Structural and Functional Studies of Bacterial Adhesion Proteins Staphylococcus aureus immunoglobulin-binding proteins Sbi and SpA and their interactions with serum proteins

Atkins, Karen

Award date:
2006

Awarding institution:
University of Bath

Structural and Functional Studies of Bacterial Adhesion Proteins

Staphylococcus aureus immunoglobulin-binding proteins Sbi and SpA and their interactions with serum proteins

Karen Atkins

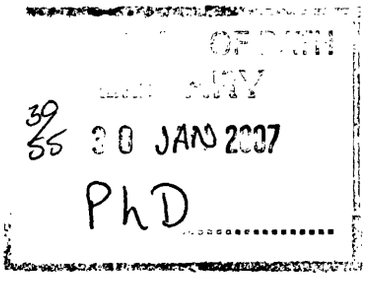
A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

July 2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346


30 JAN 2007
PhD

1 Acknowledgements

I would like to thank my supervisor Dr J. van den Elsen for the opportunity to study in his lab.

I would also like to thank Dr A.T. Jenkins in the Department of Chemistry for his help and assistance with the SPR experiments, and for the use of the SPR machine.

Dr. D. Svergun and Pau Bernado from EMBL, Hamburg Outstation assisted with the SAXS measurements and for processing the scattering data.

Dr E. Feil, Dr J. Cooper and E. Deavin carried out the sequencing and evolution work of the SpA and Sbi genes.

Final year project students including Joe Dukes and Aimee Smith for carrying out cloning and purification projects on various regions of the Sbi gene.

Dr de Groot, University of Utrecht for supplying β_2GP1 and carrying out the ELISA studies of β_2GP1 and Sbi.

The Medical Research Council (MRC) funded this PhD studentship.

Dr Susan Crennell for technical assistance with software used in homology modelling including CCP4 programs and O.

2 Contents

3 Summary

Staphylococcus aureus cell-surface proteins have been proposed to play important roles in bacterial virulence. The immunoglobulin (Ig)-binding protein A (SpA) has been well characterised and it is known to bind most mammalian IgG Fc and Fab fragments, thereby inhibiting phagocytosis and acting as a B cell superantigen. Recently, a second *S. aureus* Ig-binding protein has been discovered, Sbi, which also binds a second serum protein β_2-glycoprotein 1 (β_2GP1). In this thesis the Ig-binding characteristics of Sbi are analysed and compared to those of SpA. Both Sbi and SpA were found to bind all mammalian IgGs tested, but significant differences were found between the two proteins in their ability to form high molecular weight complexes with some mammalian IgGs. Sbi has one functional Ig-binding domain and sequence alignment predicts a second putative Ig-binding domain. Here we show that both Ig-binding domains of Sbi bind IgG with high affinity and with specificity for the Fc fragment only. Homology modelling of the two Ig-binding domains, based on the structure of domain D of SpA, supports our observation that Sbi has conserved Fc binding characteristics but does not bind Fab fragments. Structural models of Sbi also reveal that the variation in the affinity values of Sbi and SpA for IgG is mainly due to small variations in IgG sequence on certain loops at the IgG-ligand interface and at the C_H2-C_H3 interface. Multidomain models of Sbi and SpA are used to explain why SpA forms large soluble complexes with Fc and insoluble complexes with intact IgG, while Sbi forms insoluble complexes with both. Structural studies using small angle X-ray scattering, reveal that the N-terminal region of Sbi is composed of 4 globular domains, the two Ig-binding domains and two other three-helix bundle domains with novel functions.

4 Abbreviations

β_2GP1 β_2-glycoprotein 1

BLAST Basic Local Alignment Search Tool

Cna Collagen-binding protein

C_H the constant region of the Ig heavy chain

C_L the constant region of the Ig light chain

ClfA/B Clumping factor A and B

***coa*⁻** coagulase negative mutant of *S. aureus*

Ebps Elastin-binding protein

ECM Extracellular matrix

FnBPA/B Fibronectin binding proteins A and B

GISA glycopeptide intermediately susceptible *S. aureus*, also referred to as VISA, vancomycin intermediately susceptible *S. aureus*

***hla*⁻** haemolysin negative mutant of *S. aureus*

Ig Immunoglobulin

IgA Immunoglobulin A

IgD Immunoglobulin D

IgE Immunoglobulin E

IgG Immunoglobulin G

IgM Immunoglobulin M

Igs Immunoglobulins

MOE Molecular Operating Environment

MRSA Methicillin resistant *S. aureus*

MSCRAMM Microbial surface components recognising adhesive matrix molecules

MSSA Methicillin susceptible *S. aureus*

NMR Nuclear Magnetic Resonance Spectroscopy

PDB Protein Database

SAXS Small-angle X-ray scattering

Sbi *S. aureus* Ig-binding protein Sbi

SCCmec ***staphylococcal* cassette chromosome *mec***

SpA *S. aureus* Ig-binding protein SpA

TEV Tobacco Etch Virus

TNFR1 tumour-necrosis factor-α receptor 1

V_H	the variable region of the Ig heavy chain
V_L	the variable region of the Ig light chain
vWF	von Willebrand factor

5 Chapter 1

5.1 Cell adhesion properties of proteins from *Staphylococcus aureus* and their role in virulence

In 1884 Rosenbach described the isolation of two species of *Staphylococcus*: *Staphylococcus aureus* and what is now termed *Staphylococcus epidermidis.* This genus of bacteria is spherical and forms typical grape shaped clusters as the bacteria divides on two planes. Currently, more than 20 species and subspecies of this genus have been described, of which *S. aureus* is the most common. *S. aureus* derives its name from the fact that under certain conditions, the colonies are a gold colour (aureus = gold). The bacterium is non-sporing, non-motile, and generally a capsule-less cell of 1μm in diameter. It is generally found on the skin and in the nasal passages of its host. It does however cause a large range of infections in its host. These range from harmless infections, such as pimples, impetigo, boils, to serious wound infections and abscesses, septicaemia, osteomyelitis, endocarditis, food poisoning and pneumonia (Todar 2005, Prescott, Harley and Klein 2002 Ch 39, Greenwood *et al* 1997 Ch 15). It has been determined that 20% of individuals are always colonised, 60% are intermittent carriers, and 20% are never carriers (Foster 2004).

S. aureus has hit the headlines in the last few years due to the increase in infections by MRSA – **m**ethicillin **r**esistant *Staphylococcus aureus.* The first antibiotic, penicillin, was introduced in the early 1940's, with the first cases of resistance recorded in 1942 (Lowy 2003). Methicillin, the first of the semi synthetic penicillinase-resistant penicillins, was introduced in 1959 to treat infections caused by the penicillin resistant *staphylococci.* However by 1961, there were reports of resistance to this antibiotic, in what are now known as MRSA strains (Enright *et al* 2002). Since the 1960's, other antibiotics have been developed, although there is a lack of novel drugs currently in development (Foster 2004). Today, the drug used to treat MRSA infections is the glycopeptide antibiotic vancomycin, however in 1997 the first cases of vancomycin resistance were reported. These *Staphylococci* strains with decreased susceptibility to this drug (**g**lycopeptide **i**ntermediately susceptible *S. aureus* GISA) are all MRSA

strains, and are of great concern as there are no drugs currently available that can target such strains (Enright *et al* 2002, Lowy 2003).

How have antibiotic resistant strains arisen?

S. aureus resistance to methicillin occurs when the *S. aureus* genome contains the *mecA* gene. Katayama, Ito and Hiramatsu (2000) determined that this gene is carried on a novel genetic element: the ***staphylococcal*** **c**assette **c**hromosome ***mec*** (SCC*mec*); currently five different cassettes have been identified (Foster 2004). The resistance was thought to have occurred once, with a single strain acquiring the SCC*mec* cassette (Kreiswirth *et al* 1993), however it has now been established that the different MRSA strains have evolved from MSSA (**m**ethicillin-susceptible ***S. aureus***) strains (Enright *et al.*, 2002). According to this study, the major MRSA clones have all evolved from the successful MSSA strains and the GISA strains are arising from the major MRSA strains. In order to tackle this resistance problem it is necessary to develop novel drugs that will be effective to those strains that are already resistant to the currently available drugs.

5.1.1 How does *S. aureus* interact with its host?

One step toward the development of novel drugs against *S. aureus* is to understand how the bacterium interacts with its host, in the case of normal carriage as well as infection. It has been determined that *S. aureus* can interact with a number of host proteins though cell-surface adhesion molecules, these include components of the host's **e**xtracellular **m**atrix (ECM), host serum proteins, or proteins that are present on the surface of the host's cells (i.e. receptors). *S. aureus* binds to the ECM protein elastin via an elastin-binding protein (EbpS) and fibronectin binding proteins (FnBPA and FnBPB). Examples of *S. aureus* proteins that interact with serum proteins include fibrinogen, via clumping factors A and B (ClfA and ClfB), and immunoglobulin G, via by *S. aureus* Ig-binding protein A (SpA) and a recently discovered second Ig-binding protein Sbi. The adhesion proteins involved are often referred to as **m**icrobial **s**urface **c**omponents **r**ecognising **a**dhesive **m**atrix **m**olecules (MSCRAMM) (Foster and Hook 1998, O'Brien *et al* 2002, Zhang *et al* 1998). Adhesion proteins are not exclusive to *S. aureus*, there are other known examples in the same genus, and also in other infective bacteria. For example the adhesion protein SdrX found in *Staphylococcus capitis* binds collagen (Liu *et al* 2004), and *Staphylococcus*

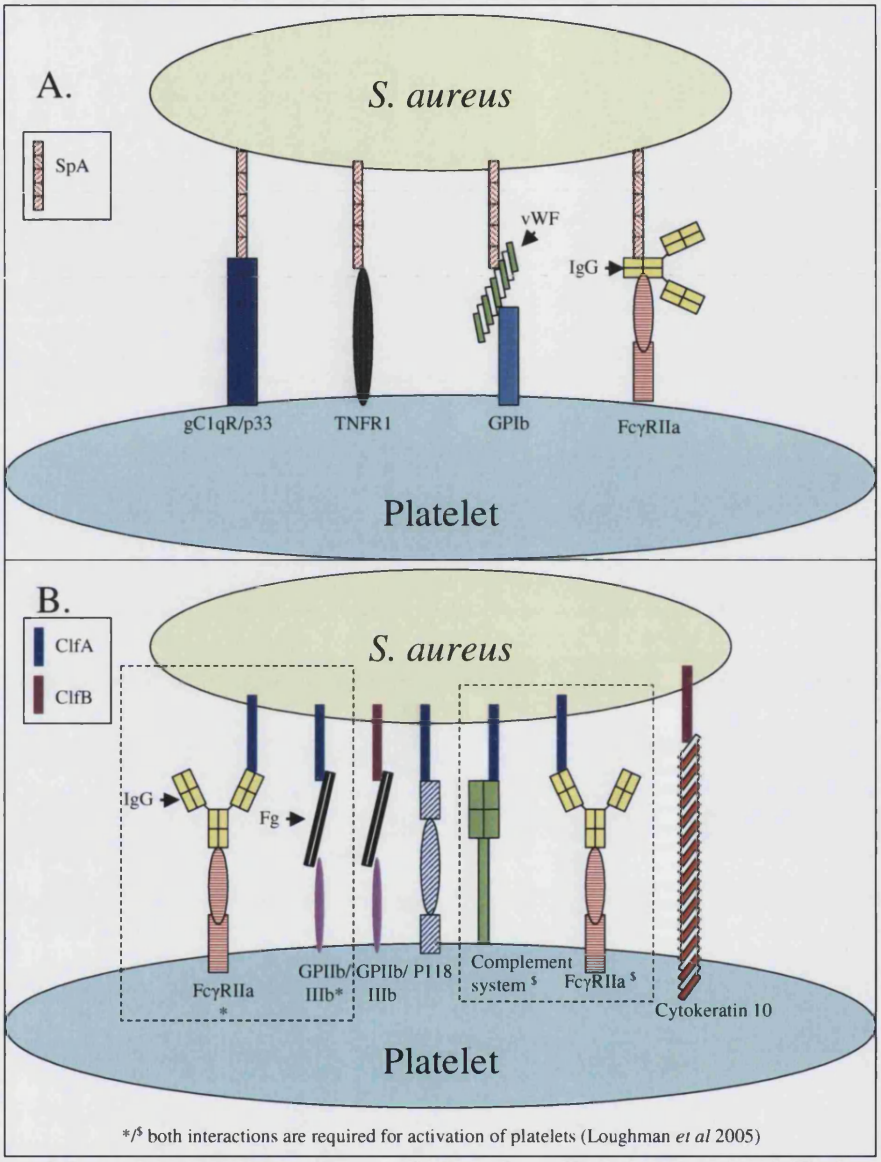


Figure 1 Interactions between S. aureus and platelet cells mediated by **(A)** SpA and **(B)** clumping factors ClfA and ClfB (adapted from O'Brien *et al* 2002).

lugdunensis contains a fibrinogen-binding protein (Nilsson *et al* 2004). *Streptococcal* protein G (SpG) and *Peptostreptococcal* protein L both bind immunoglobulins

(Tashiro and Montelione 1995). Binding to receptor proteins on host cells generally occurs through bridging between the bacterium and the host cell utilising the host serum proteins binding between both cells, but some *S. aureus* proteins have been

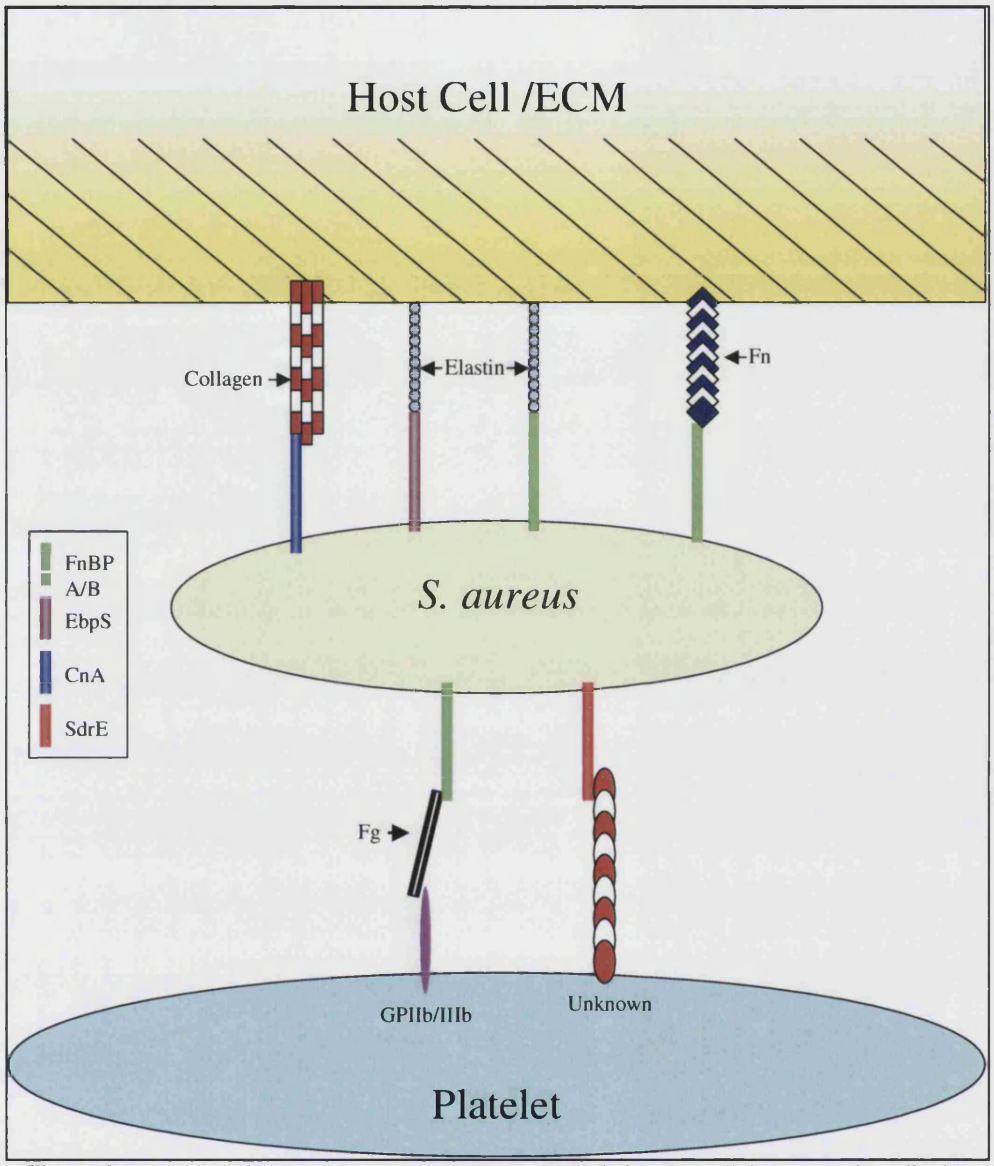


Figure 2 Interactions between *S. aureus*, platelet cells and the extracellular matrix by adhesion proteins: fibronectin binding proteins FnBPA and FnBPB; Elastin binding protein EpbS, collagen binding protein Cna, and SdrE (references in text).

identified that can bind directly to host cells. Examples of *S. aureus* interactions with host ECM and serum proteins are shown in Figure 1 and Figure 2, (also see Appendix 1), and are described in more detail below.

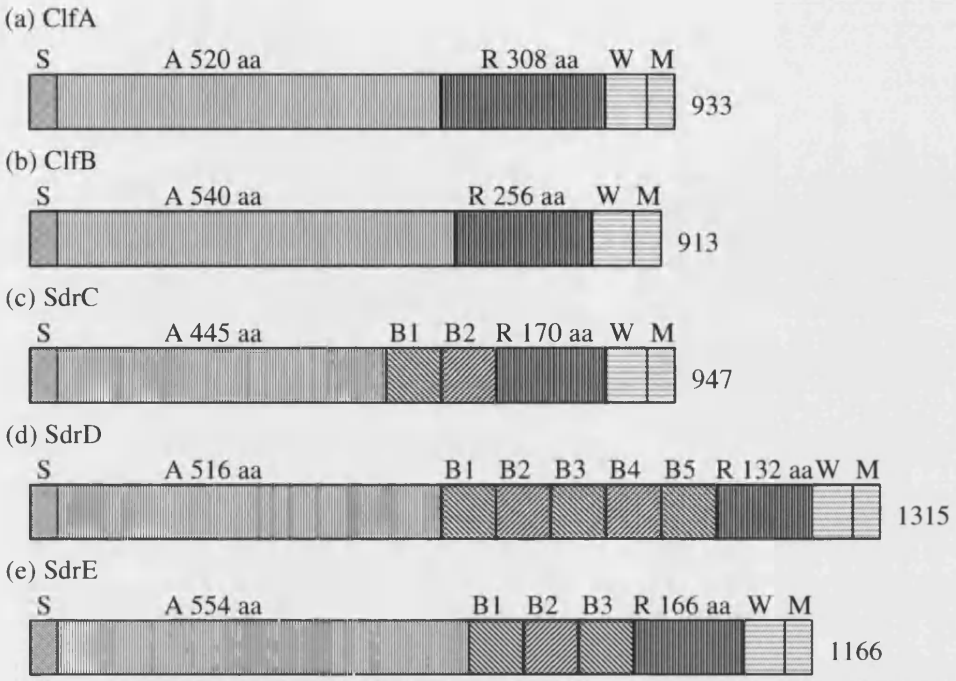


Figure 3 Structural arrangement of the serine-aspartate repeat proteins of *S. aureus* showing their relative arrangement of the (S) signal domain, the A domain, the B repeats, the (R) serine-aspartate repeat region, the (W) cell-wall spanning region, and the (M) membrane spanning region, which contains the LPXTG motif. The lengths of the A domain, the R region, and the total length of the protein is indicated (Josefsson *et al* 1998).

5.1.1.1 MSCRAMM adhesion proteins of *S. aureus*

The mechanism of attachment by *S. aureus* to its host is very complex. There are at least nine different adhesion proteins with a known ligand or function, most of which are known to bind more than one ligand. It has also been found that a number of *S. aureus* ligands interact with more than one adhesion protein. A feature common to most *S. aureus* adhesion proteins is the LPXTG sequence motif. This sequence is located in the C-terminal of the immature protein, and is cleaved between amino acids threonine and glycine by the enzyme sortase. The threonine is then amine-linked to the cell wall via the carboxyl of the threonine to the free amino group of the pentaglycine crossbridge in the staphylococcal cell wall (Schneewind, Fowler and Foull 1995). I will now describe some examples of adhesion proteins of the MSCRAMM family.

5.1.1.1.1 The SD-repeat family of S. aureus

One family of *S. aureus* cell surface proteins is characterised by the presence of serine-aspartate dipeptide repeats (sdr). The family members also contain a signal peptide, an 'A' domain, an 'R' domain, composed of SD-repeats, and an 'M' domain, containing a proline-rich membrane spanning region, the LPXTG-motif, and the membrane anchor (Figure 3). In *S. aureus* there are 5 members of this family of proteins, including the clumping factors, ClfA and ClfB, and SdrC, SdrD, and SdrE (Josefsson *et al* 1998, Foster and Hook 1998, Liu *et al* 2004).

5.1.1.1.1.1 The serine-aspartate repeat protein SdrE

SdrE, along with SdrC and SdrD, were first characterised by Josefsson *et al* (1998), using Southern blot analysis with a probe against the SD repeat region. All three proteins have the typical gene structure described above, with the addition of 'B' repeats not found in ClfA and ClfB (see Figure 3). To date, however, no ligand has been identified for these proteins, although the role of SdrE in platelet activation has been investigated. O'Brien *et al* (2002) has shown that, while the absence of the SdrE gene alone had no effect on platelet aggregation, in a double mutant with ClfA and a triple mutant of ClfA, SpA, and SdrE, increased lag time for aggregation was observed, indicating it does have a role in the activation of platelets.

5.1.1.1.2 Clumping factors ClfA and ClfB

The observation that *S. aureus* cells clump together in plasma was first described in 1908. Later it was found that *S. aureus* cultures can also induce coagulation of plasma (Boden and Flock 1989). The first protein identified to play a role in this process is coagulase, an extracellular product of *S. aureus* that converts fibrinogen into fibrin. However, it was determined that other factors also play a role in clumping and coagulation; the staphylococcal clumping factors. Recently, *S. aureus* has been shown to be the dominant cause of infective endocarditis (Loughman *et al* 2005) and a study by Moreillon *et al* (1995) revealed that clumping factor defective mutants of *S. aureus* causes a 50% reduction in cases of endocarditis in a rat model compared to the wild type strain.

ClfA and ClfB are both known to bind fibrinogen, which can bind to host platelet cells via the GPIIb/IIIa receptor. Fibrinogen is the precursor of fibrin, which forms

blood clots, and it is cleaved by thrombin to form the activated state. The binding of *S. aureus* to fibrinogen on the surface of platelets has been shown to trigger platelet activation and leads to platelet aggregation (Siboo *et al* 2001, O'Brien *et al* 2002, Loughman *et al* 2005). This binding is important for *S. aureus* virulence as was shown in a ClfA gene knockout strain, where the absence of the ClfA gene increased the lag time for aggregation of platelets (O'Brien *et al* 2002). However, fibrinogen is not the sole ligand for either ClfA or ClfB. In the case of ClfA, Loughman *et al* (2005) showed that fibrinogen alone is insufficient for platelet activation. They found that for activation to occur, the serum must also contain anti-ClfA antibodies. These bind along with the fibrinogen, thereby attaching the *S. aureus* bacterium to the platelet cells via the fibrinogen receptor and the IgG FcγRIIa receptor (Figure 1b). In addition, Loughman and co-workers (2005) have shown that activation can also occur via the complement pathway. This second mechanism of activation is slower due to the time necessary for the assembly of the complement component complexes and it also requires IgG. Siboo *et al* (2001) discovered that ClfA could also bind directly to platelets, triggering platelet activation, via a novel cell surface receptor, P118.

ClfB was identified via the same method described for SdrC, SdrD, and SdrE. ClfB has 26% sequence identity with the A domain of ClfA, and like ClfA it also binds fibrinogen (Ni Eidhin *et al.,* 1998). Unlike ClfA, which binds to the γ-chain of fibrinogen, ClfB binds to the α- and β-chains. Walsh *et al* (2004) showed ClfB has a second ligand in the form of cytokeratin 10, an essential component of the cell cytoskeleton. Cytokeratin 10 is also found on the surface of squamous cells located in the nasal passages, and as such, has been suggested to be a mechanism of attachment in nasal colonisation.

5.1.1.1.2 Elastin binding protein EbpS

Elastin is a component of the mammalian ECM, and is found in those tissues that require elasticity, such as the lungs, skin and blood vessels (Park *et al* 1991). Park *et al* (1996) identified and cloned the gene in *S. aureus* responsible for elastin binding: Elastin-binding protein (EbpS), a 23kDa protein that has the ability to inhibit *S. aureus* binding to purified elastin. EbpS, however, lacks the LPXTG cell wall anchoring motif characteristic for the MSCRAMM family of proteins.

Figure 4 The arrangement of the domains of the fibronection-binding proteins showing the (S) signal domain, the A domain, the B repeats, the D repeats, the (W) cell-wall spanning region, and the (M) membrane-spanning region containing the LPXTG motif (Wann, Gurusiddappa and Hook 2000).

5.1.1.1.3 FnBPA and FnBPB

Fibronectin is another component of the ECM, and is also found in blood clots. Two *S. aureus* proteins, **f**ibro**n**ectin-**b**inding **p**roteins **A** and **B** (FnBPA and FnBPB), were shown to bind to coverslips coated with fibronectin (Greene *et al* 1995). Both are cell surface proteins containing the LPXTG-cell wall-anchoring motif. The A regions of these proteins (shown in Figure 4) share 45% sequence identity while the W and M domains (cell-wall spanning, and membrane anchor) show much higher homology (Foster and Hook 1998, Greene *et al* 1995). It is thought that fibronectin binding is important in virulence as most invasive strains have the ability to bind fibronectin and the binding of *S. aureus* to fibronectin is required for the internalisation of the bacterium by epithelial cells and adhesions to plasma clots (Dziewanowka *et al* 1999, Roche *et al* 2004). However, Flock *et al* (1996) showed using a rat model of endocarditis, that the elimination of the bacteria's fibronectin binding properties had no effect on virulence.

Like the clumping factors, the fibronectin-binding proteins have multiple ligands. FnBPA and FnBPB can also bind fibrinogen, like the clumping factors ClfA and ClfB, and supports clumping, although FnBPA has a lower affinity for fibrinogen than ClfA (Wann, Gurusiddappa and Hook 2000). FnBPA/B can also bind elastin and this activity is present in the N-terminal 'A' domain of the two proteins (Roche *et al* 2004). In knockout strains, elastin binding is eliminated suggesting that FnBPA/B mediates elastin binding and that EbpS is not necessary for this binding.

5.1.1.1.4 Collagen-binding protein Cna

S. aureus is the major agent of bacterial arthritis or septic arthritis, causing 80% of cases. Strains isolated from such infections are able to bind collagen via the collagen-binding protein Cna. Unlike the other adhesion proteins, Cna is not found in most strains of *S. aureus*, however it is found in most strains isolated from patients suffering from septic arthritis. The protein has been shown to mediate attachment to the cartilage. This attachment via collagen and cartilage is important as was shown in a mouse model, where the loss of Cna reduced the rate of arthritis and the degree of damage to the joints (Patti *et al* 1994, Foster and Hook 1998).

5.2 Bacterial immunoglobulin-receptors and their ligands

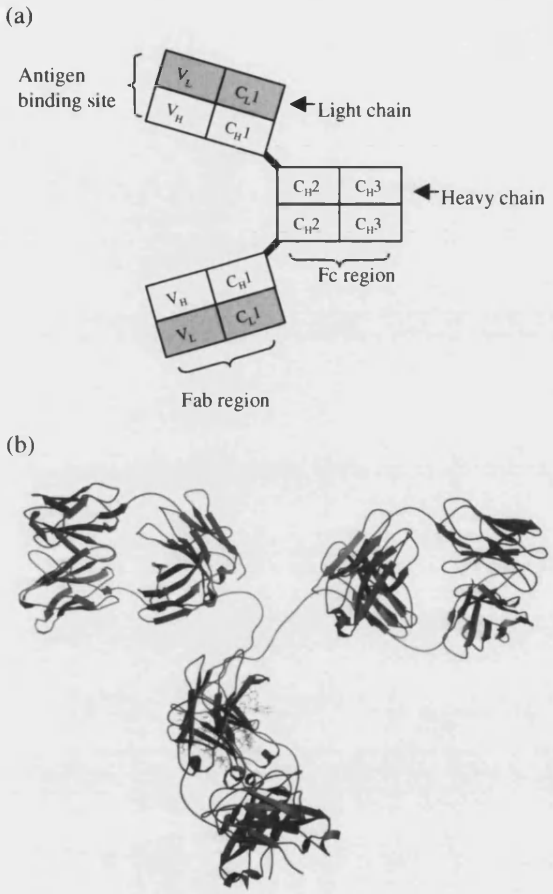


Figure 5 (a) showing the arrangement of the domains comprising IgG, also indicating the major IgG fragments, and the composition of peptide chains. **(b)** showing the crystal structure of the intact human IgG b12 (PDB code 1HZH: Saphire *et al* 2001)

In mammals the FcγIIR receptor plays a central part in the adaptive immune response. A number of bacterial species also express Ig-receptors, examples include *S. aureus, Streptococci, peptrostreptococcus magus and haemophilus somnus.* The *S. aureus* Ig-receptor SpA, for instance, binds IgG via both the Fc and Fab portions of the molecule. It is thought that these bacterial Ig-receptors are involved in masking the pathogen from the host's immune system (Wilson, McNab and Henderson 2002 Ch 10). The presence of SpA on the surface of the bacterium, coating the cell with immunoglobulins, reduces the susceptibility of the bacterial cells to phagocytosis by human neutrophils. The absence of IgG in the serum also increases the rate of phagocytosis when SpA is expressed by the bacterium. It is thought that this is due to activation of the classical complement pathway through the Fab-binding property of SpA to IgM (Palmqvist *et al* 2002).

5.2.1 Immunoglobulins

Most immunoglobulins (Ig) share a characteristic 'Y' shaped structure, consisting of two heavy chains, and two light chains, as shown in Figure 5. The heavy chain

consists of 4 globular domains, constant domains 1 to 3 (C_H1, C_H2, C_H3) and a variable domain (V_H), while the light chain is comprised of two domains: a constant domain (C_L) and a variable domain (V_L). The combination of the V_H and the V_L domains form the antigen binding site. This antigen-recognition site is limited to the hyper variable region located on the molecule's loops. Any Ig contains either a kappa (κ) or a lambda (λ) light chain, and has one of several different types of heavy chain, gamma ($\gamma1$, $\gamma2$, $\gamma3$, $\gamma4$), mu (μ), alpha ($\alpha1$, $\alpha2$), epsilon (ε), or delta (δ). The different heavy chains give rise to the five main classes of Igs: IgG (γ), IgM (μ), IgA (α), IgE (ε), and IgD (δ). IgM and IgE have slightly different heavy chains with an additional constant domain (C_H4). IgM and IgA also contain an extra protein chain, termed the J-chain, which links the Igs together to form a pentameric (IgM) or a dimeric (IgA) structure. The dimeric form of IgA is only found in mucosal secretions, in serum it exists as a monomer. The secreted form contains another protein termed the secretory fragment. The different types of Igs are shown in Figure 6.

IgG is the most abundant Ig found in the serum, and exists in 4 subclasses (IgG1, IgG2, IgG3, IgG4) all of which have homologous sequences and structure. IgM is the first antibody produced in response to a novel antigen. As mentioned before IgA is found in the serum, but is also found in secretions such as mucus, saliva, and breast milk. The two IgA subclasses (IgA1 and IgA2) are thought to have the same function. IgE is involved in immunity against parasitic worms and it is important in hyper reactive responses to allergens. The function of IgD is unknown (Parham 2000 Ch2, Wood 2001 Ch3).

5.2.2 Bacterial immunoglobulin receptors

5.2.2.1 Streptococcal protein G (SpG)

SpG is a cell wall associated protein found on the surface of bacteria from the genus *Streptococcus*, with the ability to bind IgG and serum albumin. The protein contains either two or three albumin binding domains at the N-terminus, followed by either two or three IgG binding domains; the specific number of both varies between

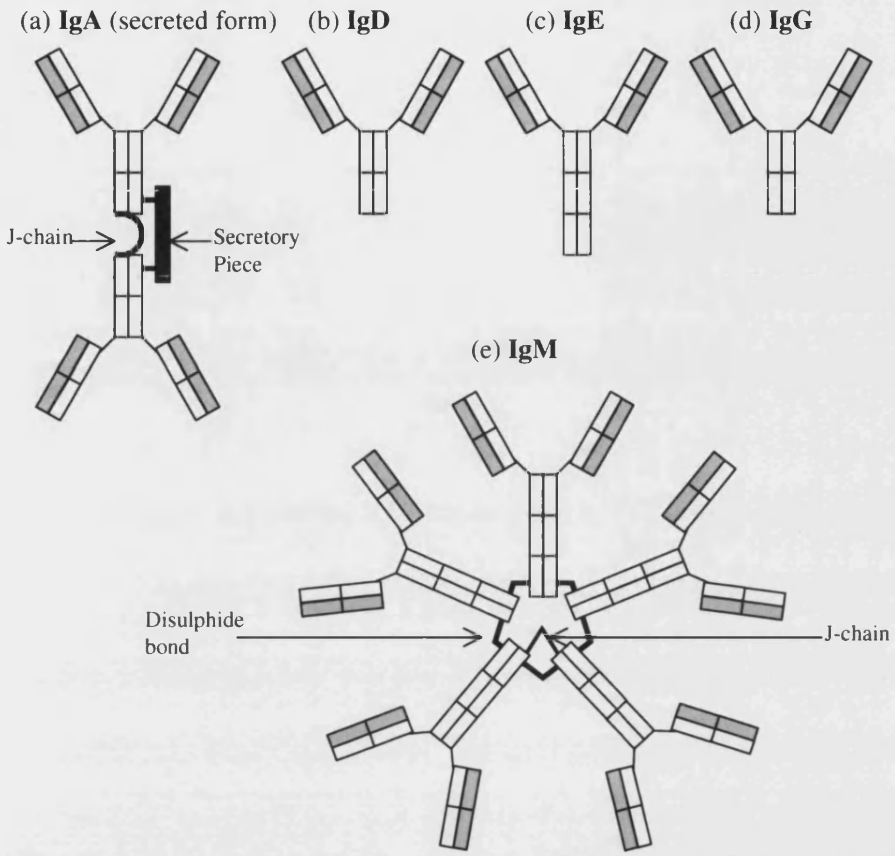


Figure 6 showing the structures of **(a)** IgA – secreted form, **(b)** IgD, **(c)** IgE, **(d)** IgG, and **(e)** IgM. IgE and IgM have additional C_H4 domains, compared to IgG, while IgA and IgM have additional peptide chains involved in multimerisation (labelled as J-chain and secretory piece, Wood 2001 Ch3).

different strains. The C-terminal is composed of a cell wall and membrane binding domains (Figure 7a, Tashiro and Montelione 1995).

Aybay (2003) investigated the binding of IgG by SpG and compared it to SpA binding. While it was known that both SpG, like SpA, could bind both Fab and Fc fragments, Aybay (2003) showed that both proteins have unique binding sites on each of the fragments.

5.2.2.2 Peptostreptococcal protein L

Protein L is found on the surface of certain strains of *Peptostreptococcus magus*, and binds to the κ chain variable domain of human IgG, IgA and IgM. The protein

contains 5 Ig-binding domains like SpA (Figure 7b), however the structure of Ig-binding domains is the same as SpG: a 4-stranded β-sheet with a single α-helix (Figure 7dii, Tashiro and Montelione 1995).

5.2.2.3 *S. aureus* Ig-binding protein A (SpA)

SpA is a well-characterised *S. aureus* cell-surface protein, with a MW of 57kDa, (in *S. aureus* strain 8325-4; Tashiro and Montelione 1995). It contains 4 or 5 highly homologous Ig-binding domains in addition to a cell-wall binding region, a cell membrane binding region, a transmembrane region in the C-terminal portion of the protein, and an LPXTG- cell membrane anchoring motif (see Figure 7c for the arrangement of SpA domains, Tashiro and Montelione 1995, Foster and Hook 1998). An artificial Z domain has also been generated based on the B domain, differing only at a single amino acid position: the B domain to Z domain substitution G22A (Tashiro and Montelione 1995, Jansson, Uhlen, and Nygren 1998, the sequences of SpA domains are shown in Figure 46).

Early studies found that SpA is specific for mammalian IgGs (Kronvall *et al.,* 1970). No IgGs from fish, amphibians, reptiles or birds (except *Rhea americana*) serum tested positive for IgG binding by SpA. This suggested that the interaction is wide spread among mammals and that the ability to immobilise IgG could be important for virulence.

Later, Jansson, Uhlen and Nygren (1998) showed that all five SpA Ig-binding domains bind to both the IgG Fc fragments and Fab fragments (the engineered Z domain binds Fc but lacks the ability to bind Fab). While SpA binds the Fc fragment IgG subclasses 1, 2 and 4, no binding has been detected for IgG3; Surolia, Pain and Khan 1982. Fab binding is limited to the variable region of heavy chains from gene family 3 (V_H3, Sasso, Silverman and Mannik 1989). It is interesting to note that it only takes a single residue change in SpA to eliminate either Fab or Fc binding. The sole difference in domain Z compared to domain B is the substitution of a glycine to an alanine (Jansson, Uhlen and Nygren 1998), while inactivation of tyrosine residues by iodination eliminates Fc binding in SpA (Romagini *et al* 1982; Nguyen,

Ghebrehiwet and Peerschke 2000). In addition to binding IgG, SpA also binds IgM (via Fab V_H3 chains) and IgA (subclass 2) (Surolina, Pain and Khan 1982).

