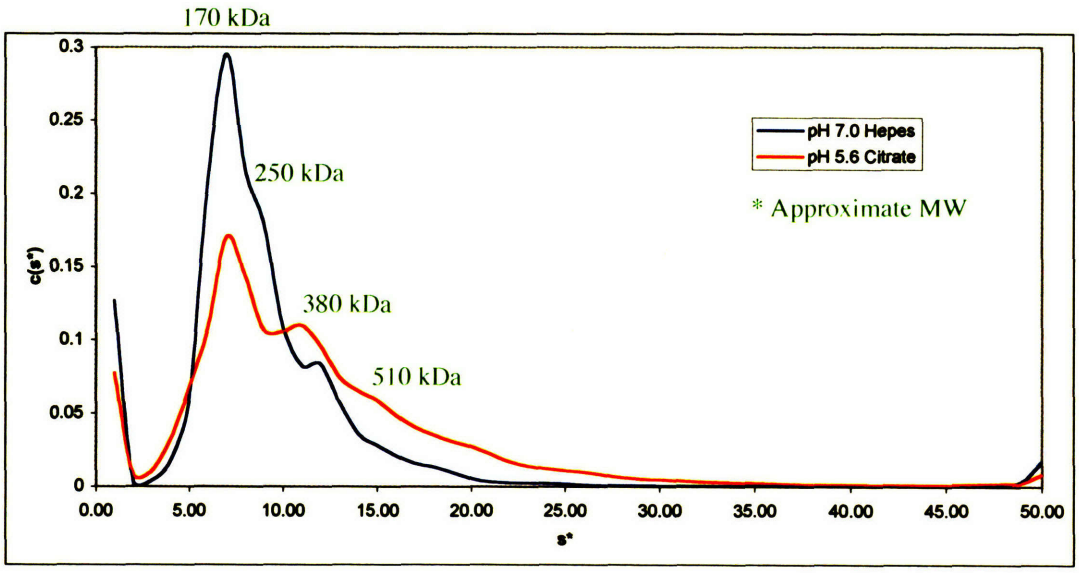


Figure 4. Sedimentation velocity analytical ultracentrifugation of Mo-MLV ΔRBD and ΔC

(A) Sedimentation velocity analytical ultracentrifugation of ΔRBD was performed. Data were collected at 20 °C at 36,000 rpm and fit with SEDFIT [36].

(B) Sedimentation velocity analytical ultracentrifugation of ΔC at pH 5.6 and pH 7.0 are shown. Data were collected at 20 °C at 30,000 rpm and fit with SEDFIT [36].

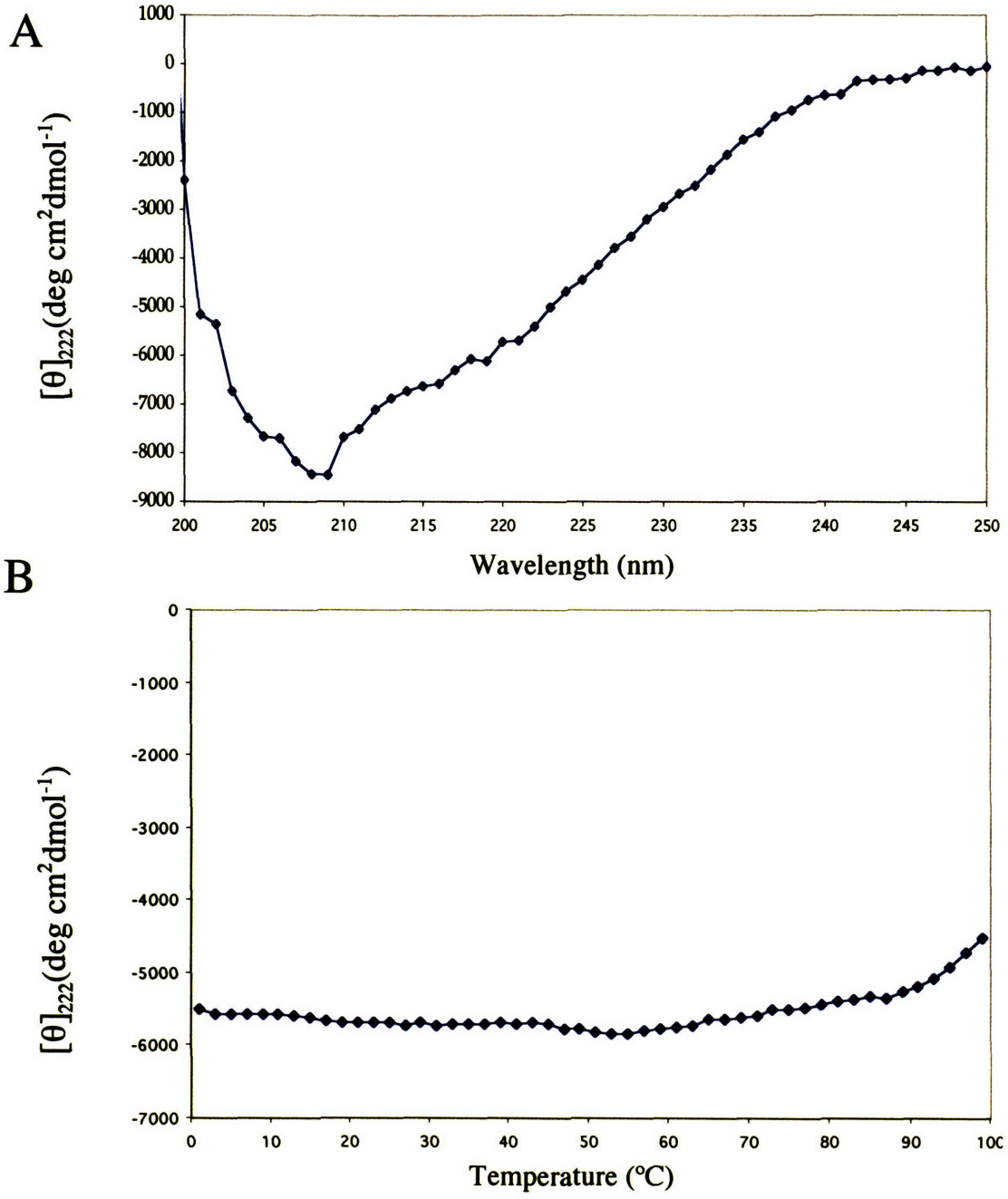


Figure 5. Circular dichroism analyses and thermal denaturation of Mo-MLV ΔC

(A) CD spectrum of a 4.5 μM solution of ΔC in HBS was collected at 25 °C. A baseline spectrum recorded from buffer alone was subtracted from the signal.

(B) Thermal melt of ΔC (4.5 μM in HBS) is shown. CD signal was measured at 220 nm in 2 °C increments from 1°C to 99 °C.

Figure 6. Viral infectivity and inhibition

(A) Schematic representation of infectivity and inhibition assays. Human 293T cells were cotransfected with pCL-Eco and pMX-IRES-GFP to generate replication deficient virions, Mo-GFP. NIH/3T3 mouse cells were infected with Mo-GFP in the absence or presence of inhibitor. Infected cells express GFP and can be analyzed by FACS.

(B) Schematic representation of the Mo-MLV TM subunit. Mo-36 (equivalent of the N peptide) and the Mo-MLV C peptide are shown with the sequences specified.

(C) Infection of mouse NIH/3T3 cells with Mo-GFP. Percentage of infected cells is plotted against the volume of virus added. Blue box indicates the amount of virus added to the viral infection inhibition experiments in part (D), based on the viral titer.

(D) Results from individual viral infection inhibition experiments. Increasing concentrations of Mo-MLV C peptide (blue) or an α-Mo-92 antibody (red) were added to NIH/3T3 cells at the time of infection with 4 μL Mo-GFP based on the titer of the virus in part (C). Infection was detected by FACS, monitoring for GFP expression. The percentage of infected cells is plotted against the concentration of inhibitor added.

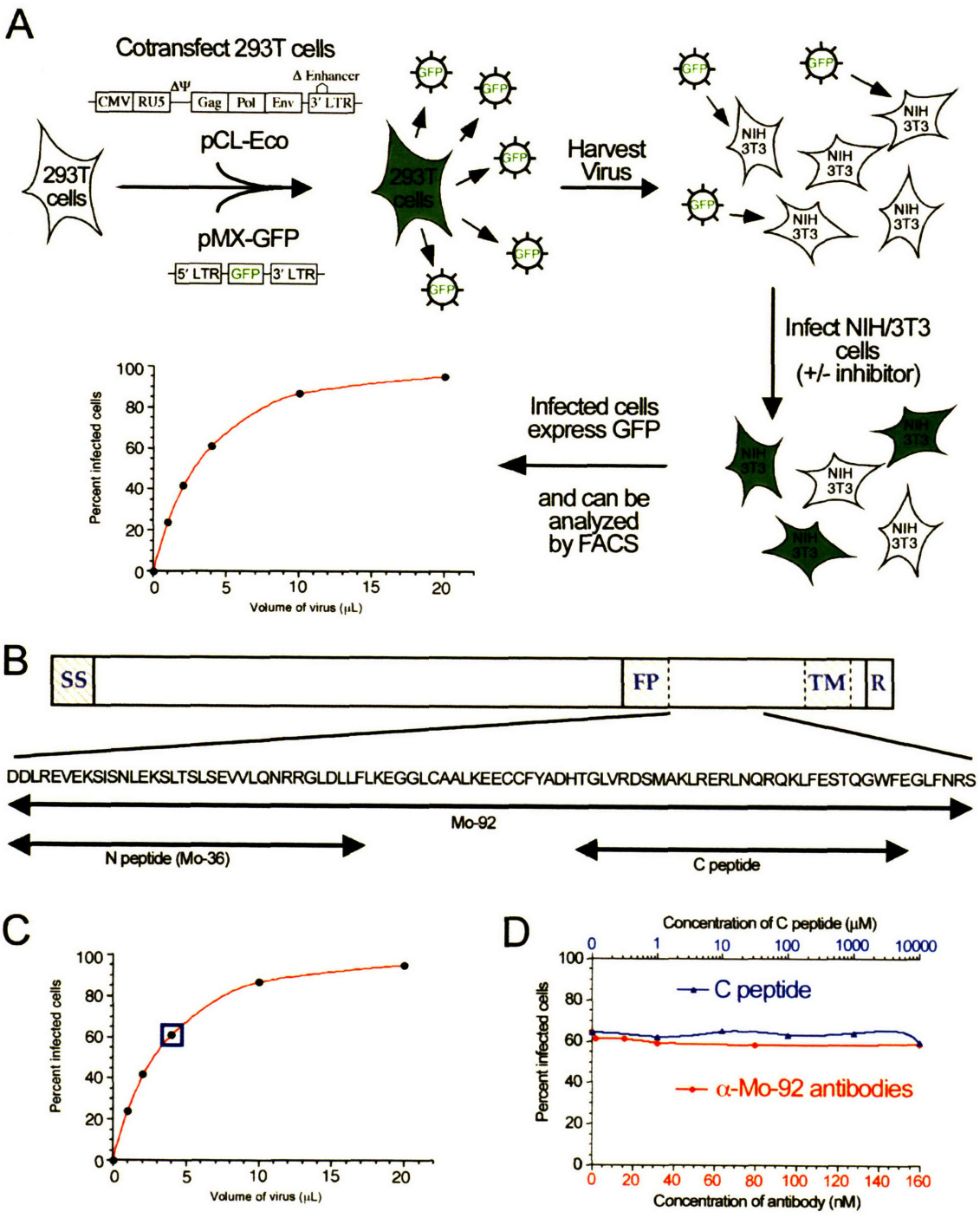


Figure 6. Viral infectivity and inhibition

APPENDIX II

Crystal Structure of a Secreted Insect Ferritin Reveals a Symmetrical Arrangement of Heavy and Light Chains

Appendix II has been submitted for publication. This project started when yellow-tinted ferritin crystals grew during crystallization trials of the Moloney Murine Leukemia virus envelope glycoprotein (Appendix I). My contributions to this work include "purification" and crystallization of ferritin, identification of the two subunits, isolation of the heavy and light chain genes, collecting the first data set and writing part of the manuscript. Anthony West Jr. helped me find the molecular replacement solution, did the structure refinement, collected the high-resolution data set and wrote the manuscript. Zsuzsa Hamburger taught me how to handle and freeze the crystals and we collected the initial data set together. Peter Hamburger contributed to the mathematical discussion.

Crystal Structure of a Secreted Insect Ferritin Reveals a Symmetrical Arrangement of Heavy and Light Chains

Agnes E. Hamburger[1,3,5], Anthony P. West, Jr.[1,5], Zsuzsa A. Hamburger[1,6], Peter Hamburger[4], and Pamela J. Bjorkman[1,2,*]

[1]Division of Biology 114-96
[2]Howard Hughes Medical Institute
California Institute of Technology
Pasadena, California 91125

[3]Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142

[4]Department of Mathematical Sciences
Indiana University - Purdue University Fort Wayne
Fort Wayne, Indiana 46805

[5]These authors contributed equally to this work.

[6]Present address: Department of Structural Biology
Stanford University School of Medicine
Stanford, CA 94305

*Correspondence: Tel: 626-395-8350; Fax: 626-792-3683; E-mail: bjorkman@caltech.edu.

Running Title: Structure of insect ferritin, a prettier shell

Summary

Ferritins are iron storage proteins made of 24 subunits forming a hollow spherical shell. Vertebrate ferritins contain varying ratios of H and L chains, however, known ferritin structures include only one type of chain and have octahedral symmetry. Here we report the 1.9 Å structure of a secreted insect ferritin from *Trichoplusia ni*, which reveals equal numbers of H and L chains arranged with tetrahedral symmetry. The H/L-chain interface includes complementary features responsible for ordered assembly of the subunits. The H chain contains a ferroxidase active site resembling that of vertebrate H chains with an endogenous, bound iron atom. The L chain lacks the residues that form a putative iron core nucleation site in vertebrate L chains. Instead, a possible nucleation site is observed at the L chain 3-fold pore. The structure also reveals inter- and intra-subunit disulfide bonds, mostly in the extended N-terminal regions unique to insect ferritins. The symmetrical arrangement of H and L chains and the disulfide crosslinks reflect adaptations of insect ferritin to its role as a secreted protein.

Introduction

Iron is an essential element in biological systems whose levels are strongly regulated to ensure availability but prevent toxicity (Arosio and Levi, 2002). Key to this regulation are the two stable oxidation states of iron. In the ferrous (Fe^{2+}) state, iron is soluble and readily used, but it is quickly oxidized under aerobic conditions, potentially producing harmful reactive species. The ferric (Fe^{3+}) ion resulting from oxidation has extremely low solubility under physiological conditions, rendering it inaccessible for incorporation into metalloproteins. Ferritin, which functions as a ferroxidase to control the transition between hydrated Fe^{2+} and insoluble Fe^{3+} and provides a site for storage of Fe^{3+}, represents a solution to this chemical dilemma that is utilized by bacteria, fungi, plants, insects, and vertebrates (Harrison and Arosio, 1996).

Ferritin functions as a spherical assembly of one or more types of individual subunits. In most cases, 24 subunits come together to form a hollow sphere with octahedral (*O* or 432) symmetry (Arosio and Levi, 2002). Up to 4500 Fe^{3+} atoms can be stored in a mineralized form inside the protein shell. Mammalian ferritins contain a variable mixture of two types of subunits: heavy (H) and light (L) chains. The H subunit contains a ferroxidase center and the L chain contains a site for nucleation of the mineral core. Ferritins isolated from different tissues contain varying and characteristic ratios of H and L chains, thought to reflect tuning of ferritin function for specific organs (Arosio and Levi, 2002).

An unusual property of ferritin is the existence of functional forms in both cytoplasmic and vacuolar/extracellular environments. In bacteria and vertebrates, ferritin is predominantly a cytoplasmic protein. However, a secreted form of ferritin is found in mammals, and serum ferritin is a clinically important measure of iron stores (Harrison and Arosio, 1996). The mechanism by which mammalian ferritin is secreted is unclear, since no conventional

hydrophobic leader peptide sequence is present. A recent study provides evidence that cytosolic ferritin can be directed to the secretory pathway through a regulated process (Ghosh et al., 2004). Insect ferritins are primarily extracellular or vacuolar, and their sequences include recognizable hydrophobic leader peptides (Nichol et al., 2002). An interesting contrast in iron regulation between mammals and insects is provided by the roles of ferritin and transferrin. In mammals, transferrin is the primary iron transport protein, and serum ferritin is an acute phase protein whose concentrations increase dramatically during infection. In insects, secreted ferritin is a major iron storage/transport molecule, while transferrin is strongly upregulated by infection (Nichol et al., 2002).