In humans the V_H3 family comprises about half of all inherited V_H genes, and leads to about 30% of peripheral B-cells binding SpA via their Fab domains in the surface Igs (Palmqvist *et al* 2005). It has been suggested that SpA, by binding to the Fab

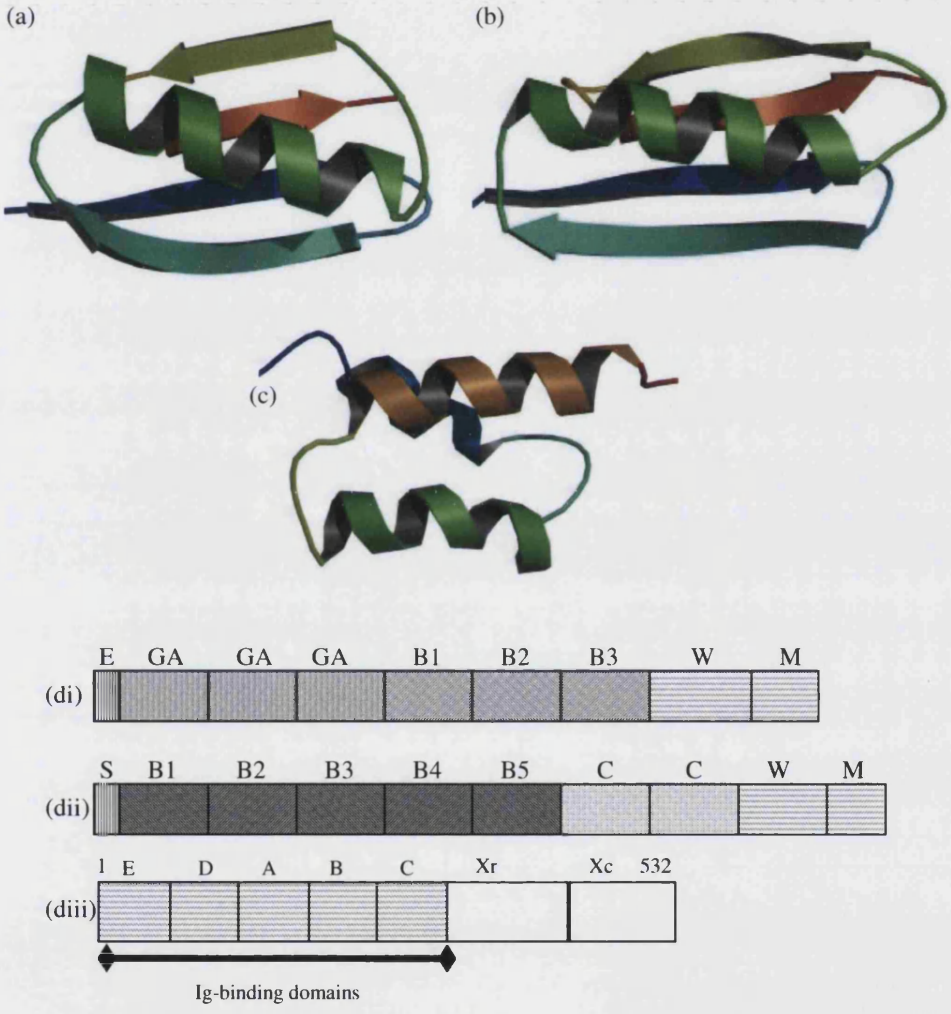


Figure 7(a) Crystal structure of SpG Ig-binding domain (PDB code 1IGD: Derrick and Wigley 1994) **(b)** NMR structure of protein L (PDB code: 2PTL: Wikstrom *et al* 1993) (c) NMR structure of SpA Ig-binding domain (PDB code 1BDD: Gouda *et al* 1992) (d) the arrangement of domains in (i) SpG, (ii) protein L and (iii) SpA (Tashiro and Montelione 1995)

fragment of these B-cell surface Igs can trigger stimulation of B-cells similar to the superantigens in T-cells. It has been demonstrated that administration of SpA to mice causes the long lasting loss of 80% of splenic B-cells reactive to SpA and reduces the levels of circulating SpA-reactive IgM and IgM-secreting cells though supraclonal deletion and apoptosis of B-cells (Silvermann and Goodyear 2002, Palmqvist *et al* 2005, Silverman and Goodyear 2006). It has also been shown that SpA-specific memory B-cells are not formed; however antibodies for conventional antigens form normally (Palmqvist *et al* 2005).

5.2.2.3.1 Other ligands of SpA

SpA has been found to bind a number of other ligands, present both in the serum and on the surface of host cells. One of the serum proteins that SpA is now known to bind to is von Willebrand factor (vWF; Hartleib *et al* 2000). vWF is a large multifunctional glycoprotein released by activated platelets. It is cleaved after release, and binds to various sites to allow platelet adhesion at sites of endothelial damage. As vWF also has a platelet bound receptor, it provides a bridging mechanism for attachment of SpA to platelet cells and a mechanism to allow the triggering of endovascular disease (Hartleib *et al* 2000, O'Brien *et al* 2002).

SpA can also bind directly to the surface of platelets via the platelet cell-surface receptor gC1qR/p33 (Nygren, Grebrehiwet and Peerschke 2000; Peerschke and Ghebrehiwet 2001). Its ligand is C1q, which is part of the complement factor C1 complex, involved in the classical complement pathway (Kishore and Reid 2000). The receptor is a multifunctional cellular protein that has a role in modulating platelet function at sites of vascular injection and inflammation. Iodination of SpA eliminates gC1qR binding to SpA as well as IgG binding, suggesting that SpA has similar structural requirements to bind both ligands (Nygren, Grebrehiwet and Peerschke, 2000).

A second receptor has also been found to bind SpA, the tumour-necrosis factor-α receptor (TNFR1), and this binding was shown by Gomez *et al* (2004) to induce inflammation in the airway epithelium. As TNFR1 is found on the surface of epithelial cells, in particularly those of the airways, it gives *S. aureus* a mechanism to

attach to these surfaces. It was shown with a *spa*⁻ mutant of *S. aureus*, that there is reduced virulence in a mouse model of pneumonia, suggesting that the TNFR1-SpA interaction is important in virulence.

5.2.2.3.2 The role of SpA in virulence

Mastitis is of concern in the dairy industry, due to its impact on milk production. *S. aureus* is known to cause this disease. Jonsson *et al* (1985) investigated the role of three *S. aureus* proteins, and compared their effects on this disease with alpha-haemolysin negative (*hla*⁻), coagulase negative (*coa*⁻) and SpA negative (*spa*⁻) mutants in a mouse model. The *hla*⁻ and *coa*⁻ mutants showed reduced virulence compared to the wild type strain, however the level of SpA production did not have any observed effect on virulence.

The role of SpA, along with alpha-haemolysin (α-haemolysin, also referred to as α-toxin), has also been investigated in subcutaneous lesions in a mouse model. Patel *et al* (1987) injected mice with *S. aureus*, using a wild type strain, along with mutant strains lacking either the α-haemolysin gene or the SpA gene, or both in a double mutant. This work found notable physiological differences in the appearance of lesions between mice injected strains lacking α- haemolysin and the wild type strain. α-Haemolysin expressing strains were at least 10-fold more virulent: α-haemolysin was shown to be a major virulence factor. However comparison of strains lacking SpA showed only differences in the size of the lesion, not its appearance. They found that SpA negative strains do have slightly reduced virulence, although the difference is not as marked as in the case of α-haemolysin.

This pattern, in which loss of SpA causes reduced virulence, but to a much lesser degree to the loss of a *S. aureus* toxin has also been found in models of other bacteria diseases. For example Callegan *et al* (1994) used a rabbit model to compare the effects of SpA and α-haemolysin in keratitis. After injection of the rabbit corneas with either wild type *S. aureus* or strains lacking SpA, α-haemolysin or both genes showed greater reduced pathology in strains lacking α-haemolysin, however, similar to the mouse lesions little difference was observed between *spa*⁻ or wildtype strains.

The role of SpA in arthritis and septic death was investigated by Palmqvist *et al* (2002). They showed that the loss of SpA produced a less severe arthritic phenotype, and a reduced mortality rate in the mice. However they also showed that injecting the SpA protein itself did not trigger arthritis in the animal joints. This suggests other components expressed by *S. aureus* are also involved in causing disease. This is supported by the work of O'Brien *et al* (2002). They looked at the role of SpA in platelet aggregation in combination with ClfA. In the presence of a *S. aureus* mutant-lacking SpA, the lag time observed for aggregation was the same as the wild type. However in the double knock out of ClfA and SpA, an increased lag time was observed when compared to the single ClfA negative mutant.

While the importance of the immunoglobulin binding activity mediated by SpA remains unclear the recent identification of a second staphylococcal Ig-binding protein may help to clarify the role of Ig-binding in *S. aureus* virulence.

5.2.3 *S. aureus* Ig-binding protein Sbi (Sbi)

Zhang *et al* (1998) used phage display techniques to identify a unique polypeptide from *S. aureus* strain 8325-4 that shared SpA's ability to bind IgG. They went on to clone the gene, and from it expressed the protein, comprised of 436 amino acids. Sequence analysis of this gene revealed the gene to contain an Ig-binding domain, a predicted Ig-binding domain, a β_2GP1-binding domain and a proline rich region (Zhang *et al* 1999). The proline-rich sequence is composed of a proline repeated every five residues. Proline rich sequences are commonly found in cell-wall spanning regions in bacteria proteins, although Sbi lacks the LPXTG motif also found in these proteins. The arrangement of domains in Sbi is shown in Figure 8.

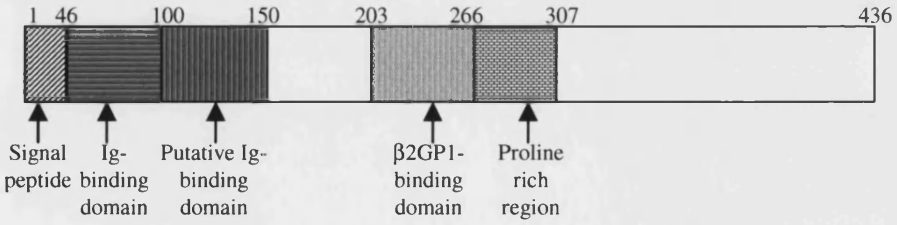


Figure 8 Sbi domain structure (adapted from Zhang *et al.*,1998 and 1999).

The shotgun phage display method used by Zhang *et al* (1998) suggested the presence of one functional IgG-binding domain, with homology to the Ig-binding domains of SpA. The highest homology is in the region of Fc binding determined by Deisenhofer (1981). The amino acid sequence suggested the presence of a second Ig-binding domain, however Zhang *et al* (1998) were not able to find any clones that bound IgG that mapped to this region of the protein.

When they examined the IgG-binding ability, they used the clone termed IG4, comprised of residues 24 to 115. They showed strong binding for human serum IgG, pig IgG, guinea pig IgG, and rabbit IgG. The Sbi fragment showed no binding for human IgG3, rat IgG, goat IgG and chicken IgG. It did show some binding to human IgM and IgA.

Zhang *et al* (1999) went on to use their phage display against newborn calf serum to detect novel interactions. Most clones were IgG-binding or fibrinogen-binding domains. However, they also identified overlapping regions of the Sbi gene that does not encode for either of the proposed Ig-binding domains, suggesting the ability of Sbi to bind a second serum protein. Affinity chromatography was used to isolate this protein, and it was identified as β_2-glycoprotein I (β_2GP1). β_2GPI is found in blood serum and is a membrane-adhesion glycoprotein. It consists of 5 homologous domains, with the fifth domain containing an additional loop that can be inserted into membranes. The protein has been implicated in the anti-phospholipid syndrome, which is characterised by thrombo-embolic complications and the production of anti-phospholipid autoantibodies. However, it has been realised that the autoantibodies bind, not to the lipids, but to β_2GPI (Bouma *et al* 1999). The minimal β_2GP1 binding sequence in Sbi consists of amino acids 204 to 261 (see Figure 8).

Zhang *et al* (2000) also examined the effect of the growth conditions on Sbi. Sbi was shown to be expressed in all tested *S. aureus* strains, but its expression level is very low. It was found that the level of Sbi expression is up regulated by the presence of IgG in the growth media. Zhang *et al* (2000) suggested that this up regulation increases the capacity of Sbi to bind β_2GPI and that this could be important in virulence. They also suggested that while Sbi contains a characteristic polyproline

repeat region for cell-wall binding but lacks the LPXTG motif, it is probably not proteolytically processed by sortases. And although they note that no membrane-spanning region was found, they suggest the C-terminal part of Sbi maybe localised intracellularly and could play a role in IgG-mediated cell signalling.

5.2.4 Summary

In this thesis, I will study both structural and functional aspects of Sbi in order to understand its role in *S. aureus* virulence. I start by predicting the structure of Sbi via homology modelling, based on its similarity to SpA (Chapter 2). Then I compare the binding characteristics of Sbi to those of SpA for a selection of serum IgGs from different animal species and for the human IgG subclasses (Chapter 3). In Chapter 4, I attempt to explain the species and subclass specificity in the IgG-binding between Sbi and SpA, observed in Chapter 3, by analysing the sequence variations between the IgGs and via the generation of homology models of Sbi-IgG and SpA-IgG complexes. In Chapter 5 I further analyse the differences in species specificity in IgG-binding by comparing the sequences of the Sbi and SpA genes present in *S. aureus* strains isolated from different animal species, to determine whether these strains have evolved specificity for IgGs from their host species. Finally, I use small angle X-ray scattering to determine a low-resolution solution structure of the extracellular region of Sbi (Chapter 6). In Chapter 7, I will discuss future directions for the continuation of this work.

6 CHAPTER 2

6.1 3-dimensional structure prediction of the N-terminal domains of the *Staphylococcus aureus* protein Sbi by homology modelling

6.1.1 Introduction

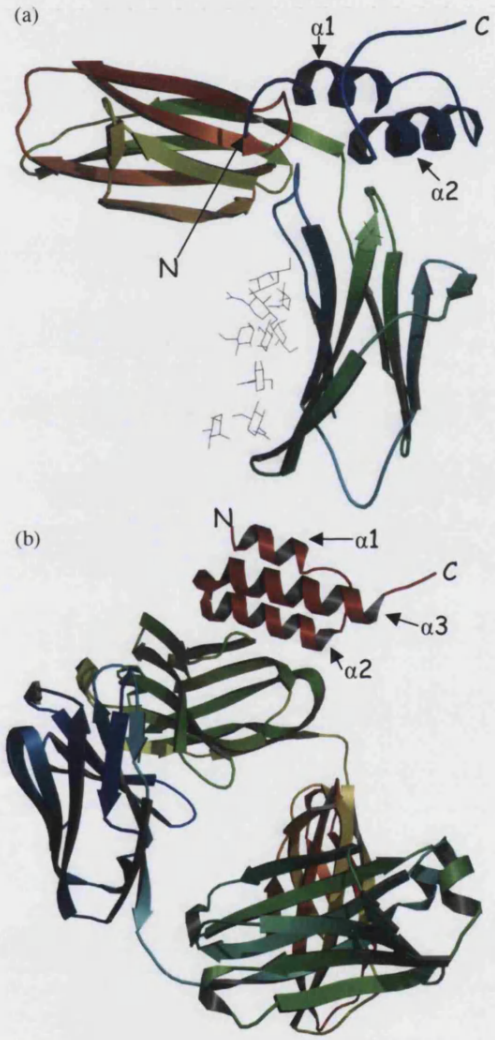


Figure 9 (a) Crystal structure of SpA in complex with IgG Fc (Deisenhofer 1981). **(b)** Crystal structure of SpA in complex with IgM Fab (Graille *et al* 2000)

S. aureus contains two Ig-binding proteins, SpA and Sbi. Sbi is a recent discovery and little is known about it other than its homology in the N-terminal region to the Ig-binding domains of SpA (Zhang *et al* 1998, 1999 and 2000). Zhang *et al* (1998) characterised its binding with IgG, discussed further in Chapters 1 and 3. SpA on the other hand has been well characterised in both in terms of its binding characterisitics for immunoglobulins (see Chapters 1 and 3 for more details) and the structure of its Ig-binding domains.

The first three-dimensional structure determined of SpA was of the B domain in complex with a human IgG Fc fragment. This showed the SpA molecule as two anti-parallel helices and a C-terminal unstructured region. The SpA molecule is seen to interact with the Fc fragment, via the two helices, in a region located

between the C_H2 and C_H3 domains (Figure 9a, Deisenhofer 1981). Later, Gouda *et al* (1992) determined the structure of the uncomplexed B domain, in solution, by NMR. This structure is different from Deisenhofer's in that SpA displays a three-helix bundle fold (Figure 9b). At the time it was suggested that the third helix could be unwound on binding. However Jendeberg *et al* (1996) demonstrated that this is not the case, and suggested that the difference observed by Deisenhofer (1981) could be due to poor structural resolution.

SpA also binds to the Fab fragment of immunoglobulins with a certain type of heavy chain (the V_H3 gene family) And in 2000 the crystal structure of the SpA D domain was determined in complex with an Fab fragment from a human IgM antibody (Graille *et al* 2000). This structure revealed that SpA has two distinct binding sites for Fc and Fab fragments. While SpA binds Fc via helices 1 and 2 (Deisenhofer 1981), it binds Fab via helices 2 and 3 (Figure 9b). In the Graille *et al* (2000) structure SpA exists as a dimer, with the dimer interface formed by helices 1 and 2 on each monomer. It is unknown whether this dimerisation occurs in vitro or in vivo or whether this is an artifact of crystallisation. Today several additional structures of the SpA domains exist that were determined by NMR. So far, there are no structures for SpA with multiple Ig-binding domains.

6.1.1.1.1 Aim of this study

Phage display and sequencing analysis by Zhang *et al* (1998) revealed the presence of one Ig-binding domain in the Sbi sequence, and predicted the presence of another. In this chapter, I attempt to predict the structure of the Ig-binding protein Sbi based on its sequence and the differences therein when compared to SpA. I will use the above-mentioned structures of SpA and its complexes to envisage the role of Sbi in *S. aureus* immunoglobulin binding. Furthermore, I will use sequence alignment and homology modelling to assign structure and function to those Sbi regions with unknown structure and functions.

6.1.1.1.2 Homology modelling

Homology modelling is used to predict a protein's 3D structure based on its amino acid sequence and the 3D structure of proteins with a similar sequence (Kirton, Baxter and Sutcliffe 2002). The first stage in this process is the identification of the structure

with similar sequence to the unknown target. If multiple structures can be identified, it can aid in the additional modelling of loop regions, which tend to be more flexible and less conserved in structure and sequence. The generation of a homology model of a protein involves a systematic replacement of residues in the template structure with those of the target sequence. Insertions and deletions should preferably occur in loop regions rather than in regions with a conserved secondary structure. There are programs that can be used for manual model building such as O (Jones *et al* 1991) while there are others that will build models automatically based on the structural templates, such as the **M**olecular **O**perating **E**nvironment (MOE; Chemical Computing Group Inc.: http://www.chemcomp.com/). In the final stages of modelling the structural model is validated in order to determine whether it meets the allowed rules of protein folding by checking whether its backbone angles are within the allowed regions of the Ramachandran plot and that it is does not have residues exposed to unfavourable environments, i.e. large numbers of exposed hydrophobic side chains (Kirton, Baxter, and Sutcliffe 2002).

The first step is to identify homologous sequences through a BLAST search (**B**asic **L**ocal **A**lignment **S**earch **T**ool; Altschul *et al* 1997), which is a "statistically driven search method." It is possible to restrict such a search to entries in the **P**rotein **D**ata**b**ase (PDB; Bernstein *et al* 1977; http://www.rcsb.org/pdb/), and avoiding those matches that do not have known structures. However there are many cases where it is not possible to identify a suitable structure though this method. For such proteins there may be other suitable templates available, because certain sequences that are only distantly related can also have conserved structure (Kirton, Baxter, and Sutcliffe 2002). These sequences can be identified via a number of structure prediction servers on the web that are designed to detect homologous structures that are more remotely related than the ones detected by BLAST (see Methods for examples). This method can in some cases provide suitable templates for model building.

6.1.2 Methods

6.1.2.1 Calculation of protein hydrophobicity

Pepplot+ (Wisconsin Package Version 10.2) was used determine hydrophobic regions in the Sbi protein sequence (PubMed accession number: BAB96206). To do this two methods were used. Kyte and Doolittle (1982) calculates an average of a residue-

specific hydrophobicity index over a window of nine residues, while the Goldman, Engelman and Steitz curve is an average of a residue-specific hydrophobicity scale over a window of 20 residues. For both methods, a curve in the upper half of the plot indicates hydrophobicity, while the curves in the lower half indicates a hydrophilic region (see Figure 10).

6.1.2.2 Prediction of protein secondary structure

PEPTIDESTRUCTURE (Wisconsin Package Version 10.2) was used to calculate secondary structure of Sbi with two different methods; the Chou-Fasman method (CF-pred, Chou and Fasman 1978), and the Robson Garnier (GORpred, Garnier *et al* 1978) method. These prediction methods generate their prediction through statistical methods (Lin *et al* 1989). In addition the protein prediction server 3D-PSSM was used (see below) which also predicts the protein's secondary structure.

6.1.2.3 Identification of sequences homologous to Sbi

BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) matches a query sequence to those in a database though a "statistically driven search method" (Altschul *et al* 1997). The complete Sbi protein sequence was input against the non-redundant GenBank database and against the database of sequences from the RCSB **p**rotein **d**ata **b**ank (http://www.rcsb.org/pdb/ Bernstein *et al* 1977). In addition, shorter fragments of Sbi were also input against the same databases for the fragments consisting of residues 145 to 270 (incorporating the β_2GP1 binding domain), 260 to 315 (the polyproline region) and 305 to 436 (the C-terminal region of Sbi).

Hits produces in BLAST all have an Expect (E) value. This described the number of hits occurring by chance. The closer the value is to "0", the more significant the match is (http://0-www.ncbi.nlm.nih.gov.mill1.sjlibrary.org/blast/blast_FAQs.shtml)

Protein prediction servers can also be used to determine structures that are homologous to a protein sequence with an unknown structure. Three different servers were used on the complete Sbi protein. One of these (3D-PSSM) was also used on the three shorter fragments described above.

6.1.2.3.1 3D-PSSM

Three-dimensional **p**osition-specific scoring **m**atrix (3D-PSSM) is a method that detects remote protein sequence homology though structural alignments (Kelley *et al* 2000) available at http://www.sbg.bio.ic.ac.uk/~3dpssm/. The database used for this server is SCOP version 1.53 (Murzin *et al* 1995).

6.1.2.3.2 GenTHREADER

GenTHREADER is a fast and reliable method of fold recognition though sequence alignment, calculation of threading potentials, and completed by evaluation by a neural network (Jones 1999, McGuffin, Bryson, Jones, 2000) at http://bioinf.cs.ucl.ac.uk/psipred/psiform.html. The method uses the PDB (updated Thu May 26 01:07:51 2005).

6.1.2.3.3 SAM-T99

SAM-T99 predicts protein structure though an "iterative hidden Markov model-based method for constructing protein family profiles," available at http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-model-library-search.html (Karplus *et al* 1998). This server searches the PDB and SCOP databases (SCOP available at http://scop.mrc-lmb.cam.ac.uk/scop/index.html, Murzin *et al* 1995).

Multiple methods were used in the identification of protein hydrophobicity, secondary structure and tertiary structure, so results could be compared and allow the determination of similarities between results for the different methods. For example the identification of the same fold in different proteins by the various prediction servers for a query protein sequence would suggest the identification of the correct fold opposed to the identification of different folds by each of the prediction servers.

6.1.2.4 Generation of protein models

Sequences were aligned with the alignment program ClustalX (Thompson *et al* 1997) in the multiple alignment mode.

Protein structures used for building models on were obtained from the Protein Data Bank (PDB). Structural alignments were carried out with the program LSQMAN, and

pdb files modified with MOLEMAN, both programs were supplied by the Uppsala Software Factory (http://xray.bmc.uu.se/usf/, Kleywegt 1996).

Protein modelling was carried out in O (Jones et al 1991), a software program initially designed to allow manipulation of electron density maps produced by macromolecules. For homology modelling, a known protein structure is taken and manipulated using the @mutator function to replace the residues in the template structure (obtained from the PDB database) to the corresponding residue in target sequence. Residues can also be added or removed in loop regions (or at the end of secondary structure elements, to maintain the secondary structure backbone of the template structure), and the loops redrawn with the Lego-loop function. Lego can also be used to change the orientation of residue side-chains to avoid side chain clashes and optimise interactions.

6.1.2.4.1 Determination of interactions in complexes

The amino acid residues in the SpA complexes with IgG Fc (1FC2, Deisenhofer 1981) and with IgM Fab (1DEE Graille *et al* 2000) involved in interactions between the SpA molecule and its ligand were determined by analysing all atoms in the SpA molecule at a distance between 1.8Å and 5.2Å from residues in the IgG Fc fragment , using the CCP4 program CONTACT (Collaborative Computational Program 1994). The Sbi homology models were superimposed onto the structures of the SpA-Fc and Fab complexes to generate models of Sbi in complex with IgG.

6.1.2.4.2 Validation of structures

Two methods were used to check models generated. Ramachandran plots were calculated in O (Jones *et al* 1991) and using PROCHECK (Laskowski *et al* 1993, Morris *et al* 1992), and any unfavourable backbone angles were adjusted.

Verify_3D was used to analyse the environment of amino acids, and to determine those "sequences that are most compatible with the environment of the residues in the 3D structure" (Bowie *et al* 1991). Verify_3D uses a 21-residue sliding window to calculate the average 3D-1D score, the centre of which is indicated on the x-axis. The

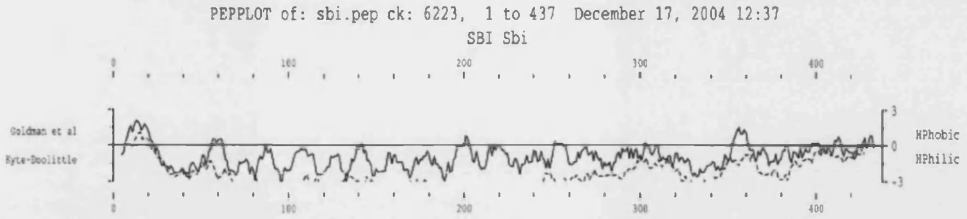


Figure 10 Hydrophobicity plot of the Sbi protein determined by Pepplot+.

sliding window is renders the scores for the first nine and last nine residues in the plot meaningless. The output is produced in a graph format (see Figure 13c for an example). A line above the axis indicates correctly modelled segments, while lines under the axis indicate a segment that is incorrectly modelled. This is a web-based program at http://www.doe-mbi.ucla.edu/Services/Verify_3D/ (Luthy *et al* 1992, Bowie *et al* 1991).

6.1.2.4.3 Generation of protein structure images

All images of molecular models were generated using Molscript (Kraulis 1991) and subsequently rendered using Povray™.

6.1.3 Results

6.1.3.1 Calculation of protein hydrophobicity of Sbi

Zhang *et al* (2000) suggested that Sbi contains a transmembrane region and cytoplasmic domain based on the fact that its expression is upregulated when *S. aureus* is grown in the presence of IgG. It was suggested that the C-terminal region of Sbi could be located in the cytoplasm and may be involved in protein signalling. The hydrophobicity plots shown in Figure 10, generated with both Goldman *et al* and Kyte-Doolittle techniques, show that the protein is very hydrophilic. The only region of the protein where both plots agree on the hydrophobic nature of the protein is a stretch of 20 residues at the N-terminus of the protein, which comprise the signal peptide. Only the Goldman method identified short hydrophobic regions at the C-terminal of Sbi. These regions may be involved in membrane crossing, but could also be part of the hydrophobic core of a globular domain.

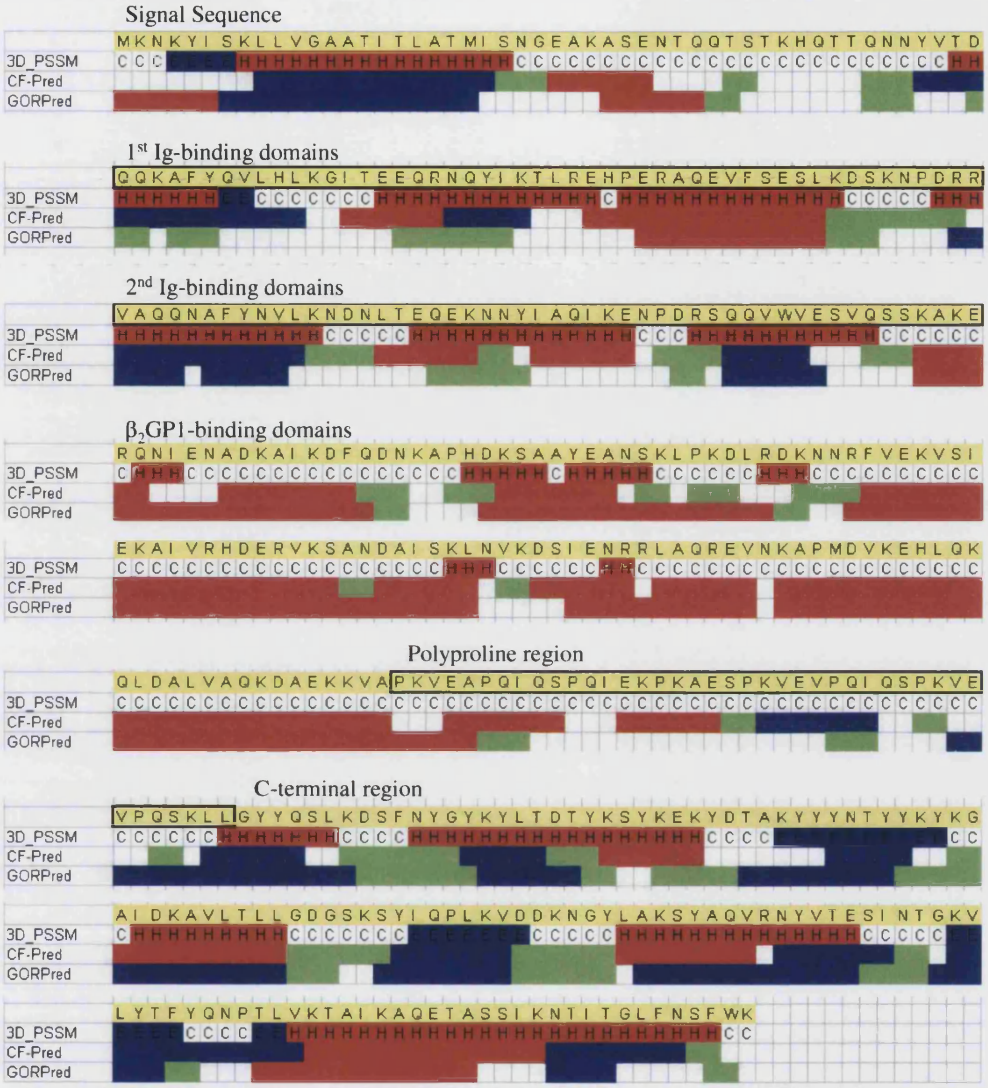


Figure 11 Secondary structure prediction of Sbi, indicating the regions with predicted functionality marked. Predicted helices (H) are shown in red and beta sheets (E) are shown in blue. Coils are depicted with 'C' and turns are highlighted in green. The 3D-PSSM prediction identified coils while the other methods identified turns.

6.1.3.2 Prediction of Sbi's secondary structure

Three different secondary structure predictions were obtained for Sbi (Figure 11) and the results aligned. The alignments show notable differences, but at least two plots were in agreement for each region of Sbi. Both Ig-binding domains were predicted as helical by 3D-PSSM, and a mixture beta-strand and alpha-helices by the other

programs. The β_2GP1-binding region is predicted to be α-helical by all methods, while all predictions produced different results for the poly-proline region. The C-terminal region is predicted to consist of a α-β-structure by two methods, although the GORpred method expects it to adopt an all beta structure.

6.1.3.3 Identification of sequences homologous to the full length Sbi protein sequence

Sbi was identified by Zhang *et al* (1998) through its IgG-binding ability. Sequencing of the gene revealed the presence of sequence homology in the N-terminal region of Sbi with the Ig-binding domains of SpA. By searching the BLAST database with the full length Sbi sequence we looked to identify other proteins with similarity to all or part of the Sbi protein sequence.

Searches against the non-redundant database in BLAST identified hits against Sbi sequences present in the database and also SpA sequences, all with E values of less than $1.00E^{-10}$. Other sequences identified to have homology to Sbi all had E values of $5.00E^{-06}$ or more. These hits with high E values include a *Drosophila* protein CG32793-PA, a neurofilament protein from *Canis familiarlis*, *Xenopus* protein Flj25286-prov, and a tail protein from Methanobacterium phage psiM2. The top ten hits after removal of all Sbi and SpA sequences are shown in Table 1, along with the region of Sbi for which the sequence shows homology.

As a method to identify functional homology of Sbi other than the Ig-binding function already identified in SpA, the BLAST search does not appear to be successful as there is no clear pattern of conserved structure or function in the homologous proteins identified.

Subsequently, the Sbi sequence was used to search structure prediction servers for possible conserved structures. However these were not able to provide better templates for model building. All results (except for those matches with SpA) had E values of more than $6.00\ E^{-09}$, and all contained a large number of gaps. While most of the structures found were α-helical, their tertiary structures did not show any

Table 1 Top 10 matches for the sequence of the Sbi protein in the BLAST search after removal of Sbi matches and of multiple matches for the same sequence (mainly SpA).

Sequences producing significant alignments:		Score (bits)	E Value	Identity	Positives	Gaps	Sbi regions which the BLAST hit has homology to						
							S	1	2	3	B	P	C
AAW38738	immunoglobulin G binding protein A precursor [Staphylococcus aureus subsp. aureus COL]	82	3.00E-14	27%	49%	10%							
				24%	42%	15%							
				31%	51%	9%							
NP_726835	CG32793-PA [Drosophila melanogaster]	55	5.00E-06	29%	52%								
NP_001003352	heavy neurofilament protein [Canis familiaris]	53	1.00E-05	29%	51%	0%							
AAH55988	Flj25286-prov protein [Xenopus laevis]	52	3.00E-05	29%	51%	0%							
T12737	tail protein - Methanobacterium phage psiM2	52	3.00E-05	19%	44%	6%							
				20%	43%	14%							
				22%	44%	4%							
AAC52610	osmotic stress protein 94	51	6.00E-05	32%	59%	1%							
				54%	78%								
				52%	72%								
				45%	70%								
NP_703354	Mature parasite-infected erythrocyte surface antigen (MESA) or PfEMP2 [Plasmodium falciparum 3D7]	51	7.00E-05	23%	42%	17%							
				19%	44%	10%							
				19%	42%	12%							
				24%	43%	15%							
				17%	38%	9%							
AAC13303	mature parasite-infected erythrocyte surface antigen [Plasmodium falciparum]	50	1.00E-04	23%	43%	17%							
NP_035150	osmotic stress protein [Mus musculus]	50	2.00E-04	19%	44%	8%							
				21%	39%	15%							
				20%	39%	10%							
				18%	39%	3%							
BAA08446	APG-1 [Mus musculus]	50	2.00E-04	19%	39%	6%							
				22%	40%	16%							

Table 2 Results of structure prediction server 3D-PSSM for Sbi sequence fragments encoding the (a) β_2GP1 binding domains, (b) the polyproline repeat and the (c) C-terminal region.

Table 2 (a)	Sequences producing significant alignments:	Score (bits)	E Value	Identities	Positives	Gaps
1LRZ	Chain A, X-Ray Crystal Structure Of Staphylococcus Aureus Fema	32	0.034	27%	54%	
1LT8	Chain B, Reduced Homo Sapiens Betaine-Homocysteine S-Methyltransferase In Complex With S-(Delta-Carboxybutyl)-L- Homocysteine	27	1.1	35%	60%	15%
1SMQ	Chain D, Structure Of The Ribonucleotide Reductase Rnr2 Homodimer From Saccharomyces Cerevisiae	26	1.4	39%	52%	8%)
1JK0	Chain A, Ribonucleotide Reductase Y2y4 Heterodimer	26	1.4	39%	52%	8%)
1U5P	Chain A, Crystal Structure Of Repeats 15 And 16 Of Chicken Brain Alpha Spectrin	26	1.9	22%	42%	
1U4Q	Chain B, Crystal Structure Of Repeats 15, 16 And 17 Of Chicken Brain Alpha Spectrin	26	1.9	22%	42%	
1UMY	Chain D, Bhmt From Rat Liver	25	3.2	32%	60%	15%
1E24	Chain A, Lysyl-Trna Synthetase (Lysu) Hexagonal Form Complexed With Lysine And Atp And Mn2+	24	5.5	23%	48%	
1BBW	Chain A, Lysyl-Trna Synthetase (Lyss)	24	7.1	22%	48%	
1KRT	Mol_id: 1; Molecule: Lysyl-Trna Synthetase (Product Of Lyss Gene); Chain: Null; Domain: Anticodon-Binding Domain (Residues 40 - 149); Ec: 6.1.1.6; Engineered: Yes	24	7.1	22%	48%	
1J7G	Chain A, Structure Of Yihz From Haemophilus Influenzae (Hi0670), A D- Tyr-Trna(Tyr) Deacylase	23	9.3	34%	48%	
1US8	Chain B, The Rad50 Signature Motif: Essential To Atp Binding And Biological Function	23	9.3	27%	69%	3%)
1II8	Chain B, Crystal Structure Of The P. Furiosus Rad50 Atpase Domain	23	9.3	27%	69%	3%)
1F2T	Chain B, Crystal Structure Of Atp-Free Rad50 Abc-Atpase	23	9.3	27%	69%	3%)

Table 2 (b)	Sequences producing significant alignments:	Score (bits)	E Value	Identities	Positives	Gaps
1PX5	Chain B, Crystal Structure Of The 2'-Specific And Double-Stranded Rna-Activated Interferon-Induced Antiviral Protein 2'-5'- Oligoadenylate Synthetase	24	7.2	43%	69%	
1OE6	Chain B, Xenopus Smug1, An Anti-Mutator Uracil-Dna Glycosylase	24	7.2	37%	58%	3%
1T11	Chain B, Trigger Factor	23	9.4	32%	61%	

Table 2 continued overleaf

Table 2 (c)	Sequences producing significant alignments:	Score (bits)	E Value	Identities	Positives	Gaps
1OSM	Chain C, Osmoporin (Ompk36) From Klebsiella Pneumoniae	27	0.85	44%	55%	
1OPF	The Structure Of Ompf Porin In A Tetragonal Crystal Form	27	1.5	43%	59%	6%
1HXX	Chain A, Ompf Porin Mutant Y106f	27	1.5	43%	59%	6%
1HXU	Chain A, Ompf Porin Mutant Kk	27	1.5	43%	59%	6%
1HXT	Chain A, Ompf Porin Mutant Nqaaa	27	1.5	43%	59%	6%
1BT9	Chain A, Ompf Porin Mutant D74a	27	1.5	43%	59%	6%
1GFQ	Ompf Porin (Mutant R82c)	27	1.5	43%	59%	6%
1GFP	Ompf Porin (Mutant R42c)	27	1.5	43%	59%	6%
1GFO	Ompf Porin (Mutant R132p)	27	1.5	43%	59%	6%
1GFN	Ompf Porin Deletion (Mutant Delta 109-114)	27	1.5	43%	59%	6%
1GFM	Ompf Porin (Mutant D113g)	27	1.5	43%	59%	6%
1MPF	Matrix Porin (Ompf) Mutant With Gly 119 Replaced By Asp (G119d)	27	1.5	43%	59%	6%
1H6V	Chain F, Mammalian Thioredoxin Reductase	25	3.2	68%	78%	5%
1MIQ	Chain B, Crystal Structure Of Proplasmepsin From The Human Malarial Pathogen Plasmodium Vivax	25	4.2	26%	50%	
1W0P	Chain A, Vibrio Cholerae Sialidase With Alpha-2,6-Sialyllactose	24	7.2	36%	63%	
1EW9	Chain B, Alkaline Phosphatase (E.C. 3.1.3.1) Complex With Mercaptomethyl Phosphonate	24	7.2	26%	45%	3%
1I7A	Chain D, Evh1 Domain From Murine Homer 2bVESL 2	24	7.2	37%	49%	5%
1B8J	Chain B, Alkaline Phosphatase Complexed With Vanadate	24	7.2	26%	45%	3%
2ANH	Chain B, Alkaline Phosphatase (D153h)	24	7.2	26%	45%	3%
1URB	Chain B, Alkaline Phosphatase (N51mg)	24	7.2	26%	45%	3%
1KIT	Vibrio Cholerae Neuraminidase	24	7.2	36%	63%	
1HQA	Chain B, Alkaline Phosphatase (H412q)	24	7.2	26%	45%	3%
1HJK	Chain B, Alkaline Phosphatase Mutant H331q	24	7.2	26%	45%	3%
1ELZ	Chain B, E. Coli Alkaline Phosphatase Mutant (S102g)	24	7.2	26%	45%	3%
1ELY	Chain B, E. Coli Alkaline Phosphatase Mutant (S102c)	24	7.2	26%	45%	3%
1ELX	Chain B, E. Coli Alkaline Phosphatase Mutant (S102a)	24	7.2	26%	45%	3%
1ANJ	Chain B, Alkaline Phosphatase (K328h)	24	7.2	26%	45%	3%
1ANI	Chain B, Alkaline Phosphatase (D153h, K328h)	24	7.2	26%	45%	3%
1ALJ	Chain B, Alkaline Phosphatase Mutant (H412n)	24	7.2	26%	45%	3%
1EHI	Chain B, D-Alanine:d-Lactate Ligase (Lmddl2) Of Vancomycin-Resistant Leuconostoc Mesenteroides	24	9.4	50%	66%	
1MCW	Chain W, Immunoglobulin Heterologous Light Chain Dimer (MCG-WEIR Hybrid)	24	9.4	26%	43%	1%

similarity. From these results we can conclude that it is not possible to predict the structures for any portion other than the Ig-binding region of Sbi. The level of sequence similarity of the domains with the Ig-binding domains of SpA is very high, with E-values of less than 5.00 E^{-11} (E-values obtained from the BLAST search).

In order to identify proteins homologous to regions other than the Ig-binding domains we searched the databases using several truncated Sbi sequences.

6.1.3.3.1 Identification of sequences homologous to the β2GPI-binding domain, the polyproline region and the C-terminal region of Sbi.

In order to identify sequences homologous to the β2GP1-binding domain, the polyproline region and the C-terminal region of Sbi, truncated Sbi sequences of each of these domains were used to search the structure prediction server 3D-PSSM for homologous structures. However, all the structures identified with homology to these regions of Sbi had E values of less than 1.00E^{-02} (see Table 2). In addition, although many of the structures homologous to these regions of Sbi had structures composed of α-helices, none of the structures identified by the searches showed a conserved fold. It is therefore impossible to use these results alone to predict a fold motif for the β2GP1-binding domain, the polyproline region or the C-terminal region of Sbi. Other information would be required to make a prediction of the structure of these regions of the Sbi protein.

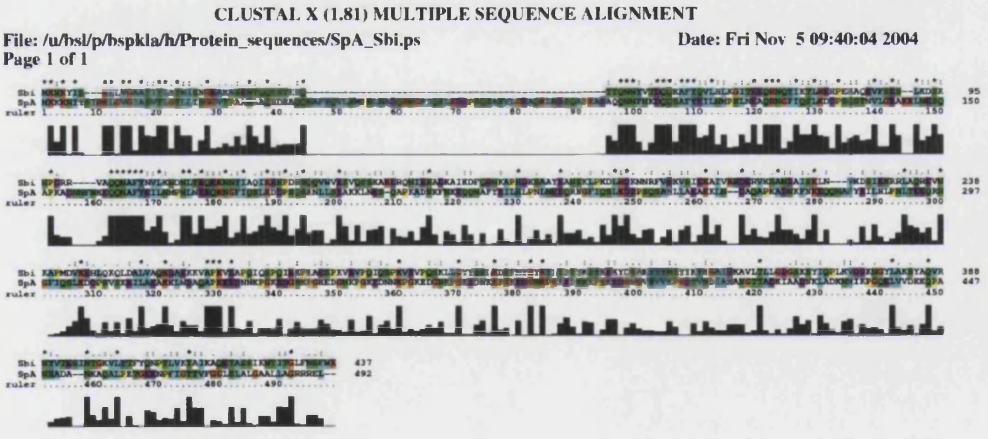


Figure 12 Sequence alignment of Sbi and SpA generated by ClustalX.

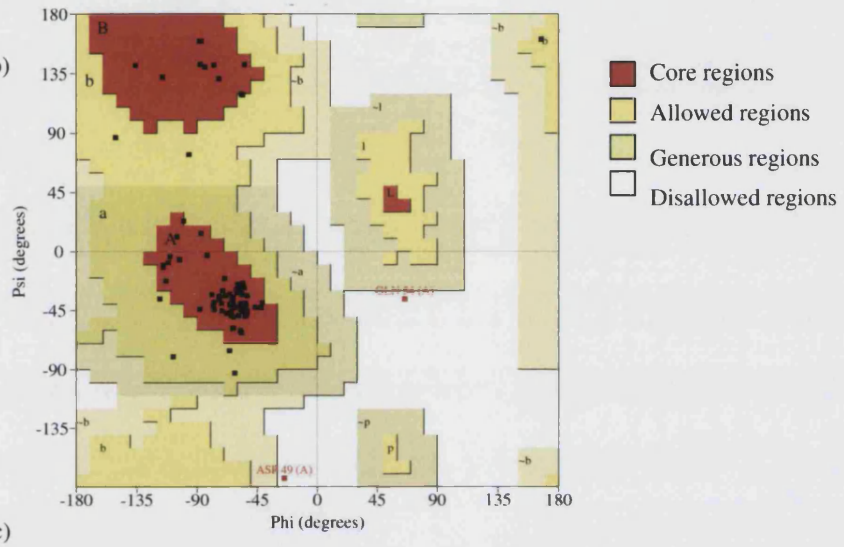
6.1.3.4 Homology modelling of Sbi's immunoglobulin binding domains

In Figure 12 the full-length protein sequences of both Sbi and SpA are aligned. SpA consists of five Ig-binding domains (known as E, D, A, B, and C) in addition to two domains involved in cell wall anchoring. Sbi consists of two Ig-binding domains, a β2-glycoprotein binding domain, a proline-rich region, and a proposed intracellular C-terminal domain. The sequence alignment most notably aligns the first Ig-binding domain of Sbi to the second Ig-binding domain of SpA (D). This alignment shows conservation in the first four domains of Sbi, but at a much higher level in the Ig-binding domains. On this basis, the Sbi Ig-binding domains are expected to adopt the three-helix bundle fold typical for the SpA domains. The regions C-terminally from the Ig-binding domains (including the β2-glycoprotein binding domain) show a certain level of homology with SpA, suggesting that they may have some structural similarity with SpA's Ig-binding domains.

As the Sbi Ig-binding domains share high sequence similarity to those of SpA, the structure of SpA in complex with the Fab fragment of IgM (1DEE) (Graille *et al* 2000) was selected as a structural template for homology modelling. This structure was chosen rather than the structure of SpA in complex with Fc (1FC2) determined by Deisenhofer (1981), in which the C-terminal third of the molecule is unstructured. In 1DEE (and other the SpA structures solved by NMR) all 3 helices are well defined. Both Sbi Ig-binding domains were modelled using 1DEE as a template and subsequently linked together, using an extended spacer peptide (Figure 13). Analysis of the model with both a Ramachandran plot and Verify 3D showed that all backbone angles were in allowed conformations and that all side chains are in favourable environments (Figure 13).

(a)

(b)



(c)

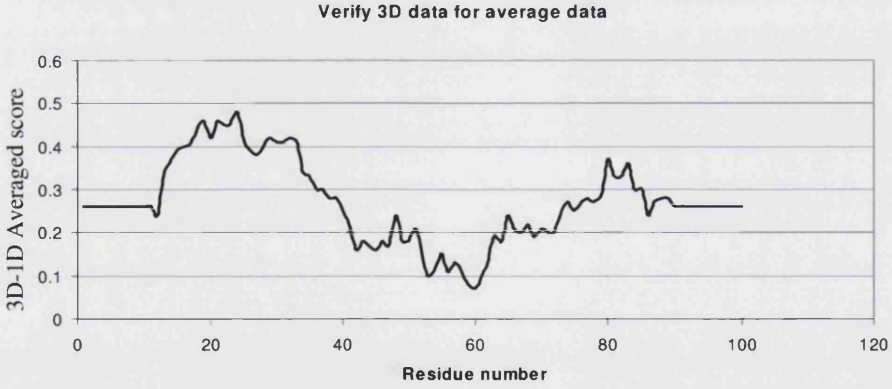


Figure 13(a) Predicted structure of Sbi's two Ig-binding domains **(b)** Ramachandran plot of the two Ig-domain model, produced by Procheck (Laskowski *et al* 1993), showing the majority of residues (black squares) fall within core and allowed regions. Residues falling in generous and disallowed regions are labelled in red. **(c)** Verify 3D analysis of the two-domain structure. The result for the Sbi model is entirely in the upper portion (above the x-axis) of the plot, suggesting it has been correctly modelled.

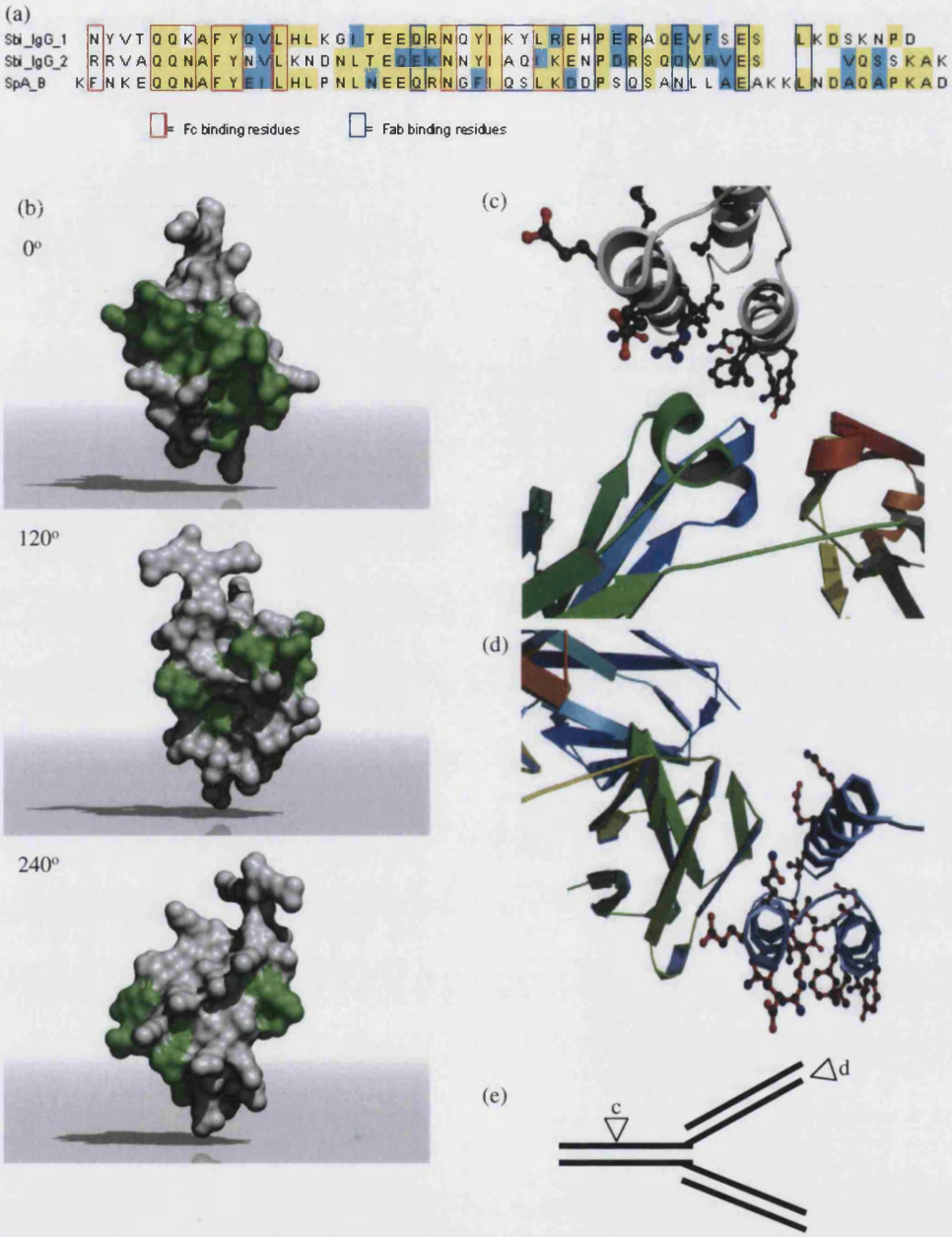


Figure 14 (a) Alignment of the Sbi Ig-binding domains to the B domain of SpA, with the residues implicated in binding indicated. **(b)** Surface structure of the first Sbi Ig-binding domain with conserved residues highlighted in green, forming a stripe on the surface of the molecule; the structure is shown in 120° rotations. The structure of the first Sbi Ig-binding domain superimposed on the **(c)** SpA-Fc complex (Deisenhofer 1981) and the **(d)** SpA-Fab complex (Graille *et al* 2000), with those residues that are identical between the two sequences represented by ball-and-stick side chains. **(e)** shows the location on intact IgG of the complexes shown in (c) and (d).

6.1.3.4.1 Prediction of Fc/Fab binding characteristics of Sbi

Superimposing the Sbi structures onto the SpA part of the SpA/IgG Fc complex and the SpA/IgG Fab complex allows comparison of the residues of SpA involved in the interaction with IgG to those of in the corresponding location in the Sbi structure (see Figure 14c&d).

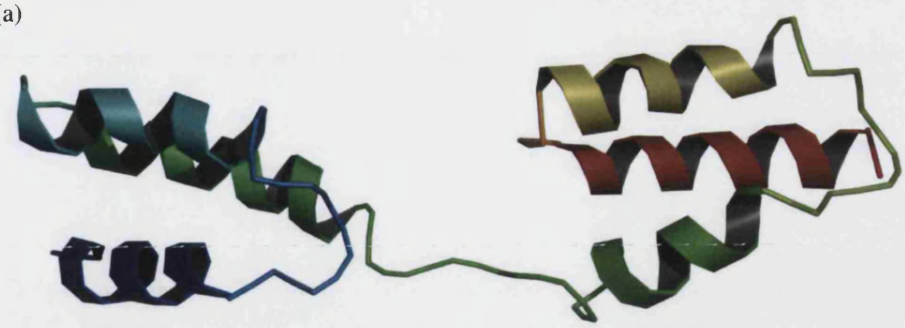
In the X-ray structure of SpA bound to the Fc fragment of IgG, the Fc fragment only interacts with two of these three helices. The SpA residues involved in binding Fc show high conservation in Sbi, as is indicated in Figure 14b. This suggests that Sbi very likely binds to the Fc region of the IgG molecule.

The structure of the SpA/IgG Fab complex and Sbi were superimposed in a similar fashion. This revealed that of the large number of residues in SpA involved in Fab binding only three were conserved in Sbi (Figure 14a, c & d). On this basis, we suggest that is unlikely that Sbi interacts with Fab in a fashion similar to SpA.

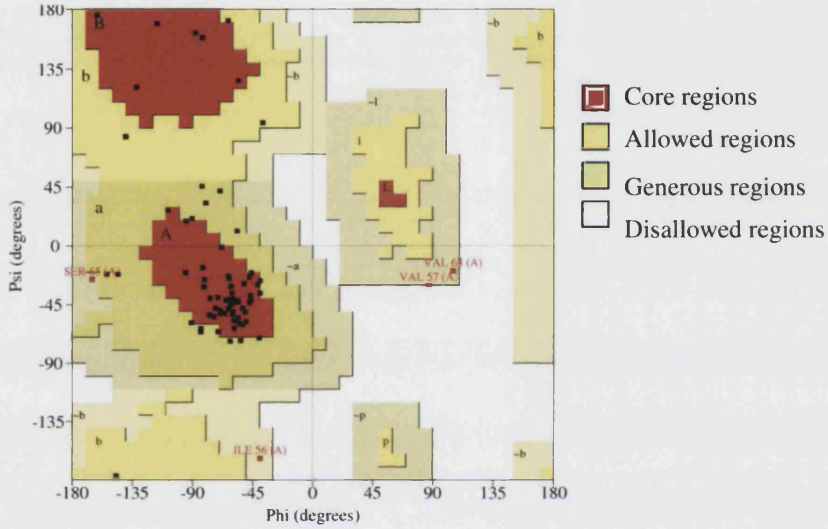
6.1.3.4.2 Predicted structure for the β_2GP1-binding region.

While the structure prediction servers did not detect any good matches for the Sbi fragments containing the β2GP1-binding domain, the sequence alignments show that it does have some similarities to the SpA Ig-binding domains. We therefore attempted to model the sequence of this region of Sbi onto the SpA structure 1DEE. However, the region of Sbi between the second Ig-binding domain and the poly-proline region is twice as long as the SpA Ig-binding domains, so the sequence was modelled onto two of these domains (Figure 15a) linked together. While the backbone angles all fall within allowed regions of the Ramachandran plot (Figure 15b), Verify 3D identifies several areas of the model with residues in unfavourable environments suggesting that this model may be incorrect (Figure 15c).

(a)



(b)



(c)

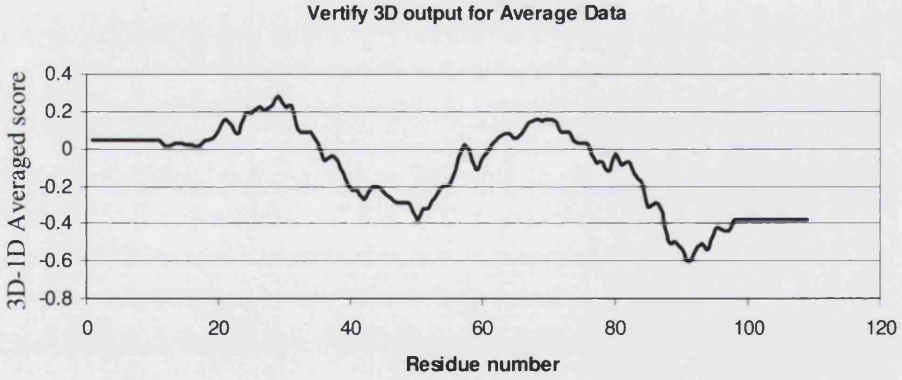


Figure 15 **(a)** Predicted structure of the β_2GP1-binding domain, including the region between this domain and the second Ig-binding domain. **(b)** The Ramachandran plot of the structure produced by Procheck (Laskowski *et al* 1993) showing the majority of residues (black squares) fall within core and allowed regions. Residues falling in generous and disallowed regions are labelled in red **(c)** Results produced by Verify_3D,. The result for this portion of the Sbi protein is almost entirely in the lower portion (below the x-axis) of the plot, indicating the poor quality of the model.

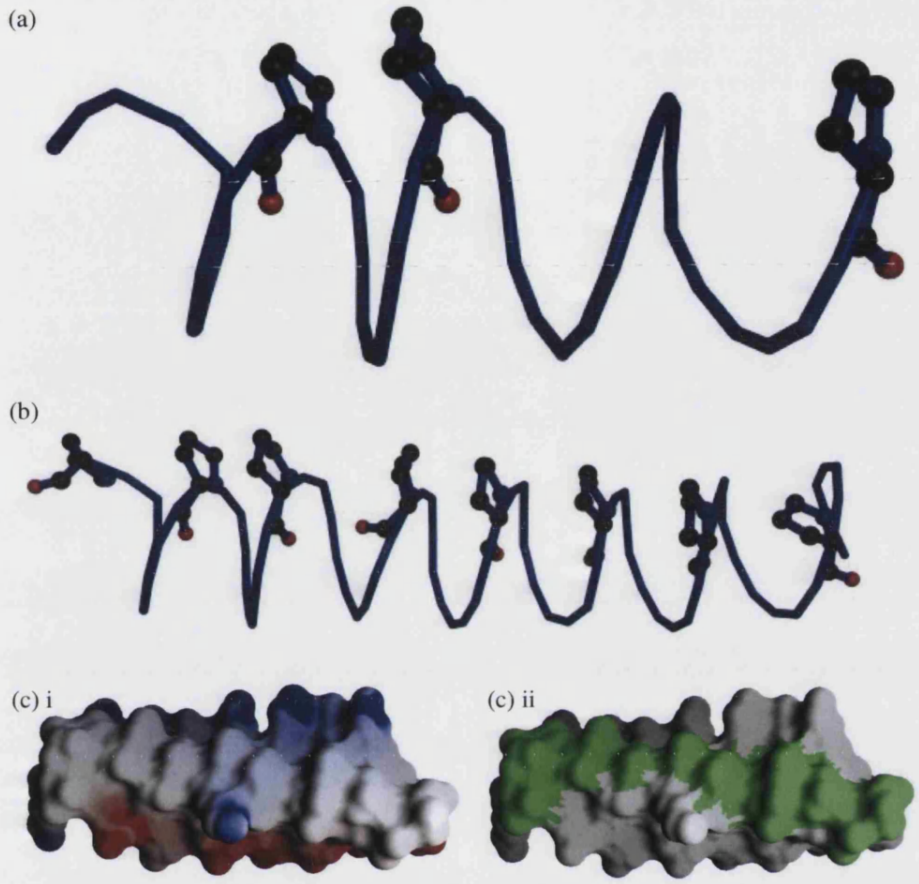


Figure 16 (a) Ribbon diagram of proline pipe-helix structure Tus (Butcher *et al* 1996, PDB accession code: 1SUT) **(b)** Backbone structure of the polyproline repeat region model from Sbi (proline side chains are shown in ball-and-stick representation). **(c)** Molecular surface representation of Sbi's polyproline repeat region homology model, coloured to indicate **(i)** surface charge (+ in blue and – in red) and **(ii)** hydrophobicity. All surfaces are generated with GRASP (Nicholls, Sharp and Honig 1991).

6.1.3.4.3 Predicted structure for the polyproline region.

A suitable template structure for the poly-proline domain was not apparent in the BLAST searches or structure prediction server results. However we were able to identify a poly-proline pipe-helix in the literature (1SUT, Figure 16a, Butcher *et al* 1996) that displays the same proline repeat unit length (a proline residue every fifth amino acid). This proline pipe-helix sequence is shorter than the region in Sbi. On this basis, we decided to use 1SUT as a template, and extend it to account for the longer Sbi sequence (Figure 16b). Most of the backbone angles do not fall within the allowed

regions of the Ramachandran plot, but neither do the angles for the template structure, because of the rare structural arrangement of the amino acids in this helix. It is not possible to analyse the structure with Verify 3D as the peptide is not long enough for such an analysis. Interestingly, the analysis of the surface hydrophobicity and surface charge distribution reveals a band of hydrophobic residues on either side of the molecule, along the longest axis of the molecule (Figure 16c), and also a band of positively changed and a band of negatively changed amino acids.

6.1.4 Discussion

6.1.4.1 Homology analysis as a basis for predicting the structure of Sbi

We used BLAST searches and structure prediction servers to identify any structures with homology to the Sbi protein sequence in order to be able to predict folds for the various Sbi regions (Table 1). We found high homology in the N-terminal region of Sbi (the Ig-binding domains) to the Ig-binding domains of SpA suggesting that these domains in Sbi form the same three-helix bundle fold that they do in SpA. However for the remainder of the Sbi protein, we were unable to determine any probable fold. None of the prediction servers predicted a common fold for the β2GP1-binding domain, the polyproline region or the C-terminal region. Therefore either Sbi contains unique folds in these regions, or is composed of an already identified fold, but that is not identifiable by the prediction servers. However, hits generated by the prediction servers did tend to favour α-helical structures, and this is in general agreement with the secondary structure predictions, suggesting that most of the Sbi protein has a mainly α-helical secondary structure.

6.1.4.2 Homology analysis as a basis for predicting the function of Sbi

BLAST database and structure prediction searches generated a large number of hits with homology to the Ig-binding domains of SpA. These hits include the region in Sbi that was shown to be a functional Ig-binding domain by Zhang *et al* (1998) and the second, putative Ig-binding domain for which the Ig-binding properties have yet to be confirmed (see Chapter 3 for studies into the binding properties of the second Ig-binding domain of Sbi).

The other regions of Sbi, the β2GP1-binding domain, the polyproline region and the C-terminal region also produce a number of hits in the database searches. However, all these hits had high E values (higher than $100.E^{-07}$). The database searches using truncated Sbi sequences also failed to produce any matches with patterns of common structure or function for either the β2GP1-binding domain, the polyproline region or the C-terminal region. These regions of Sbi may therefore display novel structural and functional features.

6.1.4.3 Predictions for the function of Sbi's Ig-binding domains

Based on the high sequence identity and functional similarity of the Ig-binding domains of SpA and Sbi we predict that both the Sbi Ig-binding domains adopt a similar structure to those of SpA. This is also important in predicting the functionality of Sbi. The early work on Sbi by Zhang *et al* (1998) showed IgG binding ability in the first Sbi Ig-binding domain but the method of phage display that they used did not provide a fragment of the second putative Ig-binding domain.

Comparison of the models of both the Sbi Ig-binding domains with the structures of SpA's Ig-binding domains (Deisenhofer 1981, Graille *et al* 2000) reveals that there are regions of sequence conservation in the 3D structure (Figure 14). The residues that are conserved between the two proteins, Sbi and SpA, are conserved in a patch on one face of the molecule, shown in Figure 14b. This conservation corresponds to the face of the SpA molecule identified to bind the Fc fragment of IgG (Figure 14c, Deisenhofer 1980), while the face on the SpA molecule that binds the Fab fragment in IgM (Figure 14d, Graille *et al* 2000) is not conserved. This suggests the Fc binding characteristics of the SpA domains is conserved in both of the Ig-binding domains of Sbi. We also predict that Sbi lacks the Fab binding characteristics observed in SpA as this region is not conserved in the corresponding Sbi domains (Graille *et al* 2000). It is possible that Sbi does not bind Fab at all or that it displays a V_H chain specificity for Fabs that is different from SpA.

6.1.4.4 Predicted structure of β_2GP1-binding region

The sequence alignment of Sbi showed homology between the β_2GPI-binding region of Sbi and the corresponding domains of SpA suggesting that there was some

conservation between Sbi and SpA in this region. Therefore we modelled this region of Sbi onto two three-helix bundle domains of SpA, linking them together. Analysis of the side-chain environment of this model suggested that this is not a favourable conformation for Sbi. Circular dichroism (CD) experiments carried out on a protein fragment containing this region of Sbi (data not shown) confirm our prediction that this region is α-helical.

6.1.4.5 The structure of the polyproline repeat region and its role in anchoring Sbi to the bacterial cell wall

SpA and Sbi share a sequence feature other than the conserved Ig-binding domains, the presence of a proline rich region. Little is known about these elements as studies on SpA have concentrated on the role of the Ig-binding domains, but it has been implicated in cell-wall binding (Tashiro and Montelione 1995). However, there are also differences in the sequence of this region between SpA and Sbi. In both proteins the proline residues form part of a proline repeat element. The main differences are that in SpA it is an octapeptide repeat and there is a large variation in the number of repeats between SpA sequences (see also chapter 5). In Sbi a proline occurs every 5^{th} position and there is no variation in the number of repeats. There are eight of these repeats in the Sbi sequences found in the database so far (see also chapter 5).

The homology model of the polyproline region of Sbi was generated based on the structure of a polyproline pipe helix from the replication arrest protein Tus (Butcher *et al* 1996). Of interest to note, is that when the hydrophobicity of the model was examined, two hydrophobic bands were found on either side of the protein as well as two bands composed of charged residues. Could this be of any functional importance? In other *S. aureus* cell surface adhesion proteins, the proline-rich regions are known to be a feature of the cell-wall binding regions (Zhang *et al* 2000). MSCRAMM proteins such as the Sdr proteins, ClfA and ClfB, FnBPB, Cna, and of course SpA, all contain proline rich regions associated with cell wall binding. However in these proteins the proline-rich region is followed by the LPXTG motif, which is proteolytically cleaved for anchoring to the cell-wall. Sbi lacks this motif, and is probably not proteolytically cleaved. The question if and how Sbi is covalently attached to the cell-wall and cell membrane therefore remains although it seems likely that it is attached to the cell wall

via the proline rich region. Is it perhaps linked to another membrane protein; or is it directly linked to the membrane though a novel mechanism?

The C-terminal region of Sbi is also very different from other *S. aureus* cell-surface adhesion proteins and our database searches did not identify any matches with proteins of known function in this region of Sbi. It is therefore difficult to identify the role of the C-terminal domain or its structure. Interestingly, this region of Sbi contains a large number of tyrosine and threonine residues, which can perhaps function as a target for phosphorylation (Zhang *et al* 2000).

The only way forward to find and answer these questions is to clone, express and characterise the Sbi domains both structurally and functionally. In the next chapter I will describe the cloning, expression, and purification of the N-terminal domains of Sbi and compare their immunoglobulin binding characteristics with SpA.

7 Chapter 3

7.1 A comparison of the immunoglobulin binding properties of *Staphylococcus aureus* adhesion proteins SpA and Sbi

7.1.1 Introduction

Ig-binding proteins expressed by bacteria on their surface are thought to be important in masking the bacterium from the host's immune system. *S. aureus* is known to express two Ig-binding proteins on its surface. These are the proteins SpA and Sbi. SpA has been shown to have a role in infectious diseases including mastitis, subcutaneous lesions and arthritis (Jonsson *et al* 1985, Patel *et al* 1987, Palmqvist *et al* 2002). In addition SpA has been shown to be important, through its role in binding IgGs in phagocytosis, as the absence of IgG in the serum increases the rate of phagocytosis. The Ig-binding characteristics of SpA have previously been characterised by Kronvall *et al* (1970), Jansson, Uhlen, and Nygren (1998), and Oda *et al* (2003) and its interactions with serum IgGs and the IgG fragments described in detail. Little is known about the IgG-binding properties of Sbi. Zhang *et al* (1998) have described the binding of the first Ig-binding domain of Sbi with serum IgG. Here we will further investigate and compare the interactions of the Sbi Ig-binding domains with both serum IgGs, IgG fragments and subclasses to those observed with SpA. This comparison is of particular interest as SpA and Sbi have a different number of Ig-binding domains (4 or 5 for SpA against 2 for Sbi) as described in Chapter 1, and we will investigate if this results in a difference in avidity for IgG molecules in the binding characteristics of the two proteins.

7.1.1.1.1 Immunoglobulin binding characteristics of SpA

Early studies characterising the binding of SpA to Ig, present in the blood sera, used Ouchterlony assays (Kronvall *et al* 1970). This method was used to study both precipitation of IgG by SpA and inhibition of the human IgG-SpA reaction. This showed that SpA binds to most mammalian IgGs, although not all mammalian IgGs are precipitated by SpA. Analysis of non-mammalian serum Igs showed no reaction with SpA.

Later the enzyme-linked immunosorbant assay (ELISA) method became the most popular technique to study binding of SpA with Igs. The information supplied with the commercially sold SpA from Pierce (http://www.piercenet.com/products/browse.cfm?fldID=01021504) lists the binding of SpA for the serum IgG of several species, and includes binding results with some of the IgG subclasses, and the other types of Ig.

More recent binding studies with SpA concentrate on measuring the affinity of SpA for human or mouse immunoglobulins (Jansson, Uhlen, and Nygren 1998, Oda *et al* 2003). Jansson Uhlen and Nygren (1998) used surface plasmon resonance (SPR) to determine the relative binding of SpA to human IgG fragments, showing that all SpA Ig-binding domains display near equal relative binding with the Fc fragment, while the SpA domains A, B, and C bind the Fab fragment better compared to the SpA E and D domains (Jansson, Uhlen, and Nygren 1998, Oda *et al* 2003). The artificial Z domain, based on the B domain, differing only by a single amino acid substitution G29A (Tashiro and Montelione 1995, Jansson, Uhlen, and Nygren 1998), binds only Fc fragments. Oda *et al* (2003) used SPR to measure the affinity of SpA for mouse monoclonal antibodies, of two IgG subclasses, IgG1 and IgG2. Using full-length recombinant SpA and a construct containing only the SpA B domain, they found similar affinities for both forms of SpA for each monoclonal IgG, although full-length SpA had a greater affinity for mouse IgG2 than for IgG1.

7.1.1.1.2 Immunoglobulin binding characteristics of Sbi

Zhang *et al* (1998, 1999) characterised the structure of the Sbi gene with a phage display shotgun approach. This method identified functional domains of IgG binding and β2GP1 binding. Analysis of the gene sequence identified a second possible IgG binding domain, although the phage display method only detected the first one. The data available on the binding of the N-terminal Ig-binding domain of Sbi with Igs was obtained using an ELISA-based method. Sbi displayed the highest relative binding for human, pig, guinea pig and rabbit serum IgGs. Sbi does not bind the human subclass IgG3, rat, goat or chicken IgG. Sbi does show some binding for IgA and IgM (see Figure 17).

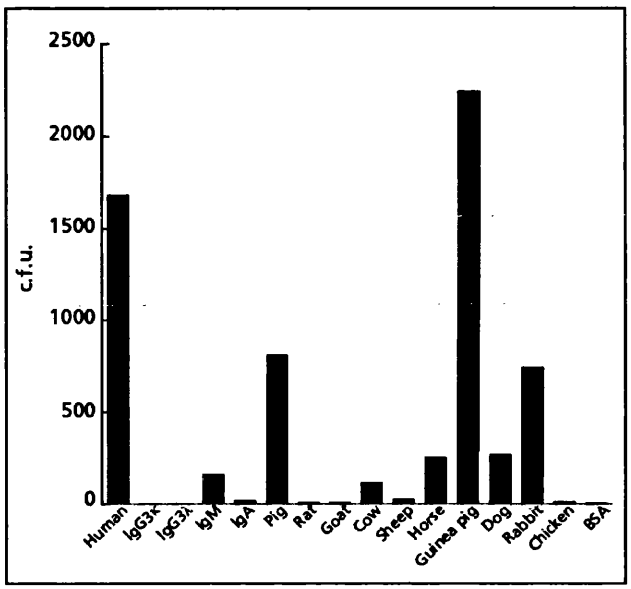
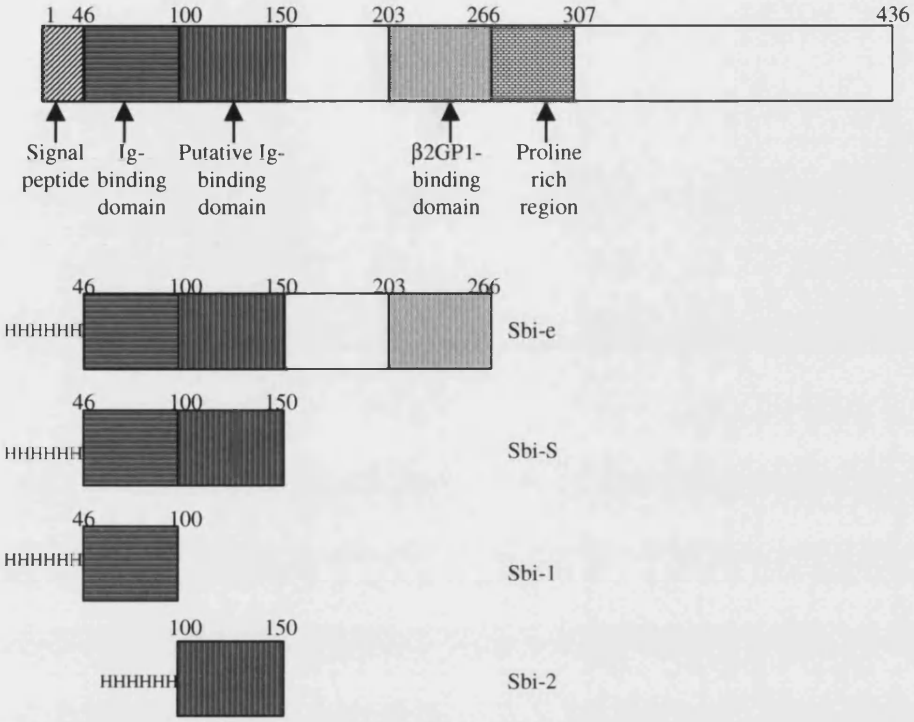


Figure 17 ELISA results of Sbi with IgG, taken from Zhang *et al* 1998.

7.1.1.1.3 Aim of this study

Here we analyse the interactions between Sbi and immunoglobulins from mammalian and avian sources, species that are like humans affected by *S. aureus* infections. In addition, the interactions between Sbi and human IgG are studied in more detail by analysing IgG subclass specificity and the interaction with Fc and Fab fragments. These results are than compared to our results from the interaction between IgG and SpA. In order to study the types of complexes formed between Sbi and immunoglobulins, the stoichiometry of the interaction, the role of multiple Ig-binding domains (avidity) and individual Ig-binding domains (affinity) in the formation of these complexes we generated four Sbi protein constructs. These constructs are comprised of an N-terminal construct (referred to as Sbi-e) containing the Sbi region adjacent to the poly-proline repeat region, a construct containing the two proposed Ig-binding domains (Sbi-S) and two constructs containing each of the Ig-binding domains (Sbi-1 and Sbi-2). The organisation of these constructs is shown in Figure 18. To analyse all these different aspects of complex formation we used size exclusion (SE) chromatography in addition to agar diffusion studies and surface plasmon resonance (SPR).

Figure 18 The structure of the Sbi gene, and the 4 constructs cloned from it.



While the ELISA method gives us the relative affinity of the two proteins for Igs, it does not produce a complete picture of the complex formation as can be gained though use of several different methods. In the original results from Kronvall *et al* (1970), the inhibition method gives only qualitative results. Agar diffusion (or Ouchterlony assay) methods can only gives qualitative results by indicating whether or not a complex forms an insoluble precipitate and whether this complex formation can be inhibited. It cannot be used to quantify the binding relationships, or infer complex sizes and therefore does not give any information about complex composition. And where quantitative methods such as ELISA or SPR have been used, it still tells us very little about what kind of complexes are formed, and in the case of the ELISA results it tells us only that binding occurs, it can not be determined whether there are differences in the formation of complexes in the serum IgGs of different species. There are also other limitations with the use of ELISA, particularly when polyclonal Ig is used, as was done by Zhang *et al* (1998). Friguet *et al* (1995) described how there is no simple correlation between absorbance measured by ELISA, the amount of each species of bound molecule, and the concentration of free

molecules. This means that with a mixture of IgG molecules, such as those found in serum, the different IgG subclasses will bind the ligand (for example Sbi) with different affinities. Therefore those IgG subclasses with highest affinity for the ligand will bind when there is a low concentration of ligand present, while with high concentrations of ligand, a disproportionally high fraction of IgG molecules with high affinity will bind to the ligand.

7.1.1.1.4 Size Exclusion Chromatography

The principle of SE chromatography is the separation of particles based on their size. For our experiments, it allowed us to measure the sizes of complexes composed of components of known size. Once insoluble components of the mixture have been removed, it is possible to estimate the size of the peaks eluted from the column, using the elution times of proteins of known molecular weight (MW) and predict their composition. A limitation of this quantitative method is that it does not give any kinetic information for the eluted complexes, and it also has limitations in the separation resolution of the sepharose column, meaning, it is difficult to estimate the size of peaks outside the calibration range of the column, and overlapping peaks can cause difficulties in obtaining an exact elution volume (Goetz *et al* 2004, Stulik, Pacakova, Ticha 2003).

7.1.1.1.5 Surface Plasmon Resonance

In order to study binding kinetics of the complexes between IgGs and Ig-binding proteins, SPR was employed. This is an optical technique that uses the evanescent wave phenomenon to measure changes in the refractive index very close to a sensor surface (McDonnell 2001). SPR measures the reflected light intensity as a function of incident light angle (see Figure 19).