Here we report the structure of a secreted ferritin from the cabbage looper/tiger moth *Trichoplusia ni* (*T. ni*), which is the first structure of a secreted ferritin, and the first structure in which two different subunit types are resolved. *T. ni* ferritin contains an equal number of H and L chains arranged with tetrahedral (*T* or 32) symmetry, as compared with previous bacterial and vertebrate ferritin structures, which have octahedral (*O* or 432) symmetry and only a single type of resolved subunit. The tetrahedral symmetry arrangement of *T. ni* ferritin creates two distinct types of 3-fold axis pores. The 3-fold pore made by the L chain subunits contains several coordinated ions. The innermost of these, modeled as Fe^{3+}, is a potential site for nucleation of the iron core. The pore created by three H chain subunits is relatively open, providing for passage of ions in and out of the ferritin cavity. The H chain ferroxidase sites of *T. ni* ferritin each contain a single Fe^{3+} ion. The structure also reveals both intra- and inter-subunit disulfide bonds, consistent with insect ferritin being a secreted protein.

Results and Discussion

Isolation of Insect Ferritin Protein and Genes

Secreted ferritin was isolated from the supernatants of baculovirus-infected *T. ni* cells as a contaminant during purification of an unrelated recombinant 6xHis-tagged protein on a Ni-agarose column. The affinity of ferritin for the Ni-agarose may reflect binding of the Ni-NTA to an iron-binding site. Crystallization trials of the recombinant protein yielded yellow-tinted crystals in several different conditions, both with and without added divalent ions. The largest crystals, which grew in 20% Jeffamine, 0.1 M Hepes (pH 7.5), were in space group C2 (a = 203.8 Å, b = 146.6 Å, c = 206.9 Å, β= 92.7°) with one complete ferritin shell per asymmetric unit. Analysis of the crystals by SDS-PAGE under reducing conditions revealed the presence of two bands of equal intensity migrating at 22 and 27 kDa rather than the 45 kDa expected for the recombinant protein. N-terminal sequencing of the two bands and a BLAST search revealed sequence similarity to insect ferritin H and L chains, but the sequence of the *T. ni* ferritin L chain had not been previously determined and only a partial EST was available for the H chain. Using the N-terminal protein sequence information, degenerate oligonucleotides were designed, and the genes for the two subunits were isolated to obtain the sequence of the entire protein for use in the crystallographic structure determination.

The derived amino acid sequences for the two *T. ni* ferritin chains are shown in Supplementary Figures 1A-1B. The smaller of the two chains (191 residues not including the hydrophobic leader peptide) includes seven conserved residues that define the ferroxidase center (Lawson et al., 1989) that are not present in the other chain (212 residues not including the hydrophobic leader peptide). By analogy with vertebrate ferritin subunits, we refer to the insect subunit containing the ferroxidase site as the H chain, although it is the smaller of the two *T. ni*

subunits, and the subunit lacking the ferroxidase site as the L chain. Several other nomenclatures have been used to designate insect ferritin subunits. The subunit containing the ferroxidase site has been called the heavy chain homolog (HCH), the S (small) subunit, FER1, or Fersub1 (Nichol et al., 2002). The other subunit is known as the light chain homolog (LCH), the G (glycosylated) subunit, Fer2LCH, or Fersub2. LCH is a misleading name for the non-ferroxidase subunit because it is not more similar to vertebrate L than to vertebrate H chains (Nichol et al., 2002).

The *T. ni* H chain shares a high level of sequence identity with previously reported Lepidoptera sequences: *Manduca sexta* (84%), *Galleria mellonella* (83%), and *Calpodes ethlius* (77%). The mature L chain sequence includes a potential N-linked glycosylation site (Asn115) observed in other Lepidoptera L chains. The L chain sequence identity with other butterfly and moth L chain sequences is also high: *Manduca sexta* (80%), *Galleria mellonella* (78%) and *Calpodes ethlius* (70%). The *T. ni* H and L chains are related by only 22% sequence identity. While the *T. ni* H chain has 30% sequence identity with the human H chain, the *T. ni* L chain has low identity with both vertebrate H and L chains (20% with the human H and 17% with the horse L chain).

Structure Determination

The structure of *T. ni* ferritin was solved at 3.1 Å by molecular replacement using the bullfrog M ferritin structure (PDB code 1MFR) (Ha et al., 1999) as a search model. Since the search model contained 24 copies of one chain type, it was initially unclear whether the H and L chains in the *T. ni* structure were randomly distributed or symmetrically arranged. Therefore, we repeated the molecular replacement using a search model with tetrahedral symmetry (constructed from 1MFR) containing conserved H chain residues, which yielded a structure with 12 copies of two

apparently distinct chains (Figures 1A-1B). Confirmation of this symmetric arrangement was provided by an anomalous difference Fourier map (Figure 1C), which revealed 12 strong peaks in the H chain ferroxidase sites but no significant peaks in the corresponding regions of the L chains. A 1.9 Å native data set was later collected from crystals grown in 20 mM Tris (pH 8.0), 150 mM NaCl at 4°C (Table 1) and used for subsequent refinement.

H and L Chains are Packed in a Symmetrical Arrangement

Although vertebrate ferritins are composed of a mixture of H and L subunits, structures have been determined only for recombinant ferritins made of a single chain type (amphibian H and L, human H, horse L, and mouse L) (Granier et al., 2001; Ha et al., 1999; Hempstead et al., 1997; Trikha et al., 1994), and for native ferritins modeled with a single chain type representing the predominant chain (horse spleen ferritin (Granier, 1997)). Ferritins containing a single type of subunit have octahedral (432) symmetry. When 24-mer shells are assembled from two different chain types, the highest possible symmetry is tetrahedral, as seen in the *T. ni* ferritin crystal structure. Another ferritin structure with tetrahedral symmetry was found for a dodecameric ferritin from *Listeria innocua* (Ilari et al., 2000), although this ferritin includes 12 copies of a single subunit (rather than 12 copies of a heterodimer), which form a spherical molecule with an inner cavity capable of storing only 500 iron atoms (Bozzi et al., 1997) as compared with 4500 for conventional ferritins.

Octahedral symmetry ferritin shells contain three 4-fold symmetry axes, eight 3-fold axes, and six 2-fold axes (not counting 2-fold axes along the 4-fold axes). A consequence of the tetrahedral symmetry of the *T. ni* structure is that the 2-fold axes, about which homodimers of ferritin subunits are arranged in an octahedral structure, are lost because the 2-fold symmetric homodimers are replaced by heterodimers of H and L chains. Likewise, the exact 4-fold axes in

octahedrally-symmetric ferritins become pseudo 4-folds or true 2-fold axes in tetrahedrally-symmetric ferritin. The 3-fold axes remain, although there are two distinct local 3-folds: one made of three H chains and one made of three L chains.

Folding Topology of Insect Ferritin H and L Chains

The overall structures of insect ferritin H and L subunits closely resemble the structures of mammalian ferritin subunits – each contains a core of four helices (A, B, C, and D) and a fifth helix (E) that is involved in intersubunit interactions near the pseudo 4-fold axis (Figure 2A). A notable difference from mammalian ferritins is an approximate 10 or 20 residue N-terminal extension in the *T. ni* H and L subunits, respectively. These extensions, which are unique to secreted insect ferritins (Nichol et al., 2002), form extended loops that bridge adjacent subunits on the outside of the ferritin shell (Figure 2A). Excluding the N-terminal extensions, the *T. ni* ferritin H and L chains are structurally very similar to mammalian ferritins: the root mean square (rms) deviation of the H chain Cα positions from those of the human H chain (PDB code 2FHA) is 1.08 Å (calculated for 160 residues) (Figure 2B). Other comparisons reveal comparable degrees of similarity: rms deviations for the following pairs are all close to 1 Å: insect L chain versus human H chain (1.11 Å for 159 Cα positions), insect H chain versus insect L chain (1.27 Å for 156 Cα positions) (Figure 2C), insect L chain versus horse L chain (PDB code 1AEW) (1.08 Å for 159 Cα positions). The largest difference between the *T. ni* ferritin chains and previously reported ferritin structures is in the L chain loop connecting helix B to helix C (residues 99 to 121). In mammalian ferritin structures, short segments of the BC loops from adjacent subunits form antiparallel, hydrogen bonded β-strands. In *T. ni* ferritin, the L chain BC loop, which is longer than its mammalian counterpart, extends away from the subunit-subunit interface, with the side chain of Tyr116 inserted into a hydrophobic region between helices A

and C. This loop is relatively disordered, perhaps due to the N-linked glycosylation site at Asn115 (although no ordered carbohydrate is observed in the crystal structure).

The H Chain Ferroxidase Site Contains a Bound Iron Atom

The initial difference Fourier and anomalous difference Fourier maps revealed that a dense ion was bound at the H chain ferroxidase site in the *T. ni* ferritin (Figure 3). The ferroxidase center is located within the hydrophilic core of the four-helix bundle of the H subunit and resembles those of vertebrate ferritins. Divalent ions (mainly Cd^{+2} or Mg^{+2}), which compete for the iron-binding site, have been critical in the crystallization solutions for vertebrate ferritins. Hence, the only ferritin structures containing native Fe^{3+} are of bacterial origin: the bacterioferritin from *Desulfovibrio desulfuricans* (Coelho et al., 2001; Macedo et al., 2003) and the unusual dodecomeric ferritin from *Listeria innocua* (Ilari et al., 2000). Because *T. ni* ferritin crystallized in the absence of added divalent ions and showed a very strong anomalous signal (over 35σ in the 12-fold NCS-averaged map) at this site, we have modeled the ion as Fe^{3+}. The ion-ligand distances (Figure 3) are similar to those seen in other ferritin ferroxidase sites (Macedo et al., 2003; Stillman et al., 2001). The low B value of the ion (19 $Å^2$ compared with 14-22 $Å^2$ for 29 Cα atoms within approximately 10 Å of the ion) supports the Fe^{3+} assignment. The Fe^{3+} ion is coordinated in an approximately octahedral geometry by Glu32, Glu67, His70 and two water molecules with one coordination site unoccupied (Figure 3). In the ferroxidase centers of previous ferritin structures, two metal ion-binding sites (called A and B) were observed (Langlois d'Estaintot et al., 2004). The conserved residues that normally coordinate the two metals in mammalian ferritins (Hempstead et al., 1997) are also conserved in insects (Supplementary Figure 1A). The *T. ni* ferritin contains one metal ion at site A as observed in previous ferritin structures that contain a single bound ion (Langlois d'Estaintot et al., 2004).

Potential coordinating residues at the B site include Glu67 (shared with site A), Gln153, and Glu116. Although site B is unoccupied in the 1.9 Å structure, a weak anomalous signal was seen at this site in the 3.1 Å resolution structure and in data sets from crystals that were soaked in ferrous ammonium sulfate prior to cryopreservation (data not shown).

Inter- and Intra-Subunit Disulfide Bonds

In common with other secreted proteins but in contrast to cytoplasmic ferritins, the *T. ni* H and L chains both contain intrasubunit disulfide bonds. In the H chain, a disulfide bond between residues 21 and 130 connects the N-terminal region of helix A to the C-terminus of helix C. In the L chain, a disulfide bond between residues 4 and 24 contributes to forming the relatively compact N-terminal region of this subunit (Figure 2A). This domain is on the outside of the ferritin shell, and it makes contact with an L chain related by a 3-fold. Specifically, the helical portion on this domain contacts the N-terminal region of the adjacent subunit's C helix.

The *T. ni* ferritin also includes intersubunit disulfide bonds (Figure 2A) connecting H chain residue 3 to residue 12 of a non-adjacent L chain (for example, subunits I and VIII of Figure 1B). Although intersubunit disulfide bonds have not been observed in previous ferritin structures, such crosslinking has been suggested to occur in artemin, a ferritin homolog in brine shrimp (Chen et al., 2003). Confirmation of the crystallographically-observed interchain disulfide bond is provided by SDS-PAGE analysis of *T. ni* ferritin: a single band of apparent molecular weight ~50 kDa is observed under non-reducing conditions as compared with two bands of molecular weights of 22 kDa and 27 kDa under reducing conditions (data not shown). With the exception of the ferritin sequences of *Anopheles gambiae* and *Aedes aegypti*, which have divergent N termini, the three cysteines in the N-terminal region of the L chain and the cysteines at positions 3, 21, and 130 in the H chain are conserved among insect ferritins (Nichol

et al., 2002) (Supplementary Figures 1A-1B), suggesting that the pattern of disulfide bonds observed in the *T. ni* ferritin structure is common to secreted insect ferritins. Others have speculated that fatty acylation of the conserved cysteines in the N termini allows ferritin to be retained in the endoplasmic reticulum (ER) and in vacuolar compartments (Nichol and Locke, 1999), which are major sites of localization of insect ferritin (Locke, 2003)). Although the *T. ni* structure does not show evidence of fatty acylation, an alternative hypothesis is that the "knobs" formed by the N-terminal extensions on the surface of insect ferritin bind an ER-resident receptor.

The Major H and L Chain Interface and Subunit Assembly

Assembly of ferritin shells with octahedral symmetry is thought to begin with the formation of homodimers of ferritin subunits (Ford et al., 1984; Hempstead et al., 1997). The corresponding dimer for insect ferritin is the H/L heterodimer (molecules I and II from Figure 1B). If subunits first associate at the I-II interface to exclusively form H/L heterodimers, and 12 dimers assemble, the ferritin shell will contain a 1:1 ratio of H:L subunits. The *T. ni* ferritin structure suggests a molecular basis for the specificity of heterodimer versus homodimer assembly. Examination of the I/II interface suggests that two tryptophan residues contribute to the specificity of heterodimer assembly (Figure 4). L chain Trp84 is located near the pseudo-2-fold axis relating the H and L chains at the "heterogen" binding site (where bacterioferritins (Coelho et al., 2001) and horse spleen ferritin (Michaux et al., 1996) bind hemes). The L chain Trp84 position is usually an arginine in mammalian ferritins and a methionine in bacterioferritin, and both of these residues play a role in heme binding (Frolow et al., 1994; Precigoux, 1994). The corresponding residue on the *T. ni* H chain is Arg68, which makes cation-pi interactions (Gallivan and Dougherty, 1999) with L chain residues Trp84 and Tyr53 and forms a salt bridge with L chain

Asp81. The arginine at this position in mammalian ferritin usually forms an intra-helix salt bridge with a glutamate at position *i*+4, but the L chain *i*+4 residue is an isoleucine in *T. ni* and most other insects (Supplementary Figure 1B). Hence, the L chain intrasubunit salt bridge in mammalian ferritins is replaced by an intersubunit salt bridge in insect ferritins. The second tryptophan at the I/II interface of *T. ni* ferritin is H chain residue Trp14. Together with H chain Ile10 and Ile15, this tryptophan forms a hydrophobic binding pocket for L chain Tyr65. Computer models of insect H/H and L/L homodimers reveal that the described interactions around both tryptophans are possible only in the heterodimeric H/L structure (data not shown), contributing to the preferential formation of heterodimers from a mixture of H and L chains. The conservation of these key residues (Supplementary Figures 1A-1B), in addition to the cysteines involved in intersubunit disulfide bonds, suggests that *T* symmetry ferritin is common in the lepidopteran, coleopteran, hemipteran, and, to a lesser extent, dipteran orders.

Although stepwise assembly beginning with H/L heterodimers assures a 12:12 mixture of H and L chains, this is not enough to ensure tetrahedral symmetry because there are 186 distinct shells that can be formed from 12 heterodimers (Appendix). To obtain a tetrahedral symmetry shell, additional specificity must occur at either the 3-fold subunit-subunit interfaces (to create two distinct 3-folds), at the pseudo 4-fold subunit-subunit interfaces, or at both. The significant differences between the H and L chain 3-fold pores (see below) suggest that assembly may be more favorable with three identical chains. Also, around the pseudo 4-fold axis a salt bridge between adjacent H and L chains involving His174 (H) and Asp196 (L) may contribute to a preference for alternating H and L subunits rather than forming a 4-fold with identical subunits.

The 3-fold and Pseudo 4-fold Symmetry Axes

The hydrophilic channels located along the 3-fold axes of ferritins are believed to be the major site of iron entrance into the central cavity (Arosio and Levi, 2002). The *T. ni* ferritin structure contains two significantly different types of 3-fold pores. The pores formed from H chains are relatively open and contain a number of ordered water molecules, while the pores formed by adjacent L chains contain several coordinated ions. Three peaks in the electron density map are observed along the L chain 3-fold pore (Figure 5A). The middle peak is the strongest, and the corresponding ion is fully coordinated by the side chains of Gln161 and Glu164 from the three adjacent L chains. Previous ferritin structures have also contained ions along the 3-fold pores, although these ions were partially or fully hydrated (Granier et al., 2003; Langlois d'Estaintot et al., 2004). The outermost peak is a hydrated ion coordinated by Glu164 and potentially His144. The innermost peak is a partially hydrated ion and is coordinated by the side chains of the three symmetry-related Glu165 residues in the 3-fold pore.

Along the pseudo 4-fold axis, the E helices of two H and two L chains form a hydrophobic interface without any large cavities or bound ions. Beneath the polar surface residues (H chain His174 and L chain Asp196) at the pseudo 4-fold, two hydrophobic layers consisting of L chain Leu199 and H chain Leu177 and, more inward, L chain Val203 and H chain Ile181, make it unlikely for ions to permeate along this axis.

Possible L Chain Ferrihydrite Nucleation Site at the 3-fold Pore

L chain is thought to assist formation of the iron core by providing a nucleation site on the inside of the ferritin shell. For mammalian L chains this nucleation site has been proposed to consist of a cluster of four glutamates directed into the cavity (Chasteen and Harrison, 1999). A high-resolution structure of mouse L chain ferritin has revealed details of this site: four cadmium ions

are coordinated by Glu57, Glu60, Glu61 and Glu64 (Granier et al., 2003). These glutamates are not conserved in bacterial or insect ferritins, and the corresponding residues in the *T. ni* L chain are Lys78, Asp81, Glu82, and Ser85. The side chain of Asp81 points away from the other three residues and is involved in a salt bridge to H chain residue Arg68. We observed no non-protein electron density in this region, further suggesting that these residues are unlikely to provide a ferrihydrite nucleation site.

There is a possible ferrihydrite nucleation site on the inside of the ferritin shell at the L chain 3-fold pore, where we observed an ion coordinated by Glu165 (Figures 5A-5B). We modeled this peak as an Fe^{3+} ion because it showed a strong anomalous peak in the original 3.1 Å resolution data set, while we assigned the two outer peaks as Ca^{2+} ions due to their weak anomalous signals. Additional density in F_o-F_c and anomalous difference Fourier maps around the Fe^{3+} ion suggest the presence of other metal ions and/or water molecules (data not shown), however the density is not sufficiently strong to model more ions. Ordered water molecules and a small cluster of metal ions were also observed at the analogous site in the horse L chain structure, although in this case the ions were thought to derive from the crystallization solution (Hempstead et al., 1997). In addition to the presence of metal ions near the *T. ni* L chain 3-fold pore, the region surrounding the pore seems particularly suited for serving as the ferrihydrite nucleation site. The surface around the pore is highly charged: the inner surface within 12 Å of the 3-fold axis is covered by 12 negatively-charged residues (Glu165, Glu166, Glu169 and Asp170 from each of the three L chains) (Figure 5B).

Mathematical Basis of Tetrahedral Symmetry

An equal number of H and L chains can be assembled into ferritin shells in 113,048 distinct configurations (Appendix). We consider two questions: what is unique about the *T* symmetry

arrangement found in *T. ni* ferritin, and is this the only conceivable symmetrical arrangement? Two types of subunits could potentially be assembled into ferritin shells with point groups T, D_4 (requiring a 16:8 subunit ratio), D_3, D_2, C_4, C_3, C_2 or C_1. Besides T being the largest proper subgroup of O, T symmetry has another unique property: only with the tetrahedral arrangement will each of the 12 subunits (of both types) always be in the same environment. A closely related property is that the T configuration is the only perfect 2-coloring of a polyhedron representing ferritin (see Appendix). A coloring of a polyhedron is 'perfect' if each symmetry operation of the underlying polyhedron causes a permutation of the colors. The molecular consequence is that this is the only arrangement in which each interface surface of each subunit will always contact one type of subunit (*e.g.*, the I-II interface is always heteromeric and the I-III interface is always homomeric). For all other arrangements, there will be at least one class of interface (*e.g.*, the I-II interface of H chains) that joins subunits of the same type at one position of the ferritin shell while joining different types elsewhere.

Conclusions

The structure of the *T. ni* secreted ferritin reveals, for the first time in a ferritin structure, the molecular basis by which two different subunits are assembled into a ferritin shell. The *T. ni* structure contains 12 H and 12 L chains arranged with tetrahedral symmetry. An advantage of ferritin with two subunits arranged with tetrahedral symmetry, versus two subunits arranged randomly, could be a greater tendency to crystallize. For example, we observed that *T. ni* ferritin formed crystals readily – during storage in a saline buffer and as a minor contaminant during crystallization trials of an unrelated protein. The tendency of native *T. ni* ferritin to crystallize may reflect a general property of secreted insect ferritin. Crystals of insect ferritin have been observed *in vivo* (Locke and Leung, 1984), and crystallization appears to assist in the formation

of protein storage granules in fat bodies of insects (Locke, 2003). For example, crystals of secreted ferritin are found in the storage granules of the premetamorphic fat body of *Calpodes ethlius* (Larsen, 1976).

The existence of homopolymeric ferritins in bacteria and the structural similarity of H and L chains in eukaryotes strongly suggest that the earliest ferritins were homomeric. The presence of heteromeric ferritins in animals suggests that there is an adaptive value in altering ferritin function by mixing two types of subunits. The random mixture of H and L subunits in vertebrate ferritin compared to the symmetrical arrangement in insect ferritin may be a reflection of the localization of these proteins. Vertebrate ferritin is principally cytoplasmic, and the ability to vary the ferritin H/L ratio in different cell types permits ferritin function to be tuned to different tissues (Arosio and Levi, 2002). For secreted insect ferritins that enter a common extracellular space, this regulatory mechanism would not be applicable. What might be the adaptive value of a non-random arrangement of subunits? If a particular subunit arrangement creates a favorable feature (e.g., a specialized pore), then a symmetric configuration would ensure that all assembled shells contained the maximum number of these features. As an example of an activity that depends on the specific arrangement of subunits, one study showed that the redox activity of recombinant human ferritin heteropolymers reached a maximum in molecules that contained a 1:1 ratio of H:L chains, which was suggested to be a result of specific interactions at the H/L interfaces (Johnson et al., 1999). In addition to the distinct H and L 3-fold pores, the tetrahedral arrangement in *T. ni* ferritin also permits all of the subunits to be involved in intersubunit disulfide bonds increasing the stability of the ferritin shell. Further studies will be required to fully understand how the structural features of insect ferritin relate to its function as a secreted protein.

Experimental Procedures

Purification and Identification of *T. ni* Ferritin

Secreted ferritin was harvested from supernatants of baculovirus-infected *T. ni* (Tn5-B1-4, High Five) cells, exchanged into 20 mM Tris (pH 8.0), 0.3 M NaCl and subjected to Ni-NTA chromatography (Ni-NTA Superflow, Qiagen). Protein from a 250 mM imidazole elution was further purified by gel filtration chromatography on a Superdex 200 column (Amersham Biosciences). The H and L chains were resolved by SDS-PAGE and transferred to PVDF membrane. The bands were excised and sequenced by automated Edman degradation, yielding the sequences ADTCYNDVALDC (L chain) and TQCNVNPVQIP (H chain).

Isolation of Ferritin H and L Chain cDNAs

mRNA was prepared from baculovirus-infected *T. ni* cells using the QuickPrep™ mRNA Purification Kit (Amersham Biosciences). cDNA was generated from the mRNA with murine reverse transcriptase and random hexadeoxyribonucleotides (Amersham Biosciences). Degenerate 5' primers were designed based on the N-terminal protein sequences and a 3' primer was designed based on a partial EST sequence of the *T. ni* H chain (accession code CF258131). A 3' poly dT oligonucleotide was used for the L chain. DNA fragments were amplified by PCR with Taq DNA polymerase (Boehringer Mannheim) and cloned into pCR®2.1- TOPO® (Invitrogen) and sequenced with M13 forward and reverse primers.

Crystallization, Data Collection, and Processing

Initial crystals (space group *C*2; one 24-mer per asymmetric unit) of "as isolated" *T. ni* ferritin were grown at 18 °C in hanging drops containing 20% v/v Jeffamine M-600, 0.1 M Hepes (pH 7.5), and flash frozen in 30% v/v Jeffamine M-600, 0.1 M Hepes (pH 7.5). Data were collected at -150°C to 3.1 Å resolution at the Stanford Synchrotron Radiation Laboratory beamline 9.1

with λ=0.95369 Å. Large crystals of *T. ni* ferritin also appeared after several months at 4 °C in the gel filtration column elution buffer (20 mM Tris [pH 8.0], 150 mM NaCl, 0.05% sodium azide). These crystals were transferred stepwise into 2.6 M sodium malonate (pH 7.0), prior to flash cooling and data were collected at -150°C to 1.9 Å resolution at the Advanced Light Source beamline 8.2.2 with λ=1.1 Å. Data were processed and scaled with DENZO and SCALEPACK (Otwinowski and Minor, 1997).

Structure Solution and Refinement

The structure was determined by molecular replacement using AMoRE (Navaza, 1994) and a polyglycine version of the 2.8 Å structure of bullfrog M(H) ferritin (PDB code 1MFR (Ha et al., 1999)) as a search model. The *R* value of the initial molecular replacement model was 49.3%. The assignment of H and L chains was determined by repeating the molecular replacement with a tetrahedral symmetry model of 1MFR (polyglycine model except the conserved H chain ferroxidase residues were retained). Rebuilding was performed using the program O (Jones et al., 1991) with maps calculated by solvent flattening, histogram matching, and 12-fold non-crystallographic symmetry (NCS) averaging using the program DM in the CCP4 suite (Collaborative Computational Project, 1994). Anisotropy and bulk solvent corrections were applied, and the model was refined with 12-fold NCS constraints and individual temperature (*B)* factors using the program CNS (Brünger et al., 1998). The NCS constraints were retained since nearly all subunit-subunit contacts should obey the NCS (except those at crystal contacts). The final model (1/12 of the asymmetric unit) contains residues 1 to 212 of the L chain and 1 to 191 of the H chain, 375 water molecules, 2 Fe^{3+} ions, and 2 Ca^{2+} ions. The average B factors for the H chain, L chain, and water molecules are 25.9, 29.8, and 39.1 $Å^2$, respectively. The ferroxidase site Fe^{3+} has a B factor of 18.9 $Å^2$. The L chain pore ions were modeled as two Ca^{2+} ions (with B

factors of 18.5 and 63.9 $Å^2$) and an Fe^{3+} ion (with a B factor of 34.0 $Å^2$ and occupancy of 0.78). Buried surface areas were calculated using a 1.4 Å probe using CNS. Figures were generated with MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1999), and Raster3D (Merritt and Bacon, 1997).

Acknowledgements

We thank Peter Snow of the Caltech Protein Expression Center for supernatants of baculovirus-infected insect cells, the Caltech Protein/Peptide MicroAnalytical Laboratory for N-terminal protein sequencing, Matthew Walsh at IPFW for helpful mathematical discussions, Elizabeth Sprague for technical assistance, and members of the Bjorkman laboratory for critical reading of the manuscript. This work was supported by a career research award from the Burroughs-Wellcome Fund (A.P.W.) and the Howard Hughes Medical Institute (P.J.B.). A.E.H. was supported by the Whitehead Institute for Biomedical Research.

Appendix

Calculation of the number of distinct arrangements of 12 H and 12 L subunits

This calculation is equivalent to finding the number of ways of 2-coloring a polyhedron with 24 faces and octahedral symmetry. We have modeled ferritin as a pentagonal icositetrahedron (Figure 1B); it has previously been modeled as a "split" rhombic dodecahedron (Hempstead et al., 1997), where each face is bisected by a line parallel to one pair of its edges, in such a way that each edge meets a single bisector. These two models have identical symmetries, and are therefore interchangeable for our purposes. Two colorings are considered the same if one can be obtained from the other via some rotation of the polyhedron. This calculation can be performed using the Pólya-Burnside method of enumeration (based on an extension of the Cauchy-Frobenius lemma). The lemma, as applied in our case, states that the average number of fixed elements (colorings left unchanged by a rotation) of a group (O) acting on a set (of all possible colorings) is equal to the number of orbits (distinct colorings taking rotations into account). This results in the following expression (see Table 2) for the number of distinct arrangements (Gilbert and Nicholson, 2004)

$$\frac{1}{24}\left[\binom{24}{12}+9\binom{12}{6}+8\binom{8}{4}+6\binom{6}{3}\right]=113{,}048$$

The calculation can be understood as follows: if we do not allow the polyhedron to rotate, then we have $\binom{24}{12}=\frac{24!}{12!12!}=2{,}704{,}156$ different colorings. Many of these are only rotations of each other; in fact, if one considers a coloring with no symmetry at all (one which looks different after any non-identity rotation), then it will be counted 24 different times. Colorings with

symmetry, however, are going to be counted fewer times. Hence, 1/24th of $\binom{24}{12}$ is almost correct, but we need to add back terms for symmetrical colorings.

To account for the symmetrical colorings, for each rotation we will count the number of colorings that are fixed by that rotation (i.e., that look the same after rotation). For a rotation about a particular 3-fold, the 24 faces divide into 8 groups of 3 faces related by the rotation. In this case, symmetrical colorings are those in which each group of 3 faces has a single color. Since there are an equal number of faces of both colors, there are $\binom{8}{4}$ different symmetrical arrangements for this rotation. There are 8 distinct 3-fold rotations, so we need to add this term 8 times. Similar terms need to be added for rotations about 2-fold and 4-fold axes. By the end, we have counted every coloring 24 different times, so we need to divide by that number to get our final answer.

Formally, the Burnside-Cauchy-Frobenius lemma can be stated as follows (Gilbert and Nicholson, 2004). Let G be a group that acts on the elements of a set X. For each $g \in G$, let Fix g = $\{x \in X \mid g(x) = x\}$, the set of elements of X left fixed by g. If N is the number of orbits of X under a group G, then

$$N = \frac{1}{|G|}\sum_{g \in G} |Fix\ \ g|$$

(where $|G|$ = number of elements of **G**). When applied to our case (Table 2) we have

$$N = \frac{\Sigma}{|O|} = \frac{2{,}713{,}152}{24} = 113{,}048$$

Table 2. Colorings of a pentagonal icositetrahedron with octahedral symmetry

Type of element, g_i, of the octahedral group	Order of g_i	Number, s, of such elements	Number of colorings left fixed by g_i, $\lvert \text{Fix } g_i \rvert$	$s \cdot \lvert \text{Fix } g_i \rvert$
Identity	1	1	$\binom{24}{12}$	$\binom{24}{12}$
90° rotation	4	9	$\binom{12}{6}$	$9\binom{12}{6}$
120° rotation	3	8	$\binom{8}{4}$	$8\binom{8}{4}$
180° rotation	2	6	$\binom{6}{3}$	$6\binom{6}{3}$
		$\lvert O \rvert = 24$		$\Sigma = 2{,}713{,}152$

The number of distinct ways of assembling 12 heterodimers is calculated in a similar manner. Without considering rotations, there are 2^{12} distinct colorings, with asymmetric colorings counted 24 different times. For symmetric colorings we find that 2^4 colorings are fixed by a given rotation around a degree 3 axis, 2^3 colorings are fixed by a 90° rotation around a degree 4 axis, 2^6 colorings are fixed by 180° rotation around a degree 4 axis, and that no colorings are fixed by rotations about the 2-fold axes. Thus the number of distinct arrangements of heterodimers is

$$\frac{1}{24}\left(2^{12} + 8 \cdot 2^4 + 6 \cdot 2^3 + 3 \cdot 2^6\right) = 186$$

Perfect Coloring

When considering the different ways that a polyhedron could be colored, perfect coloring is defined as when each symmetry operation causes a permutation of the colors. In other words, perfect coloring means when each symmetry operator of the underlying polyhedron is applied, all faces of a particular color are rotated onto a single color, which could be the same or different (Cromwell, 1997). Consider the 113,048 different ways that a 24 sided polyhedron representing ferritin (e.g., the *O* symmetry pentagonal icositetrahedron shown in Figure 1B) could be colored with 12 green and 12 blue faces. Of these possibilities, the only one with perfect coloring is the one with *T* symmetry.

Sketch of proof that there is only one perfect 2-coloring of the O symmetry pentagonal tetraicosahedron

Consider three adjacent faces related by a 3-fold axis: if two faces are colored blue, and the third face is green, then a rotation about this axis will take one blue face to blue and the other blue face to green. Hence, perfect coloring requires the three adjacent faces to have the same color. This is true of all eight "trimers". Now consider a 90° rotation about a 4-fold axis. Two adjacent trimers related by this rotation have either the same or different colors. If they have the same color, perfect coloring requires a set of four trimers related by rotations about this 4-fold to have the same color. Consider a perpendicular 4-fold: two adjacent trimers related by this 4-fold have already been assigned the same color, hence the other two trimers related by this 4-fold also must have this color. Continuing in this manner, we see that all eight trimers would be required to have the same color – this is not a 2-coloring. Therefore, two adjacent trimers related by a 90° rotation about a 4-fold axis must have different colors. Considering the three 4-fold axes, this

alternating coloring completely determines the coloring: it is the T symmetry coloring depicted in Figure 1B.