A polarised light beam is reflected though a prism, onto a gold covered surface. Only at a certain incident light angle is the light refleated back into the prism (see Figure 19); this is known as total internal reflection and is the SPR angle. Some energy in the light beam is lost as an energy wave (surface plasmon) upon interaction with the gold surface, resulting in a reduction of intensity of the reflective light. If the surface of the gold changes, i.e. a protein layer is absorbed, changes in the refractive index occur on the surface and results in a change of the SPR angle. There is a linear relationship

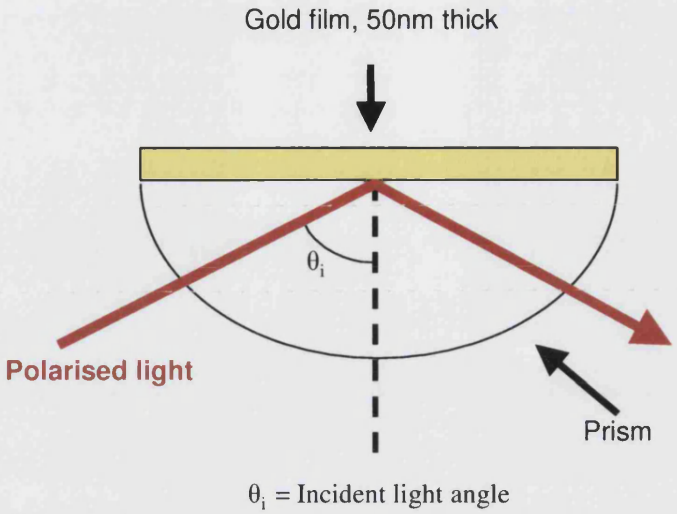


Figure 19 Polarised light is introduced into a gold film (θ_i), at an angle for total internel reflection to occur. Changes in the refractive index on the gold surface due to binding of protein molecule to the surface result in a change in θ_i needed for this to occur.

between the amount of bound protein and the shift of the SPR angle (http://www.xantec.com/html/spr.html). This shift allows the measurement of real time changes in the gold surface as protein molecules are bound to the surface (Green *et al* 2000). Measurement of the real time response (R_t) allows the rate of k_s (or k_{on}) to be determined. k_s is the pseudo 1st order rate constant. The relationship between R_t and k_s is:

$$R_t = R_{eq} (1 - \exp -(k_{ass} [P] + k_{dis}) t) \qquad (1)$$

$$R_t = R_{eq} (1 - \exp (-k_s t)) \qquad (2)$$

$$k_s = k_{ass} [P] + k_{dis} \qquad (3)$$

Therefore with at least R_t vs. time curves for three different analyte concentrations, the real time response curves can be fitted to the above equation (2) to produce a value for k_s (see Figure 22a). Chi square (χ^2) indicates the quality of the fit of the data to the equation. A plot of k_s vs. analyte concentration [P] gives the association and dissociation rate constants from the gradient – k_{ass} and the y-axis intercept - k_{dis} (see Equation (3) above and Figure 22b). The K_A and K_D equilibration constants are determined from k_{ass} and k_{dis} (see Figure 22c).

Table 3 Concentration of IgG stock solutions determined by BIORAD assay and sizes of serum IgGs as determined by size exclusion chromatography. The human IgG subclasses[&] concentrations for SE experiments were taken as given by Sigma (1mg/ml). *The SE profile of Human IgG1 appear abnormal, indicated partly by the lower than expected molecular weight, however the protein sample appeared normal on an SDS-PAGE.

Species of IgG Origin	IgG peak sizes as determined by SE (kDa)	Concentration for SE and diffusion experiments (mg/ml)	Concentration for SPR experiments (mg/ml)	Molar Concentration for SPR experiments (M)
Chicken serum IgG	590 227	5.32	9.56	6.37E-05
Cow serum IgG	417 169	5.86	18.87	1.26E-04
Goat serum IgG	379 169	6.52	7.99	5.33E-05
Guinea Pig serum IgG	462 192	9.89	2.01	1.34E-05
Horse serum IgG	430 178	6.68	7.38	4.92E-05
Human serum IgG	285 155	6.26	6.69	4.46E-05
Mouse serum IgG	212	7.90	7.90	5.27E-05
Pig serum IgG	358 151	6.33	5.87	3.91E-05
Rabbit serum IgG	407 186	5.84	6.5	4.33E-05
Sheep serum IgG	474 193	9.61	6.51	4.34E-05
Human IgG1	42*	1.0[&]	0.12	8.00E-07
Human IgG2	187	1.0[&]	0.29	1.93E-06
Human IgG4	479 214	1.0[&]	0.54	3.60E-06
Human IgG3	268	1.0[&]	N/A	N/A
Human Serum Albumin (HSA)	Not determined	13.34	N/A	N/A

By combining these three methods we aim to get a much clearer understanding of the interactions of Sbi with IgG, and also how these differ in the complexes formed with SpA.

7.1.2 Materials and methods

Serum IgGs were supplied as lyophilized powder by Sigma for several mammalian species and for chicken as a representative of a non-mammalian species. Those IgGs used were dissolved in PBS buffer, and the species used are listed in Table 3, along with their concentrations as determined by the Bio-Rad Protein Assay method (with

BSA as the standard). 25µl of the stock solutions of the serum IgG (50µl for chicken IgG) were supplemented with equilibration buffer (20mM Tris-HCl, 200mM NaCl pH7) to a volume of 100 µl and spun at maximum speed in the micro-centrifuge for 10 minutes to remove precipitated protein. The sample was then injected into the size exclusion column to determine the IgG size in kDa (see below for size exclusion method, determined sizes and concentrations of serum IgGs are shown in Table 3).

The four human κ IgG subclasses were also obtained from Sigma, provided at a concentration of 1mg/ml in Tris buffered saline pH8. 100µl of this was added to 100µl of 2x PBS and injected into the size exclusion column to obtain an elution profile. IgG1κ produced an atypical profile for an IgG (as shown in Figure 31a, IgG1 is eluted much later than the other IgGs in Figure 28 and Figure 31), but as it appeared normal on a SDS-PAGE, this was taken as standard for this IgG. The concentrations were determined for SPR experiments using BIO-RAD assay to allow the determination of molar concentration of the protein, shown in Table 3 along with size exclusion sizes. A calibration curve for the BIORAD assay is shown in Figure 20

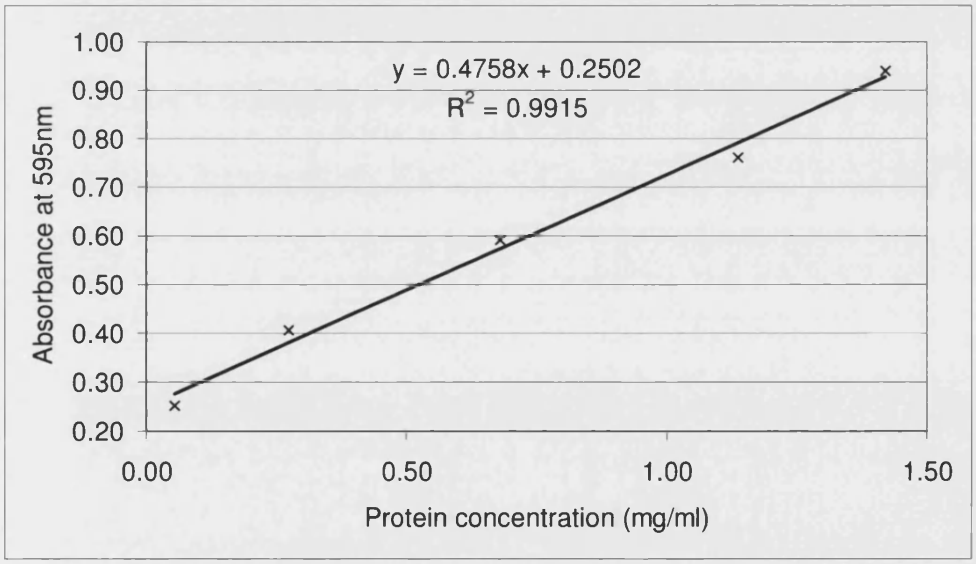


Figure 20 Calibration curve for the BIORAD assay, using a BSA standard. The spectrophometer was zero-ed against water.

The molar concentrations for all IgGs were calculated based on their expected molecular weights of 150kDa.

β_2GP1 was supplied in PBS at a concentration of 7.5mg/ml.

Human serum albumin (HSA) was obtained from Sigma, and was diluted in PBS. 25µl was supplemented with equilibration buffer to a volume of 100µl and injected into the size exclusion column to determine its elution profile (see Table 3 for concentration).

BSA was obtained from BIORAD protein assay kit. This was stored in 1ml aliquots at 1.42mg/ml. This was diluted 1:10 with running buffer for SPR experiments.

Recombinant SpA was obtained from Pierce and dissolved in 1x PBS to give a concentration of 10mg/ml.

7.1.2.1 Cloning of Sbi fragments

7.1.2.1.1 PCR amplification of the Sbi gene, nucleotides 86 to 799 (Sbi-e)

The Sbi gene was cloned from nucleotide 86 to 799; to be referred to as Sbi-e. The primers used were 5' – CAT GCC ATG GCG AGT GAA AAC ACG CAA CAA – 3' (forward primer **A**) and 5' – CCG CTC GAG TCA TTA CGC CAC TTT CTT TTC AGC– 3' (reverse primer **B**) incorporating NcoI (forward primer) and XhoI (reverse primer) restriction sites. All primers were obtained from MWG biotech AG. Template DNA was from *Staphylococcus aureus* strain Mu50. The PCR reaction mixture was composed of 0.5nmol of each of the primers, 10nmol of each dNTP (Roche), Pwo polymerase buffer, 4u of Pwo polymerase (both from Roche), and 1ul of template DNA per 100µl (made up to volume with sterile water). Reactions were heated at 95°C for 7 minutes and than subjected to 35 cycles of 45 seconds at 95°C; 1 minute at 63.4°C; and 2 minutes 72°C. The reactions were held at 72°C for a further 10 minutes and than stored at 4°C.

7.1.2.1.2 PCR amplification of the Sbi gene, nucleotides 86 to 469, 127 to 283 and 278 TO 469 (Sbi-S, Sbi-1 and Sbi-2)

N-terminal regions of the Sbi gene were cloned, Sbi-S (nucleotides 86 to 469) containing both the proposed Ig-binding domains; Sbi-1 (127 to 283) the first proposed Ig-binding domain; and Sbi-2 (nucleotides 278 to 469) the second proposed Ig-binding domain. Sbi-S was cloned with forward primer **A** and reverse primer **C** (5' – CCG CTC GAG GAG TCA TTA ATT TTC AAT ATT TTG ACG – 3'). Sbi-1 was cloned with forward primer **D** (5' – CAT GCC ATG GGA ACT CAA AAC AAC TAC GTA ACA – 3') and reverse primer **E** (5' – CCG CTC GAG TCA CTA GCT GTC TTT AAG TGA TTC AGA – 3'). Sbi-2 was cloned with forward primer **F** (5' – CAT GCC ATG GAC AGC AAG ACC CCA GAC CGA – 3') and reverse primer **C**. all forward primers contained an NcoI site, and the reverse primers an XhoI site. PCR reactions were set-up up as described above, and heated at 95°C for 7 minutes, then subjected to 35 cycles of 45s at 95°C, 1 minute at 47°C (for Sbi-S and Sbi-1) or 60°C (for Sbi-2), and 45s at 72°C. Reactions were held at 72°C for a further 10 minutes and than held a 4°C. A schematic representation of the constructs is shown in Figure 18.

7.1.2.1.3 Insertion into the His-tagged parallel vector

PCR fragments were purified with the QIAquick purification kit (QIAGEN) using the micro-centrifuge method and eluted into 30µl of buffer EB. Purified fragments were digested in a volume of 50µl containing 30µl of PCR fragments, Nebuffer 2, BSA (5µg), NcoI (40u), XhoI (15u) (all from New England Biolabs) and sterile water, and incubated overnight at 37°C.

pHIS parallel vector 1 was obtained from Peter Sheffield (described by Sheffield *et al* 1999). The vectors were grown in *E. coli* strain XL1-Blue™ (Stratagene) and purified from the cells with the QIAprep Spin Miniprep kit (QIAGEN) using the micro-centrifuge method. The vectors were digested in a volume of 20µl containing 5µl of purified vector, Nebuffer 2, BSA (2µg), NcoI (16u), XhoI (8u) and sterile water and digested overnight at 37°C.

2µl of digested vector and 6.5µl of digested PCR fragment were added to 200u T4 ligase and T4 ligase buffer (both New England Biolabs) (final volume 10µl) and incubated at 16°C overnight.

2µl of the ligation mixture was added to 100µl of component cells (*E. coli* XL1-blue™). After ½ hour on ice, the cells were heat shocked at 42°C for 30 seconds. After incubation for 1 hour at 37°C in Luria-Bertani (LB) media (10g Tryptone (Fischer Biotech), 5g yeast extract (Merck) and 10g sodium chloride per litre), the cells were spread on LB agar plates (15g agar per litre of LB media) containing ampicillin (1mg per litre of LB media). The plates were incubated overnight at 37°C. Colonies were used to inoculate 10ml LB media containing ampicillin (100 mg/l) and cultured overnight at 37°C.

The vectors were purified, and those containing the correct inserts identified with PCR, restriction digests, and sequencing. The Sbi-e and Sbi-S vectors were then transformed into BL21 star™ cells (Invitogen), and the Sbi-1 and Sbi-2 vectors were transformed into Rossetta™ cells (Novagene). These clones were used to generate glycerol stocks (1ml overnight culture plus 0.5ml 50% sterile glycerol)

7.1.2.1.4 Expression in *E. coli*

1 litre *E. coli* cultures were grown in LB media with ampicillin at 100 mg per litre, inoculated with 10ml of starter culture (10ml LB media plus 10µl ampicillin grown at 37°C overnight, inoculated from either transformations on a LB ampicillin plate or glycerol stocks) at 37°C. Upon reaching an optical density (OD) of 0.6 at 600nm, the Sbi-e and Sbi-S cultures were induced with 1ml IPTG (0.5M). Sbi-e was grown for 3 hour at 28°C, while Sbi-S was grown at 25°C. Sbi-1 and Sbi-2 cultures were grown at 37°C for 4 to 6 hours, and induced with IPTG overnight at 25°C for agar diffusion experiments, and at 20°C to improve expression and solubility for the SPR experiments. Cells were collected by centrifugation (7,000rpm for 20 minutes) and subjected to a freeze-thaw cycle at -20°C.

The pelleted cells were resuspensed in lysis buffer (20mM Tris-HCl, 500mM NaCl, 10% glycerol, 1mM PMSF, 10mM β-mercaptoethanol), pH 7 for Sbi-e and pH8 for

all other clones and sonicated according to the following protocol: a cycle of 0.3 second pulse followed by 1.0 second cooling, for one minute; this is repeated 5 times with one minute of cooling between each cycle. This was carried out on ice. Soluble and insoluble fractions were separated by centrifugation.

All samples taken from cultures were analysed by SDS-PAGE on 15% glycine gels using Novex gel system (run at 180 volts until the dye front reach the base of the gel).

7.1.2.2 Purification of Sbi-e by ion exchange chromatography

Sbi-e was purified via ion-exchange chromatography as this gave purer protein than affinity chromatography on a nickel column did. However all the other constructs produced purer protein by affinity chromatography (described below). The soluble fraction of the Sbi-e cell lysate was loaded onto a Resource S column (Amersham Biosciences) connected to an AKTA purifier machine (Amersham Biosciences). The soluble fraction was passed though the column and the column washed with 2ml 10mM Tris-HCl pH7, collected in a single fraction. Bound protein was eluted with a gradient of 10mM Tris-HCl pH7 1M NaCl in two steps; a target of 40% over 5ml and 100% over a further 5ml. The eluate was collected in 1ml fractions and the Sbi-e containing fractions were determined by SDS-PAGE. Protein fractions were concentrated with Centricon-10 concentrators (Amicon)

Ion exchange chromatography gave pure protein (>95%), eluted at approximately 40% salt concentration. Sbi-e is highly soluble and could be concentrated up to 30mg/ml. However the protein was not stable and prone to degradation. Aliquots were kept at –80°C for long-term storage, and at –20°C for short-term storage. Purity and stability of the protein were determined by SDS-PAGE.

A stock solution of Sbi-e was concentrated as described above to a concentration of 23.7mg/ml, measured by UV absorption at 280nm (extinction co-efficient for Sbi-e at 280nm = 20400 M^{-1} cm^{-1}; Abs 0.1% (=1 g/l) 0.6; calculated by ProtParam: http://us.expasy.org/tools/protparam.html Gasteiger *et al* 2005). This was aliquoted into tubes containing 42µl, each containing 1mg of protein.

7.1.2.3 Affinity purification of Sbi-1 and Sbi-2

Sbi-1 and Sbi-2 were both purified by affinity chromatography with a chelating HiTrap column (Amersham Biosciences) connected to the AKTA purification system. The column was loaded according to the manufacturer's instructions with nickel sulphate and equilibrated in binding buffer (10mM Tris-HCl pH8, 300mM NaCl). The column was loaded with the soluble protein fraction and washed with binding buffer. The protein was eluted with elution buffer (10mM Tris-HCl pH8, 300mM NaCl, 1M Imidazole) on a 0% to 25% gradient over 60min at a flow rate of 0.5ml/min. This yielded pure protein for both constructs (>90%), both eluted at about 15% imidazole. The imidazole was removed with a PD-10 column (Amersham Biosciences) to leave the proteins in binding buffer.

Both Sbi Ig-binding domains were used for SPR; Sbi-1 at a concentration of $4.45e^{-5}$ M (0.43mg/ml); Sbi-2 = $3.50e^{-5}$ M (0.35mg/ml). The concentrations of the Sbi domains were determined by BIORAD assay, for which a calibration curve is shown in Figure 20.

See Chapter 6 for purification of Sbi-S.

7.1.2.4 Binding studies

7.1.2.4.1 Double diffusion assay

Double immunodiffusion experiments were performed on Petri dishes containing a 1% agarose gel. Wells were punched in the agar and individual wells filled with 100 µl of Sbi-E 1mg/ml in PBS; Human serum IgG (Sigma) or IgG Fc (Bethyl Laboratories) 1mg/ml in PBS; or recombinant SpA (Pierce) 1mg/ml in PBS and left to incubate for 48 hours at room temperature. Insoluble 'immune complexes' formed at the zone of equivalence were visualised by Coomassie staining.

7.1.2.4.2 Size exclusion chromatography

IgG and HSA samples were prepared as described before for obtaining an elution profile with either 1µl of the Sbi-e stock ($7.41e^{-4}$ M) or 2µl of the SpA stock solutions ($2.38e^{-4}$ M) added. This was incubated on ice for 10 minutes and then spun at maximum speed for 10 minutes in the microcentrifuge. Samples were injected into the

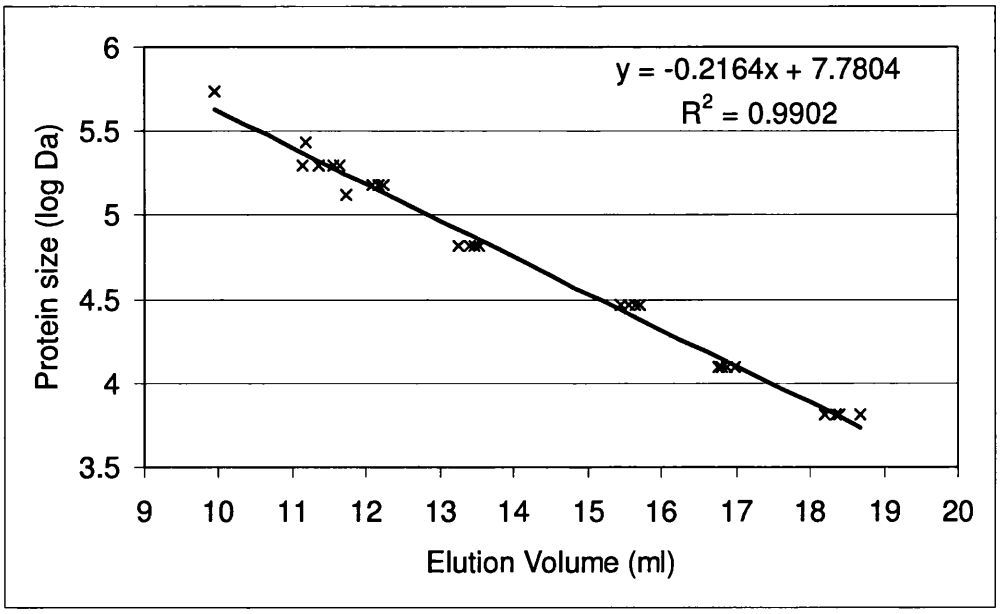


Figure 21 Calibration curve for the Superdex HR 200 10/30 column, using the protein standards aprotinin, cytochrome C, carbonic anhydrase, albumin, alcohol dehydragenase, β-amylase, and urease proteins (all from Sigma).

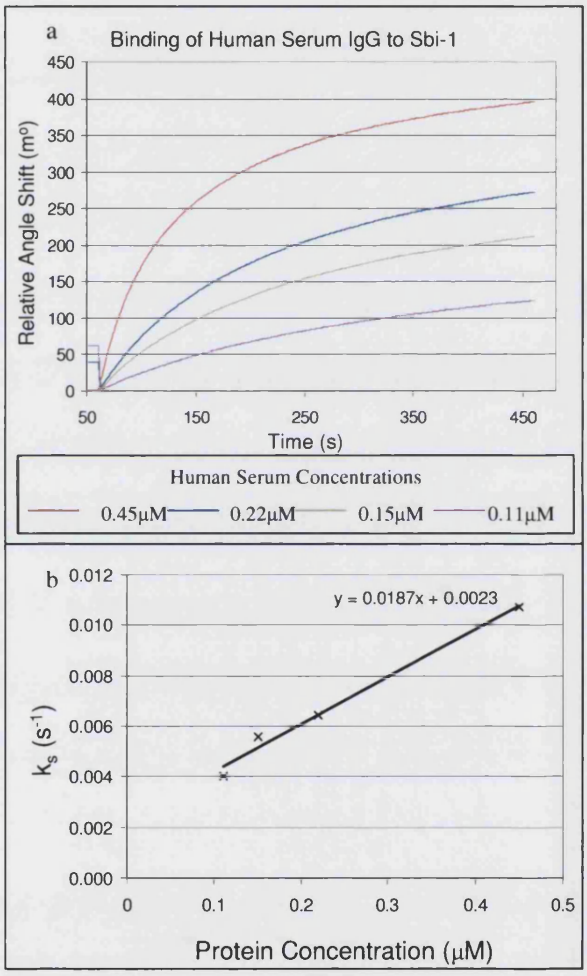
Superdex HR 200 10/30 column (Amersham Biosciences) connected to the AKTA purifier, equilibrated in 20mM Tris-HCl pH7, 200mM NaCl. The column was washed with 1.5 column volumes (approximately 37ml) of equilibration buffer, and the peak size(s) measured from the standard curve calculated using the standard proteins: aprotinin, cytochrome C, carbonic anhydrase, albumin, alcohol dehydragenase, β-amylase, and urease proteins (all from Sigma). The void volume of the column was determined with blue dextran (Sigma). The calibration curve is shown in Figure 21.

A sample of Sbi-1 was collected from the protein purification. To test this product for binding ability, SE experiments were carried out. The concentration of the eluted protein was not determined. It was added to human or goat IgG to determine the effect on the IgG elution profile. A set amount of eluate was added to the serum IgG prior to loading the sample on the SE column. A fixed volume Sbi-1 was used in the experiments, or a multiple thereof.

7.1.2.4.3 Surface plasmon resonance (SPR)

All previously listed IgGs were used (Table 3). The IgG was dissolved in 1x PBS and further dilutions for experiments were in running buffer.

SPR experiments were carried out using a two-channel Autolab ESPRIT machine and SPR sensor discs coated with Ni^{2+} ions complexed on a 2D chelating surface (Winsdor Ltd/Xantec). Measurements were carried out with the Electrochemical



c

$$A + B \underset{k_{dis}}{\overset{k_{ass}}{\rightleftharpoons}} AB$$

$$K_A = \frac{[AB]}{[A][B]}$$

$$K_A = \frac{k_{ass}}{k_{dis}}$$

Figure 22. Soluble serum IgG is added to Sbi-1 bound to a gold sensor disc at different concentrations. The relative refractive angle of the laser beam was then measured. This gives a set of binding curves, one for each concentration of IgG used. **(a)** shows a typical set of SPR sensograms for the binding of human serum IgG to Sbi-1. **(b)** Typical example of a plot of k_s vs protein concentration, where k_{ass} is the gradient and k_{dis} is the intercept with the y-axis. In experimental data, the k_s valves are determined from the curves in (a) by the fitting the data, using the kinetic evaluation program ESPRIT. **(c)** Kinetic equations used in the determination of the kinetic constants K_A and K_D.

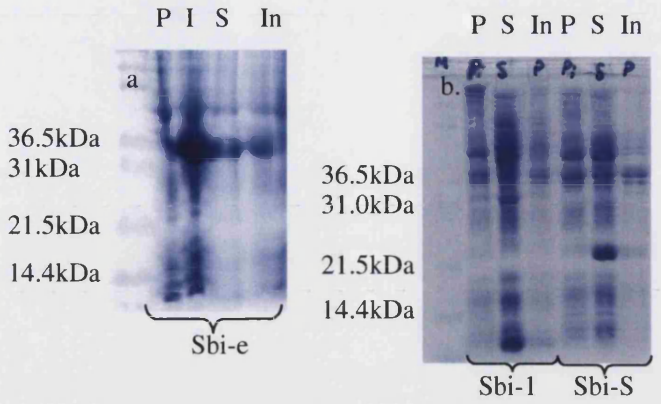


Figure 23 SDS-PAGE analysis of the expression of recombinant Sbi constructs, with the pre-induction sample (P), post-induction sample (I), soluble fraction (S) and insoluble fraction (In) for **(a)** Sbi-S and **(b)** Sbi-e and **(c)** Sbi-1.

Surface Plasmon Resonance Interaction Technology Data Acquisition program (Eco Chemie B.V. The Netherlands, version 4.1.0 (2004)). Experimental design was based on the work by Nieba *et al* (1997). Running buffer was 10mM Tris-HCl pH8.0, 300mM NaCl, 0.005% v/v Tween 20. The surface was activated with 300μM $NiSO_4$, 10mM Tris-HCl pH8, 300mM NaCl, 0.005% v/v Tween 20. Regeneration buffer was 0.35M EDTA pH8.5.

A baseline was run for 60s, the Sbi construct was then added, and association monitored for 400s, and dissociation measured for 20s. A further baseline was run for another 60s, and than the IgG sample was added in one channel, the control BSA was added to the other channel. Association was measured for 400s and dissociation for 20s. EDTA was than added to regenerate the surface.

Kinetics were determined with the ESPRIT Kinetic Evaluation program (Eco Chemie B.V. The Netherlands), version 4.1.0 2004). The differential data (the IgG channel minus the BSA channel) was analysed using a monophasic-fitting model. The fitting program was used to determine a value of k_s for all concentrations. The k_s value was then plotted against IgG concentrations in Microsoft Excel, and a trendline calculated. The k_{ass} value was taken from the gradient of this line, and the k_{dis} value from the intercept with the y-axis. K_A is k_{ass} divided by k_{dis} (Figure 22). χ^2 values give an

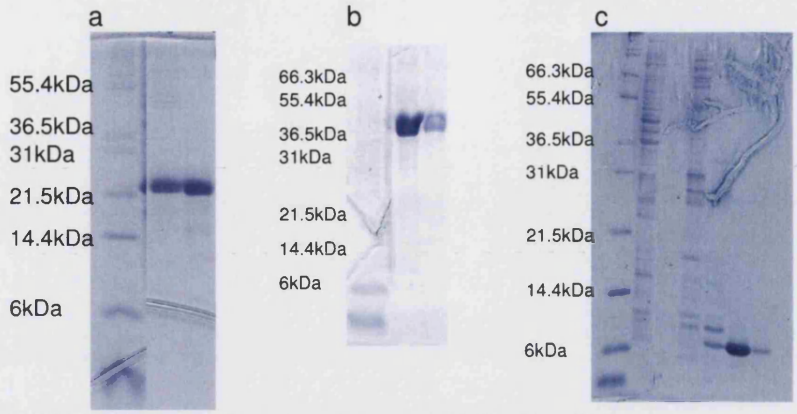


Figure 24 SDS-PAGE analysis of protein purification of **(a)** Sbi-S by affinity chromatography, **(b)** Sbi-e by ion exchange chromatography and **(c)** Sbi-1 by affinity chromatography.

indication of the quality of the fitting of the response data to the the real time response equation, calculated by the ESPRIT Kinetic Evaluation program, while R^2 is an indication of the fitting of the trendline to the data by Microsoft Excel, and of the reliability of k_{ass} and k_{dis}.

Sbi-1 and Sbi-2 under went some dissociation from the nickel surface in the absence of EDTA after long periods of time on the surface. Therefore the association phase measurements were limited to 400s to minimise the loss of Sbi-1 and Sbi-2. The dissociation phase was not used to determine a value of kdiss as after 400s Sbi-1 and Sbi-2 appeared dissociate from the nickel disc as evidence in association phase readings were the amount of protein on the surface started to decrease from its peak level, suggesting the dissociation of Sbi in addition to its IgG ligand.

7.1.3 Results

7.1.3.1 Cloning and expression of Sbi constructs

PCR amplification of the fragments of *sbi* gave single bands for the constructs Sbi-e, Sbi-S, Sbi-1 and Sbi-2 (data not shown). These fragments were inserted into the pHIS parallel vector (Sheffield *et al* 1999), and transformed into XL1-Blue *E. coli* cells. The presence of the correct insert was confirmed by sequencing of the plasmid. The plasmid containing the Sbi-e and Sbi-S inserts were then transformed into BL21 star™ *E. coli* cells, while the ones containing the Sbi-1 and Sbi-2 inserts were

transformed into Rossetta™ *E. coli* cells. All clones showed good expression at optimum temperatures of 28°C (Sbi-e) and 25°C (other clones). Reducing the induction temperature of Sbi-1 and Sbi-2 to 20°C further improved the amount and solubility of the purified protein. There was no evidence of leaky expression in pre-induction samples; induction with IPTG gave good levels of expression (Figure 23) at the expected protein sizes. Affinity (for the Sbi-1 and Sbi-2 constructs) or ion exchange chromatography (for sbi-e only) with all proteins gave pure protein (Figure 24).

The purified Sbi-e fragment could be concentrated up to 30mg/ml. Initial binding studies were carried out to both check functionality and to optimise the method for size exclusion chromatography with a mouse monoclonal antibody, and a sample of β_2GP1 (a kind gift from Ph. de Groot). This showed Sbi-e binds both proteins (Figure 25a&b, the calibration curve is shown in Figure 21) with a shift of 30kDa when Sbi-e is in the presence of the mouse IgG, and a shift of 60kDa when Sbi-e is added to β_2GP1. The profile of Sbi-e in the absence of either ligand is shown in Figure 26. ELISA-based binding studies with an Sbi construct and truncated β2GP1 constructs carried out in the de Groot lab, showed that Sbi only binds β2GP1 constructs composed of domains I-II-III and domains I-II-III-IV as well as the full-length β2GP1 (Figure 25c). Size exclusion chromatography with IgG in the presence of Sbi-e showed a shift in the peak size of IgG corresponding to 30kDa in the presence of Sbi-e, while the β2GP1 peak shifted by 60kDa. These results indicate that the recombinant Sbi-e construct produces a functional protein product with both IgG-binding and β2GP1-binding characteristics.

7.1.3.2 Binding studies

7.1.3.2.1 Analysis of insoluble SpA-IgG and Sbi-IgG complexes

It is know that in the presence of certain species of IgGs, SpA precipitates the IgG. This can be tested by placing a sample of SpA in one well of an agar plate, IgG in another, and allowing diffusion to occur. Precipitation can then occurs between the two wells, as shown in Figure 27. To determine whether Sbi shows the same patterns of precipitation as SpA, both Sbi and SpA where placed in wells in an agar plate along with samples of human serum IgG and the human IgG Fc fragment in a double

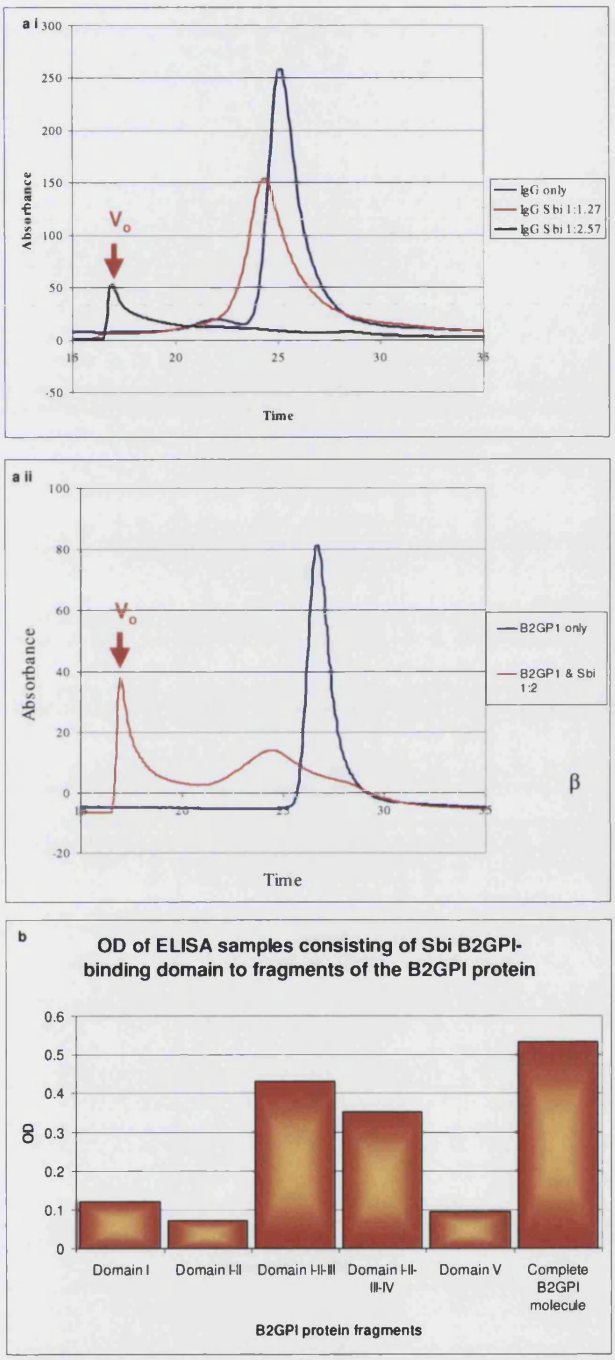


Figure 25 Size exclusion chromatography elution profiles of **(ai)** mouse monoclonal IgG and **(aii)** β2GP1 in the presence and absence of sbi-e. **(b)** ELISA results (carried out by de Groot; Utrecht) with β2GP1 and the β2GP1-binding domain of Sbi.

diffusion assay, shown in Figure 27. The arch-shaped pattern of precipitation in the diffusion equivalence zone between SpA, Sbi, and human serum IgG indicates that the two Ig-binding proteins share the same epitope on human IgG presumably the Fc fraction. The line of precipitation produced by the SpA-IgG complex is broader compared to the line of precipitation of the Sbi-IgG reaction and a spur formation of precipitate towards the Sbi well (highlighted by an arrow in Figure 27) indicates that SpA possesses a binding specificity for a human IgG epitope that is not shared by Sbi (presumably the Fab binding ability of SpA). In the absence of the Fab fragment SpA's capacity to precipitate human IgG Fc was completely eliminated, while precipitation with Sbi remains unchanged (see Figure 27).

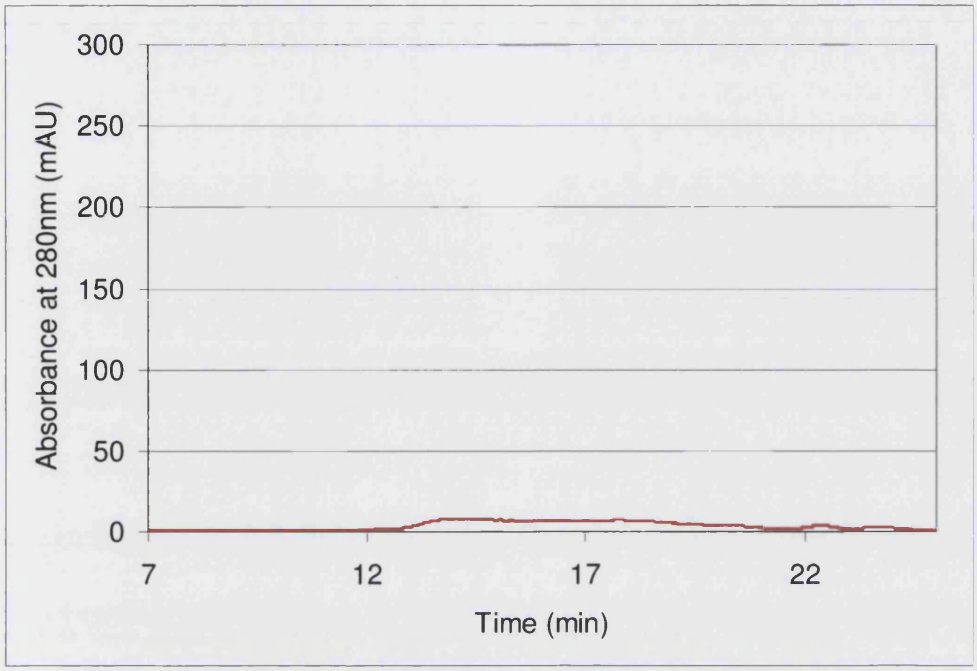


Figure 26 Size exclusion elution profile of Sbi-e in the absence of IgG or β_2GP1

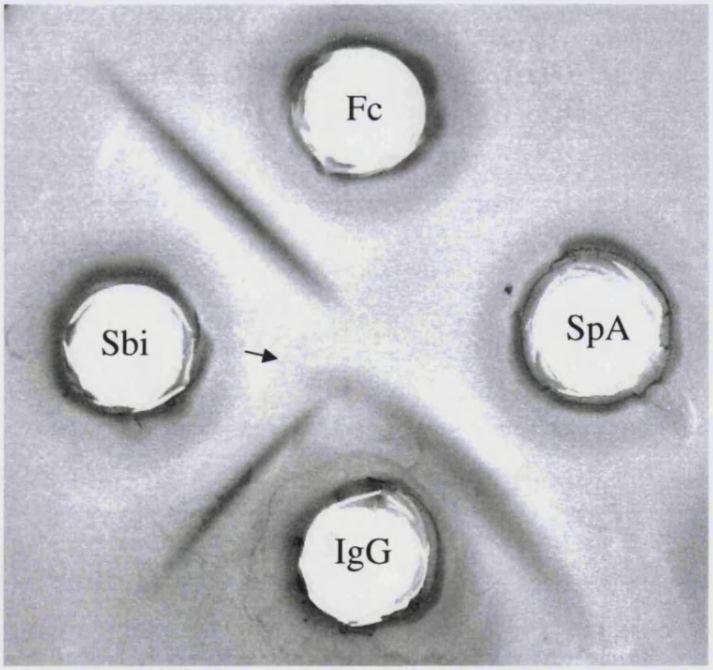


Figure 27 Double immunodiffusion assay with recombinant SpA and Sbi against human serum IgG and IgG Fc fragment. The spur of precipitation, indicating partial identity between the IgG binding epitopes of SpA and Sbi, is indicated by an arrow.