Proposition: For a 2-colored polyhedron with identical faces, consider the edges of the polygon that makes up each face. A 2-coloring is perfect if and only if each type of edge always joins faces of the same color or always joins faces with opposite colors.

Sketch of Proof

Consider a perfect 2-coloring. Taking any two adjacent faces, the edge between these faces either joins the same color or different colors. A symmetry operation will either leave the colors of these faces unchanged or reverse them. In either case the edge will still join faces of the same color or different colors. This proves the "if" part of the proposition. Consider a non-perfect 2-coloring. There is a symmetry operation that leaves some faces the same color, while reversing the colors of other faces. For this symmetry operation, we have two sets of faces: "reversing" faces and "non-reversing" faces. Choose an edge joining a reversing face to a non-reversing face. Before the symmetry operation, this edge joins faces of either the same color or different colors; after the operation the edge joins faces in the other manner.

References

Arosio, P., and Levi, S. (2002). Ferritins: Structural and Functional Aspects. In Molecular and Cellular Iron Transport, D. M. Templeton, ed. (New York, Marcel Dekker, Inc.), pp. 125-154.

Bozzi, M., Mignogna, G., Stefanini, S., Barra, D., Longhi, C., Valenti, P., and Chiancone, E. (1997). A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in Listeria innocua. J Biol Chem *272*, 3259-3265.

Brünger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., *et al.* (1998). Crystallography and NMR system: A new software system for macromolecular structure determination. Acta Cryst D *54*, 905-921.

Chasteen, N. D., and Harrison, P. M. (1999). Mineralization in ferritin: an efficient means of iron storage. J Struct Biol *126*, 182-194.

Chen, T., Amons, R., Clegg, J. S., Warner, A. H., and MacRae, T. H. (2003). Molecular characterization of artemin and ferritin from Artemia franciscana. Eur J Biochem *270*, 137-145.

Coelho, A. V., Macedo, S., Matias, P. M., Thompson, A. W., LeGall, J., and Carrondo, M. A. (2001). Structure determination of bacterioferritin from Desulfovibrio desulfuricans by the MAD method at the Fe K-edge. Acta Crystallogr D Biol Crystallogr *57*, 326-329.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D *50*, 760-763.

Cromwell, P. R. (1997). Polyhedra (Cambridge, United Kingdom, Cambridge University Press).

Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta Crystallogr D Biol Crystallogr *55 (Pt 4)*, 938-940.

Ford, G. C., Harrison, P. M., Rice, D. W., Smith, J. M., Treffry, A., White, J. L., and Yariv, J. (1984). Ferritin: design and formation of an iron-storage molecule. Philos Trans R Soc Lond B Biol Sci *304*, 551-565.

Frolow, F., Kalb, A. J., and Yariv, J. (1994). Structure of a unique twofold symmetric haem-binding site. Nat Struct Biol *1*, 453-460.

Gallivan, J. P., and Dougherty, D. A. (1999). Cation-pi interactions in structural biology. Proc Natl Acad Sci U S A *96*, 9459-9464.

Ghosh, S., Hevi, S., and Chuck, S. L. (2004). Regulated secretion of glycosylated human ferritin from hepatocytes. Blood *103*, 2369-2376.

Gilbert, W. J., and Nicholson, W. K. (2004). Modern Algebra with Applications, Second edn (Hoboken, New Jersey, John Wiley & Sons, Inc.).

Granier, T. (1997). Comparison of the structures of the cubic and tetragonal forms of horse-spleen apoferritin. Acta Crystallogr D Biol Crystallogr *53*, 580-587.

Granier, T., Gallois, B., Langlois d'Estaintot, B., Dautant, A., Chevalier, J. M., Mellado, J. M., Beaumont, C., Santambrogio, P., Arosio, P., and Precigoux, G. (2001). Structure of mouse L-chain ferritin at 1.6 A resolution. Acta Crystallogr D Biol Crystallogr *57*, 1491-1497.

Granier, T., Langlois d'Estaintot, B., Gallois, B., Chevalier, J. M., Precigoux, G., Santambrogio, P., and Arosio, P. (2003). Structural description of the active sites of mouse L-chain ferritin at 1.2 A resolution. J Biol Inorg Chem *8*, 105-111.

Ha, Y., Shi, D., Small, G. W., Theil, E. C., and Allewell, N. M. (1999). Crystal structure of bullfrog M ferritin at 2.8 A resolution: analysis of subunit interactions and the binuclear metal center. J Biol Inorg Chem *4*, 243-256.

Harrison, P. M., and Arosio, P. (1996). The ferritins: molecular properties, iron storage function and cellular regulation. Biochem Biophys Acta *1275*, 161-203.

Hempstead, P. D., Yewdall, S. J., Fernie, A. R., Lawson, D. M., Artymiuk, P. J., Rice, D. W., Ford, G. C., and Harrison, P. M. (1997). Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. J Mol Biol *268*, 424-448.

Ilari, A., Stefanini, S., Chiancone, E., and Tsernoglou, D. (2000). The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. Nat Struct Biol *7*, 38-43.

Johnson, J. L., Norcross, D. C., Arosio, P., Frankel, R. B., and Watt, G. D. (1999). Redox reactivity of animal apoferritins and apoheteropolymers assembled from recombinant heavy and light human chain ferritins. Biochemistry *38*, 4089-4096.

Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst *A47*, 110-119.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr *24*, 946-950.

Langlois d'Estaintot, B., Santambrogio, P., Granier, T., Gallois, B., Chevalier, J. M., Precigoux, G., Levi, S., and Arosio, P. (2004). Crystal structure and biochemical properties of the human mitochondrial ferritin and its mutant Ser144Ala. J Mol Biol *340*, 277-293.

Larsen, W. J. (1976). Cell remodeling in the fat body of an insect. Tissue Cell *8*, 73-92.

Lawson, D. M., Treffry, A., Artymiuk, P. J., Harrison, P. M., Yewdall, S. J., Luzzago, A., Cesareni, G., Levi, S., and Arosio, P. (1989). Identification of the ferroxidase centre in ferritin. FEBS Lett *254*, 207-210.

Locke, M. (2003). Surface membranes, Golgi complexes, and vacuolar systems. Annu Rev Entomol *48*, 1-27.

Locke, M., and Leung, H. (1984). The induction and distribution of an insect ferritin--a new function for the endoplasmic reticulum. Tissue Cell *16*, 739-766.

Macedo, S., Romao, C. V., Mitchell, E., Matias, P. M., Liu, M. Y., Xavier, A. V., LeGall, J., Teixeira, M., Lindley, P., and Carrondo, M. A. (2003). The nature of the di-iron site in the bacterioferritin from Desulfovibrio desulfuricans. Nat Struct Biol *10*, 285-290.

Merritt, E. A., and Bacon, D. J. (1997). Raster3D: Photorealistic molecular graphics. Meth Enzymol *277*, 505-524.

Michaux, M. A., Dautant, A., Gallois, B., Granier, T., d'Estaintot, B. L., and Precigoux, G. (1996). Structural investigation of the complexation properties between horse spleen apoferritin and metalloporphyrins. Proteins *24*, 314-321.

Navaza, J. (1994). AMORE- An automated package for molecular replacement. Acta Cryst *A50*, 157-163.

Nichol, H., Law, J. H., and Winzerling, J. J. (2002). Iron metabolism in insects. Annu Rev Entomol *47*, 535-559.

Nichol, H., and Locke, M. (1999). Secreted ferritin subunits are of two kinds in insects molecular cloning of cDNAs encoding two major subunits of secreted ferritin from Calpodes ethlius. Insect Biochem Mol Biol *29*, 999-1013.

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

Precigoux, G. (1994). A crystallographic study of haem binding to ferritin. Acta Crystallogr D Biol Crystallogr *50*, 739-743.

Stillman, T. J., Hempstead, P. D., Artymiuk, P. J., Andrews, S. C., Hudson, A. J., Treffry, A., Guest, J. R., and Harrison, P. M. (2001). The high-resolution X-ray crystallographic structure of the ferritin (EcFtnA) of Escherichia coli; comparison with human H ferritin (HuHF) and the structures of the Fe(3+) and Zn(2+) derivatives. J Mol Biol *307*, 587-603.

Trikha, J., Waldo, G. S., Lewandowski, F. A., Ha, Y., Theil, E. C., Weber, P. C., and Allewell, N. M. (1994). Crystallization and structural analysis of bullfrog red cell L-subunit ferritins. Proteins *18*, 107-118.

Accession Numbers

Atomic coordinates and structure factors will be deposited in the Protein Data Bank. The sequences of the *T. ni* H and L chain cDNAs will be deposited in GenBank.

Table 1. Data Collection and Refinement Statistics

Space group	*C*2
Unit cell (Å)	203.8, 146.6, 206.9; β = 92.7 °
Data collection	
Resolution (Å)	30.0-1.90 (1.97-1.90)
Number of observations	875,805
Unique reflections	463,465
% Complete*	97.0 (98.7)
$I/\sigma I$	11.7 (2.2)
R_{merge}(%)‡	6.5 (34.4)
Refinement	
Resolution (Å)	20.0-1.90
Reflections in working set $\|F\|>0$	440,103
Reflections in test set $\|F\|>0$	23,175
R_{cryst} (%)¶	18.9
R_{free} (%)	19.4
Number of non-hydrogen atoms	
Protein (403 residues)§	3,225
Water	375
Other (2 Fe^{+3}, 2 Ca^{+2})	4
Model geometry	
Rmsd bond length (Å)	0.009
Rmsd angles (deg)	1.26
Ramachandran plot	
Most favored region (%)	91.7
Additional allowed region (%)	8.3
Generously allowed region (%)	0.0
Disallowed region (%)	0.0

Values in parentheses indicate the high-resolution shells. *Complete is the number of independent reflections/total theoretical number. ‡R_{merge} $(I)= (\Sigma|I(i) - \langle I(\boldsymbol{h})\rangle|/\Sigma I(i))$, where $I(i)$ is the ith observation of the intensity of the hkl reflection and $\langle I\rangle$ is the mean intensity from multiple measurements of the h,k,l reflection. ¶R_{cryst} $(F) = \sum_h||\boldsymbol{F}_{obs}(\boldsymbol{h})| - |\boldsymbol{F}_{calc}(\boldsymbol{h})||/ \sum_h|\boldsymbol{F}_{obs}(\boldsymbol{h})|$, where $|\boldsymbol{F}_{obs}(\boldsymbol{h})|$ and $|\boldsymbol{F}_{calc}(\boldsymbol{h})|$ are the observed and calculated structure factor amplitudes for the h,k,l reflection. R_{free} is calculated over reflections in a test set not included in atomic refinement. §Number of atoms in strict NCS refinement.

Figure Legends

Figure 1. Structure of Secreted *T. ni* Ferritin Reveals Symmetric Arrangement of H and L chains.

(A) Ribbon diagram showing the H (green) and L (blue) chains. Native Fe^{3+} ions at the ferroxidase sites are shown as red spheres.

(B) Schematic diagram illustrating the tetrahedral symmetry of insect ferritin. Each face of this pentagonal icositetrahedron represents a subunit. The total buried surface areas at the subunit-subunit interfaces are: I/II (heterodimer) 3250 $Å^2$, I/III (H chain trimer) 1690 $Å^2$, I/VI (pseudo 4-fold) 1250 $Å^2$, I/VII (pseudo 4-fold) 1525 $Å^2$, II/VII (L chain trimer) 2560 $Å^2$.

(C) Stereoview of a 4 Å anomalous difference Fourier map (not NCS averaged; contoured at 3 σ). Each of the 12 peaks results from an iron atom bound at the ferroxidase site in a ferritin H subunit. When this map is contoured at a lower level additional peaks appear along the L chain 3-fold pores, but not in the L chain four helix bundle. The superimposed chiral, non-regular icosahedron whose 12 vertices correspond to the peaks in the anomalous difference Fourier map is the simplest polyhedron with exact *T* symmetry.

Figure 2. Structures of *T. ni* Ferritin H and L Chains.

(A) Ribbon diagrams of the H (green) and L (blue and light blue) subunits. The Fe^{3+} ion at the ferroxidase site is shown in red. Also shown are the approximate locations of the 3-fold axes (blue and green triangles) and the pseudo 4-fold axes (black ovals). Disulfide-bonded cysteines are shown in yellow in ball-and-stick representation. The N-terminal region of each subunit extends on the outside of the ferritin shell and forms a disulfide bond with a non-adjacent subunit of the opposite type. Helices A-E are labeled in green for the H chain and in blue for the L chain.

(B) Stereoview of a superposition of the *T. ni* ferritin H chain (green) with the human ferritin H chain (magenta, PDB accession code 2FHA). Rms deviation for the superposition is 1.08 Å calculated for 160 Cα atoms.

(C) Stereoview of a superposition of the *T. ni* ferritin H (green) and L (blue) chains. The rms deviation value for the superposition is 1.27 Å calculated for 156 Cα atoms.

Figure 3. Stereoview of a Close Up of the H Chain Ferroxidase Site.

Residues at or near the Fe^{3+} binding site are shown in ball-and-stick representation. The Fe^{3+} ion is shown as a dark red sphere and the two water molecules are shown as small red spheres. Coordinating side chains and their distances (Å) are shown.

Figure 4. Packing of the Major H/L Interface.

Residues involved in key intersubunit interactions are shown in ball-and-stick representation. The H and L chains are colored green and blue, respectively. Hydrogen bonds and salt bridges are shown as dashed lines. The location of the pseudo 2-fold is shown as an open oval.

Figure 5. Close Up of the L Chain 3-fold Pore.

(A) A side view of the L chain 3-fold pore shown with the inside of the shell at the bottom. A 12-fold NCS-averaged annealed omit F_O-F_C electron density map (cyan mesh) is contoured at 8σ. The NCS-averaged anomalous Fourier map (red mesh) is contoured at 13σ. The F_O-F_C electron density was calculated excluding the coordinates of the ions coordinates. The top two ions (cyan) were modeled as Ca^{2+}, while the lower ion (red) was assigned as Fe^{3+}, based in part on the stronger anomalous signal at the Fe^{3+} location.

(B) The L chain 3-fold pore viewed from the inside of the shell. A large cluster of negatively charged residues line the inner surface near the 3-fold axis.

Supplementary Figure 1. Sequence Alignment of Insect H and L Chains.

(A) Sequence alignment of insect ferritin H chains. Residues at the ferroxidase center are highlighted in yellow. Cysteines involved in disulfide bonds are shown in red.

(B) Sequence alignment of insect ferritin L chains. Cysteine residues involved in disulfide bonds are shown in red and asparagines with potential N-linked oligosaccharides are shown in blue. Pairs of cysteines in disulfide bonds are marked with symbols (* for L chain 4 and L chain 24, # for L chain 12 and H chain 3, and @ for H chain 21 and H chain 130).

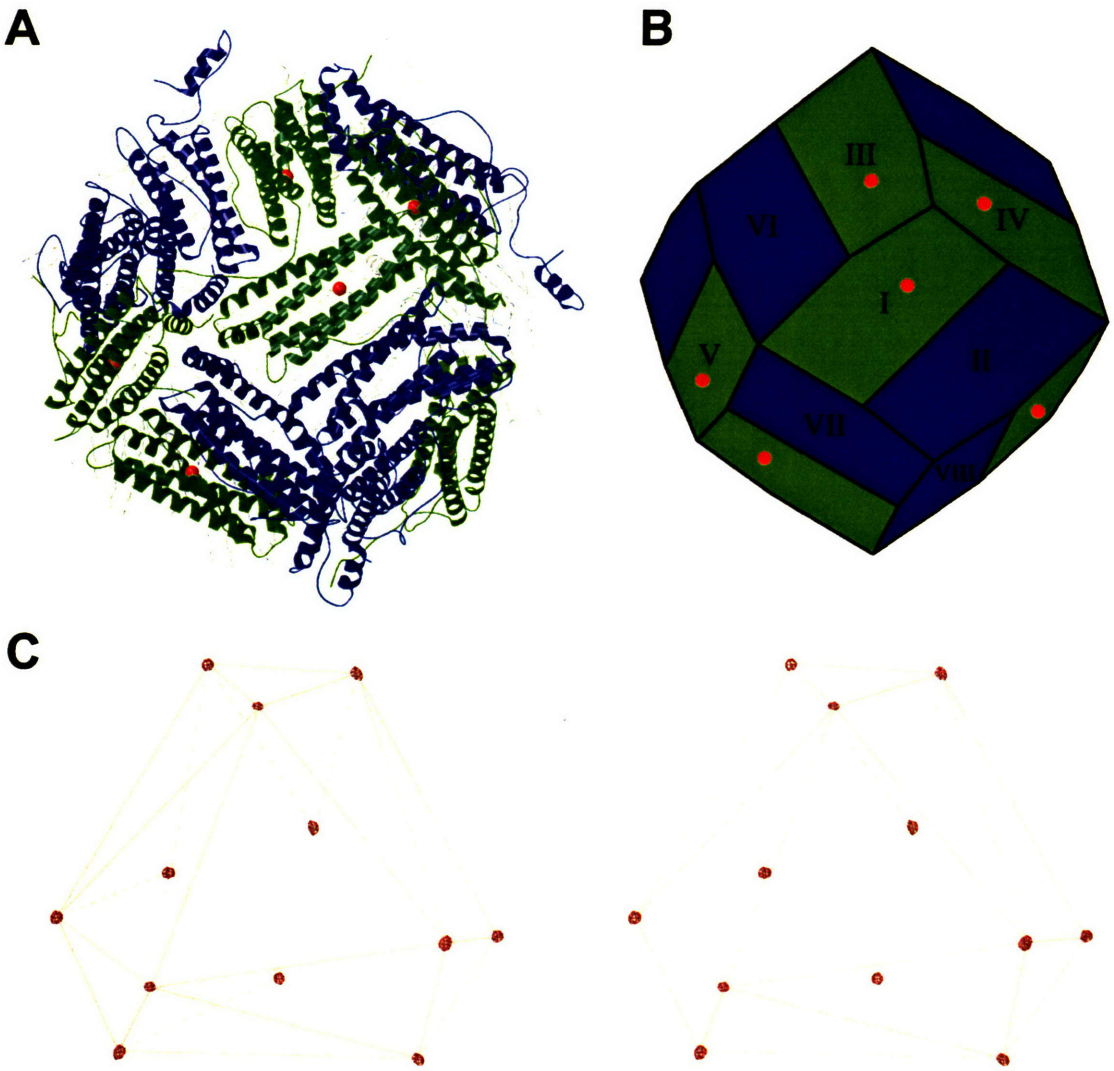


Figure 1. Structure of Secreted *T. ni* Ferritin Reveals Symmetric Arrangement of H and L chains.

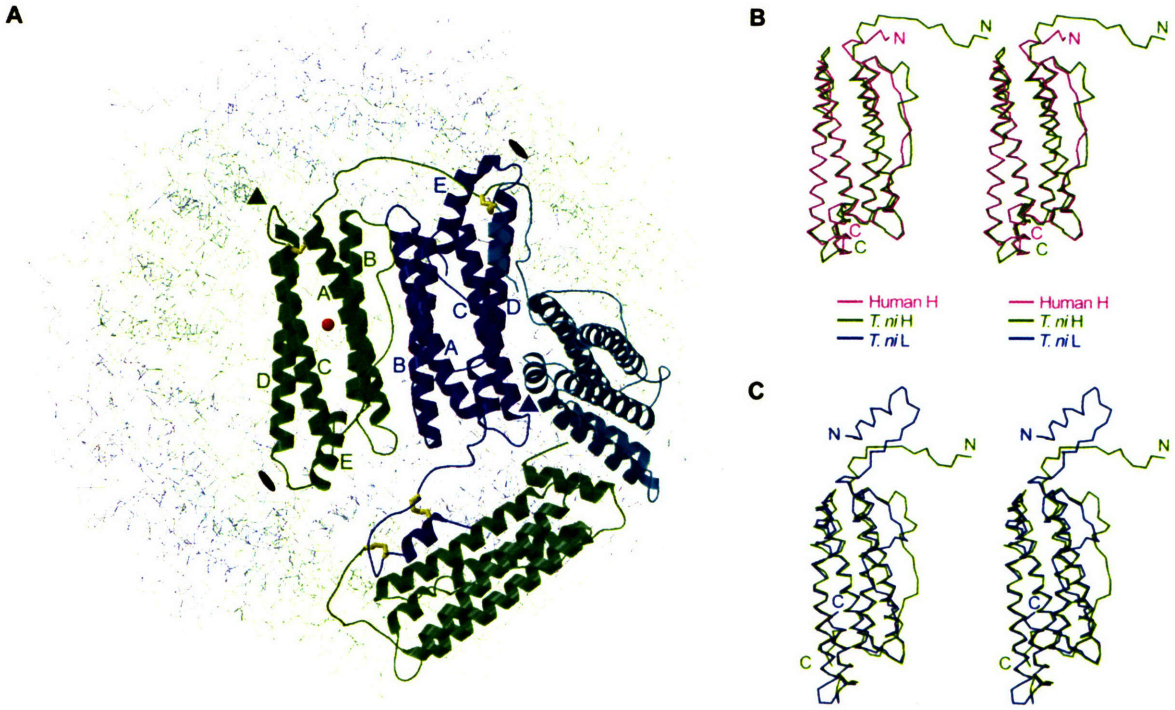


Figure 2. Structures of *T. ni* Ferritin H and L Chains.

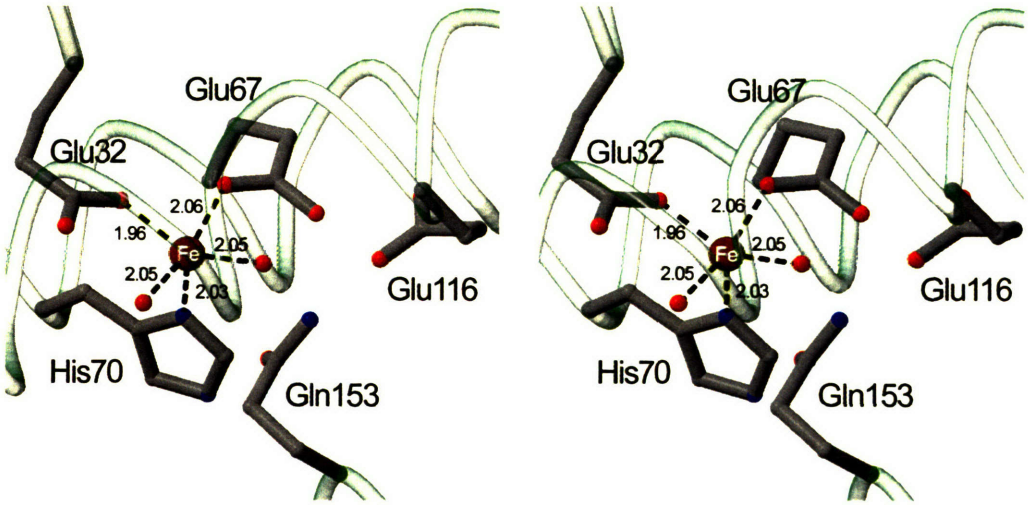


Figure 3. Stereoview of a Close Up of the H Chain Ferroxidase Site.

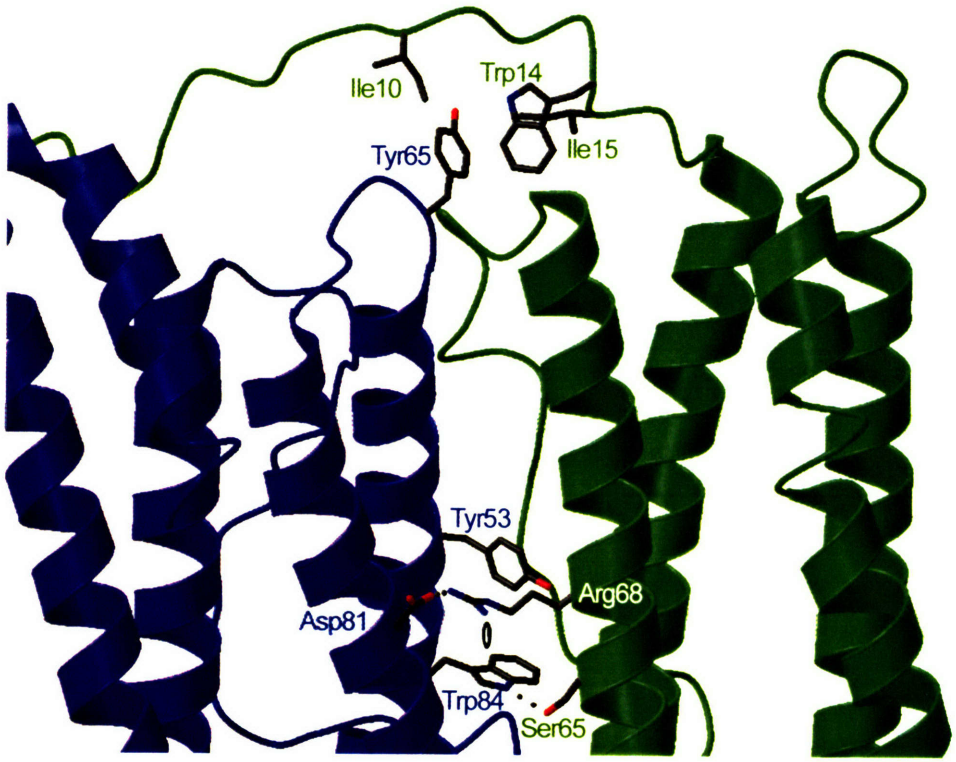


Figure 4. Packing of the Major H/L Interface.

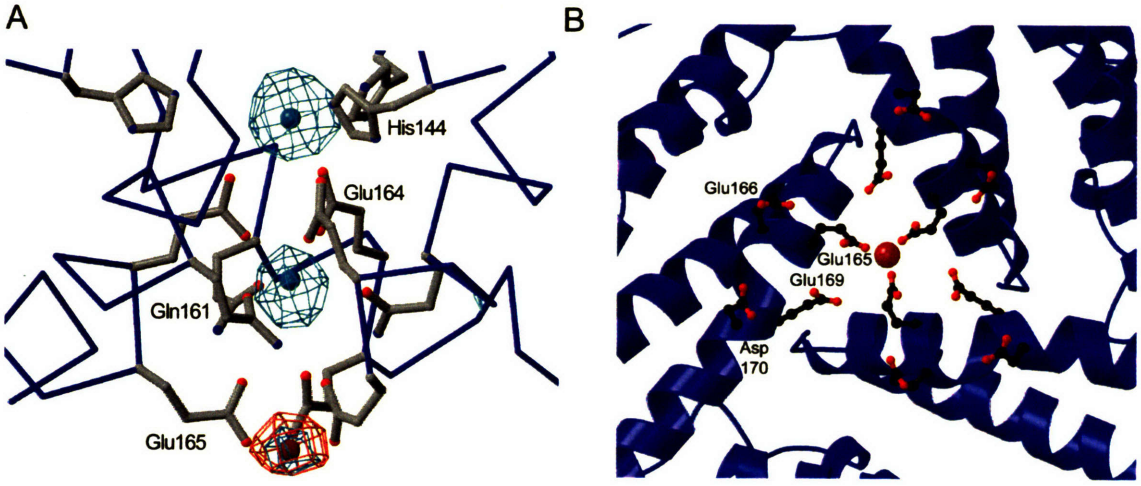


Figure 5. Close Up of the L Chain 3-fold Pore.

```
                                    #                   @
Trichoplusia ni H chain  :     TQCNVN PVQIPK DWITMHRSCRNSMRQQIQMEVGASLQYLAM  42
Manduca sexta            :     TQCHVN PVNIQR EWITMHRSCRDSMRRQIQMEVGASLQYLAM
Galleria mellonella      :     TQCTVN PVNIPK EWITMQRPCRDSMRRQIQMEVAASLQYLAM
Calpodes ethlius         :     TQCNVN PVTIPT DWITMTSGCRTSVRHQIQMEVAASLQYLAM
Apriona germari          :   E IECRGE NVNVPT DWLDMQQFCVSSVRNQIEEELKAAMQYMAM
Leptinotarsa decemlineata:   A LECSYK ELDIPK DWIDMEKACVKKMRAQVEDELKAAMQYMAM
Drosophila melanogaster  :    DFKCSLA VPEITK  WVDMKDACIKGMRNQIQEEINASYQYLAM
Nilaparvata lugens       : QNANDRCSIDMDDTLEKVEWKTMHSNCTLEVKDQIKMEYNAAMIYLSL
Anopheles gambiae        :      QVT DTDAPSSTDEWNYMNRSCSAKLQDQINKEFDAAIFYMQY
Aedes aegypti            :        QEQTVGATDNYQWDSVDDQCLAALHRQINKEFDASIIYLKY

Trichoplusia ni          : GAHFSKDVVNRPGFAQLFFDAASEEREHAMKLIEYLLMRGELTNDVSS  90
Manduca sexta            : GAHFSKDKINRPGFAKLFFDAAGEEREHAMKLIEYLLMRGELTNDVTS
Galleria mellonella      : GAHFSKDTINRPGFAKLFFDAGSEERGHAMKLIEYLLMRGELTSDVTS
Calpodes ethlius         : GAHFSRDGINRPGFAKLFFDASSEERGHALKLIEYLLMRGELTSNISS
Apriona germari          : GAHFSKDIVNRPGFAKMFFEAASEERQHAIKLISYLLMRGELTSKVSE
Leptinotarsa decemlineata: GAHFSKDTVNRPGFAEIFFKSASEEREHAIKLISYLLMRGELTSKVSS
Drosophila melanogaster  : GAYFSRDTVNRPGFAEHFFKAAKEEREHGSKLVEYLSMRGQLTEGVSD
Nilaparvata lugens       : GVHFSRDFVNRPGFAKFFFESASEERQHAIKLIEYLSMRGESVTDIAK
Anopheles gambiae        : GAYFAQYQVNLPGFEKFFFNAASEEREHGMKLIEYALMRGQKPIDRNT
Aedes aegypti            : AAYFAQEKINLPGFEKFFFHAAAEEREHGIKLIEYALMRGKAPAD KH

                                                                    @
Trichoplusia ni          : L LQVRP   PTRSSWK GGVEALEHALSMESDVTKSIRNVIKACEDD 133
Manduca sexta            : L IQVRA   PQRNKWE GGVDALEHALKMESDVTKSIRTVIKACEDD
Galleria mellonella      : L IQIRP   PERKSWS SGVEALEHAVKMESDVTKSIRTVISDCESD
Calpodes ethlius         : L ITIRP   PERKSWE SGQEALEHALRMETAVTKSIKNVIVNCEHD
Apriona germari          : L IRSRK LVPQKTYWD SGVEALKDALNLEASVTKKIRKVIKNCEED
Leptinotarsa decemlineata: L IK RN LMPSQTTWT NGVSALKDALKLEASVTRKIRDVIKVCEEA
Drosophila melanogaster  : L INVPT V  AKQEWT DGAAALSDALDLEIKVTKSIRKLIQTCENK
Nilaparvata lugens       : L VKLDPETMPGMASVSLNGKEALEKALQQEVLVTNNILKVMKACENE
Anopheles gambiae        : FSLNFANPAARVDAEQGSVALTALKAALAKEQEVTKSIRELIKICEED
Aedes aegypti            : FKLNYDHEV PTVTT    GESALETALQKEVEVTKSIRGVIKACEDG

Trichoplusia ni          :   SEFN     DYHLVDYLTGDFLEEQYKGQRDLAGKASTLKKLMDRH 174
Manduca sexta            :   PEFN     DYHLVDYLTGEFLEEQYKGQRDLAGKASTLKKMLDRN
Galleria mellonella      :   PNFN     DYHLVDYLTGEFLEEQYKGQRDLAGKASTLKKMMDRH
Calpodes ethlius         :   REANGRDDNDYHLVDYLTGEFLDEQYKGQRDLAGKAATLKKMMDRH
Apriona germari          :   S FN     DYHIVDYLTGDFLTEQYQGQRDIAGKVSTLEKLKVKH
Leptinotarsa decemlineata:   KSFN     DYHLVDYLSGDFLGEQYQGQRDIAGKISTLEKMTEKH
Drosophila melanogaster  :   P YN     HYHLVDYLTGVYLEEQLHGQRELAGKLTTLKKMMDTN
Nilaparvata lugens       : EVKDAAWTLPNDYHLVDWLTAEFLDEQYKGQRDIAGKLSTLLKMGSSN
Anopheles gambiae        :     HN     DYHLVDYLTGEFLEEQHQGQRDLAGKITMLSKLLRTN
Aedes aegypti            :     SN     DFHLADYLTGEYLDEQHKGQRELAEKIATLKKMKKSA

Trichoplusia ni          : EALGEFIFDKKLLGIDV                                 191
Manduca sexta            : SALGEFIFDKKLMGMDI
Galleria mellonella      : ASLGEFIFDKKLLGIDV
Calpodes ethlius         : SALGEFIFDKRLLGMDI
Apriona germari          : GALGEFLFDKKLLNGEL
Leptinotarsa decemlineata: GALGEWLFDKKLLKGEL
Drosophila melanogaster  : GELGEFLFDKTL
Nilaparvata lugens       : YHLGEFLFDKKLLSNEA
Anopheles gambiae        : PKLGEFMFDKQNM
Aedes aegypti            : PKLGEFLFDKNHM
```