When tested against both the animal serum IgGs and the human subclasses, Sbi-e formed insoluble complexes with mouse, rabbit, Guinea pig, and human serum IgGs in addition to the human subclasses IgG1 and IgG2. SpA differs from Sbi in that, in addition to human and Guinea pig serum IgG, it also precipitates very strongly with pig serum. The results with Sbi-S show the same precipitation patterns as found with Sbi-e, indicating that the insoluble complexes formed by Sbi are caused solely by presence of Sbi's two immunoglobulin binding domains. Although the formation of insoluble complexes was expected to require the presence of multiple binding domains, the single Ig-binding constructs Sbi-1 and Sbi-2, also produce insoluble complexes but only with Guinea pig serum IgG (see Table 4 for precipitation results).

7.1.3.2.1.1 Analysis of soluble SpA-IgG and Sbi-IgG complexes

Before analysing the binding characteristics of Sbi and SpA with size exclusion chromatography, elution profiles for each of the IgG samples were obtained first. Subsequently a fixed concentration of either Sbi or SpA was added to the same amount of IgG used to generate the IgG profile. From the difference between the two elution profiles we analysed the size (molecular weight) of the complexes formed. The elution profile of the purified Sbi-e at the concentration used in the size exclusion experiments is shown in Figure 25a. Precipitation of the serum IgGs in the size exclusion chromatography experiments was deteated by the addition of SDS buffer to the microcentrifuge tube and heated to dissolve any precipitated or pelleted protein. This mixture was than loaded an SDS-PAGE gel for separation and detection of protein bands.

We found that both Sbi and SpA bind human serum IgG, with most protein precipitating before injection (Figure 28a). In addition, a small amount of residual soluble protein was observed, as both unbound IgG and large complexes eluted at the void volume. Sbi and SpA also bound most of the tested mammalian serum IgGs, with different binding characteristics (Figure 28b to Figure 28g). With rabbit IgG, most of the IgG in complex with SpA was eluted as a large soluble complex at the void volume; no insoluble complexes were detected by SDS-PAGE analysis on the residual protein complexes that precipitated in the centrifugation step preceding the protein injection. In contrast, all rabbit IgG in complex with Sbi-e was lost as

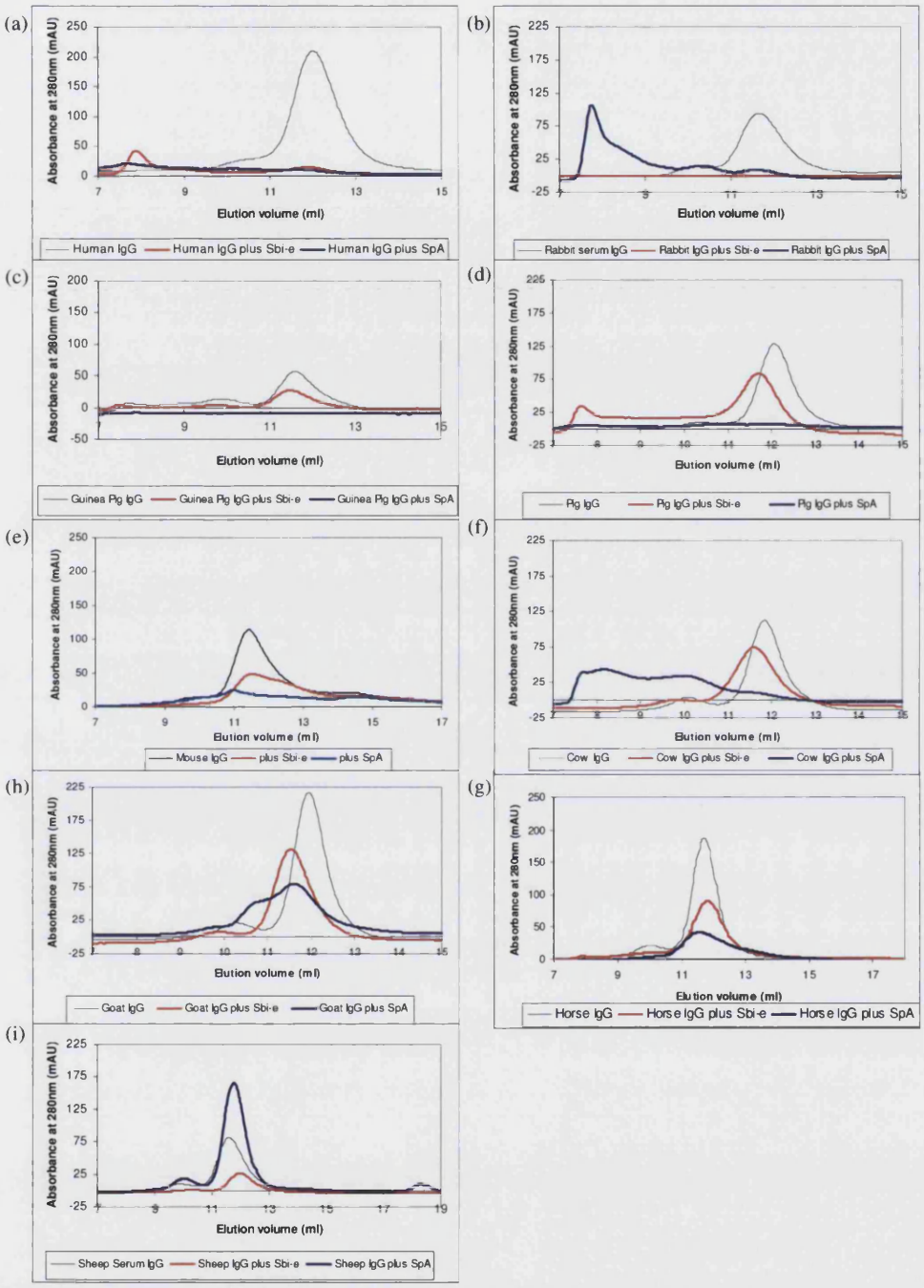


Figure 28 Size exclusion chromatography elution profiles for the serum IgGs from **(a)** human, **(b)** rabbit, **(c)** Guinea pig, **(d)** pig, **(e)** mouse, **(f)** cow, **(g)** goat, **(h)** horse and **(i)** sheep in the presence and absence of Sbi-e and SpA.

Table 4 Insoluble complexes form by serum IgGs in the presence of the Ig-binding proteins SpA and Sbi. + indicates the formation of white precipitation on the agar gel, while ++ and +++ indicate increasing amounts of precipitation.

Serum IgG	Ig-binding protein or domain				
	SpA	Sbi-e	Sbi-S	Sbi-1	Sbi-1
Mouse		++	++		
Rabbit		++	++		
Goat					
Chicken					
Cow					
Guinea Pig	+	++	++	+	+
Sheep					
Horse					
Human	++	++	++		
Pig	+++				

precipitate. However when with SpA in complex with Guinea pig serum IgGs, all protein is lost as precipitate, while in complex with Sbi-e, only about half the IgG is lost as insoluble precipitate, the remainder was present as a soluble complex, with anapproximate size of 201kDa (see Table 5), a size very close to the size of Guinea pig IgG (192kDa). While complexes of pig IgG and Sbi show no formation of insoluble precipitate, it does form a number of different soluble complexes (with an approximate size of 181kDa for the major complex peak), which are eluted between unbound pig IgG and the void volume. This peak indicates that in the presence of Sbi-e, most of the pig serum IgG is forming a 1:1 complex with Sbi-e. In the presence of SpA, virtually all the pig serum IgG is lost as insoluble complexes. In the presence of mouse serum IgG, Sbi-e causes precipitation. A soluble peak is also detected, however, of a similar size as the mouse serum IgG peak detected in the absence of an Ig-binding protein, suggesting some components in the mouse serum IgG fraction does not bind to Sbi-e. Mouse serum IgG, in the presence of SpA, is eluted in a complex of about 400kDa, approximately twice the size of the mouse serum IgG, suggesting that SpA causes mouse IgG dimerisation. In the case of both cow and goat serum IgGs, the complex formed with Sbi-e resulted in a shift of the IgG peak by

about 30kDa, suggesting a 1:1 complex formation. However, in the presence of SpA, larger complexes were also observed (Table 5). Complexes of SpA with cow IgG were much larger than the complexes with goat IgG, with the largest eluted at the void volume. The size exclusion chromatography profile of Sbi-e-goat IgG complexes contained one major peak, with an approximate size of 197kDa. Only for the complexes with the serum IgGs from pig, cow and goat it is possible to estimate the complex composition; for these species in the presence of Sbi a 1:1 complex is formed.

Two other mammalian IgGs were also examined with different results. Horse IgG in complex with both Sbi-e and SpA initially did not appear to cause the elution peaks to shift, but resulted in the broadening of the peaks (Figure 28h). Addition of ten times the original amount of Ig-binding protein resulted in a shift of the peaks eluted, with an SpA-IgG complex of about 320kDa and a larger Sbi-e peak complex of about 335kDa. Sbi-e in complex with horse IgG produces a broad peak suggesting the presence of more than one type of complex (Table 5).

Sheep IgG produced peaks of identical size both in the presence and absence of the Ig-binding proteins, and thereby suggesting that there is no binding of sheep protein by Sbi-e and SpA (Figure 28i).

The only non-mammalian serum IgG tested, that of chicken, differences shown in the elution profiles when the Ig-binding proteins were added were not larger enough to correspond to Sbi or SpA binding (Figure 29). Human serum albumin (HSA) was used as a control. This again showed no shift in the elution profile in the presence of either Ig-binding protein (data not shown).

Elution profiles of human and goat serum IgGs were also determined in the presence of Sbi-1. Human and goat serum IgGs were use to compare the complex formation of Sbi-1 with serum IgGs predicted to form complexes of high and low molecular weight complexes in its presence, respectively. In the case of goat serum IgG, the major peak shifted about 18kDa. This suggests that there are two molecules of Sbi-1 bound to a single IgG molecule. In the elution profile of human serum IgG with the same amount

Table 5 The peak sizes determined by size exclusion chromatography for IgG molecules and the Sbi and SpA IgG complexes, calculated from elution volumes **(a)** serum IgGs and **(b)** human IgG subclasses. The presence of precipitated protein was confirmed through the addition of SDS-PAGE buffer to the microcentrifuge tube, and heating it to dissolve any precipitated protein present. This mixture was then loaded onto an SDS-PAGE gel for separation and straining of protein bands. Refer to Figure 21 for the size exclusion chromatography calibration curve

(a) IgG species	IgG peak sizes as determined by SE (kDa)	Sbi-e results			SpA results		
		Major Peak Size (kDa)	Unbound IgG present?	Precipitation present?	Major Peak Size (kDa)	Unbound IgG present?	Precipitation present?
Human	285 155	No major peak	Yes	Yes	No major peak	Yes	Yes
Pig	358 151	181	No	No	314	Yes	Yes
Rabbit	407 186	No major peak	No	Yes	Void volume	Yes	No
Guinea pig	462 192	201	Yes	Yes	No major peak	No	Yes
Mouse	212	206	Yes	Yes	413	Yes	Unclear
Sheep	474 193	176	Yes	No	159	Yes	No
Horse	430 178	171 335	Yes	No	198 320	Yes	No
Cow	417 169	188	No	No	Void volume	Yes	No
Goat	379 169	197	No	No	190	No	No
Chicken	590 227	239	Yes	No	169	Yes	No

(b) IgG species	IgG subclasses peak sizes as determined by SE (kDa)	Sbi-e results			SpA results		
		Major Peak Size (kDa)	Unbound IgG present?	Precipitation present?	Major Peak Size (kDa)	Unbound IgG present?	Precipitation present?
IgG1	235	None	No	Yes	832 235	No	No
IgG2	187	1335	No	Yes	381	No	No
IgG3	268	298 270	Yes	No	271 311	Yes	No
IgG4	479 214	1652	No	No	1448 417	No	No

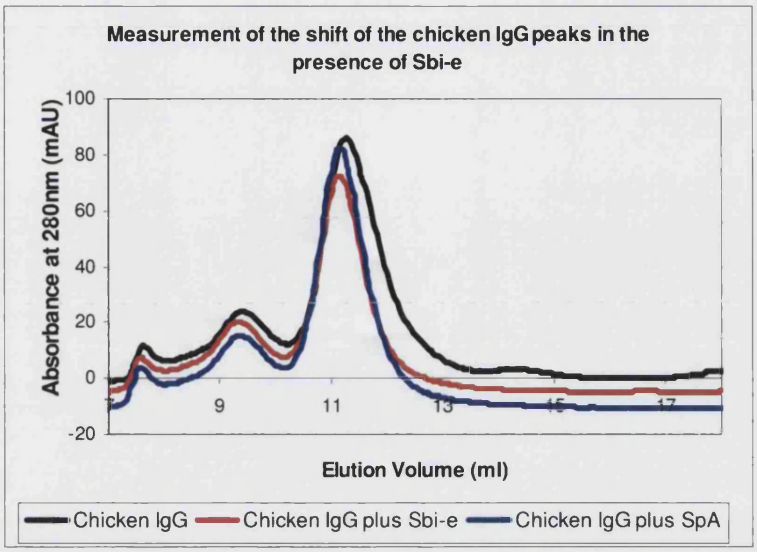


Figure 29 Size exclusion chromatography elution profiles for chicken serum IgG in the presence and absence of Sbi-e and SpA.

Table 6 Peak sizes for the complexes formed by Human and Goat serum IgGs in the presence of Sbi-1

IgG species	IgG peak sizes as determined by SE (kDa)	Major Peak Size (kDa)	Unbound IgG present?	Precipitation present?
Goat	373 160	434 179	No	No
Human	155	725 178	No	No
Human	155	1,255 501 213	No	No

of Sbi-1, the major peak showed a similar shift in size (23kDa), again suggesting the possibility of two Sbi-1 molecules being bound to the IgG molecule. However some protein was eluted at an earlier time, with a calculated size of 726kDa, suggesting the formation of a large soluble complex. An additional elution profile of human IgG with five times the amount of Sbi-1 showed an increase in the amount of protein in the larger complex, with an additional increase in the size of this complex. (see Table 6 for peak sizes, and Figure 30 for elution profiles).

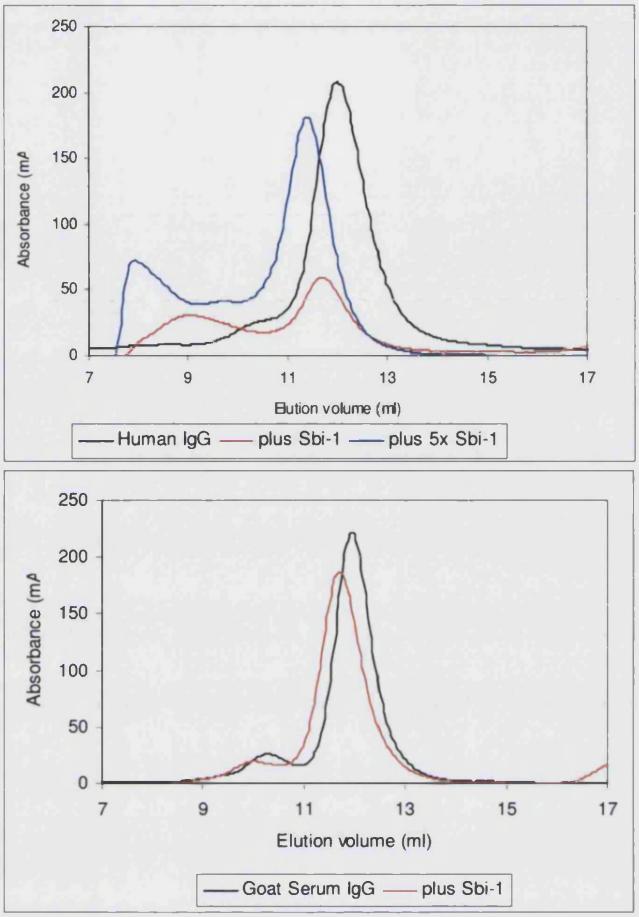


Figure 30 Elution profiles for human serum IgG and goat serum IgG in the presence and absence of Sbi-1

7.1.3.2.1.2 Analysis of soluble complexes of SpA and Sbi with Human IgG subclasses

Sbi-e interactions with the four human IgG subclasses were compared to those of SpA. With IgG1κ, both Ig-binding proteins formed complexes that precipitated; SpA also formed some soluble complexes (Figure 31a). Sbi-e in complex with IgG2κ and IgG4κ formed large soluble complexes, eluted at the void volume (Figure 31b and Figure 31d). SpA also formed soluble complexes in the presence of IgG2κ and IgG4κ, but after the void volume (see Table 5 for peak sizes and Figure 31b & d for elution profiles). There was no precipitate with SpA in complex with IgG4κ, but a

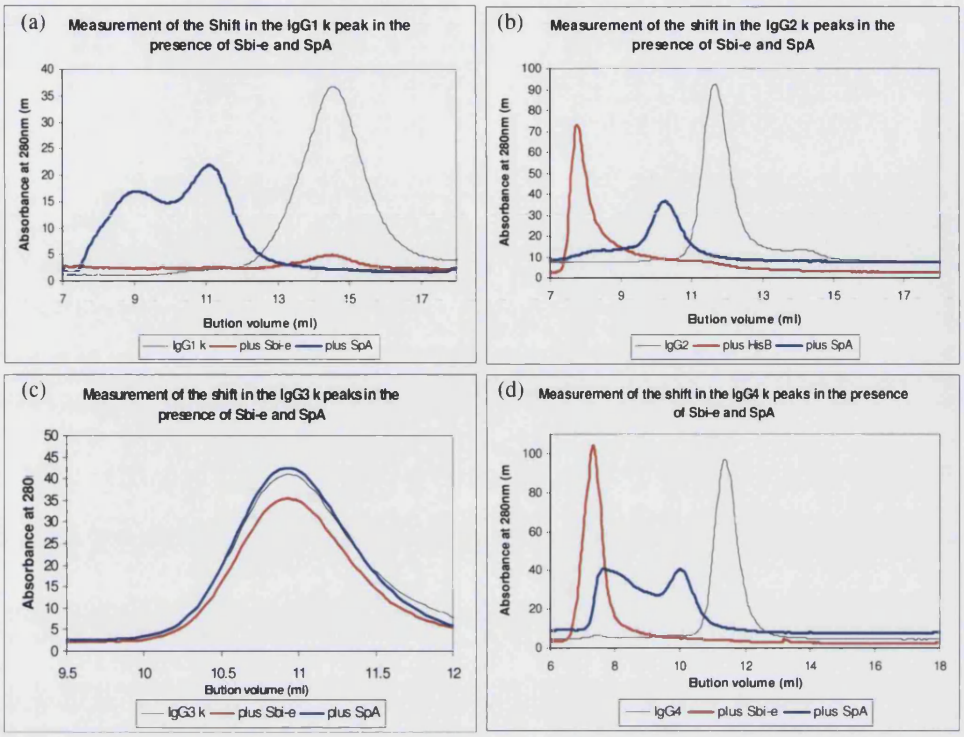


Figure 31 Size exclusion chromatography elution profiles for the human IgG subclasses **(a)** IgG1, **(b)** IgG2, **(c)** IgG3, and **(d)** IgG4 in the presence and absence of Sbi-e and SpA.

large amount with IgG2κ (data not shown). Both Sbi-e and SpA failed to show any binding for IgG3κ, indicated by the absence of a shift in the IgG peak, (Figure 31c).

7.1.3.2.2 Surface plasmon resonance studies

7.1.3.2.2.1 Immunoglobulin-binding characteristics of Sbi's individual Ig-binding domains

Rate constants (k_{ass} and k_{dis}) and equilibration constants (K_A and K_D) were determined for all antibodies, and compared. These binding results are shown in Figure 32 and Table 7a.

The quality of data fitting varies between species, as is indicated by the chi square (χ^2) values. We believe this discrepancy is due to using serum IgGs in our experiments, rather than monoclonal antibodies, as the different IgG subclasses present in the serum are expected to give different k_{ass} and k_{dis} rate constants, and therefore affect the determination of equilibrium constants. χ^2 are given in Appendix 2. The

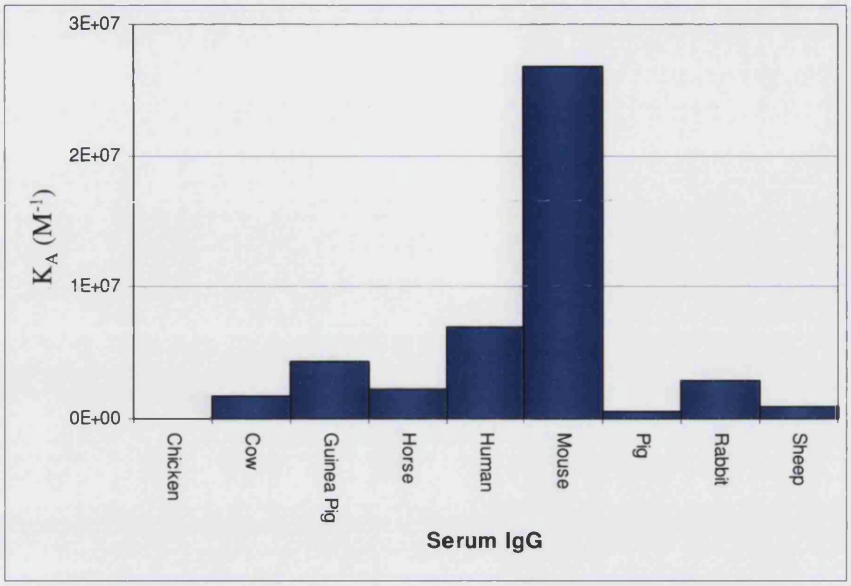


Figure 32 Association constants for the Sbi Ig-binding domain Sbi-1 in the presence of the serum IgGs.

heterogeneity of the IgG means that the binding curve produced and the resulting value of k_s (or k_{on}) is an average for the binding and dissociation of the different IgG subclasses on the surface. Therefore the values produces for k_{ass}, k_{dis}, K_A, and K_D give the affinity for a defined collection of IgG subclasses (i.e. the serum IgG) for the Sbi-1 ligand.

The values for the association rate constant (k_{ass}: the rate at which the components in the reaction combine to form the product) indicate binding occurs with all antibodies. (see Figure 22 for the theoretical reaction equation and the calculation of rate and equilibration constants). Chicken IgG has the lowest association rate ($k_{ass} = 3.54e^{1}$ s^{-1} M^{-1}) and mouse serum IgG has the highest ($k_{ass} = 7.81e^{4}$ s^{-1} M^{-1}). The IgGs of species that show poor binding with Sbi-e (goat, horse, sheep, and chicken) all have a k_{dis} value of less than $1.0e^{-3}$ s^{-1}. Guinea pig IgG has the highest dissociation rate of $7.3e^{-3}$ s^{-1}. The values of equilibrium constant K_A are the lowest for chicken, cow, horse, pig, and sheep serum antibodies (K_A less than $2.16e^{6}$ M^{-1}; K_D greater than $4.64e^{-7}$ M). Goat serum IgG occurs at the other end of the spectrum, with very high affinity ($K_A = 4.07e^{7}$), but it shows virtually no dissociation ($k_{dis} = 1.00e^{-05}$).

Table 7 Affinity values for the complexes of **(a)** Sbi-1 with the serum IgGs (all R^2 values were calculated from three data points on a $k_s=k_{ass}[P]+k_{dis}$ line, except for human serum IgG which used five) and **(b)** Sbi-1 and Sbi-2 with the Human IgG subclasses, calculated from SPR measurements (the number of data points for R^2 values are in brackets). R^2 is the value produced by Microsoft Excel© to show the quality of fitting a set of data points to a line.

(a)	k_{ass} ($M^{-1}s^{-1}$)	k_{dis} (s^{-1})	K_A (M^{-1})	K_D (M)	R^2
Chicken serum IgG	3.53E+01	9.00E-04	3.93E+04	2.55E-05	1.0000
Cow serum IgG	3.98E+03	2.30E-03	1.73E+06	5.78E-07	0.9561
Goat serum IgG	4.07E+02	1.00E-05	4.07E+07	2.46E-08	0.8868
Guinea Pig serum IgG	3.13E+04	7.30E-03	4.28E+06	2.34E-07	0.9601
Horse serum IgG	2.16E+02	1.00E-04	2.16E+06	4.64E-07	0.9371
Human serum IgG	1.88E+04	2.70E-03	6.95E+06	1.44E-07	0.8717
Mouse serum IgG	9.88E+04	3.70E-03	2.67E+07	3.74E-08	0.9549
Pig serum IgG	3.11E+03	5.80E-03	5.36E+05	1.87E-06	1.0000
Rabbit serum IgG	1.09E+04	3.80E-03	2.86E+06	3.49E-07	0.9854
Sheep serum IgG	8.39E+01	9.00E-05	9.32E+05	1.07E-06	0.9966

(b)	k_{ass}		k_{diss}		K_A	
	Sbi1	Sbi2	Sbi1	Sbi2	Sbi1	Sbi2
IgG1	2.03E+05	2.07E+05	5.80E-03	4.80E-03	3.51E+07	4.30E+07
IgG2	7.18E+04	1.42E+04	1.80E-03	4.50E-03	3.99E+07	3.15E+06
IgG4	3.94E+04	2.51E+04	5.20E-03	4.00E-03	7.59E+06	6.27E+06

(b) cont.	K_D		R^2	
	Sbi1	Sbi2	Sbi1	Sbi2
IgG1	2.85E-08	2.32E-08	0.8853 (5)	0.9422 (3)
IgG2	2.51E-08	3.17E-07	0.8825 (6)	0.9351 (6)
IgG4	1.32E-07	1.59E-07	0.9504 (7)	0.999 (4)

7.1.3.2.2.2 Affinity of Sbi-1 and Sbi-2 for human IgG subclasses

Rate and equilibration constants were also determined for the human IgG subclasses against both Sbi Ig-binding domains. As expected, the fitting of the data for the human IgG subclasses is generally better than for the serum IgGs, with χ^2 values for all three measured IgG subclasses and two Sbi Ig-binding proteins of less than 12000 (A limit of 12000 for χ^2 was used as all values for k_s with a χ^2 below this could be fitted onto a straight line. k_s values with a χ^2 above this tend to be outliers). Results for the human subclasses are shown in Figure 33 and Table 7b. The k_{ass} values for IgG1 are similar for both Sbi-1 and Sbi-2, while the k_{ass} values for IgG2 and IgG4 are similar to each other in complex with both Sbi-1 and Sbi-2. All three IgG subclasses have similar k_{dis} values with either Sbi domain (except the combination of Sbi-1 and

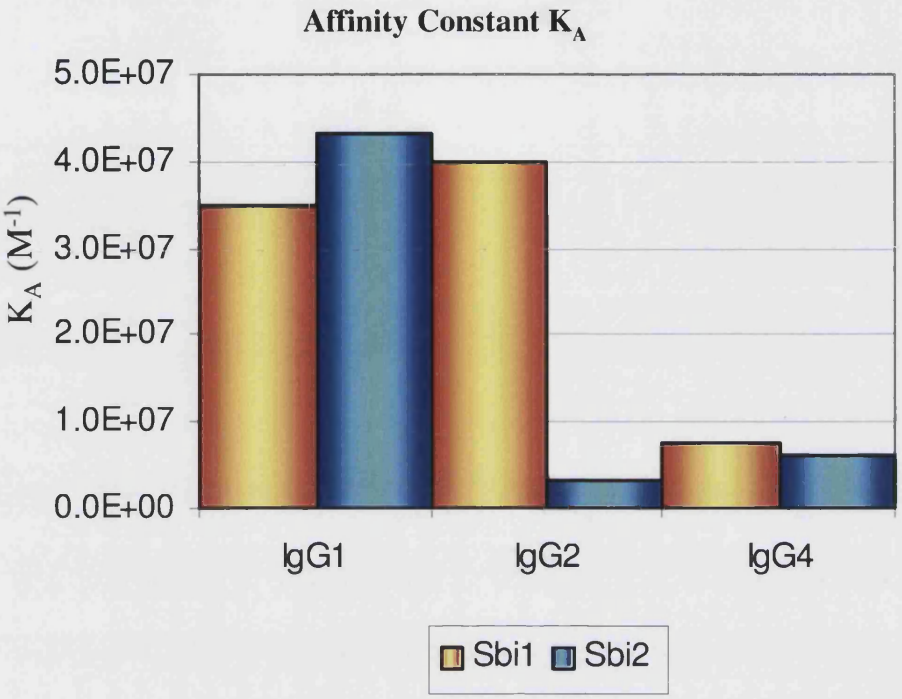


Figure 33 Affinity constant for the Sbi Ig-binding domains Sbi-1 and Sbi-2 with the human IgG subclasses.

IgG2 that have a lower k_{dis}). When the affinity of Sbi-1 and Sbi-2 is compared for each of the IgGs, it was found that both Sbi domains bind IgG1 with similar affinity, as do they with IgG4. Sbi-1 binds both IgG1 and IgG2 with similar affinity, however the Sbi-2 plus IgG2 has affinity, which is significantly lower.

7.1.4 Discussion

SpA was the first of the *S. aureus* Ig-binding proteins to be discovered. The early binding studies with SpA and IgGs from different species were performed with diffusion assays. Kronvall *et al* (1970) tested a wide range of mammalian sera and we have obtained serum IgG samples for a number of these. Kronvall *et al* (1970) showed that Guinea pig, rabbit and pig serum IgGs precipitated in the presence of SpA (human myeloma precipitated as the control in all reactions), while mouse, horse, goat and sheep serum IgGs all showed inhibition of the myeloma protein reaction.

The first work on the characterisation of Sbi binding to Ig was carried out by Zhang *et al* (1998). They used ELISA to examine the binding between Sbi and mammalian serum IgGs and also the different classes of human IgGs. The highest binding was observed with human, Guinea pig, rabbit and pig IgGs. Goat, sheep and chicken did not appear to bind Sbi, while cow and horse IgGs gave an intermediate reaction.

How though can the results of SpA obtained via diffusion assay and Sbi obtained by ELISA be compared? It would appear to be necessary to compare the reactions of IgGs with either SpA or Sbi under the same conditions. To understand this we analysed the binding patterns of both SpA and Sbi under the same conditions, both proteins were subjected to diffusion assays and size exclusion (SE) chromatography to determine the size and stochiometry of the complexes formed, and SPR to study the binding kinetics.

The Sbi protein was cloned into 4 different constructs, one containing the N-terminal region domains (Sbi-e), one containing both the proposed Ig-binding domains (Sbi-S), and two containing the individual Ig-binding domains (Sbi-1 and Sbi-2). All four constructs showed binding to IgG. In the agar diffusion assays, Sbi-e and Sbi-S showed the same pattern of precipitation, while in the SPR experiments Sbi-1 showed the highest affinity for those IgGs that were precipitated in the agar diffusion assays. This suggests that, although we did not confirm the Sbi domains in the different constructs maintained the same structure through structural methods, as the binding ability is conserved, either the correct structure is maintained in the Ig-binding domains in the different constructs, or the Ig-binding domains fold correctly on binding to the IgG molecules. Evidence to support the hypothesis of the Ig-binding domains of Sbi binding corrently folded when in a construct as a single domain is available in the structure of the SpA Ig-binding domain structure solved by Gouda *et al* (1992). In this structure, solved by NMR, a single SpA Ig-binding domain is folded into the typical three-helix bundle expected for Ig-binding domain in the absence of either its ligand or the other SpA Ig-binding domains in the protein.

The sizes of the IgG molecules used in the binding studies were measured by size exclusion chromatography. This revealed variation from the expected size of an IgG molecule of 150kDa. The size of different IgG molecules are difficult to calculate as

the complete sequences of all the polypeptide chains are required, and many of the IgG sequences available are not complete. It was possible to use the sequences for the human IgG structure solved by Saphire et al (2001) to calculate a size for the protein of 148kDa. However this did not include the size of the carbohydrate attached to the IgG molecule. This is the only example where we were able to use sequences for a complete IgG molecule. It is reasonable to assume that IgGs from different species have different carbohydrates attached to them, and therefore affecting the IgG molecule size in addition to any variations in the lengths of the protein chain sequences themselves. In addition the differences in the protein chains between the species of IgGs would affect the extinction co-efficents and absorbance of the proteins at 280nm, meaning that this in addition to the varations in the concentrations of the protein stock solutions (discussed below) result in variation of the height of the IgG peaks formed in the elution profiles in Figure 28.

One of the factors influencing the height of the IgG peaks in Figure 28 is the protein concentration. The stock solutions of IgG were prepared in a fixed volume of buffer, however concentration measurements revealed variation in the concentration of the different stock solutions produced. This gives rise to the variation in the height of the IgG peaks in Figure 28. However this does not affect the parameter measured in the size exclusion results, where the change in the IgG profile after the addition of Sbi or SpA was determined. Molar concentrations of the IgG stock solutions were determined for use in the SPR experiments for determination of the affinity constants from protein concentrations determined by the BIORAD assay .

Different buffer conditions were used for the SPR experiments compared to the size exclusion experiments (a change of pH7 to pH8 in the 10mM Tris-HCl 300mM NaCl buffer plus the addition of 0.005% v/v Tween 20). This was due previously work had optimised the SPR conditions with this buffer (Nieba *et al* 1997) and the Sbi-1 and Sbi-2 constructs proved to be more stable under these conditions than the conditions used in the size exclusion experiments where Sbi-e was used.

Three different binding studies methods were used, the agar diffusion assay identified which combinations of IgG, Sbi and SpA formed insoluble complexes, size exclusion chromatography allowed measurement of the sizes of any soluble complexes present,

and SPR gave data on the affinity of Sbi for IgG. Insoluble complexes were also detected through size exclusion chromatography as the insoluble protein was collected in microcentrifuge tubes prior to the injection of the soluble complexes into the column. Such protein was run on SDS-PAGE gels, however the results (other than the presence or absence of precipitated protein) where unclear due to the large amounts of protein present. In addition analysis of soluble complexes eluted from the column showed on SDS-PAGE that not all the protein bands expected for IgG and Sbi or SpA could be detected due to low amounts of protein present and the dilute nature of the eluted fractions collected.

7.1.4.1 Implications of the differences observed in the binding of Sbi and SpA to IgG

The double diffusion assay with SpA and Sbi against human serum IgG showed the formation of an arch around the IgG well (Figure 27), indicating identity in the binding site on the IgG molecule for both proteins. This supports our previous prediction that both SpA and Sbi bind to the same site on the Fc fragment (see Chapter 2). The formation of a spur in the reaction in the direction of the Sbi well, however, indicates that there is an additional binding component in SpA. Our agarose diffusion experiments using human IgG Fc fragments indicate that Fab-binding by SpA plays a major role in the formation of insoluble complexes with human IgG. The absence of Fab, as predicted in Chapter 2, does not affect the formation of insoluble Sbi-IgG complexes. From these results we conclude that it is therefore unlikely that Sbi has a V_H family Fab specificity different from SpA, ruling out its possible role as a B cell super antigen. It may well be that the helix 2-helix 3 face of Sbi binds to other yet to be identified ligands.

As precipitation of IgG by SpA requires the intact IgG (Kronvall and Frommel 1970, Figure 27), it can aid in explaining the formation of the very large soluble complexes formed by some serum IgG such as rabbit serum IgG (Figure 28 and Table 5). The formation of these large soluble complexes suggests that, in the case of rabbit IgG, SpA cannot bind to the Fab regions of the IgG, thereby reducing its ability to crosslink between IgG molecules and subsequent precipitation. In the examples where SpA does form precipitation, as is the case with human serum IgG, SpA is known to bind certain subclasses of V_H chains (V_HIII, Roben, Salem, Silverman 1995). This

indicates that Fab-induced crosslinking may also play a role in other serum IgGs that also precipitate in the presence of SpA (pig and Guinea pig serum IgG).

Comparison of the SpA and Sbi binding reactions by both the diffusion assay and SE chromatography revealed significant differences between the two proteins. SpA appears to react most strongly with pig IgG while Sbi reacts strongly with the human and rodent IgGs. Both precipitate human IgG. Interestingly, while no precipitation is observed for the rabbit IgG in the presence of SpA, it is entirely eluted at the void volume as very large soluble complexes in the SE chromatography experiments. Also of interest is that of those serums IgGs tested, all those that have strong affinity for Sbi in the SPR experiments are precipitated by at least one of the Ig-binding proteins.

We also showed that the individual Sbi Ig-binding domains, Sbi-1 and Sbi-2, both bind to IgGs. Sbi-1 formed large soluble complexes in the presence of human IgG, and both domains are capable of precipitating Guinea pig IgG. As SpA required the presence of Fab binding for precipitation to occur, it is not clear how Sbi is triggering precipitation. This phenomenon is discussed further in Chapters 4 and 6.

7.1.4.2 Heterogeneity in immunoglobulin binding

The elution profiles of the tested IgGs indicate that many of the IgG complexes consist of components with a variety of molecular weights, including, in some cases insoluble complexes that accumulated in the microcentrifuge vial before injection. What is the origin of this variation in the SE results? One likely explanation is that the serum IgGs tested, and the different IgG subclasses therein, have different individual affinities for the two Ig-binding proteins. For example, for both Sbi and SpA in the presence of human IgG, most protein is precipitated, however a small amount remains unbound in solution. This is in agreement with the SPR results with the human IgG subclasses, where IgG 1, 2 and 4 all bind strongly (Figure 31), while no binding was detected for IgG3. Apart from the difference in valency between SpA (four Ig-binding domains in the construct used in the experiments described here) and Sbi (two Ig-binding domains), there are also differences between the Ig-binding domains themselves. In the case of SpA, where the domains are very highly conserved, this results in small differences in the binding of the Fc fragments (Jansson, Uhlen, Nygren 1998). Similarly, the individual Ig-binding domains of Sbi in our SPR

experiments showed that the two domains do not have equal affinity for all the human IgG subclasses (Figure 33).

The Guinea pig elution profile in the presence of Sbi-e shows that, in addition to a large amount of precipitation, there is unbound IgG present. Guinea pig serum IgGs are known to exist in two subclasses (1 and 2) and it is possible that only one of these can bind Sbi-e. In the case of SpA they both appear to be able to bind, as we see no unbound IgG in the presence of SpA. This variation in affinity between the different IgGs in the serum is supported to a degree by the difficulty to fit the SPR data acquired for some serum IgGs. However when the human IgG subclasses were used, a consistently better fit was obtained. It was noted by Friguet *et al* (1995) in ELISA experiments where serum IgGs were used opposed to monoclonal antibodies that measured binding is an average of the different IgGs. Of the serum IgGs with higher affinity, the results of the fitting data for rabbit serum IgG is better than most of the other serum results. We were only able to find references to a single rabbit IgG class in the databases, appearing to support the view that the more consistent results are obtained in the absence of multiple IgG subclasses.