Supplemental Figure 1A. Sequence alignment of insect H chains

```
                                    *       #           *
Trichoplusia ni L chain:        ADTCYNDVALDCGITSNSLALPRCNAVYGEYGSHGNVATELQA  43
Manduca sexta           :       ADTCYQDVSLDCSQVSNSLTLPNCNAVYAEYGHHGNVAKEMQA
Galleria mellonella     :       EDACYNDVSLQCAQASNNLGLAHCNSIYGEYGRHGNVATEMQA
Calpodes ethlius        :        DVCYQDASMECGLASNSLELSNCNAVYGNYGRHGNVASEMQA
Apriona germari         : QVEDHLSKSCYNDIDTICKHSKLSPKDSYCSAKYGGIN   KVQEGLQK
Drosophila melanogaster: KD    DEYCQNTVITACSTSAFS GNSICNARFAGIDH   IEPEIQS
Nilaparvata lugens      : IKPDAEKGACVKSVANFCHATEQ  KISDCNAQYSGF HH  VHSDLQQ
Anopheles gambiae       :   TDLSANDCEINVE ECSPTYSSF LSRSGKT         VENDLKQ
Aedes aegypti           :         DNNNSTVSFT     AQFSSIAH           IGNDLQT

Trichoplusia ni         : YAKLHLERSYDYLLSAAYFNNYQTNRAGFSKLFKKLSDEAWSKTIDIIK  92
Manduca sexta           : YAALHLERSYEYLLSSSYFNNYQTNRAGFSKLFRKLSDDAWEKTIDLIK
Galleria mellonella     : YANLHLERSYEYLLSAAYFNNYQTNRDGFSKLFKKLSDQAWEKTIELIK
Calpodes ethlius        : YANLHIERSYQYLLSPAFFDNYNTNRKGFSALFKKLSDHAWSKSIELIQ
Apriona germari         : FVNDHFTLSFHYLLMATHFDNYNKNRPGFEKLFRGLSDDTWEDGIELIK
Drosophila melanogaster: YINANLAKSYDYLLLATHFNSYQKNRPGFQKLYQGLSDRSFEDSIALIK
Nilaparvata lugens      : FVVTQIEQSFQFLTMATKFGNYKSNRPGFEKLYRGLADKSWEESIELMK
Anopheles gambiae       : YTSQLVDKSFHFLMMSSAFNKHSLDRPGFEKLYRKISDKAWADAIELIK
Aedes aegypti           : FTSQQLEKSFDFLLLAFNFDQYMIDRPGFEKLYRKISDKAWEDTEKLIK

Trichoplusia ni         : HVTKRGDKMNFD  QHSTMK TERKNYTA      ENHELEALAKALDTQ 133
Manduca sexta           : HITMRGDEMNFA  QRSTQKSVDRKNYTV      ELHELESLAKALDTQ
Galleria mellonella     : HITKRGGEMNFA  QRSTQQPAERKNYTV      ELHELESLAKALDTQ
Calpodes ethlius        : HITKRGDVMDFS  RRSTLASTA KNVTL      ELPELESLAHALDTQ
Apriona germari         : YITKRGGEMNFN LQ   SYFNETKP DA      ELYEYYAVGKALDNH
Drosophila melanogaster: QVTRRGGIVDFN TRHESSGSVSTKRVTL      EVDELHSLALALDTE
Nilaparvata lugens      : YITSRGYDVNLK IT PYQYSNNTKSLTEISTYPEISELKSLSMALEMN
Anopheles gambiae       : YQSRRGSFGHL  VQPS  KGENYGKVL       DVQELSSLQFALDYE
Aedes aegypti           : YQSKRGLTVELKDLKGGVIGQLNDGKVGGSISLLDSDEISSLKVALGYE

Trichoplusia ni         : KELAERAFYIHREATR NSQ H       LH DPEIAQYLEEEFIEDHA 172
Manduca sexta           : KELAERAFFIHREATR NSQ H       LH DPEVAQYLEEEFIEDHA
Galleria mellonella     : KEIAERAFYIHREATR NSQ H       LH DPEVAQYLEEEFIEDHS
Calpodes ethlius        : KEMAERAFYIHREATR NSQ K       TH DPEIAQYLEEEFVEYQA
Apriona germari         : KKLALEAFEVQKEAAN KAK D       YH DPEITSYLEHEFMHKHR
Drosophila melanogaster: KQLATGATHVHSRATH ATD A       ER DPELAHYFEENFLGKQA
Nilaparvata lugens      : KFLAEKAHDIHHNAAS HSKDK       PH DAEVMSFLENTYVHKHA
Anopheles gambiae       : KQMAKEAHAIHRKISHAHSKAGSNGSDDVYHYDPDAAHYLDENIIEYQS
Aedes aegypti           : KILAEESHHIHKKISHAHDNKA       TY DPDVAHFLDEEIIEYQS

Trichoplusia ni         : EKIRTLAGHTSDLKKFITANNGHDLSLALYVFDEYLQKTV          212
Manduca sexta           : KTIRNLAGHTTDLKRFVSGDNGQDLSLALYVFDEYLQKTV
Galleria mellonella     : KTIRELAGHTTDLKSFITVNNGQDKSLAFYLFDEYLQKTV
Calpodes ethlius        : KVIRDLAGHTTDLKKFVVSNKGQDLSLALYLFDEYLQKSV
Apriona germari         : DIVK LAGYTSDLNKIL   DGPDSSLSLYLFDEYLQKQ
Drosophila melanogaster: ESVRKLSGYANDLAKLM   KVPDPSLSVYLFDEYLQKQ
Nilaparvata lugens      : DTIRTLTGHVNDLHK ITQTRGVDANLATFMFDEFLLKA
Anopheles gambiae       : GVVRDLAGYVHNLKHFTSAKHAAN DLGNHVFDEFLAKVE
Aedes aegypti           : GTIRKLTGYIYNLDSIIKEDKTKD  LGIHMFDEYLDKVE
```

Supplemental Figure 1B. Sequence alignment of insect L chains

APPENDIX III

Steric Accessibility of the HIV-1 gp41 N-Trimer Region

Appendix III has been accepted for publication in the *Journal of Biological Chemistry*. I started this project in Peter Kim's lab at MIT. My contribution to this work includes coming up with the initial ideas for the study, cloning and expressing most of the proteins, performing the initial syncytia assays and assisting in the editing of the manuscript. The initial viral infectivity assays were done by Heng Chhay in the Kim lab. After the Kim lab closed, a former postdoc from the Kim lab, Michael Kay, finished the project in his lab at the University of Utah in Salt Lake City. Michael Kay's contribution includes supervising the research done in his lab, data analysis and writing the manuscript. The viral assays, binding studies, Ub data collection, and the proteolysis and precipitation studies were carried out by Brett Welch. Sunghwan Kim did the syncytia assays, expression of IQN36, refolding and characterization of BPTI, and performing the extended linker studies.

STERIC ACCESSIBILITY OF THE HIV-1 gp41 N-TRIMER REGION

Agnes E. Hamburger[*†‡], Sunghwan Kim[†§], Brett D. Welch[†§], and Michael S. Kay[§]

From the [§]Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT, 84132 and [*]Department of Biology, 31 Ames Street, Room 68-132, Massachusetts Institute of Technology, Cambridge MA 02139

Running Title: Steric Accessibility of the HIV-1 gp41 N-trimer Region

[‡]Current Address: Department of Biology, 114-96, California Institute of Technology, Pasadena, CA 91125

[†]A.E.H., S.K., and B.D.W. contributed equally to this work.

Address correspondance to: Michael S. Kay, Department of Biochemistry, MREB 211, 50 N. Medical Drive, Salt Lake City, Utah 84132, Tel. 801-585-5021; Fax 801-581-7959; E-mail: kay@biochem.utah.edu

During HIV entry, gp41 undergoes a series of conformational changes that induce membrane fusion. Immediately prior to fusion, gp41 exists in a pre-hairpin intermediate, in which the N-peptide and C-peptide regions of gp41 are exposed. Rearrangement of this intermediate into a six-helix bundle composed of a trimeric coiled coil from the N-peptide region (N-trimer) surrounded by three peptides from the C-peptide region provides the driving force for membrane fusion, while prevention of six-helix bundle formation inhibits viral entry. Because of its central role in mediating viral entry, the N-trimer region of gp41

is a key vaccine target. Extensive efforts to discover potent and broadly neutralizing Abs against the N-trimer region have, thus far, been unsuccessful.

In this study, we attach a potent C-peptide inhibitor that binds to the N-trimer region to cargo proteins of various sizes to examine the steric accessibility of the N-trimer during fusion. These inhibitors show a progressive loss of potency with increasing cargo size. Extension of the cargo/C-peptide linker partially restores inhibitory potency. These results demonstrate that HIV defends its critical hairpin-forming machinery by steric exclusion of large proteins and may explain the current dearth of neutralizing Abs against the N-trimer. In contrast, previous results suggest the C-peptide region is freely accessible during fusion, demonstrating that the N- and C- peptide regions are in structurally distinct environments. Based on these results, we also propose new strategies for the generation of neutralizing Abs that overcome this steric block.

HIV entry is mediated by the viral envelope (Env) glycoprotein. Env is initially produced as gp160, which is proteolytically cleaved into non-covalently associated transmembrane (gp41) and surface (gp120) subunits. gp120 is primarily involved in recognition of cellular receptors, while gp41 is anchored in the viral membrane and mediates membrane fusion. The gp41 ectodomain contains two helical heptad repeat sequences (N- and C-peptide regions) (1,2). Peptides corresponding to these helical regions (N- and C-peptides) are dominant-negative inhibitors of HIV membrane fusion (2,3). Isolated N- and C-peptides form a six-helix bundle (trimer-of-hairpins) when mixed in solution (4-6). In this structure, three N-peptides form a central parallel trimeric coiled coil (N-trimer) surrounded by three anti-parallel C-peptides that nestle between neighboring N-peptides.

Based largely on these inhibitory and structural data, a working model of HIV-1 membrane fusion has been proposed (Fig. 1) (3,5). Initial interaction of Env with its target cell occurs via gp120 binding to CD4 and a coreceptor (typically CCR5 or CXCR4). This binding induces a series of large conformational changes in gp120, which are propagated to gp41 via the gp41-gp120 interface. At this stage, gp41 transiently adopts an extended "prehairpin intermediate" conformation that bridges both the viral and cellular membranes. This state is believed to persist for at least 15 min (3,7,8), but eventually collapses into a trimer-of-hairpins structure, which pulls both membranes into tight apposition and induces membrane fusion (Fig. 1).

In this model, the prehairpin intermediate exposes the isolated N-trimer, while the C-peptide region exists in an unknown and possibly unstructured conformation remote from the N-trimer (3). At this stage, the prehairpin intermediate is vulnerable to binding of exogenous N- and C-peptides. Binding of these peptide inhibitors denies access of the endogenous N- or C-peptide regions to their appropriate intramolecular partner, thwarting hairpin formation and membrane fusion. This model predicts that any molecule that binds to the prehairpin intermediate and disrupts association of the N- and C-peptides will inhibit membrane fusion and has been successfully applied to the development of several potent entry inhibitors (9-11).

Additionally, the gp41 prehairpin intermediate has several promising features as an inhibitory target (12). Peptide mimics of the N-trimer region have been structurally characterized at high resolution (4-6). The interface between the N- and C-peptides is highly conserved among diverse HIV strains of both lab-adapted and clinical isolates (9). The N-trimer also presents a long (>100 Å), deep groove with an extensive binding surface (4-6). These special properties have led many groups to search for Abs that can disrupt this interface (reviewed in (13)).

C-peptide Inhibitors

Several peptide fusion inhibitors derived from the N- and C-peptide regions of gp41 have been described (2,3,12,14-16). The most potent are peptides derived from the C-peptide region (e.g., C34, DP178/T20, T1249), which have low nM IC_{50}'s against viral entry in cell-cell fusion (syncytia formation) and viral infectivity assays (reviewed in (17)). Several mutations leading to T-20 resistance have been mapped to the N-peptide region of gp41 (18), providing strong support that the N-trimer is the primary target of C-peptide inhibitors.

gp41 N-trimer as a vaccine target

As demonstrated by the efficacy of C-peptide inhibitors, the N-trimer region of gp41 is a very attractive candidate for vaccine efforts. Many such efforts have been undertaken using various peptide mimics of the N-trimer region (e.g., N-peptide, 5-helix, IZN36, and $N35_{CCG}$-N13 (17,19-21)). These efforts have produced a large number of Abs with specific and high affinity binding to their targets, but weak and/or narrow neutralizing activity in standard viral entry and spread assays. Interestingly, some of these anti-N-trimer Abs can inhibit fusion if bound to a temperature-arrested intermediate fusion state (19) or in the presence of soluble CD4 (sCD4) (21). Currently, there are only two reported anti-gp41 Abs that exhibit potent and broadly neutralizing activity, 2F5 and 4E10, which bind just outside the C-terminal border of the C-peptide region, an area with uncertain structure (reviewed in (22)).

In this study, we test the hypothesis that the N-trimer of gp41 is sterically restricted in the prehairpin intermediate, which may explain the current dearth of broadly neutralizing Abs against this target (Fig. 1). All of the known fusion inhibitors that target this structure (e.g., C34, T-20, T-1249, D-peptides) are small (<40 residue) peptides and could circumvent such a steric block. We have constructed fusions of a well characterized C-peptide inhibitor (C34) to a series of protein cargoes of varying sizes to determine if such a steric block exists and, if so, to define

its size cutoff. Our results demonstrate that C-peptide fusion proteins lose inhibitory potency with increasing size and that the N-trimer region of gp41 is likely to be poorly accessible to proteins as large as Abs. These results have important implications for gp41 vaccine design as well as for the production of second-generation C-peptide entry inhibitors. This steric restriction also helps to better define the conformation of the pre-hairpin intermediate.

Materials and Methods

Reagents

Plasmids were obtained from the following sources: pET vectors (Novagen), pMAL-c2G (NEB), pEBB-HXB2 and pEBB-JRFL (gifts from B. Chen) (23). Reverse phase HPLC (RP-HPLC) was performed using a C18 column (Vydac). All Ni affinity purifications used His-Select HC Nickel Affinity Gel (Sigma) or His-Select HC Nickel Magnetic Resin (Sigma). The NIH AIDS Research and Reference Reagent Program provided the following reagents: pNL4-3.Luc.R-E- (N. Landau), HeLa-CD4-LTR-ß-gal cells (M. Emerman), HOS-CD4-fusin cells (N. Landau).

Protein Expression, Purification, and Characterization

C37-H6 (C37), derived from the HXB2 Env sequence, was expressed and purified as previously described (9). Proteins used in this study were Bovine Pancreatic Trypsin Inhibitor (BPTI), Human Ubiquitin (Ub), Sperm Whale Myoglobin (Mb), enhanced Green Fluorescent Protein (GFP, from Clontech), and *E. coli* Maltose Binding Protein (MBP, from NEB). Linker sequences were Ser_4Gly_2 for BPTI-C37, Ub-C37, and GFP-C37 and Ser_5Gly_2 for Mb-C37 and MBP-C37 (linker sequences are slightly different for cloning reasons). The extended linker constructs had the following linker sequences: $SSS(GGGS)_3SSSGG$ (MBP1-C37) and

$SSS(GGGS)_3S(GGGS)_3SSSGG$ (MBP2-C37). The DNA encoding each protein was cloned into the following plamsids: pET9a (for BPTI-C37, Ub-C37, Mb-C37, and GFP-C37), pET20b (for BPTI-H6, Ub-H6, Mb-H6, and GFP-H6); pMAL-c2G (for MBP-H6, MBP-C37, MBP1-C37, and MBP2-C37). Proteins were expressed in BL21(DE3)pLysS (Novagen) for pET9a and pET20b vectors and XL1-Blue (Stratagene) for pMal-c2G vectors. All proteins have C-terminal His-tags (His_6) and were purified using Ni affinity chromatography.