7.1.4.3 Affinity versus avidity

SpA contains 4 Ig-binding domains while there are only two domains present in Sbi-e. Does SpA therefore form larger complexes than Sbi? With pig and cow serum IgGs, Sbi forms a 1:1 complex (discussed below), but with SpA, large insoluble (pig) or soluble (cow) complexes are formed. Although it was impossible to determine the complex size for either species (the size of the soluble complexes of cow IgG-SpA complex are outside the calibration range of the column), these results suggest that the complexes formed in the presence of SpA are larger than the complexes with Sbi. The size of such complexes could be determined by analytical ultracentifugation. In the case of human, rabbit, mouse and Guinea pig IgGs all complexes formed in the presence of Sbi or SpA are insoluble (the rabbit IgG-SpA complex is soluble, but its size could not be resolved by the SE column) and it is not possible to determine which Ig-binding protein forms the larger complex. The formation of these large complexes with Sbi and SpA are discussed further in Chapter 4.

The size exclusion profiles showed that in addition to large complexes there are also complexes that indicate the formation of 1:1 complexes with the serum IgGs from pig and cow in the presence of Sbi, and also with sheep, horse and goat IgGs with both Sbi and SpA. Possible explanations for the formation of these 1:1 complexes rather than larger complexes are: steric hindrance, interfering with the binding of more than one IgG for these serum IgG, or that only one of the Ig-binding domains is specific for a particular serum IgG present. The 1:1 complexes formed appear to be stable due to the narrow peaks produced in the elution profile (if dissociation and re-binding were occurring throughout the the elution from the column, broading of the eluted peak would be expected to occur due to acceleration and deceleration of the complexes as the size of the complexes increases and decreases).

All the other serum IgGs tested other than Chicken IgG produce complexes that are either not soluble or are above the resolution limit of the column. In order to determine the sizes of the different types of complexes formed in the presence of human, rabbit, Guinea pig (in the presence of both SpA and Sbi), cow and pig (in the presence of SpA only) requires a method such as analytical ultracentrifugation, which has a higher resolution for complex sizes.

The serum IgGs, for which we analysed the interaction with the Ig-binding proteins SpA and Sbi, are from animals that are known to acquire serious infections with *S. aureus*. Compared to the binding of human IgG, most of these IgGs (with the exception of pig IgG) bind both Ig-binding proteins to a lesser extent. What is responsible for this variability in binding and the ability to form insoluble complexes?

Are the differences observed in the binding due to changes in the amino acid sequence/structure of the IgGs from the different animal species, and if so what are these changes? The Sbi constructs, and SpA used in our experiments are both derived from *S. aureus* strains isolated from human hosts. Is it possible that the Ig-binding proteins from strains of *S. aureus* specific for these animals have acquired a high specificity for immunoglobulins from these animals? These questions will be addressed in the next two chapters of this thesis.

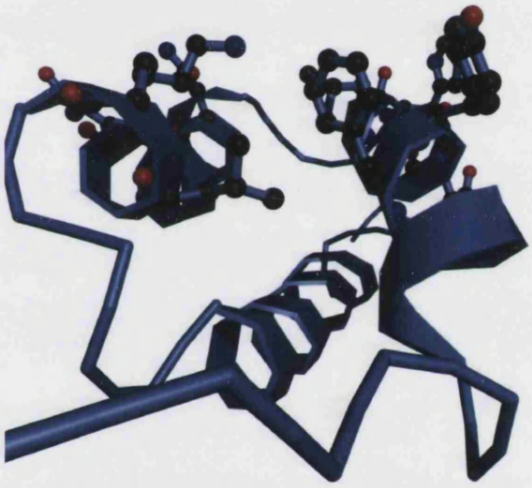
8 Chapter 4

8.1 Analysis of the interactions between Sbi and its ligands through homology models of the Sbi-IgG and SpA-IgG complexes

8.1.1 Introduction

8.1.1.1 SpA Ig-binding domain structures

Figure 34 NMR structure of SpA with the residues identified to be involved in binding IgG shown as ball-and-stick representations.



SpA binds to the Fc portion of the IgG molecule (Deisenhofer 1981, Kronvall *et al* 1970, Jansson, Uhlen, Nygren 1998) via interactions with the C_H2 and C_H3 domains of IgG. These interactions were first described by Deisenhofer (1981) from crystallographic data of SpA domain B in complex with the Fc fragment of human IgG. Two other studies helped to shed light on the importance of the interacting residues in the formation of this complex. Romangnani *et al* (1982) used iodination of SpA, a technique that specifically inactivates tyrosine residues in the protein of interest. Iodination of the single tyrosine present in every SpA Ig-binding domain eliminates Fc binding, although the Fab binding is unaffected. This tyrosine residue (Y48, Y109, Y167, Y225, Y283 in domains E, D, A, B and C, respectively) is not only conserved in all SpA Ig-binding domains but also in both of the Sbi Ig-binding domains. Other work by Gouda *et al* (1998) studied the interaction between SpA and Fc using NMR. The NMR structure of SpA is shown in Figure 34, with the residues determined by NMR to be involved in Fc binding shown. Table 8 lists the residues identified by Deisenhofer (1981) and Gouda et al (1998) to be involved in Fc binding.

Table 8 Residues identified by Deisenhofer (1981) and Gouda et al (1998) to be involved in Fc binding by X-ray and NMR studies respectively

NMR study (Gouda *et al* 1998)	X-ray study (Deisenhofer 1981)	
	Contact-1	Contact-2
	F1	
	Q5	
	Q6	
	N7	
F9	F9	
Y10	Y10	
E11		
I12	I12	
L13	L13	
		E20
	R23	
N24	N24	N24
		G25
		F26
	I27	
Q28	Q28	Q28
		S29
L30		
K31	K31	K31
		D37
		D38
		P39
		S40
		Q41

While there is structural data available for the SpA Ig-binding domain and its interaction with IgG, there is no structural information about the formation of the large SpA-IgG complexes. We have confirmed earlier work by Kronvall *et al* (1970) that certain species of serum IgG form insoluble complexes with SpA, such as human and pig serum IgGs, while others such as cow and goat serum IgGs bind SpA but do not precipitate in its presence. The structure of an Fc fragment in complex with the Z domain (the artificial SpA Ig-binding domain) shows a Z domain molecule in

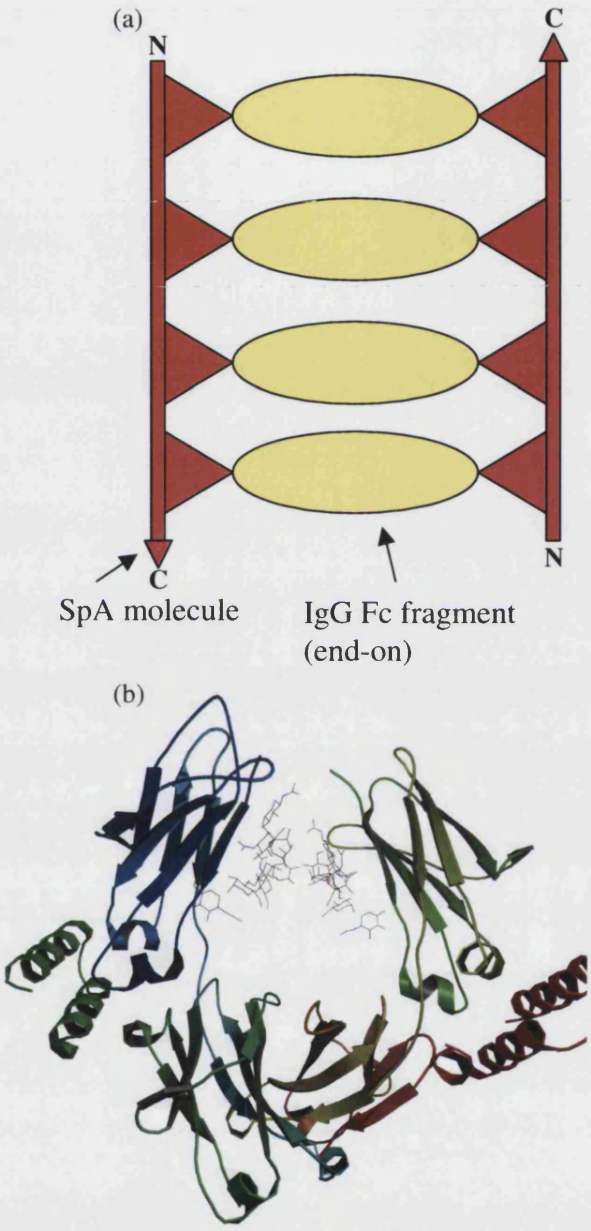


Figure 35 (a) cartoon of SpA/IgG multimer as predicted by Hanson and Schumaker (1984). **(b)** Crystal structure of an Fc fragment in complex with 2 SpA domains (Raju *et al* 2003, PDB accession code: 1OQO).

complex with each of the two Fc polypeptide chains (Figure 35a), suggesting that an IgG Fc molecule has two functional SpA binding sites (Raju *et al* (2003; PDB accession code 1OQO). This observation supports earlier work from Hanson, Phillips,

and Schumaker (1984), and Hanson and Schumaker (1984). Hanson, Phillips and Schumaker (1984) used analytical ultracentrifugation to measure the size of the complexes formed between SpA and rabbit serum IgG. They identified several different complexes with sizes: 7S, 10S, 13S, and 15S. They also showed that the formation of these complexes and their distribution was sensitive to the ratio between SpA and rabbit serum IgG. Indeed when the ratio of IgG to SpA was ≥ 2, the complexes formed were a unique 17S complex. Electron microscope analysis of this complex, suggested it is composed of four molecules of IgG and 2 of SpA. In a later study by Hanson and Schumaker (1984) they tried to explain the complexes observed in the analytical ultracentifugation. For the larger 17S complex, they predicted that the four IgG molecules are stacked on top of each other, with an SpA molecule running down each side (Figure 35a). This model by Hanson and Schumaker (see Figure 35a) is supported by the SpA structure determined by Raju *et al* (2003) which shows a single IgG Fc fragment with a SpA Z domain molecule bound to each IgG peptide (see Figure 35b).

Aim of this study

In Chapter 3, I showed that Sbi and SpA have different binding properties for a set of serum IgGs. But what causes these differences? As discussed in Chapter 2, there are differences in the sequences of Sbi and SpA, but the Fc binding face is largely conserved. There are also differences between the sequences of the IgGs used in Chapter 3. How, though, are these sequence variations arranged on the three-dimensional IgG structure, and is the binding site for SpA (and Sbi) affected. To investigate this, I will use the structure of the complex of SpA with a human IgG Fc fragment to analyse the interactions between SpA, Sbi and the set of serum IgGs used in Chapter 3 in homology models. In addition, I will present a model, based on the work by Hanson and Schumaker (1984) and Raju et al (2003), to explain the differences in immune-complex formation between SpA and Sbi.

8.1.2 Methods

8.1.2.1 Sequence alignment

Protein sequences for Sbi, SpA and the IgGs were obtained from the PubMed database (http://www.ncbi.nih.gov/entrez/query.fcgi) in fasta format, and aligned with the alignment program ClustalX (Thompson et al 1999) in the multiple alignment

mode. Accession numbers for the human, horse, cow, sheep, pig, Guinea pig, rabbit and chicken IgGs used in this modelling study are listed in Table 9a and the accession numbers for Ig-binding protein sequences are listed in Table 9b. Sequences for the structure used as templates for modelling IgG and Sbi were obtained from the RCSB Protein Database (PDB) (http://www.rcsb.org/pdb/), again in fasta format; PDB code 1FC2 (Deisenhofer 1981).

Table 9(a) Accession numbers for the IgG protein sequences. **(b)** Accession numbers for Sbi and SpA protein sequences.

(a) Species and IgG Subclass	PUBMED Accession number for protein sequence
1FC2	Taken from PDB database
Human_IgG3	A23511
Human_IgG1	GHHU
Human_IgG4	G4HU
Human_IgG2	G2HU
Horse IgG6 *	CAC86341.1
Horse_IgG4 *	AAS18415.1
Horse_IgG7	AAS18414.1
Horse_IgG3	CAC86339.1
Horse_IgG1	CAC44760.1
Cow IgG2*	S06611
Cow_IgG2a	AAB37380.1
Cow_IgG1	AAB37381.2
Cow_IgG3	AAC48762.1
Sheep IgG1*	CAA49451.1
Pig IgG4*	I47162
Pig_IgG2a	I47159
Pig_IgG2b	I47160
Guinea_pig IgG2*	P01862
Rabbit IgG*	P01870
Chicken IgG (clone 36)*	S00390

(b) Ig-binding protein	PUBMED Accession number for protein sequence
Sbi	CAG44122
SpA	AAB05743

8.1.2.2 Homology modelling

Protein structure coordinates were obtained from the PDB. The X-ray crystal structure of the SpA domain B in complex with an IgG Fc fragment (PDB accession code 1FC2) was used as a template for the models of the complexes between the Sbi Ig-binding domains and the IgG Fc fragments from seven different animal sera listed in Table 9. The models were generated by systemically replacing the residues of the template with those of the sequence being modelled in the program 'O' (see Chapter 2 for more details).

All IgG sequences except those for rabbit and chicken contained a two-residue insertion between amino acids E388 and N389 (according to the numbering of the human IgG Fc structure) in the sequence. This insertion is located in a loop region, and the lego function in 'O' was used to convert the extended loop into a single turn helix. Chicken IgG contains two insertions, 2 residues between amino acids G385 and Q386, and 9 residues between D399 and S400, according to the amino acid numbering of the human IgG Fc structure, and a single residue deletion that is not

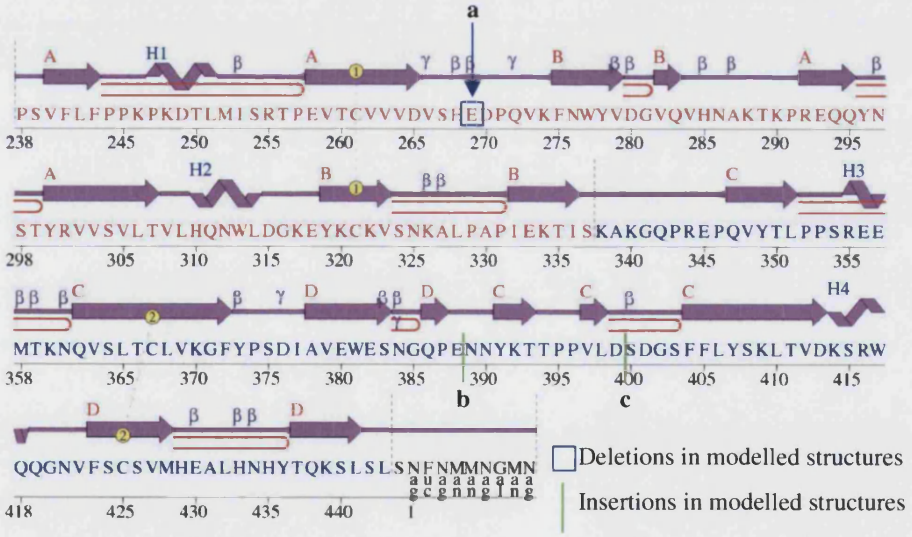


Figure 36 Sequence and secondary structure of human IgG Fc fragment for 1FC2. Marked are the locations where the insertions and deletions occur in the Fc models. The sequence for the C_H2 domain is in red while the sequence for the C_H3 domain is in blue

present in any of the other sequences (E269 in the human sequence). Figure 36 shows the human IgG Fc amino acid sequence for the structure that is used as a template for modelling, with the sites of insertions and deletions highlighted (the amino acid numbering used follows the 1FC2 sequence). The insertion and deletion sites in the homology models were refined using the lego-loop function. In cases of long insertions in loop regions, the expected level of confidence for the particular loop in the resulting model is low. With the exception of these regions in the IgG Fc homology models, there were no significant alterations to the backbone structure of the remainder of the IgG Fc structures or the Sbi models. Side-chains replacements in the Sbi Ig-binding domain models, were adjusted to a position that resembles the corresponding SpA residue side-chain as closely as possible, using the rotamers that were available in the lego loop function. In the case of the Fc molecules, alterations to the side-chain conformations were only performed in the regions of loop modelling due to insertions or deletions, unless a clash was identified by visual examination of the molecule.

8.1.2.3 Determination of potential interactions between Sbi/ SpA and the IgG Fc models

For the homology models of the complexes of IgG Fc and each of the Sbi Ig-binding domains or SpA Ig-binding domain B, the 1FC2 structure was used as a template. The resulting complexes were energy minimised using the **M**oleculer **O**perating **E**nvironment (MOE; Chemical Computing Group Inc.: http://www.chemcomp.com/) software, and all atoms of the binding partners within 1.8Å and 5.2Å were analysed using the CCP4 program CONTACT (Collaborative Computational Program 1994). The residues that comprised the interface between the IgG Fc and the Ig-binding protein were compared for all the combinations of IgG species used with the three Ig-binding protein models (SpA-B, Sbi-1 and Sbi-2).

The interacting residues at the interface of the two constant (C_H) domains of the Fc molecule were also taken into account. This was done by separating constant domains (at residues 338 and 339 according to the PDB file 1FC2). Subsequently, CONTACT was used to determine all atoms within 0Å to 5.2Å distance of an atom in the other half of the molecule. All residues that were detected in this region on both halves of the molecule were than compared to those of the other species examined.

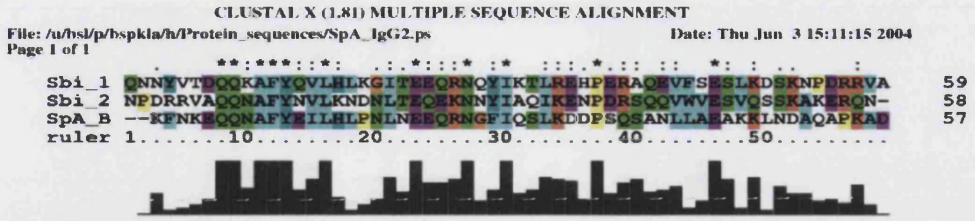


Figure 37 Sequence alignment of the SpA domain B to both the Ig-binding domains of Sbi. This alignment was used in the generation of models of the Sbi domains.

8.1.2.4 Analysis of the side-chain environment with VERITY 3D

The environments of the residue side-chains of the IgG Fc models were analysed to determine if any residues were exposed to unfavourable environments. This was done with the web-based program VERIFY_3D (http://doe-mbi.ucla.edu/Services/Verify_3D/) (Luthy *et al* 1992).

8.1.3 Results

8.1.3.1 Homology modelling of Sbi's Ig-binding domains

The two Sbi Ig-binding domains were aligned to the B domain of SpA in the 1FC2 structure (Figure 37). The Sbi domains were modelled onto this structure rather than using the previous models generated in Chapter 2 to conserve any changes in the structure that are the result of binding SpA to IgG Fc and that are present in the SpA-Fc complex, but are absent in the SpA-Fab complex.

8.1.3.2 Homology modelling of IgG Fc fragments

We used IgG sequences from eight mammalian, and a single avian species that were available in the sequence databases (corresponding to the IgGs tested in binding studies with Sbi in Chapter 2 (a sequence for goat IgG could not be found in the Pubmed database)). These sequences were aligned with the sequence of the IgG in the structure of the human IgG Fc – SpA complex (Deisenhofer 1981). The sequence identity and similarity of these IgGs with the sequence of the human Fc template (1FC2) are indicated in Table 10. The sequences for the human and chicken IgGs shared only 35.4% identity. Of the mammalian sequences analysed, cow IgG was the least similar to human IgG (with 63.9% identity and 74.0% similarity), while rabbit was the most similar (with 73.1% identity and 80.3% similarity). These data are also

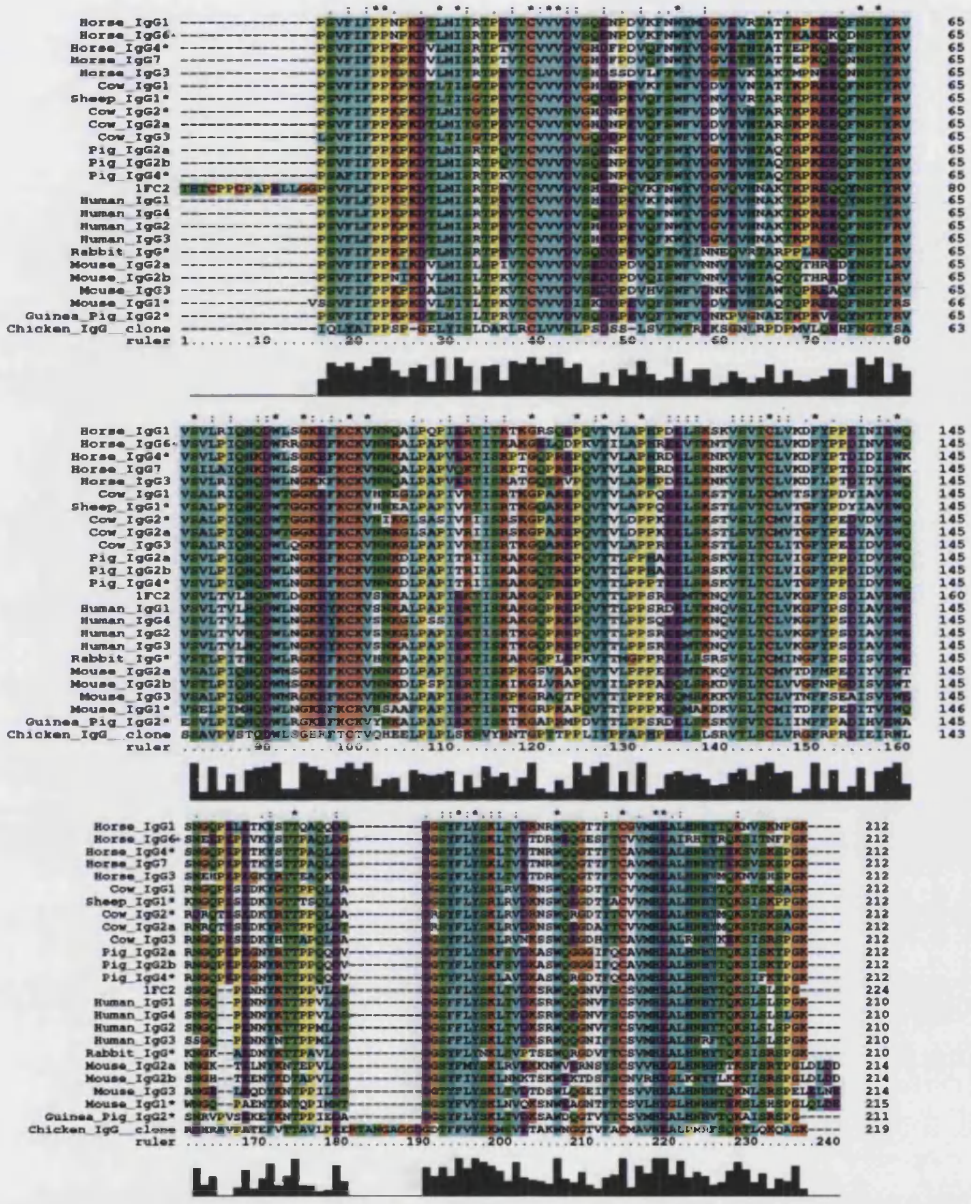


Figure 38 Alignment of IgG sequences. The sequences indicated with * were used in the generation of IgG models.

shown in Table 10. The alignment of the mammalian IgGs (Figure 38) contained a two-residue insertion in most of the sequences between residues E388 and N389 of the human sequence (rabbit IgG was the only exception). This loop, however, is located at a position in the structure that is not likely to interfere with SpA and Sbi binding.

The chicken IgG model contains a short insertion in the same loop as the mammalian insertion, and a single residue deletion in another loop (residue E269, following to the human sequence) that is also unlikely to interfere with the binding to SpA and Sbi (Figure 38). Chicken IgG also contains a longer insertion, in a loop region located near the hinge region of the two Fc domains (between residues D399 and S400 in the human sequence; Figure 38). While in human IgG Fc this loop appears not to be involved in SpA binding, the resulting loop in the chicken IgG model is lengthened, and could interfere with the angle between the two constant domains and also affect the SpA binding site.

All modelled IgG structures were analysed with VERIFY 3D. This analysis revealed that, while most structures are very similar, certain IgGs have regions that differ from the template structure (human IgG Fc: 1FC2). These differences could be related to the differences in binding, as most of these areas involve regions on the protein backbone that compose the link between the C_H2-C_H3 domains (residues S337 to R344 in the human sequence) (Figure 39).

Table 10 Relative sequence identity (as a percentage) of the IgG Fcs used in this modelling study

	Chicken IgG	Cow IgG2	1FC2	Guinea Pig IgG2	Horse IgG6	Pig IgG4	Rabbit IgG	Sheep IgG1
Chicken IgG clone36	100	36	35	37	33	38	37	34
Cow IgG2	36	100	64	63	61	74	64	80
1FC2	35	64	100	71	69	72	73	68
Guinea Pig IgG2	37	63	71	100	62	67	69	68
Horse IgG6	33	61	69	62	100	68	63	64
Pig IgG4	38	74	72	67	68	100	70	73
Rabbit IgG	37	64	73	69	63	70	100	69
Sheep IgG1	34	80	68	68	64	73	69	100

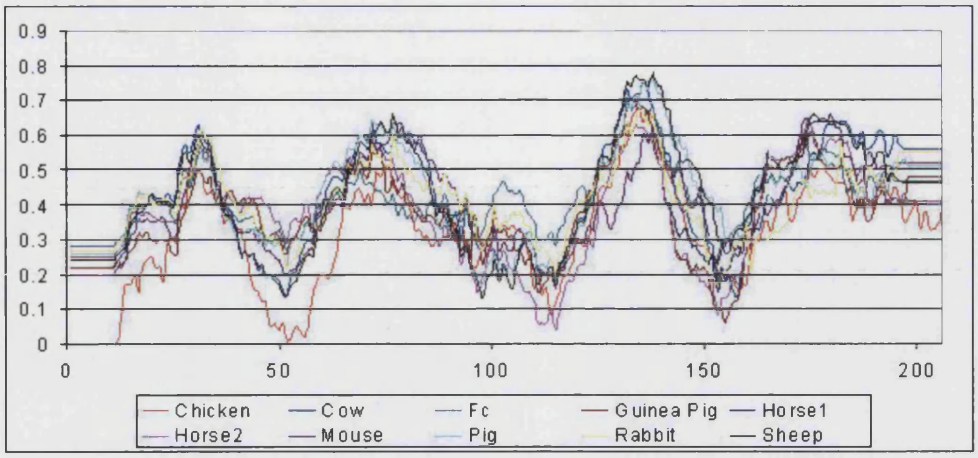


Figure 39 Verify 3D results for the IgG models. The result for all the IgG models are entirely in the upper portion of the plot, suggesting it has been correctly modelled (except for the first nine residues of the Chicken IgG sequence, however the first and last nine residues of all protein analysed by this method are meanless, see 6.1.1.1.2).

8.1.3.3 Analysis of the SpA and Sbi residues involved in the interaction with human IgG Fc

The program CONTACT was used to determine which pairs of residues interact between the Ig-binding protein and its IgG Fc ligand. The interactions detected in the SpA-Fc complex were compared to the interactions seen in the models of the Sbi-1 Fc and Sbi-2 Fc complexes, to determine whether the changes in the sequence of the three Ig-binding domains could affect the interaction between the two components of the complex. The interaction pairs are shown in Figure 40a for all three Ig-binding domain models (SpAB, Sbi-1, and Sbi-2) in complex with a human IgG Fc. The plots in Figure 40a show the interactions between the human IgG molecule and an Ig-binding domain. Residues 250 to 438 are the human IgG residues whilst residues 1 to 31 are those of the Ig-binding protein. CONTACT (see methods and Chapter 2) was used to calculate those residues in the IgG molecule that are within a fixed distance (1.8 to 5.2Å) of one in the Ig-binding domain. Residues within this distance are assumed to be interacting. The present of a green square in the horizontal row indicates that residue is interacting with the residue labelled at the top of that column.

Changes in the interaction patterns do not necessarily correspond to changes in the sequence of the Ig-binding protein; such changes could also be the result of changes

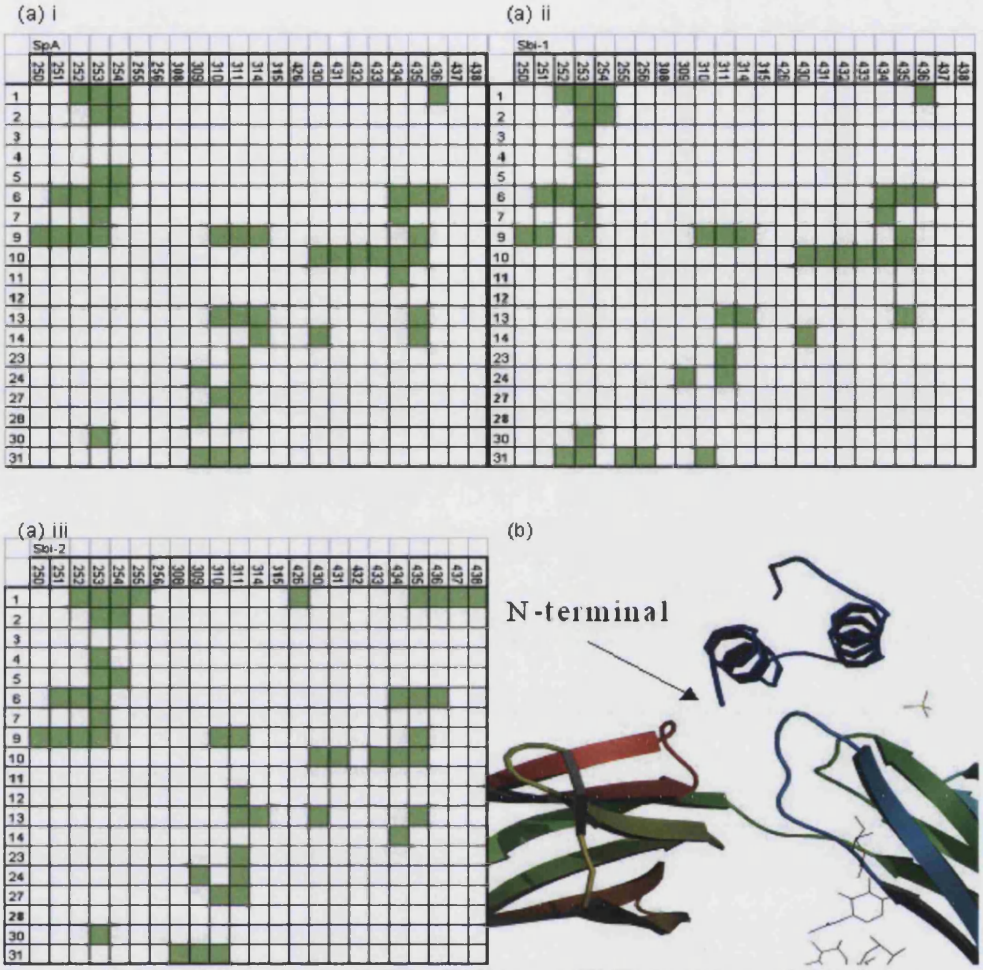


Figure 40 (a) Interaction patterns for **(i)** SpA, **(ii)** Sbi-1 and **(iii)** Sbi-2 with human IgG Fc fragment. **(b)** 1FC2 showing the N terminal region wrapped around the IgG. Residues 250 to 438 are the human IgG residues while residues 1 to 31 are those of the Ig-binding protein. A green square indicates an interaction between the IgG residue and the residue in the Ig-binding domain.

elsewhere in the Ig-binding domain, resulting in the adjustment of the backbone of both components of the complex in the MOE energy minimisation program. The substitutions that appear to influence the interaction pattern are Q28K and K31R in Sbi-1 and F1R and Q28A in Sbi-2. Of these, the F1R substitution of Sbi-2 is probably of least interest as (1) this residue was not identified to interact with IgG in the NMR study of Gouda *et al* (1998) and (2) we believe that in the full length SpA molecule, the N-terminal of the B domain will not interfere with the binding of the Fc molecule as it does in the 1FC2 structure with the N-terminal part of the SpA molecule wrapped around the Fc molecule (Figure 40b). Of the remaining residues mentioned here, all

were identified to be involved in binding Fc by Gouda *et al* (1998), however we are unsure as to how these affect the binding of Sbi to IgG in comparison to SpA.

8.1.3.4 Analysis of the IgG Fc residues involved in the interaction with the Ig-binding proteins

The same methods as described in 8.1.3.3, were used to determine where substitutions in the IgG sequence between the different IgG species resulted in changes in the contacts between the IgG molecule and the Ig-binding domain. Examination of the interaction patterns between the Ig-binding proteins and the models of the different species' IgGs reveals numerous differences (the interaction data are shown in Appendix 4to Appendix 6). Comparison of the interactions between pairs of residues reveals that, while some interactions occur in all models, there are other interactions that only occur in a few models (which I termed 'gain' of interactions), or interactions that occur in most of the models, but not all (where an interaction occurs in 7 or more cases, any model lacking this interaction is termed as having 'lost' the interaction).

When the sequences are examined for differences, not all changes in the interactions between residues can be accounted for by differences in sequence. This suggests that differences in other regions of the Fc molecule may cause small changes in conformation at the C_H2 and C_H3 domain interface that perhaps affect the interaction with Ig-binding proteins without the sequence in the binding region itself being altered. It was however possible to identify several Fc residue substitutions affecting the interactions with the Ig-binding proteins. The interactions between the Fc models and the Ig-binding proteins of interest are listed in Table 11, Table 12, and Table 13.

Changes in the contacts between the IgG Fc molecule and the Ig-binding domain include the gain of an interaction Q6-F436 in the chicken IgG model complex, compared to the human Fc sequence: Y436, and lost another N24-S309 (Q309 in human Fc sequence). In the case of the F436Y substitution, it is not clear how the substitution changes the interaction. It could be due, not to the substitution, but re-arrangement of the backbone due to the larger number of substitutions present in the chicken IgG sequence. However the same substitution (F436Y) is found in the human IgG3 sequence in addition to a H435R substitution (from the sequence of Human IgG1), which also occurs in both chicken and human IgG3. Human IgG3 is unique

among the naturally occurring human IgG subclasses in that it does not bind either SpA or Sbi (see chapter 3). The H435R substitution in the Chicken IgG model is not observed in any of the complexes with Ig-binding proteins, however it is observed in the interactions between the C_H2-C_H3 domains. Therefore, while F436Y alone probably does not prevent binding, in combination with H435R, its effect could knockout SpA and Sbi binding. Indeed, Jendeburg et al (1997) showed the introduction of the H435R F436Y mutations into human IgG1 eliminated its ability to bind SpA. The Q309S substitution probably causes the loss of the interaction with N24 in the Ig-binding domain due to the reduced size and length of the serine side chain in comparison to the glutamine side chain.

Table 11 Changes in interactions between the Fc homology models and SpA. Superscript numbers refer to interactions marked in Appendix 4.

Species	Gained mutations	Lost Mutations
Human IgG1 (1FC2)	K31-L309[4]	
Chicken_IgG_clone36	K31-S309[4] Q6-F436[19]	F9-E250[1] N24-S309[2] Q28-S309[3]
Cow_IgG2		
Guinea_Pig_IgG2	I12-L314[9] E20-R315[11] R23-R315[12] N24-R315[13] Q6-V436[19]	
Horse_IgG4	F1-A426[14]	
Horse_IgG6	E20-R314[6] E20-R314[7] N24-R314[8] F1-T435[15] L13-T435[17] H14-T435[18]	F9-T435[16]
Mouse_IgG1	I12-L314[9] Q6-H436[19]	
Pig_IgG4		
Rabbit_IgG	K31-T309[4] N19-R315[10] E20-R315[11] R23-R315[12] N24-R315[13]	N24-T309[2] Q28-T309[3]
Sheep_IgG1	H14-T314[5]	

Table 12 Changes in interactions between the Fc homology models and Sbi-1. Superscript numbers refer to interactions marked in Appendix 5.

Species	Gained mutations	Lost Mutations
Human IgG1 (1FC2)		R31-309[7]
Chicken_IgG_clone36	R31-252[2] R31-255[4] L13-310[8] Y1-426[19] Q6-436[24]	F9-250[1] N24-309[5] R31-309[7]
Cow_IgG2	Q6-254[3]	L13-314[13]
Guinea_Pig_IgG2	L13-315[15] E20-315[16] R23-315[17] N24-315[18] Q6-436[24]	
Horse_IgG4	E20-311[10] R23-311[11]	F9-311[9] Y1-434[21]
Horse_IgG6	Y10-314[12] H14-314[14] L13-315[15] E20-315[16] R23-315[17] N24-315[18]	F9-250[1] F9-435[22] H14-435[23]
Mouse_IgG1	Q6-254[3] R31-255[4] K28-310[6] L13-315[15] Y10-431[20] Q6-436[24]	
Pig_IgG4		
Rabbit_IgG	N24-309[5] E20-315[16] R23-315[17] N24-315[18]	
Sheep_IgG1	R31-255[4]	

Of all the interactions lost or gained in the Fc models, the only amino acids for which a pattern can be found between the side chains present and the level of binding observed in Chapter 3 is residue 315 (see Table 14 for the identity of this residue in all models). The role of side chain of residue 315 appears to be the strongest in determining the pattern of binding. Several of the Fc models have gained an interaction for residue 315 with at least one of the Ig-binding domains, however interactions involving residue 315 are absent in the Fc models for human (1FC2), chicken, cow, horse IgG4 and sheep. This amino acid is the most variable residue detected in the Fc-Ig-binding protein interface. We find that there is a strong

Table 13 Changes in interactions between the Fc homology models and Sbi-2. Superscript numbers refer to interactions marked in Appendix 6.

Species	Gained mutations	Lost Mutations
Human IgG1 (1FC2)	K31-V308[4]	
Chicken_IgG_clone36	K31-D256[3] R1-M426[14] Q6-F436[18]	N24-S309[5] K31-S309[7]
Cow_IgG2	R1-V426[14]	
Guinea_Pig_IgG2	K23-R315[11] N24-R315[12] Q6-V436[18]	
Horse_IgG4	F9-V250[1]	F9-K311[8]
Horse_IgG6	Y10-R314[9] K23-R315[11] N24-R315[12] I27-R315[13] R1-V426[14] R1-H434[16] R1-T435[17]	
Mouse_IgG1	F9-V250[1] K31-L255[2] A28-M309[6] L13-N315[10] K23-N315[11] R1-L428[15] Q6-H436[18] R1-E438[19]	
Pig_IgG4	K23-N315[11]	
Rabbit_IgG	K13-R315[10] K23-R315[11] N24-R315[12]	N24-T309[5]
Sheep_IgG1		

correlation between the residue type at position 315 and the binding characteristic described in chapter 3, across all ten species. We propose that large charged residues such as arginine and asparagine are required for the high binding levels while the small side chains such as glycine and serine are linked with low binding species.

In ten Fc models, residue 315 exists as five different residues. The species for which an interaction was observed in our homology models of the Ig-binding domain – Fc complexes are Guinea pig IgG2 (R315), horse IgG6 (R315), mouse IgG1 (N315), pig IgG4 (N315), and rabbit IgG (R315). For all these species high levels of binding was observed, except for horse IgG6. The only species to show high levels of binding (Chapter 3) not in this list is the human Fc model determined by Deisenhofer (1981).

Table 14 Summary of residues present at position 315 in the IgG models. The affinity of Sbi and SpA for the different serum IgGs was determined in Chapter 3. * marks the IgG sequences used in the generation of the models. [&] Human IgG3 shares the high binding residue at position 315 found in the other human IgG sequences, however it lacks SpA or Sbi binding due to the present of the substitutions H435R and F436Y.