BPTI required refolding after expression for correct formation of disulfide bonds. Briefly, after Ni affinity purification, BPTI-C37-H6 and BPTI-H6 were reduced with 100 mM ß-mercaptoethanol at pH 8 and dialysed into 5% acetic acid. The proteins were air oxidized in the presence of a 1:10 ratio of oxidized:reduced glutathione at pH 8, 4°C for 24 h. The correctly folded proteins were isolated using RP-HPLC and were confirmed by near-UV circular dichroism (Aviv 62DS) and measurement of trypsin inhibiting activity as previously described (24).

Cys-Gly-Gly-Asp-IZN36 (10) was cloned into pET14b and expressed in BL21(DE3)pLysS. IZN36 was purified from inclusion bodies (solubilized in 6 M GuHCl) using Ni affinity chromatography. The protein was then dialyzed into 5% acetic acid and purified by RP-HPLC. This material was reduced with TCEP (Pierce) and biotinylated at its unique Cys residue using Biotin-HPDP (Pierce). After biotinylation, the His-tag was removed by thrombin cleavage (Novagen) and the cleaved product was purified by RP-HPLC. The sequence of the final product is:

GSHMCGGDIKKEIEAIKKEQEAIKKKIEAIEKEISGIVQQQNNLLRAIEAQQHLLQLTVWG IKQLQARIL.

All protein masses were confirmed by MALDI or electrospray MS (U. of Utah Core Facility). All proteins were judged >98% pure by SDS-PAGE. Protein concentrations were measured by UV absorbance at 280 nm (25).

Surface Plasmon Resonance (SPR) Analysis

Binding experiments were performed using a Biacore 2000 optical biosensor (U. of Utah Protein Interaction Core Facility) equipped with research-grade CM5 sensor chips (Biacore). A standard coupling protocol was employed to immobilize streptavidin (SA, Pierce) (26). Biotinylated IZN36 was captured on a SA surface, and free SA surfaces served as references.

Binding analysis of C37 and C37-fusion proteins was performed at 25°C with a data collection rate of 2.5 Hz. The binding buffer (PBS (Gibco) + 0.005% P20 detergent (Biacore) + 1 mg/mL BSA (fraction V, Fisher)) was prepared, vacuum filtered, and degassed immediately prior to use. Stock solutions of C37, C37-fusion proteins, and corresponding control proteins (without C37) were prepared in binding buffer at 100 nM. Protein binding was analyzed by injecting samples for 1 min over the IZN36 and reference surfaces using KINJECT at a flow rate of 50-100 μL/min. The dissociations were monitored for 3 min. The IZN36 surfaces were completely regenerated using one 3 s pulse of 6 M guanidine-HCl or three 6 s pulses of 0.1% SDS.

Data from the reference flow cells were subtracted to remove systematic artifacts that occurred in all flow cells (27). The data were normalized to the highest point in the response curve to facilitate comparison. Binding at one concentration was analyzed using a 1:1 binding model in CLAMP (28) assuming enough information from the curvature of the responses to determine the approximate kinetic parameters for the reactions (29).

Cell-Cell Fusion and Viral Infectivity Assay

Cell-cell fusion was monitored as previously described (30). Briefly, HXB2 Env expressing CHO cells (gift from M. Krieger (31)) were mixed with HeLa-CD4-LTR-Beta-gal cells in the presence of inhibitors for 20 h at 37°C. Syncytia were stained with X-gal (Invitrogen) and counted.

Viral infectivity was measured as previously described (9). Briefly, pseudotyped viruses were produced by co-transfecting 293T cells using Fugene (Roche) with pNL4-3.Luc.R-E- and either pEBB-HXB2 or pEBB-JRFL. After 36-48 h, viral supernatants were collected and sterile filtered. HXB2 or JRFL pseudotyped virus was added to HOS-CD4-fusin or HOS-CD4-CCR5 cells, respectively, in the presence of inhibitors. HXB2 assays included 20 μg/mL DEAE-dextran (23). After 12 h, virus and inhibitor were removed and replaced with fresh media. Cells were lysed 40-44 h after infection using Glo Lysis buffer (Promega), and luciferase activity was measured using Bright-Glo (Promega). IC_{50} values for both assays were calculated by fitting data to the equation, $y = k/(1+ [\text{inhibitor}]/IC_{50})$, where y = normalized number of syncytia or luciferase activity and k = scaling constant (k = 1 for syncytia assay and is floated for viral infectivity assay, see Fig. 2B legend).

Assays for Inhibitor Proteolysis and Precipitation

C37 fusion inhibitors were incubated in tissue culture media (DMEM + 10% fetal bovine serum, Invitrogen) at 37°C for 20 h. Proteins were purified from the media by 1 h incubation at RT with magnetic Ni affinity beads. The resin was washed 3x with PBS and proteins were eluted by boiling in LDS sample buffer (Invitrogen). Eluted samples were separated by SDS-PAGE and visualized with SimplyBlue stain (Invitrogen). Unpurified media samples were analysed before and after centrifugation (10 min. at 18,000x g) by Western blot using polyclonal rabbit anti-His-

tag Ab (Abcam) and SuperSignal West Pico substrate (Pierce), as well as visually analysed for precipitate.

RESULTS

Production of Fusion Proteins

To test for steric constraints in accessing the gp41 N-trimer region, we have constructed a series of inhibitors containing a C-peptide attached to cargo proteins of various sizes (Fig. 1). The cargo partners used in this study were selected for the following properties: monomeric, soluble, globular, stable, tolerant to C-terminal additions, and free of non-specific peptide binding. Cargo proteins meeting these inclusion criteria and used in this study range from 6 to 41 kDa (Table 1). For these studies, C37 (9), the recombinant His-tagged version of the previously characterized synthetic peptide C34 (30,32), was used as the reference inhibitor. In each fusion protein, C37 is connected at its N-terminus to the cargo's C-terminus by a flexible six or seven residue Ser/Gly linker. This linker was designed to be long enough to allow the proper orientation of C37 as it binds to the N-trimer, but short enough for the attached cargo to prevent access to an occluded binding site. The N-terminus of C37 was chosen for attachment of cargo because this attachment site points away from the membrane (whereas, the C-terminus of C37 is expected to be near the viral membrane, and, therefore, less accessible – see Fig. 1). For each fusion protein, a matching control protein lacking C37 was also produced.

Size and inhibitory potency are inversely correlated

The inhibitory potency of each inhibitor was tested using a cell-cell fusion (syncytia) assay utilizing HXB2 env and two viral infectivity assays utilizing either HXB2 (X4) or JRFL (R5) envs (Table 1, Fig. 2). C37 shows high potency inhibition in all assays (IC_{50} = 0.85 to 8.2 nM).

Inhibition is slightly weaker than seen with C34 (30), as expected from the loss of helix-stabilizing synthetic blocking groups found in C34. For reference, the anti-gp41 Abs 2F5 and 4E10 have reported IC_{50} values of ~0.2 to 7 nM against HXB2/IIIB lab strains in cell-cell and viral infectivity assays similar to those used in this study (33,34).

The smallest fusion protein, BPTI-C37, also displays high potency in both assays, very similar to C37, demonstrating that our C37-cargo linker does not interfere with inhibitory activity. Ub-C37 is a slightly weaker (2.5 to 5.5-fold) inhibitor than C37, while Mb-C37 and GFP-C37 both show more substantial (21 to 65-fold) reductions in potency in both assays. MBP-C37 shows the most dramatic change with a 75 to 228-fold drop in potency. None of the control proteins (cargo without C37 peptide) inhibit at up to 1 μM (10 μM for MBP with JRFL env) in either assay (data not shown).

In general, the cell-cell fusion and viral infectivity assays show similar losses of activity with increasing size of the inhibitors with a slightly more pronounced effect on cell-cell fusion and JRFL-mediated viral entry. For HXB2 env we observe up to a 4-fold greater potency in cell-cell fusion vs. viral infectivity, as seen in studies of other fusion inhibitors (2,11,30,35). As expected, inhibitors were less potent against the primary isolate JRFL in the viral infectivity assay. For most of the inhibitors, the viral infectivity data shows a reproducible increase in infectivity (above the uninhibited values) at low inhibitor concentrations (see legend Fig. 2B). This "overshoot" has also occasionally been seen in other studies of fusion inhibitors (36-38), but has not been explained.

<u>The C-peptide remains accessible when linked to fusion partners</u>

To ensure that linkage of C37-H6 to each of the partner proteins did not affect the accessibility of C37 for binding to a sterically open target, the fusion proteins and C37 were

assayed for binding to IZN36, a soluble mimic of the N-trimer (10), using SPR. Each fusion protein was flowed over the control and IZN36 surfaces. C37 reversibly bound to IZN36 with a low nM K_D (Fig. 3). The calculated K_D's for the fusion proteins are clustered in a narrow range around the C37 value (2-fold lower to 2-fold higher). The estimated kinetic parameters are similarly clustered, ranging from 3.2-fold slower to 1.4-fold faster (association rate) and up to 3.2-fold slower (dissociation rate). These rates are only approximate due to small systematic deviations from the fitting model, but as expected there is a slight trend towards slower association and dissociation rates with increasing molecular weight. These small differences in binding kinetics are likely responsible for some of the variation in potency observed here but rule out distinct binding kinetics as the major contributor to the substantial differences in potency among these inhibitors. These results also show that the accessibility and affinity of C37 are not significantly altered in the context of the fusion proteins. None of the cargo proteins alone showed measurable association with IZN36 at 100 nM (inset, Fig. 3).

Partial restoration of inhibitory potency with extended Gly/Ser linkers

To test if a longer linker could overcome the steric block and restore inhibitory potency of our weakest inhibitor, we extended the flexible linker in MBP-C37 from its original length of 7 amino acids to 20 (MBP1-C37) or 33 (MBP2-C37) using Gly/Ser residues (Table 1). Both extended linker inhibitors exhibit partial recovery of inhibitory potency. Compared to MBP-C37, MBP1-C37 and MBP2-C37 are 2.3 to 2.9-fold and 2.6 to 6.1-fold more potent, respectively (Table 1). Compared to MBP-C37, MBP1- and MBP2-C37 interact similarly with IZN36 as measured by SPR (K_ds vary by <20%, k_a and k_d are <2-fold higher). In contrast to the other cargo-C37 fusions, a significant portion of the increased potency in MBP1- and MBP2-C37 may be attributable to an increased association rate.

Stability of fusion proteins during fusion assays

Inhibitors were analyzed for precipitation or extensive proteolysis to demonstrate that these processes did not cause the observed decrease in potency of the fusion proteins. C37 and the C37-fusions were incubated in tissue culture media at 37°C for 20 h to simulate the harshest conditions faced by the inhibitors during the cell-cell fusion and viral infectivity assays. We observed only trace (<2%) degradation for all of the inhibitors (data not shown), allowing us to conclude that proteolysis did not cause a significant decrease in the potency of our inhibitors. We cannot, however, rule out the contribution of minor proteolytic breakdown products to increased inhibitory potency, particularly for the least potent inhibitors (1% contamination with free C37 would result in an apparent cell-cell fusion IC_{50} value of ~100 nM for a completely inactive inhibitor). Therefore, the described potencies of the inhibitors presented in this study should be considered an upper limit. An anti-His tag Western blot comparing samples before and after high-speed centrifugation revealed no precipitation (data not shown).

DISCUSSION

In principle, the gp41 N-trimer is an especially promising inhibition target, but despite the generation of numerous Abs with tight and specific binding against various mimics of the N-trimer, none of these Abs display broadly neutralizing activity, reviewed in (17). Our results suggest that HIV may have developed a strong steric defense against immune attack for its critical N-trimer region. In this study we show that the gp41 N-trimer region has poor accessibility to large proteins. It is a logical extrapolation of the data presented here that a protein as large as IgG (150 kDa), even though it forms a somewhat elongated shape, will suffer a steric block at least as severe as we observe with our largest protein, MBP (41 kDa), which is smaller than the individual (~50 kDa) domains of an IgG. This defense may be a major factor in frustrating efforts to induce neutralizing Abs against the N-trimer region and may also explain why such neutralizing Abs against the N-trimer have not yet been observed in infected patients.

The steric restriction of the N-trimer stands in stark contrast to apparent accessibility of the extreme C-terminal region of the gp41 ectodomain (between the C-peptide region and the transmembrane domain). The only known potent and broadly neutralizing Abs against gp41 (2F5 and 4E10) target this region (22). Recent studies have suggested that this region may adopt a helical or beta-strand conformation or cycle between the two (33,39). For the most thoroughly studied Ab against this region, 2F5, a full length IgG (~150 kDa) is more potent than the Fab (~50 kDa) (33), suggesting a freely accessible site.

There is also suggestive evidence that the C-peptide region may be more accessible than the N-trimer. The designed proteins 5-helix (25 kDa) (9) and N_{CCG}-gp41 (35 kDa) (40) target the C-peptide region and are potent entry inhibitors. Recently, a Pseudomonas Endotoxin (PE) fusion with 5-helix (5-helix-PE, 65 kDa) was shown to inhibit viral entry with similar potency as 5-

helix (41), although a toxic effect from PE may mask a loss of potency. While the C-peptide region is likely accessible, it is difficult to target for vaccine studies, as it is unclear what organized structure (if any) this region adopts during viral entry.

C37 inhibits viral fusion by binding along the full length of the surface groove of the N-trimer, including the deep hydrophobic "pocket" region, previously shown to be an essential player in viral fusion. Inhibitors that specifically target this pocket have been developed ((10), B.D. Welch and M.S. Kay, unpublished results). In future studies, it will be important to test such pocket specific inhibitors to see if they can circumvent the steric block observed here. It will also be important to check if cargo fused to the C-terminus of C37 shows a similar pattern of steric blockage.

The steric block we observe in the gp41 N-trimer is reminiscent of steric restrictions observed in gp120. These restrictions have been attributed to glycosylation ("glycan shield") (42,43) and/or inaccessible antigens (38,44,45). Previous studies with several broadly neutralizing gp120 Abs have shown that smaller versions of these Abs (Fabs or scFvs) often have significantly improved potency, despite a loss of avidity (38,46). The N-trimer steric block observed here may be more strict than seen in gp120. Proteins the size of Fabs (~50 kDa) and scFvs (~25 kDa) are already too large to fully access the gp41 N-trimer. Interestingly, the N-trimer region does not contain any glycosylation sites, probably due to its ultimate complete burial in the six-helix bundle structure. The N-trimer, however, may be affected by nearby glycosylation sites in gp120 or other regions of gp41 (the C-peptide region and N/C-peptide connecting loop are extensively glycosylated). A glycosylation site near the gp120 V3 loop has been shown to affect accessibility of the 2F5 Ab to its gp41 epitope in resistant strains (43).

Implications for C-peptide inhibitors

Our results suggest that attempts to improve the longevity of C-peptide inhibitors in the bloodstream may also be frustrated by steric issues. For instance, T-20, a 36-residue peptide recently approved by the FDA, is rapidly cleared from the bloodstream by kidney filtration, dramatically increasing dosing requirements. A reasonable approach for prevention of this rapid clearance is to crosslink C-peptide inhibitors to larger proteins (e.g., albumin) or high molecular weight polyethylene glycol (PEG), which also can reduce peptide immunogenicity (47). Our results suggest that these straightforward approaches will likely reduce the potency of modified C-peptides, although use of smaller proteins or low molecular weight PEG may lessen this effect. Our extended loop constructs suggest the possibility that longer linkers between these bulking groups and the C-peptide inhibitor could improve accessibility to the N-trimer. Stiffer (e.g., helical) linkers may provide better separation from large fusion partners and restore inhibitory potency better than the flexible Gly/Ser linkers employed here.

An important caveat to applying our results to T-20 is that, compared to C34, T-20 is derived from a gp41 sequence shifted about ten amino acids towards the C-terminus and its binding site extends beyond the N-trimer region. Although they are thought to have a similar mechanism of action, T-20 and C34 (and the similar T-1249) vary in their potencies against different HIV-1 strains and their sensitivities to resistance mutations (18,48,49).

Future directions - Overcoming the steric block

We hope that the observation of this steric block can be used to improve the chances of discovering a broadly neutralizing Ab against this valuable HIV target, rather than discouraging this effort. Specifically, we suggest that a designed, sterically restricted N-trimer antigen could be used to generate, boost, or screen for potent neutralizing Abs able to overcome the steric

block. Currently used mimics of the N-trimer region (e.g., 5-helix, IZN36, N_{CCG}-gp41) could be modified by attachment to bulky proteins or large inert particles such that only Abs capable of penetrating a sterically recessed target would be selected.