IgG species and subclass	Residue 315	Relative binding affinity for serum IgGs
Horse IgG1	S	Low binding
Horse IgG6*	R	
Horse IgG4*	S	
Horse IgG7	S	
Horse IgG3	N	
Sheep IgG1	G	Low binding
Cow IgG1	G	Low binding
Cow IgG2*	G	
Cow IgG2a	G	
Cow IgG3	G	
Pig IgG2a	N	High binding
Pig IgG2b	N	
Pig IgG4*	N	
Human IgG1 (1FC2)	D	High binding
Human IgG1	N	
Human IgG4	N	
Human IgG2	N	
Human IgG3[&]	N	
Rabbit IgG*	R	High binding
Mouse IgG2a	S	High binding
Mouse IgG2b	S	
Mouse IgG3	R	
Mouse IgG1*	N	
Guinea Pig IgG2*	R	High binding
Chicken clone 36*	S	Low binding

Here, the residue 315 is an aspartate, while in the other human Fc sequences listed in Table 14, this residue is an asparagine. As N315 is also found in other IgG species, this supports Human IgG being a high binding species based on sequence alone, as shown in the binding studies (Chapter 3). It is assumed that the low binding for horse serum IgG, in the case of the IgG6 subclass, despite the presence of R315 is due to the presence of a substitution at position 314 (L314R), with the larger arginine side chain disrupting the interacting between the IgG and the Ig-binding domain.

Table 15 Differences in interactions between the C_H2 and C_H3 domains between the IgG models. Numbers in superscript refer to squares in Appendix 3.

Species	Gained interactions?	Lost interactions?
Human IgG1 (1FC2)	NONE	
Chicken_IgG_clone36	S246-E378[3] G249-E378[6] E250-D376[7] T340-P374[13] P342-P374[14] N339-R375[16] N339-D376[19] N339-F404[21] E250-R435[26] T310-R435[28]	T340-R373[10] G341-R373[11] R338-P374[12] Y252-V428[23] R338-E430[25] Y252-R435[27]
Cow_IgG2	K246-D378[3]	R338-P374[12] R338-E430[25] T314-H435[29]
Guinea_Pig_IgG2	K246-H378[3] P247-A375[4] A342-N371[8] K338-A375[15] T339-A375[16] K340-A375[17] T339-P398/G402[20] T339-S404[21] R315-E430[24]	P247-H378[5]
Horse_IgG4	D249-D378[6] V250-D376[7] T340-P374[13] P339-D376[19] P339-Y404[21]	
Horse_IgG6	K246-N378[3] E342-P374[14] A339-P375[16]	
Mouse_IgG1	T339-E375[16] T339-D376[19] V250-H435[26] L251-H436[30]	T339-F373[9] K340-F373[10] G341-F373[11] P247-L428[22]
Pig_IgG4	K246-D378[8]	
Rabbit_IgG	NONE	
Sheep_IgG1	P244-Y376[1] P244-Y376[2] T339-D375[16] S337-Y376[18] T339-Y376[19]	R338-E430[25] T314-Y435[29] T252-T436[31]

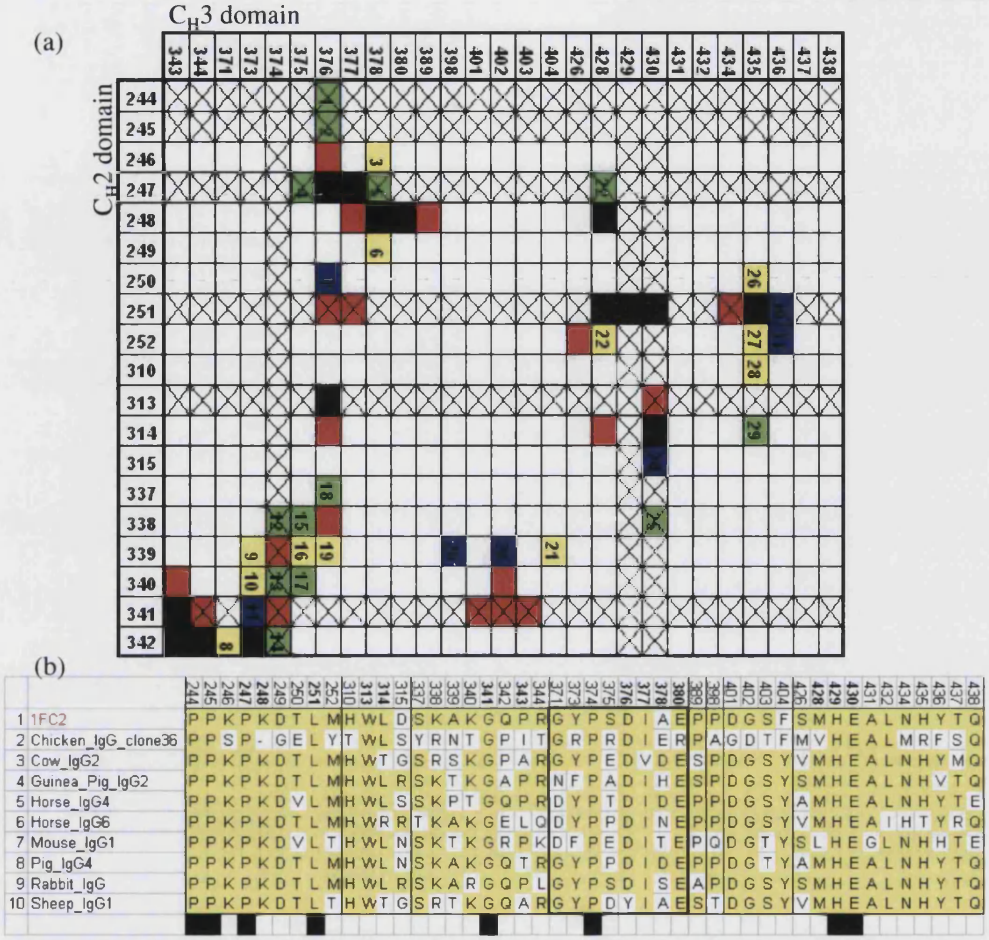


Figure 41 (a) Summary of interactions observed in Appendix 1 between the IgG C_H2 and C_H3 domains. The numbered squares refer to the interactions listed in Table 15. Interactions detected in all models are marked in black. Changed interaction pattern: marked in red do not correspond to changes in sequence; blue corresponds to changes in the C_H2 domain; green to changes in the C_H3 domain; yellow to changes in both domains. **(b)** Sequence of the IgG models in the regions identified to be involved in interacting with the other domain. Conserved residues are highlighted in yellow. Residues conserved in all sequences have a black mark under them.

8.1.3.5 Analysis of the interactions at the C_H2 and C_H3 interface of the IgG Fc models

The interactions between the C_H2 and C_H3 domains and information on the loss or gain of interactions is shown in Appendix 2.

Analysis of the interaction data showed numerous changes in the pattern of interactions between the two C_H domains in all models (listed in Table 15). However

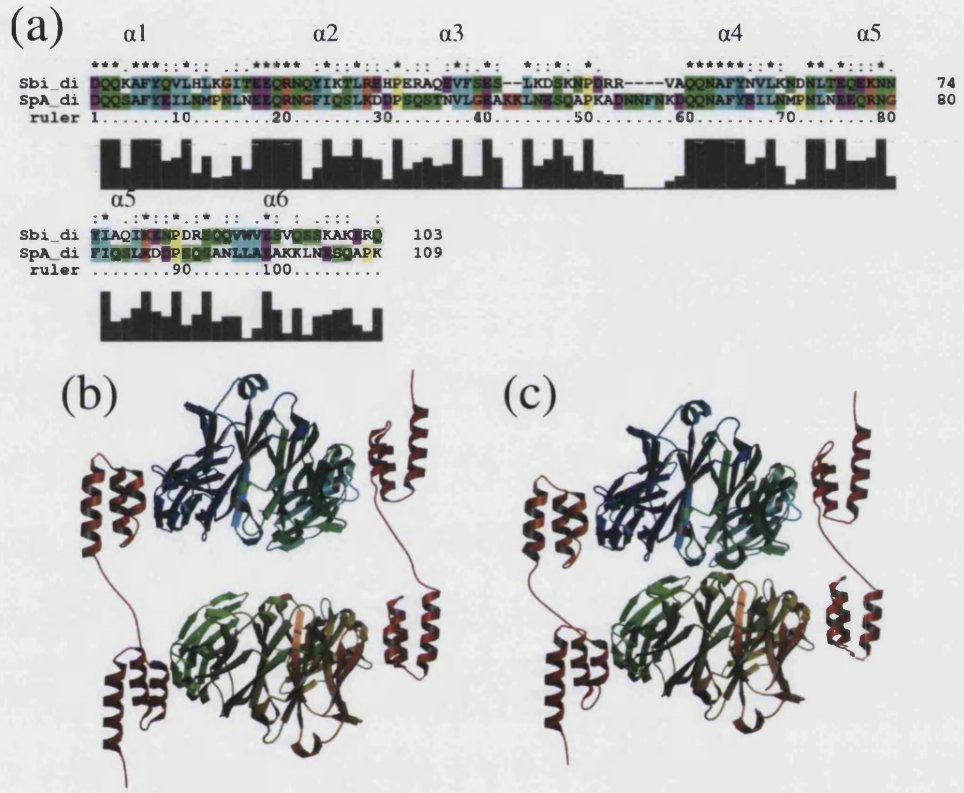
not all variations in the interactions between the C_H2 and C_H3 domains are due to changes in side chains. This can occur if a substitution in the sequence causes a change in the conformation of the backbone at a different point in the sequence. Table 15a shows all interactions between the C_H2 and C_H3 domains, and indicates which interactions are the result of residue substitution, and those interactions which are not.

There are unique sequence changes that could be affecting the interactions between the two domains. These substitutions observed in the Fc models are shown in Table 15. The Chicken Fc model contains several interactions that appear to be affected by sequence substitutions. These include the gain of interactions at E250-D476; T340-P374; E250-R435; and T310-R435, and loss of interactions at T340-R373 and Y252-V428. Interaction changes in the other models include residues T339 (Guinea pig IgG2 and Mouse IgG1), T340 (Horse IgG4), H436 (Mouse IgG1), and Y376 (Sheep IgG1).

8.1.3.6 Homology model of SpAED in complex with multiple IgG molecules

No structure of SpA determined by a structural method exists with more than one Ig-binding domain in a single polypeptide chain. We therefore generated our own using the SpA structure of the D domain in 1DEE, and linking the two identical molecules together in O. The sequence of second Ig-binding domain was altered to match that of the SpA A domain (the sequence of the model is shown in Figure 42a). The PDB structure 1OQO (Raju *et al* 2003) contains the Fc fragment of an IgG in complex with a Z domain bound to each polypeptide chain (Figure 35b: the Z domain is the artificial SpA Ig-binding domain). The two-domain SpA model was rotated onto each of the Z domains and a second copy of 1OQO moved onto the other Ig-binding domain. The finished structure is shown in Figure 42b. This structure clearly shows that it is possible to fit two Fc molecules sandwiched between two SpA molecules in a fashion that corresponds to the Hanson and Schumaker model (Figure 35).

Figure 42(a) Sequence alignment of the SpA and Sbi two domain structures used in the generation of the large IgG complexes. **(b)** Predicted structure of SpA in complex with 2 IgG molecules. **(c)** Predicted structure of Sbi in complex with two IgG molecules.



8.1.3.7 Homology model of Sbi1-2 in complex with multiple IgG molecules

A similar model of Sbi in complex with IgG was also generated. The above complex of SpA and IgG was used as a template. Sbi was built onto each of the two SpA molecules in the complex. However the region of peptide chain linking the Sbi Ig-binding domains together is shorter by four residues than the linker in SpA. Therefore the extra residues were removed, and as a consequence the two-bound Fc molecules are positioned closer together. The resulting structure (Figure 42c) shows that such a complex cannot exist without significant steric clashes between the bound Fc molecules.

8.1.4 Discussion

In this chapter we used molecular modelling to examine and predict the amino acids and their interactions at the binding interface in complexes between IgG Fc fragments

and Ig-binding proteins Sbi and SpA. Since this interaction is located near the junction between the C_H2 and C_H3 domains, we also included the residues at this domain interface and the interactions that may affect binding.

8.1.4.1 Prediction of the IgG Fc residues that are important in the complexes with SpA and Sbi

The examination of the interface of the C_H2 and C_H3 domains revealed a number of interactions that differ between the various IgG Fc models. For example the sheep IgG1 model is the only one to contain a tyrosine at position 376, all other models have an aspartate at this position. The tyrosine side chain is larger and bulkier than the aspartate side chain, and probably affects the tilt angle between the C_H domains, thereby affecting the distance between the two halves of the SpA binding site.

Two substitutions in the IgG Fc protein correspond to observed binding levels to the Ig-binding protein.

The interaction pattern analysis identified residue 315 to be important in the binding of IgG to Sbi and SpA. At position 315, an arginine was found in three models (Guinea pig, horse IgG6, and rabbit IgG) while an asparagine (pig, mouse and human) was found in three others. Serum IgGs from all these animal species were all shown to bind strongly to both Sbi and SpA. The serum IgGs that show low binding affinity for SpA and Sbi in Chapter 3 (horse, sheep, cow and chicken) have either a glycine or a serine at position 315.

Positions 435 and 436 also affect binding to Sbi and SpA. Almost all sequences have the residues tyrosine (435) histidine (436). Guinea pig IgG2 has the substitution Y435V, however it is not possible to determine from the binding studies on the serum IgG whether this substitution affects binding. A clear relationship however could be determined for both human IgG3 and chicken IgG. While a model for human IgG3 was not generated, we showed that this subclass did not bind to either Sbi or SpA. Chicken IgG also failed to show any binding to Sbi and SpA. And in its model, the substitutions Y435F H436R were shown to affect the interaction pattern. It therefore appears that these substitutions can completely eliminate all binding of an IgG to Sbi and SpA. This is supported by the work of Jendeberg *et al* (1997), who showed that

the introduction of the Y435F H436R substitution into IgG1 eliminates its ability to bind SpA. It is therefore possible that other IgGs, both mammalian and non-mammalian, that do not bind SpA (or Sbi) could share this substitution.

The type of residue present at positions 315, 435 and 436 give the only patterns that we detected that could be clearly correlated to the binding results already obtained in Chapter 3. Other substitutions could also affect binding, but to determine whether any of the identified substitutions do in themselves affect binding to Ig-binding proteins, it would be necessary to carry out site directed mutagenesis on an IgG molecule.

8.1.4.2 Predictions on the formation of insoluble complexes of Sbi and IgG

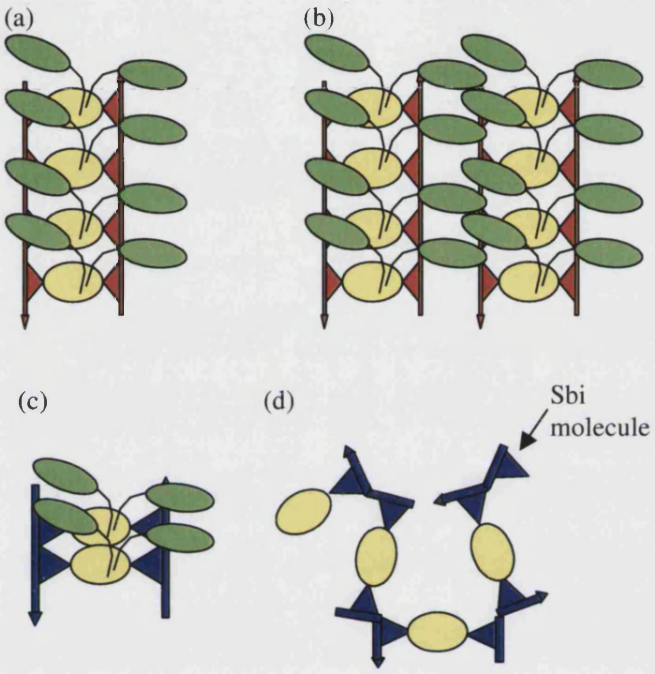
Comparisons of the SpA complexes with Fc and Fab fragments (Chapter 2) to the sequence of the Sbi Ig-binding domains strongly suggested that Sbi can only bind to the Fc portion of an immunoglobulin. This is confirmed by the binding studies described in Chapter 3 where the results of the SE experiments showed that Sbi only binds to the Fc fragments of digested IgGs. It therefore appears that Fab binding is not an important factor in the formation of large complexes as Sbi can form large insoluble complexes with only the Fc region of IgGs. SpA however required the Fab region to be present for the precipitation of IgG to occur (see Figure 27).

Hanson and Schumaker (1984) proposed a model for the formation of the large soluble complexes seen with rabbit serum IgG in the presence of SpA composed of 2 SpA molecules and 4 IgG molecules. The most likely model for such a binding ratio is a complex of four IgG molecules stacked in the same orientation, with SpA molecules running up and down each side of the Fc molecules (Figure 35).

We generated a 2-domain model of SpA, by linking adjacent SpA domains together based on the crystal structure of IgG with 2 Z-domains bound (Z-domains are the artificial protein A domains, Raju *et al* 2003). This model shows that it is possible to fit IgG molecules between the Ig-binding domains of two SpA molecules (Figure 42b).

We replaced the SpA molecules with Sbi on the 2-domain model we had generated. However, unlike SpA, it was not possible to get the Sbi domains far enough apart to

Figure 43(a) The SpA/IgG complex from Figure 35a with additional Fab regions present (green) on IgG Fc molecules. **(b)** Crosslinking of the SpA/IgG complexes via Fab. **(c)** Representation of the Sbi/IgG complex from Figure 42, showing the clash of the IgG Fc molecules. **(d)** Possible mechanism of cross-linking IgG Fc (yellow, Fab regions removed) with Sbi molecules (blue).



prevent overlapping of the IgG side chains on the adjacent IgG faces (Figure 42c and Figure 43c) strongly suggesting that Sbi cannot form such complexes, but it could form complexes with other arrangements of the subunits.

The only question remaining regarding the SpA-IgG complex is whether this is important in the formation of precipitation. This question is important as Hanson, Philips, and Schumaker (1984) carried out their experiments with rabbit IgG. Our experiments with rabbit serum IgG and SpA showed that this species is unique as it was the only example where high molecular weight complexes were formed, but remained completely soluble. In addition SpA requires the presence of the Fab region of the IgG molecule for precipitation to occur with human IgG, suggesting the SpA does not bind rabbit IgG via its Fab regions. (The previous work of Kronvall *et al* (1970) did show precipitation of rabbit sera by SpA, but as their SpA was purified from *S. aureus* cultures, it could have been contaminated with Sbi (both molecules are

of similar size, and both would bind to an affinity column based on SpA's affinity for IgG). Our SpA and Sbi are both recombinant proteins and should not contain any other *S. aureus* proteins.

Our immunodiffusion results (Chapter 2) suggest that SpA:Fc complexes indeed form the structure predicted by Hanson and Schumaker (1984, Figure 35a). This is supported by the modelling studies presented here. We predict though that the formation of this complex is not sufficient for precipitation to occur. We suggest that this complex forms with SpA in the presence of rabbit IgGs, assuming that SpA does not bind to the Fab regions of rabbit IgGs. In species such as human or pig, we think precipitation is triggered because, although the complex structure predicted here is formed between the Fc and SpA, binding also occurs between SpA and the Fab regions, cross-linking the stacked complexes, and causing precipitation (see Figure 43b).

The precipitation of IgG by Sbi could also be the result of crosslinking by IgG molecules. Rather than stacking of IgGs, and crosslinking via the Fab region, as we predicted to occur for the SpA Fc complexes, we envisage that Sbi can crosslink IgG molecules via the Fc region, as shown in Figure 43c&d.

However the individual Sbi Ig-binding domains, Sbi-1 and Sbi-2, can also precipitate IgG (see Chapter 3). Analysis of Sbi-1 by size exclusion chromatography and SAXS (Chapter 6) shows that this construct has a tendency to oligomerise. Could this have a role in the ability of the Sbi-1 and Sbi-2 to precipitate IgG?

9 Chapter 5

9.1 Sequence comparison of Immunoglobulin binding proteins SpA and Sbi isolated from different animal species

9.1.1 Introduction

As we have described in Chapter 3, Sbi and SpA show different binding characteristics to the IgGs isolated from serum of different animal species. Sbi for instance binds particularly well to human and rodent (rabbit, Guinea pig and mouse) serum IgGs, while SpA favours human, pig and Guinea pig serum IgGs and forms large soluble complexes with rabbit serum IgG. We have also found that both SpA and Sbi display poor affinity for serum IgGs from species such as goat, horse and cow. *S. aureus*, however , is known to cause infections and disease in these domestic animals. A well-known example is mastitis in dairy cattle. This raises questions about the role of Ig-binding proteins in infection of these animal species. In the previous chapter (Chapter 4), we have examined the molecular origin of the observed differences in IgG binding, but in our studies to this point, all binding experiments have been carried out with Ig-binding proteins cloned from a strain of *S. aureus* isolated from humans. In addition, the available SpA and Sbi sequences used in the analysis in Chapter 4 are obtained from *S. aureus* strains isolated from human hosts. Are the SpA and Sbi genes from strains infecting animal hosts different from the strains infecting humans? Could it perhaps be that Sbi and SpA from *S. aureus* strains specifically infecting domestic animal species have acquired specific affinity for IgGs from these animals?

In order to address these questions we have sequenced the Sbi and SpA genes from *S. aureus* strains isolated from a number of different species and analysed them for sequence changes and the functional implications of these changes. The sequence data are also used to assess the evolution of SpA and Sbi genes relative to their sequence type and host, and to investigate selective pressures that may act on these genes.

Table 16 Primers for sequencing the Sbi and SpA genes

Primer Set	Gene	Primer sequence (5' - 3')
1	Sbi	Forward - CGC TTT TTT ACA TAG TTA ACA CTA
		Reverse - GGC TTC TAT CAG GGT TTT CTT TAA
2		Forward - GAC CGA CGT GTT GCA CAA CA
		Reverse - CCT AAT AAT TTA GAT TGA GGA ACC
3		Forward - GAT GCT GAA AAG AAA GTG GCG
		Reverse - CGA TGT AAG GAT TAT TTA ACA AG
4	SpA	Forward - TTT TAG TAT TGC AAT ACA TAA TTC G
		Reverse - CTT TGG ATG AAA CCA TTG CG
5		Forward - GGT GAA GCT AAA AAA TAA AAC G
		Reverse - GCT CAC TGA AGG ATC GTC
6		Forward - GAA GAA CAA CGT AAC GGC
		Reverse - TGA TAA GAA TCA TCA AAA TGC

9.1.2 Methods

9.1.2.1 PCR amplification of the Sbi and SpA genes

The *S. aureus* strain collection used for sequencing was provided by Bernard Poutrel, INRA, Pathologie Infectieuse et Immunologie.

The PCR mix consisted of 5µl 10x Mg^{2+} free buffer, 5µl $MgCl_2$, 1µl 10mM dNTPs, 0.2µl Taq polymerase, 1µl of each of the two primers at 100pmol, 0.5µl of DNA template, and made up to a reaction volume of 50µl with SDW. A heating protocol of 5 minutes at 95°C followed by 34 cycles of 95°C for 30 seconds, 50°C for 30seconds (55°C for primer set 1 and 2), 72°C for 60 seconds, and finished with 10 minutes at 72°C.

Primers are shown in Table 16. Forward primer 4 and reverse primer 6 were used to amplify SpA, while all six SpA primers were used for sequencing it due to the repetitive nature of the gene. Sbi was amplified in three fragments with the 6 different primers. The presence of the correct fragments from the amplification reaction was checked with a 0.8% agarose-ethidium bromide gel, run for 15 to 20 minutes at 200 to

250 volts. Sbi reactions produce three fragments while the SpA reactions produced a single fragment.

The PCR products were purified by adding 50µl of 95% ethanol and 2µl sodium acetate to each sample, and frozen for one hour. The samples were centrifuged at 3500rpm for one hour to pellet the PCR products. The supernatant was removed by inverting the plates onto paper towel and centrifuged at 500rpm for one minute. The pellets were washed in 150µl of 70% ethanol and centrifuged at 3500rpm for 30 minutes. The supernatant was removed by inverting the plates onto paper towel and centrifuged at 750rpm for 1 minute. The samples were left to air dry for 10 minutes. The pellets were resuspended in 15µl of distilled water, vortexed for 2 to 3 minutes and these samples were spun at 1000rpm.

9.1.2.2 Sequencing reaction

1µl of primer (either forward or reverse) was added to 2µl of the purified PCR product and 2µl of Big dye (Version 2.0, Applied Biosystems). Each Sbi PCR product reaction required two reaction mixtures (one forward primer, and one reverse primer), while the SpA PCR products required six reaction mixtures (three forward primers, and three reverse primers). The resulting products were purified as described above, except they were not resuspended in water, but were sealed and stored at –20°C until they were loaded onto the ABI Prism 377 DNA sequencer.

9.1.2.3 Databases and programs

Seqmen program (DNASTAR, http://www.dnastar.com/web/r13.php) was used to trim and aligne the obtained sequences. Sequences were aligned using the ClustalX programme followed by manual inspection of variable sites.

MEGA (**M**olecular **E**volutionary **G**enetic **A**nalysis) version 2.1 software (Kumar S., Tamura K., Jakobsen IB., Nei M., Arizona State University, Tempe, 2001) was used to identify variable sites and calculate ds/dn ratios by the modified Nei and Gojobori method (Nei and Gojobori 1986) with Jukes Cantor correction. This program was also used to translate the sequences for use in protein modelling.

PAUP program (**P**hylogenetic **A**nalysis **U**sing **P**arismony) version 4.0b10 was used to construct Neighbour Joining Trees and maximum likehood trees (Swofford 2000).

Table 17 *S. aureus* strains from which the Sbi and SpA genes were sequenced. Those strains for which a gene was successfully sequenced are indicated (X).

		Sbi	**SpA**
8325	human	X	X
C101	human	X	X
D2-95.40	Cow	X	X
D4-121.24	Goat	X	X
D4-135.45	Cow	X	X
D8-452.11	Rabbit	X	X
D8-496.17	Rabbit	X	
D8-520.25	Pig	X	
D8-520.26	Pig	X	X
D8-522.05	Pig	X	X
D8-522.13	Poultry	X	X
D8-522.16	Poultry	X	
D8-522.18	Poultry	X	X
D8-526.12	Rabbit	X	X
D8-526.16	Pig	X	
D8-534.30	Cow	X	
D8-542.04	Ewe	X	
D8-610.13	Ewe	X	X
D8-660.22	Goat	X	
D9-786.04	Goat	X	X
D9-786.05	Ewe	X	X
D9-786.11	Ewe	X	
H783	human	X	X
MRSA252	human	X	X
MSSA476	human	X	X
MW2	human	X	X
N315	human	X	X
	Etoile cow	X	X
D8-524.04	Poultry		X
D8-526.03	Rabbit		X
D8-660.22	Goat		X
H560	human		X
H116	human		X

The Bellerophon server (http://foo.maths.uq.edu.au/~huber/bellerophon.pl) was used to identify chimeric sequences (Huber *et al* 2004).

Datamonkey (www.datamonkey.org) was used to detect any areas of gene sequence that are subjected to positive or negative selection (Kosakovsky, Pond and Frost 2005).

9.1.3 Results

9.1.3.1 Analysis of gene sequences of Sbi and SpA from *S. aureus* strains isolated from goat, cow, horse, rabbit, sheep, human and poultry

Sequences of Sbi were obtained from a collection of 23 strains, while 21 strains gave sequences for SpA. 5 additional sequences, available in sequence databases, were obtained from human *S. aureus* strains 8325, MW2, N315, MRSA252, and MSSA476. The strains used and the organisms from which they were isolated are listed in Table 17. All Sbi nucleotide sequences display conserved gene lengths of 1311 or 1314 nucleotides, while the SpA sequences showed much greater variability, with sequences between 1293 and 1624 nucleotides long.

9.1.3.1.1 Sbi

Sbi sequences show high levels of conservation in both nucleotide and protein sequence. A total number of 140 nucleotide changes were observed, giving rise to 60 amino acid changes and some synonymous (silent) substitutions. Interestingly, most of these changes are restricted to Sbi's C-terminal region. 111 nucleotide changes (47 amino acid changes) out of 140 in total (60 amino acid changes) were observed in this region. The number of substitutions in each region of Sbi are shown in Table 18. For the N-terminal part of the protein, the signal sequence of Sbi has a single residue change occurring in only one strain (S7X). The first Sbi Ig-binding domain has two residue changes, one occurring in a single strain (V48I), and the other (N96K) occurring in four of the analysed strains. For the second Sbi Ig-binding domain no substitutions were found, while the region 151 to 203 (to be referred to as domain 3) contains one substitution in a single strain (D163G). The β2GPI binding region has a number of substitutions, most of which occur only once (S218T, N222Q, E223K,

Table 18 The number of mutations at protein level in each of the Sbi domains, the length of each domain, and the number of mutations as a percentage of the domain residues.

Sbi domains	No. of mutations	No. of residues in domain	Number of mutations as a percentage of the domain residues
Signal peptide	1	45	2%
1st Ig-binding domain	2	55	4%
2nd Ig-binding domain	0	56	0%
Domain 3	1	49	2%
β2GP1-binding domain	8	55	15%
Proline rich region	2	41	5%
C-terminal	45	132	34%

K224S and Q249H), while others occur in four different strains (V244I, K245Q, and E246K). One substitution (E223V) occurs in eleven strains. The polyproline repeat has two substitutions: A284V occurs in three strains, while K288Q occurs only once. None of the analysed substitutions shown any pattern related to a specific host from which the strain was isolated.

9.1.3.1.2 SpA

As with Sbi, the most variable region of the SpA gene is located in the C-terminal part, but this variability is restricted to the first portion of the C-terminal region, comprised of the proline octa-repeat region. This region is characterised by the octapeptide repeat motif XKPGKEDZ (where X is G or N, while Z can be K or N). The number of repeats is highly variable, ranging from 4 to 16 in the strains sequenced. Other than the octapeptide repeat region, the C-terminal of SpA in the strains sequences is completely conserved (see Table 19 for the number of changes in each domain). There appears to be no relationship between the number of proline repeats and the species from which the *S. aureus* strain was isolated (Table 21). The SpA gene is well conserved in the Ig-binding domains, with between 9 and 23 nucleotide changes and 2 to 6 amino acid changes in any one domain. However, while the sequences of these domains are conserved, not all sequences contain all 5 Ig-binding domains. In several SpA sequences one domain was missing; Table 21 shows the Ig-binding domains found in each strain as well as the number of octapeptide

Table 19 The number of mutations at the protein level found in each of the SpA Ig-binding domains, the length of each domain, and the number of mutations are shown as a percentage of the domain residues.

Sbi domains	No. of mutations	No. of residues in domain	Number of mutations as a percentage of the domain residues
E	2	57	4%
D	5	61	8%
A	4	58	7%
B	5	58	9%
C	4	58	7%

repeats. Domain D is absent in strains D8-452.11, D8-526.03, and D8-610, while strains D2-95.40, D8-526.03 and H560 are missing domain B. Strain H116 is missing part of domain C.

Two strains (H116 and D8-610.13) contain a single nucleotide insertion 3' after the Ig-binding domains, resulting in a frameshift and subsequent mis-translation. Whether this is a genuine frameshift or a sequencing error needs to be clarified.

9.1.3.2 Synonymous and non-synonmous substitutions

Ds/dn ratios for the sequences were obtained. This is ratio between of the number of synonymous (silent) mutations and the number of nonsynonyous (changes in the DNA sequence that result changes in the amino acid sequence) mutations. This suggested that there is no selection pressure on either the whole genes of Sbi or SpA, or any of the domains contained within the genes (except for the β_2GP1 binding domain of Sbi for which a ds/dn ratio of 0.786 was found, as shown in Table 20).

Table 20 Ratio of synonymous and non-synonymous changes in each gene or domain. A ratio of less than one indicates positive selection, a value if 1 means there is no selection pressure.

Gene or Domain	**ds/dn**
Sbi	2.103
• B_2GP1 binding domain	0.786
• C-terminal	2.056
SpA	3.166
• Ig-binding domains	9.250

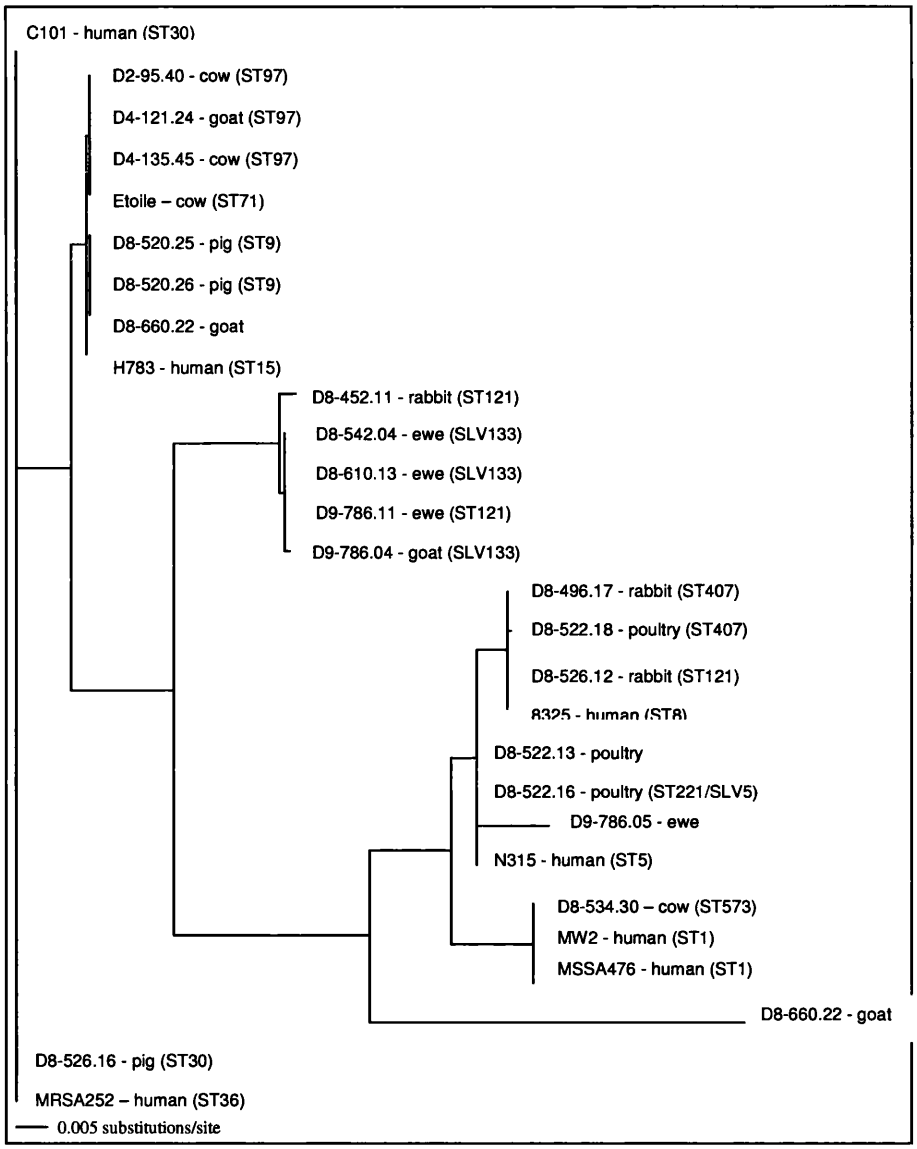
Table 21 The Ig-binding domains of SpA that are present in each strain (indicated with x), and the number of proline octapeptide repeats for each strain. H116 (&) contains part of domain C.

		SpA Ig-binding domains present					No. of Proline repeats
		E	D	A	B	C	
C101	Human	x	x	x	x	x	11
D2-95.40	Cow	x	x	x		x	12
D4-121.24	Goat	x	x	x	x	x	4
D4-135.45	Cow	x	x	x	x	x	12
D8-452.11	Rabbit	x		x	x	x	10
D8-520.26	Pig	x	x	x	x	x	10
D8-522.05	Pig	x	x	x	x	x	12
D8-522.13	Poultry	x	x	x	x	x	12
D8-522.18	Poultry	x	x	x		x	10
D8-526.03	Rabbit	x		x	x	x	8
D8-660.22	Goat	x	x	x	x	x	4
D9-786.04	Goat	x	x	x	x	x	10
D9-786.05	Ewe	x	x	x	x	x	10
	Etoile cow	x	x	x	x	x	4
H116	Human	x	x	x	x	&	7
H783	human	x	x	x	x	x	13
D8-524.04	Poultry	x	x	x	x	x	13
D8-526.12	Rabbit	x	x	x	x	x	12
D8-610.13	Ewe	x		x	x	x	16
H560	Human	x	x	x		x	12
N315	Human	x	x	x	x	x	12
MW2	Human	x	x	x	x	x	10
MSSA476	Human	x	x	x	x	x	10
MRSA252	Human	x	x	x	x	x	13
8325	Human	x	x	x	x	x	13

9.1.3.3 Phylogenetic trees

Phylogenetic trees were generated to show the evolutionary relationship between the SpA and Sbi genes in the different strains sequenced. Two types of tree were generated, a maximum likehood tree and a neighbour joining tree. Trees generated for both the Sbi and SpA genes show that the sequences group to sequence type of the *S. aureus* strain; there is no apparent pattern relationship with the host from which the bacterium was isolated. The maximum likehood tree of Sbi is shown in Figure 44 and the neighbour joining tree of SpA is shown in , and the sequence types are also indicated.

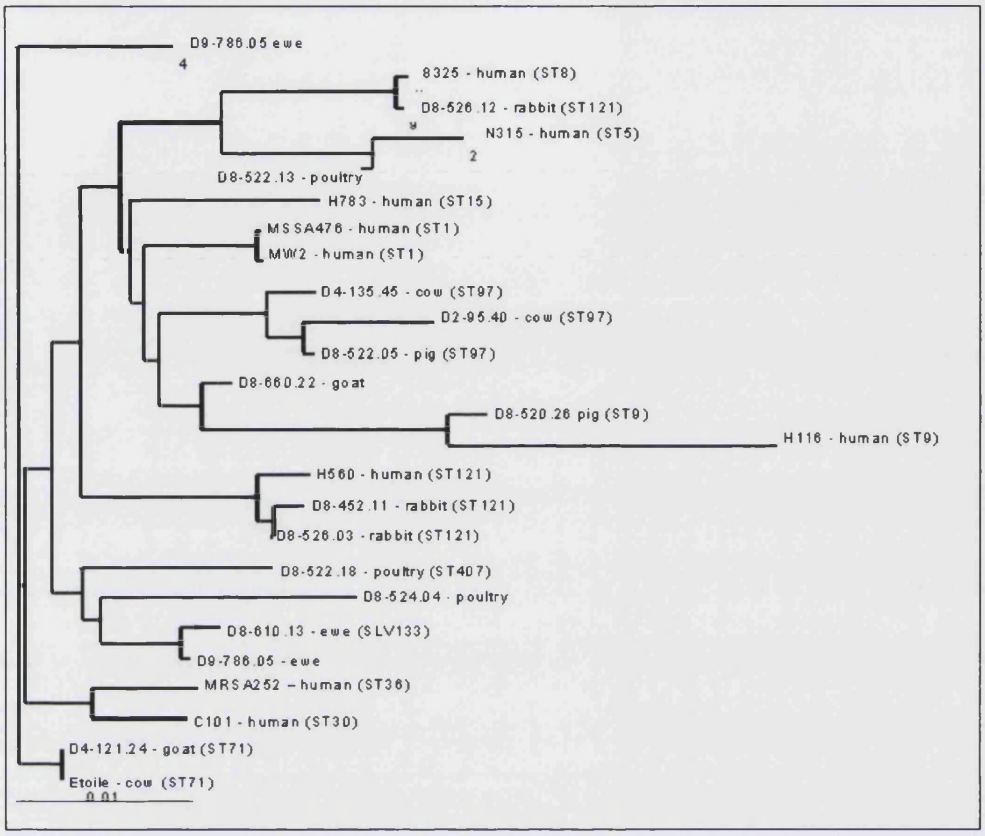
Figure 44 Maximum likehood tree of Sbi, generated by PAUP program (Swofford 2000).