Neutralizing Abs against sterically blocked gp120 targets often utilize unusually long CDR H3 loops to access recessed antigens (33,46,50). The insertion of longer linkers connecting MBP to C37 results in partial recovery of inhibitory activity, suggesting that extended CDR H3 loops may help penetrate the steric block on the gp41 N-trimer. These Abs are difficult to generate in small animals, as Abs in primates have longer CDR H3 loops on average than rodents (51). Potent N-trimer Abs may be more easily found using strategies that enrich for this type of Ab (e.g., engineered Ab libraries, Ab phage display, immunization of primates). Alternatively, very high affinity (sub-nM) Abs against the N-trimer may still be sufficiently neutralizing despite a substantial decrease in potency caused by the steric block. Recently, Merck has reported preliminary results on an antibody that binds to the N-trimer region and possesses neutralizing activity against some HIV strains (M. Miller and R. Geleziunas, Abst. 13th International HIV Drug Resistance Workshop, abstract #9, 2004). No detailed information on this Ab has yet been published, but it will be interesting to see if or how this Ab circumvents the steric block we observe here (e.g., high affinity Ab that can tolerate several hundred-fold loss in activity, extended variable loops, specific targeting of a subsite in the N-trimer).

Finally, our results suggest that the traditional depiction of the prehairpin intermediate as a symmetric structure (e.g., Fig. 1) may be inaccurate. The steric block of the N-trimer and apparent accessibility of the C-peptide region show that they reside in very different environments. Possible sources of this asymmetry include interactions with gp120, other regions

of gp41, or cell surface proteins, as well as glycosylation and differences between the curvature of the viral and cellular membranes.

Acknowledgments

We thank Peter Kim, in whose lab this work originated. We also thank D. Goldenberg (BPTI folding); D. Myszka and his Protein Interaction Group at the U. of Utah (SPR); N. Welker, B. Stadtmueller, B. Kelly, C. Kieffer, and K. Rigby (protein production); W. Sundquist and C. Hill (advice and support); U. von Schwedler, B. Chen, V. Planelles, and M. Root (virology advice and assistance); and D. Eckert (invaluable discussions and critical review of the manuscript); This work was partially supported by NIH grant P01GM066521 (to M.S.K.) and an NIH Training Grant in Biological Chemistry (to B.D.W.).

REFERENCES

1. Lu, M., Blacklow, S. C., and Kim, P. S. (1995) *Nat Struct Biol* **2**, 1075-1082
2. Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B., and Matthews, T. J. (1994) *Proc Natl Acad Sci U S A* **91**, 9770-9774
3. Chan, D. C., and Kim, P. S. (1998) *Cell* **93**, 681-684
4. Tan, K., Liu, J., Wang, J., Shen, S., and Lu, M. (1997) *Proc Natl Acad Sci U S A* **94**, 12303-12308
5. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) *Nature* **387**, 426-430
6. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) *Cell* **89**, 263-273
7. Munoz-Barroso, I., Durell, S., Sakaguchi, K., Appella, E., and Blumenthal, R. (1998) *J Cell Biol* **140**, 315-323
8. Melikyan, G. B., Markosyan, R. M., Hemmati, H., Delmedico, M. K., Lambert, D. M., and Cohen, F. S. (2000) *J Cell Biol* **151**, 413-423
9. Root, M. J., Kay, M. S., and Kim, P. S. (2001) *Science* **291**, 884-888
10. Eckert, D. M., and Kim, P. S. (2001) *Proc Natl Acad Sci U S A* **98**, 11187-11192
11. Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A., and Kim, P. S. (1999) *Cell* **99**, 103-115
12. Eckert, D. M., and Kim, P. S. (2001) *Annu Rev Biochem* **70**, 777-810
13. Burton, D. R., Desrosiers, R. C., Doms, R. W., Koff, W. C., Kwong, P. D., Moore, J. P., Nabel, G. J., Sodroski, J., Wilson, I. A., and Wyatt, R. T. (2004) *Nat Immunol* **5**, 233-236
14. Wild, C., Oas, T., McDanal, C., Bolognesi, D., and Matthews, T. (1992) *Proc Natl Acad Sci U S A* **89**, 10537-10541
15. Wild, C., Greenwell, T., and Matthews, T. (1993) *AIDS Res Hum Retroviruses* **9**, 1051-1053
16. Jiang, S., Lin, K., Strick, N., and Neurath, A. R. (1993) *Nature* **365**, 113
17. Weiss, C. D. (2003) *AIDS Rev* **5**, 214-221
18. Rimsky, L. T., Shugars, D. C., and Matthews, T. J. (1998) *J Virol* **72**, 986-993
19. Golding, H., Zaitseva, M., de Rosny, E., King, L. R., Manischewitz, J., Sidorov, I., Gorny, M. K., Zolla-Pazner, S., Dimitrov, D. S., and Weiss, C. D. (2002) *J Virol* **76**, 6780-6790
20. Opalka, D., Pessi, A., Bianchi, E., Ciliberto, G., Schleif, W., McElhaugh, M., Danzeisen, R., Geleziunas, R., Miller, M., Eckert, D. M., Bramhill, D., Joyce, J., Cook, J., Magilton, W., Shiver, J., Emini, E., and Esser, M. T. (2004) *J Immunol Methods* **287**, 49-65
21. Louis, J. M., Nesheiwat, I., Chang, L., Clore, G. M., and Bewley, C. A. (2003) *J Biol Chem* **278**, 20278-20285
22. McGaughey, G. B., Barbato, G., Bianchi, E., Freidinger, R. M., Garsky, V. M., Hurni, W. M., Joyce, J. G., Liang, X., Miller, M. D., Pessi, A., Shiver, J. W., and Bogusky, M. J. (2004) *Curr HIV Res* **2**, 193-204
23. Chen, B. K., Saksela, K., Andino, R., and Baltimore, D. (1994) *J Virol* **68**, 654-660
24. Coplen, L. J., Frieden, R. W., and Goldenberg, D. P. (1990) *Proteins* **7**, 16-31
25. Edelhoch, H. (1967) *Biochemistry* **6**, 1948-1954
26. Johnsson, B., Lofas, S., and Lindquist, G. (1991) *Anal Biochem* **198**, 268-277
27. Myszka, D. G. (1999) *J. Mol. Recognit.* **12**, 279-284
28. Myszka, D. G., and Morton, T. A. (1998) *Trends Biochem. Sci.* **23**, 149-150
29. Canziani, G. A., Klakamp, S., and Myszka, D. G. (2004) *Anal Biochem* **325**, 301-307

30. Chan, D. C., Chutkowski, C. T., and Kim, P. S. (1998) *Proc Natl Acad Sci U S A* **95**, 15613-15617
31. Kozarsky, K., Penman, M., Basiripour, L., Haseltine, W., Sodroski, J., and Krieger, M. (1989) *J Acquir Immune Defic Syndr* **2**, 163-169
32. Malashkevich, V. N., Chan, D. C., Chutkowski, C. T., and Kim, P. S. (1998) *Proc Natl Acad Sci U S A* **95**, 9134-9139
33. Zwick, M. B., Komori, H. K., Stanfield, R. L., Church, S., Wang, M., Parren, P. W., Kunert, R., Katinger, H., Wilson, I. A., and Burton, D. R. (2004) *J Virol* **78**, 3155-3161
34. Stiegler, G., Kunert, R., Purtscher, M., Wolbank, S., Voglauer, R., Steindl, F., and Katinger, H. (2001) *AIDS Res Hum Retroviruses* **17**, 1757-1765
35. Sia, S. K., and Kim, P. S. (2003) *Proc Natl Acad Sci U S A* **100**, 9756-9761
36. Wei, X., Decker, J. M., Liu, H., Zhang, Z., Arani, R. B., Kilby, J. M., Saag, M. S., Wu, X., Shaw, G. M., and Kappes, J. C. (2002) *Antimicrob Agents Chemother* **46**, 1896-1905
37. Reeves, J. D., Gallo, S. A., Ahmad, N., Miamidian, J. L., Harvey, P. E., Sharron, M., Pohlmann, S., Sfakianos, J. N., Derdeyn, C. A., Blumenthal, R., Hunter, E., and Doms, R. W. (2002) *Proc Natl Acad Sci U S A* **99**, 16249-16254
38. Labrijn, A. F., Poignard, P., Raja, A., Zwick, M. B., Delgado, K., Franti, M., Binley, J., Vivona, V., Grundner, C., Huang, C. C., Venturi, M., Petropoulos, C. J., Wrin, T., Dimitrov, D. S., Robinson, J., Kwong, P. D., Wyatt, R. T., Sodroski, J., and Burton, D. R. (2003) *J Virol* **77**, 10557-10565
39. Barbato, G., Bianchi, E., Ingallinella, P., Hurni, W. H., Miller, M. D., Ciliberto, G., Cortese, R., Bazzo, R., Shiver, J. W., and Pessi, A. (2003) *J Mol Biol* **330**, 1101-1115
40. Louis, J. M., Bewley, C. A., and Clore, G. M. (2001) *J Biol Chem* **276**, 29485-29489
41. Root, M. J., and Hamer, D. H. (2003) *Proc Natl Acad Sci U S A* **100**, 5016-5021
42. Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M., Saag, M. S., Komarova, N. L., Nowak, M. A., Hahn, B. H., Kwong, P. D., and Shaw, G. M. (2003) *Nature* **422**, 307-312
43. McCaffrey, R. A., Saunders, C., Hensel, M., and Stamatatos, L. (2004) *J Virol* **78**, 3279-3295
44. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) *Nature* **393**, 648-659
45. Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R., and Wilson, I. A. (2001) *Science* **293**, 1155-1159
46. Choe, H., Li, W., Wright, P. L., Vasilieva, N., Venturi, M., Huang, C. C., Grundner, C., Dorfman, T., Zwick, M. B., Wang, L., Rosenberg, E. S., Kwong, P. D., Burton, D. R., Robinson, J. E., Sodroski, J. G., and Farzan, M. (2003) *Cell* **114**, 161-170
47. Harris, J. M., and Chess, R. B. (2003) *Nat Rev Drug Discov* **2**, 214-221
48. Reeves, J. D., Miamidian, J. L., Biscone, M. J., Lee, F. H., Ahmad, N., Pierson, T. C., and Doms, R. W. (2004) *J Virol* **78**, 5476-5485
49. Kliger, Y., Gallo, S. A., Peisajovich, S. G., Munoz-Barroso, I., Avkin, S., Blumenthal, R., and Shai, Y. (2001) *J Biol Chem* **276**, 1391-1397
50. Zwick, M. B., Parren, P. W., Saphire, E. O., Church, S., Wang, M., Scott, J. K., Dawson, P. E., Wilson, I. A., and Burton, D. R. (2003) *J Virol* **77**, 5863-5876
51. Wu, T. T., Johnson, G., and Kabat, E. A. (1993) *Proteins* **16**, 1-7

Table 1: IC_{50} in (nM) of fusion proteins in cell-cell fusion and viral infectivity assays

Protein	Fusion partner MW	Cell-cell fusion	IC_{50} ratio (cell-cell fusion)	Viral infectivity HXB2	IC_{50} ratio HXB2	Viral infectivity JRFL	IC_{50} ratio JRFL
C37	0	0.85	1.0	2.8	1.0	8.2	1.0
BPTI-C37	6.5 kDa	1.5	1.8	3.1	1.1	4.8	0.6
Ub-C37	8.6 kDa	4.7	5.5	6.8	2.5	37.7	4.6
Mb-C37	17 kDa	30.8	36.2	58.0	21.0	414	50.5
GFP-C37	27 kDa	28.9	34.0	118	42.8	533	65.0
MBP-C37	41 kDa	192	225	206	74.8	1874	228
MBP1-C37	41 kDa	75.1	88.4	88.9	32.2	640	78.0
MBP2-C37	41 kDa	31.2	36.7	79.2	28.7	516	62.9

IC_{50} standard errors are <25% for both assays. IC_{50} ratios are relative to C37.

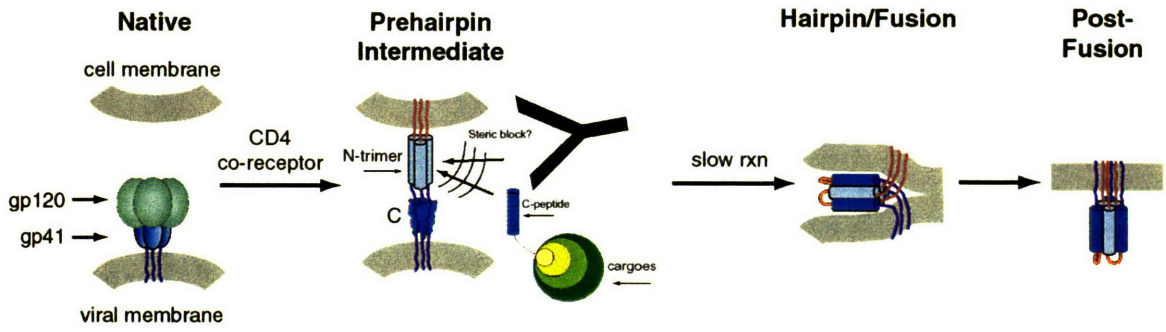


Fig. 1. Model of HIV-1 membrane fusion pathway (adapted from (9)). Formation of the trimer-of-hairpins drives the viral and cellular membranes together, leading to fusion. The N-peptide region (gray), C-peptide region (blue), gp120 (green), gp41 (light blue), gp41 fusion peptide (red), and transmembrane domain (purple) are shown. gp120 is removed from the prehairpin intermediate for clarity. Also shown are a series of C37 fusion proteins of different sizes and an anti-N-trimer Ab attempting to access the N-trimer, but potentially blocked by a steric restriction. Sizes of the Ab and fusion proteins are only approximately to scale.

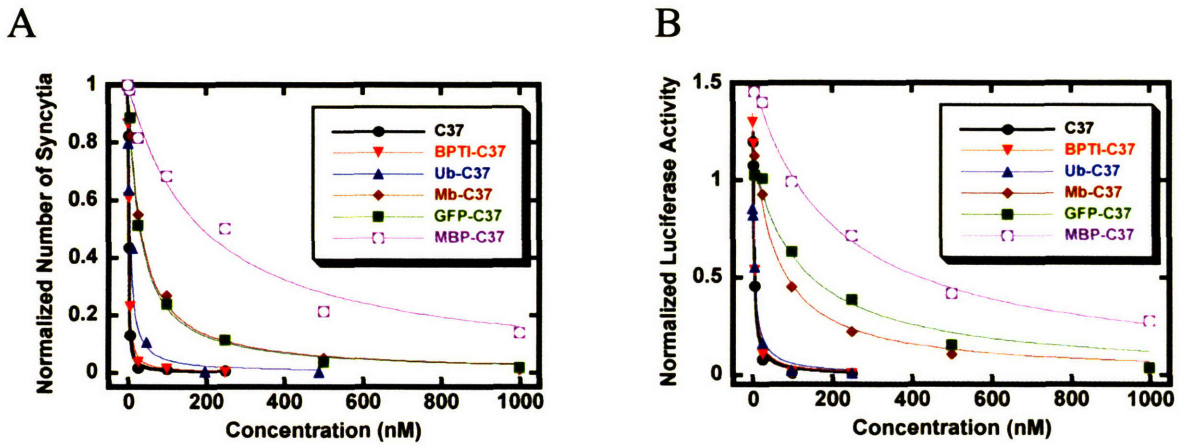


Fig. 2. Inhibitory activity of C37 and C37-fusion proteins. Data points are averages of at least quadruplicate measurements. Data are normalized to uninhibited fusion activity. A. Cell-cell fusion assay (HXB2 env). Standard errors of each point are <0.05. B. Viral infectivity assay (HXB2 env). Standard errors of each point are <0.1. An "overshoot" is observed at low inhibitor concentrations (data above 1.0). Fitting the viral infectivity data to a simple Langmuir equation with a fixed zero inhibitor point produces noticeable deviation from the data near the zero point because of this overshoot. Fitting the data without fixing the zero inhibitor point (as done in this study) improves the quality of the fit, but does not significantly affect the relative IC_{50} values of the inhibitors.

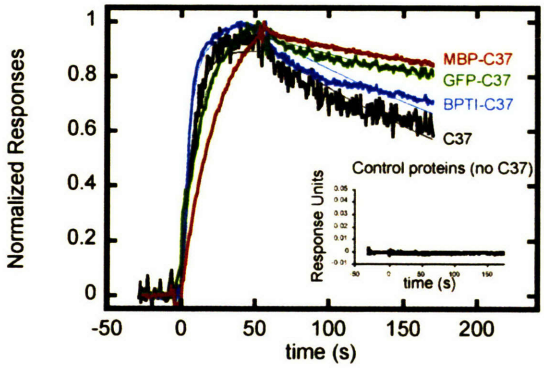


Fig. 3. Binding of C37 and C37-fusion inhibitors to IZN36 measured by SPR. Responses for representative inhibitors were normalized and overlaid to facilitate their comparison (thick traces). Fits to the interaction model are included (thin traces). Inset: Interaction of control proteins (no C37) with IZN36 surface.