9.1.3.4 Functional implications of the amino acid substitutions in the Ig-binding regions of the SpA and Sbi genes

Consensus sequences were generated for each of the SpA Ig-binding domains, and used to generate an alignment of the domains. This shows that there are amino acid substitutions at 15 positions in these domains out of 58 residues. The alignments are shown in Figure 46. Also marked are the positions of the three helices as they occur in the NMR structure of SpA domain B (PDB code: 1BDD). This alignment indicates that few of the changes found occur in the first two helices in the domain.

Figure 45 Neighbour joining tree of SpA, generated by PAUP program (Swofford 2000).



Those that do lie in this region, occur either in the loop regions or at the ends of the α-helices, substitutions N11H, M12L (both domain A) and T16N (domain C) are both predicted to be part of the loop region between helices 1 and 2 (see Figure 46) while E17D (domain C) and E18A (domain D) occur at the N-terminus of helix 2. More importantly, only one change occurs in the residues identified to be involved in

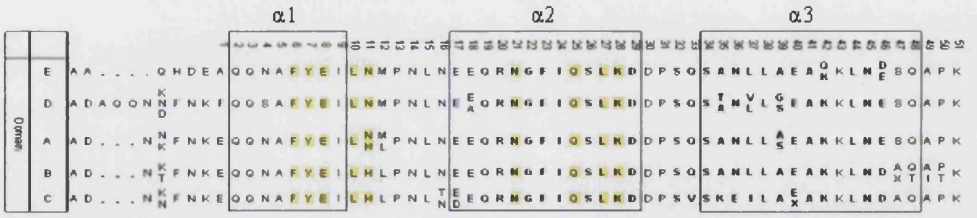


Figure 46 Alignment of SpA Ig-binding domains with the substitutions that occur in the different strains highlighted. The positions of the helices are marked.

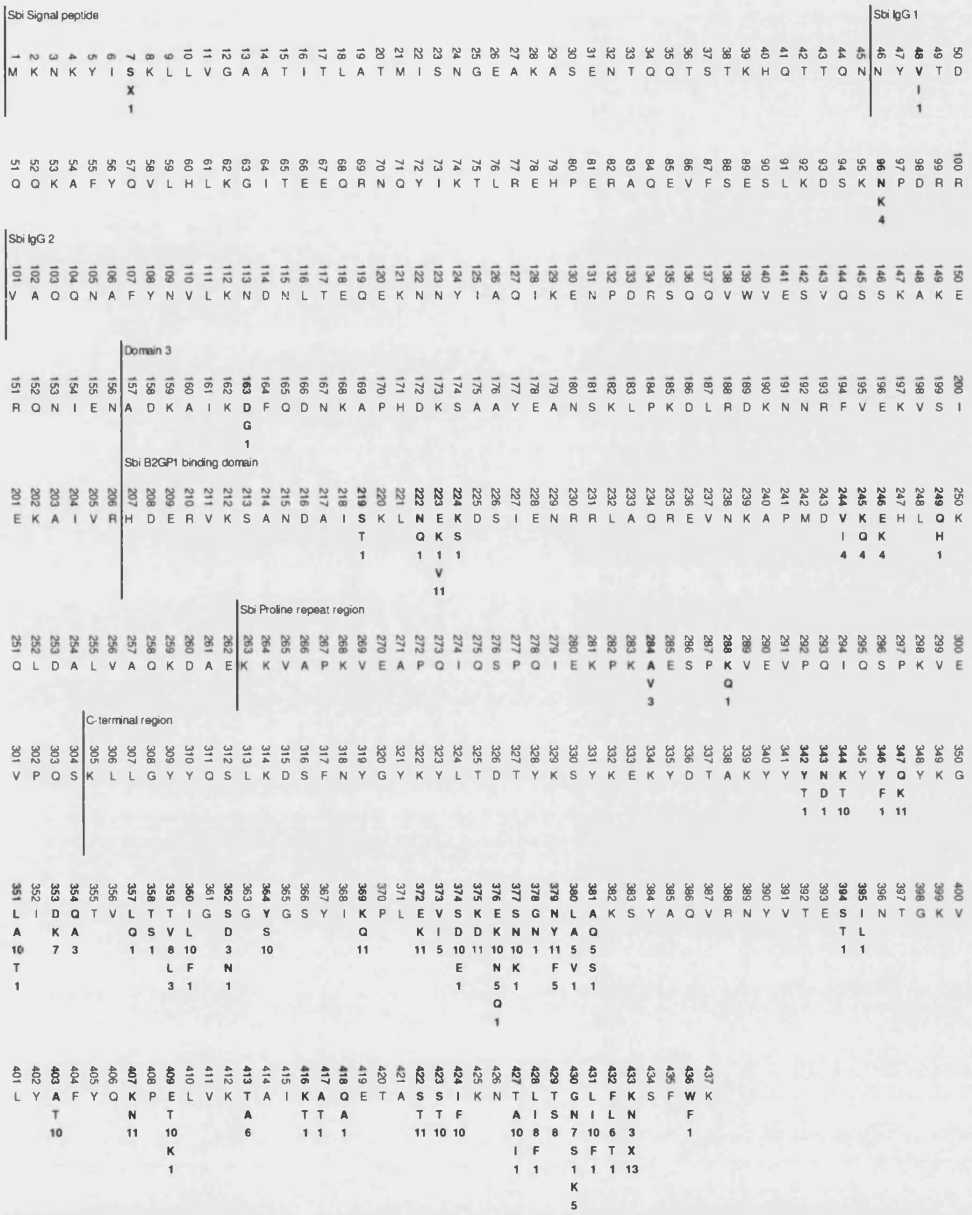


Figure 47 The sequence of Sbi with the substitutions at each position shown with the number of strains in which that that substitution occurs.

Fc-binding (as identified by Gouda et al 1998); this is in the A domain at residue N11H, and this substitution results in the presence of an histidine which occurs in domains B and C at this position. The alignment shows that the third helix is much more variable in its sequence. As a number of these changes occur in the residues involved in the binding of Fab (Meininger et al 2000), this could imply that they are

perhaps important in determining the Fab V_H chain specificity in different host species.

Compared to SpA, the N-terminal region of the Sbi sequences is much more conserved at the protein level. There are only 14 changes in the first 304 residues (including the signal peptide), none of these changes occur in more than 4 strains (except at residue 223). The C-terminal region, in contrast, is much more variable. In the region from residues 305 to 437, there are 45 changes, some of which are observed in over 10 different strains. A portion of these changes are clustered (see Figure 47), perhaps indicating the presence of loop regions in the protein's structure?

9.1.4 Discussion

Analysis has shown differences in the binding of Sbi and SpA to serum IgGs from different species. The analysis of the sequences of the IgGs involved indicates that slight differences in these regions could play a role in the binding of IgG to its ligand. All studies, however, were carried out using SpA and Sbi genes cloned from *S. aureus* strains isolated from humans. It could therefore be that strains residing in other animals differ in the sequence of their Ig-binding proteins in order to improve the ability of the bacteria to specifically bind the host's immune proteins.

Through the sequencing of the Ig-binding proteins SpA and Sbi from *S. aureus* strains isolated in different animals, the variation in sequence of these genes was examined. The results show that the variation in the Ig-binding regions in terms of substitutions at the genetic level and substitutions that appear in the protein are marginal. Those substitutions that do occur cannot easily explain the variation observed in the binding levels of IgG, especially in the binding of the Fc fragment and how it is overcome in different species through changes in the sequence of the Ig-binding proteins to produce higher affinity for the host's serum IgGs.

The variation observed in SpA cannot fully explain the variation in IgG binding via the Fc region. Of those residues in SpA identified to be involved in binding Fc by NMR (Gouda et al 1998), only the residue N11 is substituted in any of the SpA Ig-binding domains. This substitution is at a position that shows variation between the

SpA domains (Figure 46). As all SpA domains bind IgG via its Fc domain (Jansson, Uhlen, Nygren 1998), this change is unlikely to greatly affect its Fc binding ability.

What is of possible interest is the variation that does occur in SpA's Fab-binding sequences. Most of the variation observed in the Ig-binding domains is in the protein sequence for the third helix (Figure 46). From structural analysis, it is known that helices 2 and 3 are responsible for Fab binding (Graille *et al* 2000, Meininger *et al* 2000). As there is some variation in this region, it is possible that it does have some effect on Fab binding, however some of these changes normally occur in the other domains at the corresponding positions. It is therefore difficult to know exactly what, if any, effect these changes will have on functionality without assessing the effects of the substitutions in direct binding studies.

The only other region in SpA which showed variation was the number of proline repeats in the octapeptide repeat region. This region of SpA has been implicated in cell-wall binding (Zhang *et al* 2000) and has been suggested as a marker for phenotyping *S. aureus* strains (Koreen *et al* 2004), however how the variation of the number of repeats affects the cell-wall binding is unknown. The C-terminal region of SpA after the octapeptide repeat is conserved in all strains sequenced.

Sbi is generally much more conserved, with little variation in sequence in most of the protein (Figure 47). However, like SpA, Sbi also shows variation in the C-terminal region. The first 36 residues after the polyproline region, like the rest of the protein, are highly conserved. This region contains a large number of tyrosines. As mentioned before, tyrosines are known to be sites of phosphorylation (Zhang et al 2000), and this region has been suggested to play a role in regulating the expression of the protein.

It is the most C-terminal part of the protein that shows the most variability, in the last 105 residues of the protein, there are 47 amino acids substitutions. However this variability is not spread over the entire sequence, it tends to cluster into regions. One hypothesis that can explain this clustering is the presence of loop regions in the proteins' tertiary structure. One of these regions of high variability is near the end of the polypeptide (Figure 47), a region that is often flexible in proteins.

The sequence alignments of SpA and Sbi (Chapter 2) showed some homology in the N-terminal region of Sbi up to the polyproline repeat region. The highest level of sequence similarity is found in the Ig-binding domains, but in both the β_2GP1-binding region and the region flanked by the β_2GP1-binding sequence and the second Ig-binding domain (domain 3) we have also found similarity with SpA (Figure 12). The main difference in the alignment is that SpA has five N-terminal Ig-binding domains, while Sbi can be divided into four N-terminal domains. The similarities in sequence of the SpA and Sbi Ig-binding domains suggest that these could have evolved from a gene duplication event. We have found in this chapter that several of the SpA genes only have four N-terminal domains. Is it possible that SpA and Sbi have evolved from a common ancestor gene, that all the domains in Sbi could once have resembled the sequence of the SpA domains more closely, and that in some SpA strains, they have gained an extra domain, or the Sbi ancestor gene and some SpA genes have lost a domain. If the two genes have evolved from a common ancestor, it is likely the structure in this region is conserved. We have been able to predict the structure of the Sbi Ig-binding domains with a high degree of confidence, but not the other domains (chapter 2). To confirm our prediction a structural analysis of these domains is needed.

10 Chapter 6

10.1 Analysis of the Sbi domain structure by small-angle X-ray scattering

10.1.1 Introduction

There is evidence to suggest that the Sbi and SpA genes are related. The N-terminal immunoglobulin-binding domains of both show high levels of sequence similarity (chapter 1) and there are similarities in sequence for the region including the β2GP1-binding domain of Sbi with SpA's Ig-binding domains. Sequence analysis of the SpA gene revealed that in several strains this protein consists of only 4 Ig-binding domains (chapter 5). We have proposed (chapter 2) that Sbi consists of 4 domains, the 2 Ig-binding domains, the β2GP1-binding domain, and a putative third domain. Since this region of the Sbi gene is also of the same length as the 4 SpA Ig-binding domains, could it be that the structure of this portion of the two proteins is conserved as well? Our structural prediction for the N-terminal domains of Sbi is that all 4 domains adopt a three-helix bundle fold, as is predicted for SpA. This prediction is based on the three domains of SpA, for which three-dimensional structures have been determined by both NMR and X-ray crystallography (SpA domains B, D, and E; SCOP database http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.b.bd.b.b.A.html, Murzin *et al* 1995).

Although no structure exists of multiple Ig-binding domain constructs of SpA, it is assumed that the molecule is composed of an elongated chain of four (or five) three-helix bundles linked together (chapter 4). To investigate whether Sbi adopts a similar domain arrangement we attempted to determine the three-dimensional structure of Sbi's N-terminal region comprising the Ig-binding domains and the β2GP1-binding domain. Incessant attempts to crystallise Sbi-e and Sbi-S failed, likely due to the multi-domain structure and flexibility of these polypeptides. Structural analysis of these constructs by NMR proved unsuccessful as the large Sbi-e construct is close to the upper size-limit for this technique. The tendency of the smaller Sbi constructs to form multimers at high protein concentrations was also made them unsuitable for this method. Therefore, we opted to use small-angle x-ray scattering (SAXS).

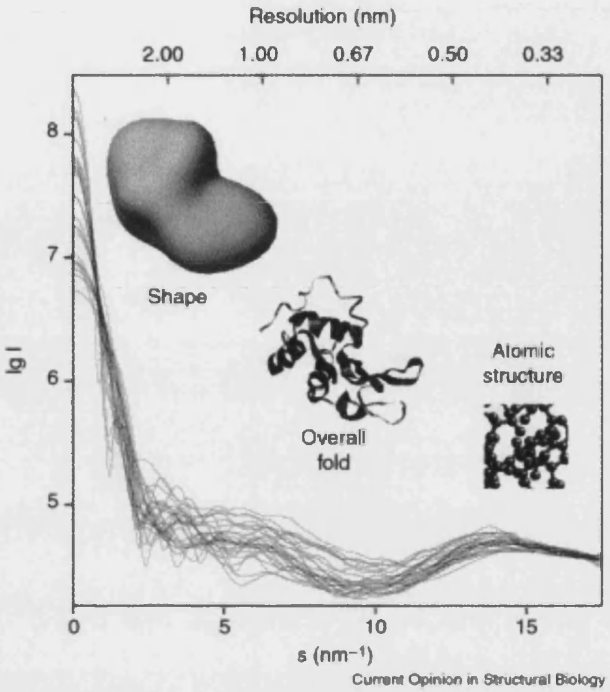


Figure 48 shows X-ray solution scattering curves for 25 different proteins. The scattering intensities are plotted on a log scale, the upper axis displays the spatial resolution ($\Delta=2\pi/s$). At low resolution angles of 2 to 3nm, the rapidly decaying curves are determined by the particle shape. At medium resolution (2 to 0.5nm), the curve is defined by the overall fold of the protein, while at low resolution (above 0.5nm) the curves are all simaler, and defined by the molecule's atomic structure. SAXS is only suitable for determining information on molecule shape, and quaternary and tertiary structure (Figure taken from Svergun and Koch 2002)

This technique, like X-ray crystallography is suitable for proteins with a high molecular weight, and like NMR, it is a solution technique and does not require the production of crystals. SAXS has been shown to be well suited to studying flexible, low compactness or even unfolded macromolecules in solution. In SAXS, a solution of macromolecules is exposed to X-rays, and the scattered intensity is measured. The random positions and orientations of particles result in an isotopic intensity distribution, which, for monodisperse solutions of noninteracting particles, is proportional to the scattering from a single particle over all orientations. At low resolution, the scattering curves of proteins with different structures are very different, determined by the shape of the particle. As the resolution increases, the curves become more similar. SAXS thus provides information on the shape, and quaternary and tertiary structure, but it is not suitable for the analysis of atomic detail (Svergun and Koch 2002). Refer to Figure 48 for an example of SAXS scattering curves.

Our aim with this technique is to gain such structure information that we can then fit the Sbi models to, to improve the reliability of these.

10.1.2 Methods

Expression of the Sbi constructs Sbi-e, Sbi-S and Sbi-1 was described in Chapter 3, as was the purification of Sbi-e and Sbi-1. Methods for SE chromatography were also described in Chapter 3.

10.1.2.1 Affinity purification of Sbi-S

Sbi-S was purified by affinity chromatography with a chelating HiTrap column (Amersham Biosciences) connected to the AKTA purification system. The column was loaded according to the column instructions with nickel sulphate and equilibrated in binding buffer (10mM Tris-HCl pH8, 300mM NaCl, 0.005% Tween 20). The sample was loaded with the soluble protein fraction and washed. The protein was eluted with elution buffer (10mM Tris-HCl pH8, 300mM NaCl, 0.005% Tween 20, 1M Imidazole) on a 0% to 15% gradient over 60min at a flow rate of 0.5ml/min. This gave pure protein (>90%), eluted at about 15% imidazole. The imidazole was removed using a PD-10 column (Amersham Biosciences) to leave the proteins in binding buffer.

10.1.2.2 Trypsin digests

Sbi-e was prepared in a solution of 2mg/ml as determined by UV absorption (see below) in 10mM Tris-HCl pH7 buffer. 1µl of prepared trypsin was added (10mg/ml in PBS prepared from a dry protein power – Sigma). The reaction was carried out on ice and samples removed to determine the apparent level of degradation at different time points. The digestions were analysed by electron spray mass spectrometry by Andrew Gill at the Institute for Animal Health, Compton.

10.1.2.3 TEV digest

A 2ml fraction of purified protein was collected from the affinity column elute for both Sbi-S and Sbi-1. To this 40U of TEV protease (Invitrogen) was added, and the reaction was allowed to occur overnight at room temperature. The digested protein was analysed by light scattering to examine protein molecule size.

10.1.2.4 Dynamic light scattering experiments

Dynamic light scattering measures fluctuations about the average intensity of light scattered by a protein solution of defined concentration in excess of that scattered by background sources. The scattered intensity fluctuations are primarily due to the random diffusive motion of the protein molecules into and out of a focused laser beam. (Wilson 2003) The size of particles in solution can be measured. This is because globular proteins are assumed to be spherical in shape. Therefore the Stokes-Einstein relation can be used to estimate the apparent hydrodynamic radius of the protein. From this, the apparent size of the protein can be estimated (Wilson 2003).

In light scattering and SAXS, it is important to have a monodisperse protein solution, this occurs when all molecules in a system are of a uniform molecular weight (http://composite.about.com/library/glossary/m/bldef-m3514.htm). Polydispersity occurs when the size of molecules in a system is not uniform (http://composite.about.com/library/glossary/p/bldef-p4086.htm).

The Sbi constructs were measured by light scattering immediately after purification to determine the effect of different components in the buffers, and to determine whether the oligomerisation of the Sbi proteins had been reduced. Proteins Sbi-e, Sbi-S, Sbi-1 and Sbi-2 were analysed in the purification buffer eluted from the column (containing Imidazole), and after a PD-10 de-salting step to exchange and remove certain components of the buffer (i.e. Imidazole). The samples analysed did not have their concentration determined. 0.5M EDTA was added to some samples to absorb any nickel released from the chelating column. Sbi-S and Sbi-1 were also subjected to TEV digestion, to remove the His-tag. The dynamic light scattering apparatus machine used was the Malvern Zetasizer (Nano S, Malvern Instruments, Malvern, UK), and all experiments carried out measured the sizes of the particles present and the estimated protein size in kDa by the crystallography screen function.

Of the data obtained for the Sbi constructs by this method, the estimated size of the molecule is larger in all cases than the theoretical size of the constructs. This can be partly explained by errors caused by the assumptions about the spherical shape of the proteins, as we know Sbi-e in particular is expected to be an elongated molecule. This can lead to errors in the results of the measurements of the molecules in solution,

however as all techniques show variability and differences to theoretical molecular sizes, it suggests that the errors that occur in light scattering data is of limited importance in this case.

10.1.2.5 Preparation of samples for small-angle X-ray scattering (SAXS)

Samples for SAXS were in 10mM Tris-HCl pH8, 300mM NaCl, 0.005% Tween 20, and stored at the concentration they were in after buffer exchange. Samples were only concentrated shortly prior to the SAXS experiment to reduce aggregation and precipitation. Protein concentrations were determined by UV absorption. The extinction coefficients at A_{280nm} are as follows: Sbi-e 20400 (Abs 0.1% (=1 g/l) 0.6), Sbi-S 18910 (Abs 0.1% (=1 g/l) 1.031), Sbi-1 10430 (Abs 0.1% (=1 g/l) 1.095).

10.1.2.6 SAXS experiments and data analysis

Synchrotron radiation X-ray scattering data was collected on the X33 beam line (Boulin *et al.*, 1988; Koch and Bordas, 1983) at the Hamburg EMBL Outstation (on the DORIS III storage ring, at DESY). Solutions of Sbi-e were measured on a MAR345 image plate detector at protein concentrations of 2.8, 8.4 and 16.7 mg/ml and sample-detector distance of 2.4m and wavelength λ = 1.5Å, covering the momentum transfer range $0.013 < s < 0.45$ Å^{-1} ($s = 4\pi \sin(\theta)/\lambda$ where 2θ is the scattering angle). To check for radiation damage, two 2-minute exposures were compared; no radiation effects were observed. The data was averaged after normalization to the intensity of the incident beam, corrected for the detector response, and the scattering of the buffer was subtracted. The difference data was extrapolated to zero solute concentration following standard procedures, and, for the gas detector, the curves measured in different angular intervals were merged. All data manipulations were performed using the program package PRIMUS (Konarev *et al.*, 2003).

The forward scattering I(0) and the radius of gyration R_g were evaluated using the Guinier approximation (Guinier, 1939) assuming that at very small angles ($s < 1.3/R_g$) the intensity is represented as $I(s) = I(0) \exp(-(sR_g)^2/3)$. These parameters were also computed from the entire scattering patterns using the indirect transform package GNOM (Svergun, 1992), which also provided the maximum particle dimension D_{max}.

The molecular mass (MW) of the solute was evaluated by comparison of the forward scattering with that from reference solutions of bovine serum albumin (MW = 66 kDa). The excluded volume of the hydrated particle (the Porod volume) was computed using the equation (Porod, 1982):

$$V = 2\pi^2 I(0) / \int_0^\infty s^2 I_{exp}(s) ds \qquad (1)$$

Prior to this analysis an appropriate constant was subtracted from each data point to force the s^{-4} decay of the intensity at higher angles following the Porod's law (Porod, 1982) for homogeneous particles. This "shape scattering" curve was further used to generate low resolution *ab initio* models of Sbi-e using the program DAMMIN, (Svergun, 1999) which represents the protein by an assembly of densely packed beads. Simulated annealing was employed to build a compact interconnected configuration of beads inside a sphere with the diameter D_{max} that fits the experimental data $I_{exp}(s)$ minimizing the discrepancy (χ^2 in SAXS indicates how well the scattering data fits the structure solution, a figure for χ^2 of around 1 indicates a correct solution):

$$\chi^2 = \frac{1}{N-1} \sum_j \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)} \right]^2 \qquad (2)$$

where N is the number of experimental points, c is a scaling factor and $I_{calc}(s_j)$ and $\sigma(s_j)$ are the calculated intensity and the experimental error at the momentum transfer s_j, respectively.

The refinement of the Sbi structure was performed with the program BUNCH (Petoukhov and Svergun, 2005). This algorithm uses three-dimensional structures or homology models for the folded domains, and adds interdomain linkers and terminal loops represented as dummy residues chains (Petoukhov *et al.*, 2002). A simulated annealing algorithm implemented in BUNCH optimises an energy function containing the discrepancy (2) to the experimental SAXS profile and force field terms to avoid steric clashes and ensure native-like conformations for the added chains.

The possibility of interdomain dynamics in Sbi-e was further explored using a new program based on a genetic algorithm. The experimental scattering profile is fitted by a combination of scattering patterns computed from an ensemble of conformations

adopted by potentially flexible protein in solution. To generate a representative ensemble, 5,000 conformers of Sbi-e were constructed using the program BUNCH without using the experimental data in the energy term. In addition, 150 conformations were generated using the initial part of the SAXS profile in order to have certain population of structures with the R_g close to the experimental one. For each of the 5,150 conformers the theoretical scattering curve was calculated with the program CRYSOL (Svergun *et al.*, 1995). The genetic algorithm aims to select an ensemble of K models such the scattering pattern from their equimolar mixture that best fits the experimental data (Bernadó, Petoukhov, Mylonas, Blackledge & Svergun, in preparation). The program was run for different K values to identify the minimum number of conformers in the sub-population still compatible with the experimental data. It was found that the runs with $K \geq 5$ yielded a good agreement with experimental SAXS curve.

10.1.2.7 Sbi homology modelling

In Chapter 2, homology models were generated of both the Ig-binding domains and the β2GP1-binding domain (see Figure 13 and Figure 15). The Ig-binding region of Sbi (composed of two three-helix bundles) and the β2GP1-binding domain (also

Table 22 Sizes of the Sbi constructs, determined by SE chromatography.

Ig-binding protein	Peak Elution volume (ml)	Protein Size (kDa)
SpA	11.756	177
	9.999	433
Sbi-e	10.401	353
	12.014	156
	13.224	84
	14.324	48
	14.324	48
	16.67	15
Sbi-S	14.504	44
	15.163	31
Sbi-1	10.844	282
	15.457	27
	17.69	8.7
	18.313	6.4

composed of two three-helix bundles) were linked together in O. This generated a model of the N-terminal region of the Sbi protein. This could be used in the refinement of the SAXS data.

10.1.3 Results

10.1.3.1 Size exclusion analysis of Sbi and SpA

Three Sbi constructs were analysed by SE chromatography: Sbi-e, Sbi-S, and Sbi-1, in addition to the commercial SpA. Theoretical masses for each are 30.9kDa (Sbi-e), 18.3kDa (Sbi-S), 9.7kDa (Sbi-1) and 46.7kDa (SpA). However the measured size of each protein in the SE chromatography analysis did not agree with these theoretical sizes. In solution, all proteins constructs behaved as particles with larger masses (Table 22), suggesting that all exist as multimers in solution. The SE chromatography of Sbi-e shows the different profiles obtained with different protein concentrations, suggesting that the size of the Sbi-e particle increases as a function of the protein concentration (see Figure 49).

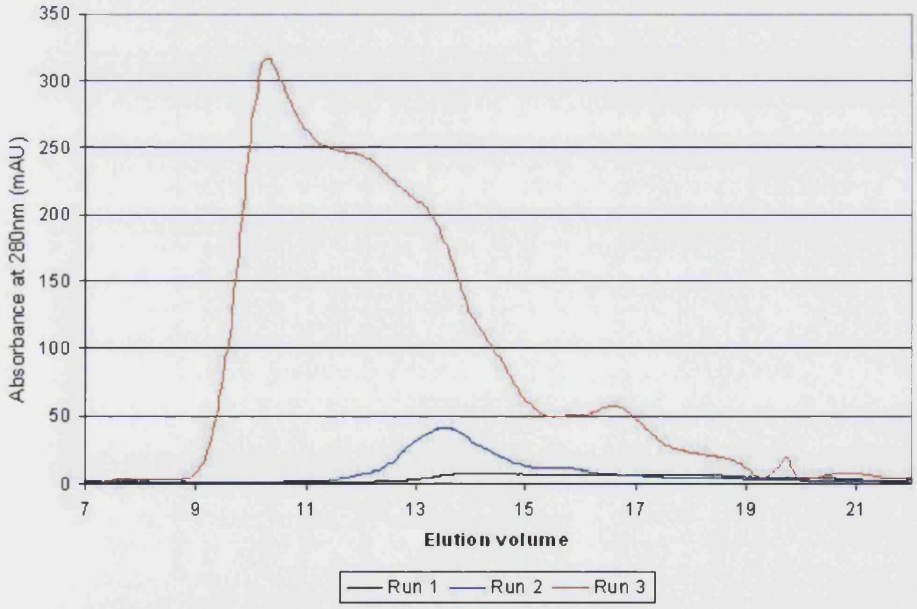


Figure 49 SE chromatography profiles of Sbi-e, with undetermined start concentrations for Sbi-e

Table 23 Sizes of Sbi constructs, determined by Light Scattering measurements, including the polydispersity index (PSI).

Sbi construct	Diam. (nm)	PSI	Est. MW	
Sbi-e	8.63	0.592	351.5	
Sbi-S	5.67	0.651	48.2	TEV digested de-salted[&]
Sbi-1	3.68	0.302	18.4	TEV digested
Sbi-2	4.32	0.491	21.1	TEV digested de-salted[&]

10.1.3.2 Dynamic light scattering analysis of Sbi

The dynamic light scattering properties of the Sbi constructs were analysed by determining which was the most populous peak in the scattering data, using intensity polydispersity (PSM) measurements. The peak nearest to the most populous peak in size in the crystallisation screen analysis (volume PSM) was than taken to give an estimated size of the component of the peak. This data is shown in Table 23. Comparisons of the estimated MW to the theoretical MW of the constructs show that all the protein sizes are much greater than would be expected for the Sbi protein monomers. Sbi-S and Sbi-1 (TEV digested) appear to exist as a dimer, while Sbi-2 behaves as a tetramer. Sbi-e and Sbi-1 (EDTA present) in solution form multimers of approximately ten times the expected size. As all samples were filtered prior to the experiment, it is unlikely to be due to aggregation or dust in the samples.

10.1.3.3 Molecular size determined by small-angle X-ray scattering

Small-angle scattering was used to measure the size and effective molecular weight of the proteins (Table 24). Sbi-e was the only construct to be measured as a monomer.

Protein	Concentration (mg/ml)	R_g (nm)	D_{max} (nm)	Effective MW (kDa)	Theoretical MW (kDa)
Sbi-1	0.6	5	16	90	9.7
Sbi-S	2.4	5.6	17	100	18.3
	0.28	3.0	17	28	
Sbi-e	8.4	4.5	15	36	30.9

Table 24 The sizes of Sbi constructs determined by SAXS. Protein concentrations were determined at 280nm (extinction coefficients are given in methods). Rg is the radius of gyration, and Dmax is the maximum paritcule diameter

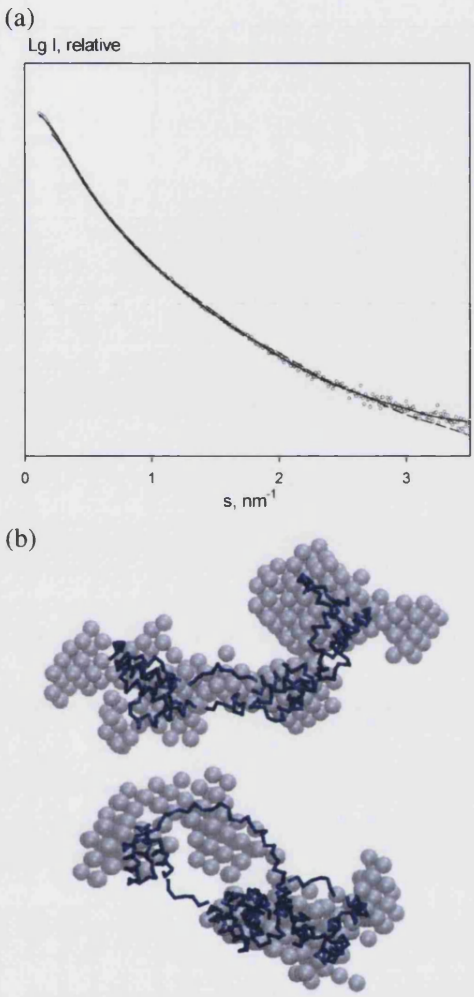


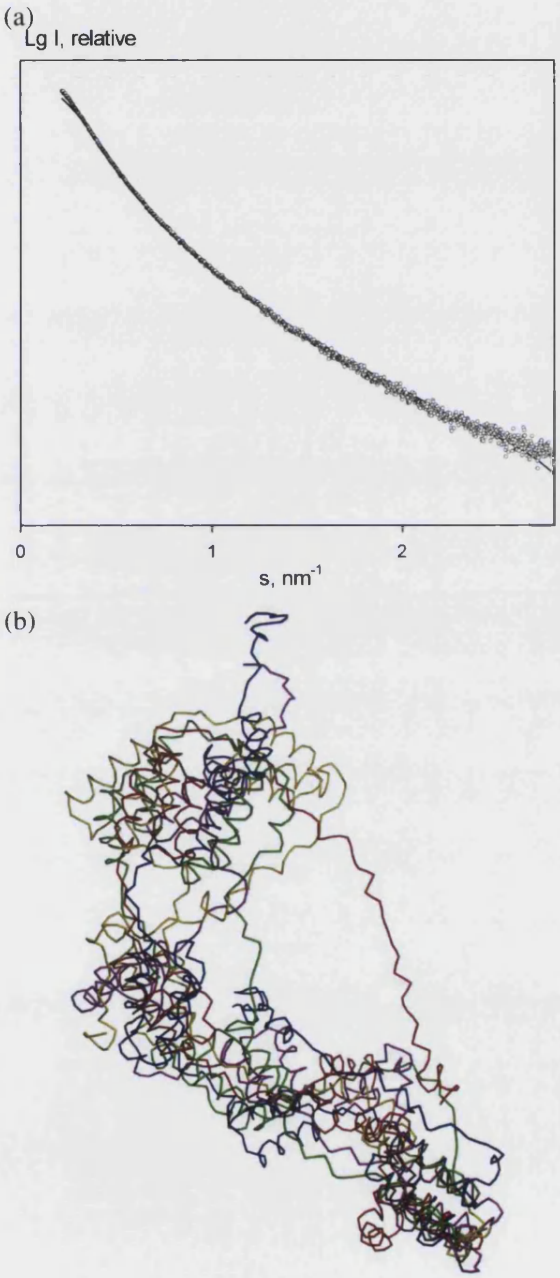
Figure 50 (a) Experimental scattering curve of Sbi-e (open circles) and the scattering from the model obtained with DAMMIN (solid line) and BUNCH (dashed line). The plot displays the logarithm of the scattering intensity as a function of momentum transfer $s = 4\pi \sin(\theta)/\lambda$ where θ is the scattering angle and $\lambda = 0.15$ nm is the X-ray wavelength. **(b)** Two different orientations of the structural models of Sbi-e. Beads show the *ab initio* shape obtained by DAMMIN, which is superposed to the trace representation of a BUNCH structural model (in blue). The bottom view is rotated by 90 degrees counterclockwise around the vertical axis.

Sbi-S was measured at two concentrations; at the lower concentration, Sbi-S appears to dimerise while at the higher concentration it appears to form a pentamer. Sbi-1 was measured at a single concentration, and its complex appears to contain nine monomers.

10.1.3.4 SAXS analysis of Sbi-e

The X-ray scattering curve in Figure 50a yields a MW estimate of 33±3kDa indicating that the protein is monomeric in solution. The experimental radius of gyration R_g and the maximum size, D_{max}, were 46±1Å and 160±10Å, respectively. These results indicate that Sbi is a highly elongated molecule in solution. The low resolution shape of Sbi reconstructed *ab initio* using the bead modelling program DAMMIN (Svergun, 1999) neatly fits the experimental data with χ=1.2 (Figure 50a). A typical

Figure 51 (a) Experimental scattering curve of Sbi-e (open dots) and the scattering obtained using the genetic algorithm protocol (solid line). **(b)** Backbone trace representation of the five conformers ensemble obtained with the genetic algorithm.



reconstructed model displayed in Figure 50b is indeed very elongated (155 Å diameter) and depicts four bead domains connected by thinner loops. This appearance is compatible with four structural domains of Sbi joined by flexible linkers. The high resolution and homology models of the N-terminal domains of Sbi were employed to further model the structure of Sbi using the program BUNCH (Petoukhov and Svergun, 2005) as described in the methods. The model was represented using the folded domains connected by dummy residue linkers, and additional chains were introduced to model the N and C-terminal tails. Different runs of BUNCH yielded reproducible results displaying similar domain organizations and fitting the experimental data with χ=1.7 (Figure 50a). A typical BUNCH result superimposed with

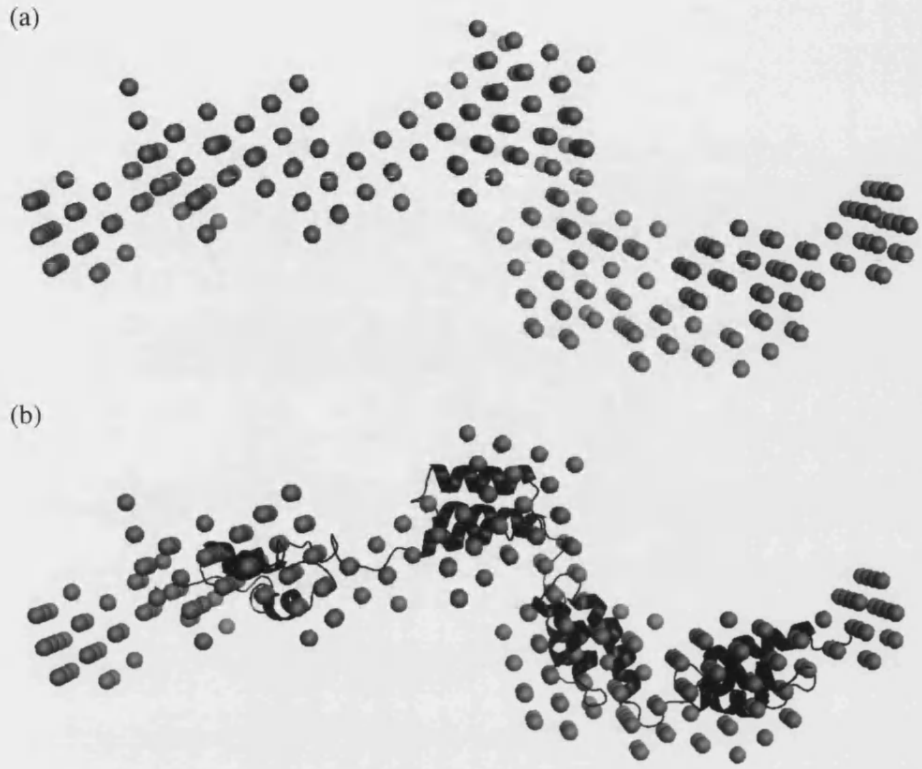


Figure 52 (a) Density obtained from SAXS density and **(b)** with the homology model fitted.

the *ab initio* shape in Figure 50b reveals a very good overlap between the two independently derived models from different programs. The overall shape of Sbi is rather extended but somewhat bent so that the C-terminal domains of the BUNCH model display some side contacts. Interestingly, the long N-terminal loop of Sbi, although extended, appears to be folded back and becomes not completely accessible to the solvent (this feature is typical for BUNCH reconstructions).

Multidomain proteins often reveal interdomain dynamics depending on the length of the linkers connecting the folded entities (Yuan *et al.*, 2002). To account for the possible dynamic effects, an alternative interpretation was used allowing for the data fitting by a mixture of conformers. Using the genetic algorithm as described in the methods, it was found that selection of a five conformer ensemble from a representative set of randomly generated structures was sufficient to fit the experimental data. A typical set of structures provided by the genetic algorithm

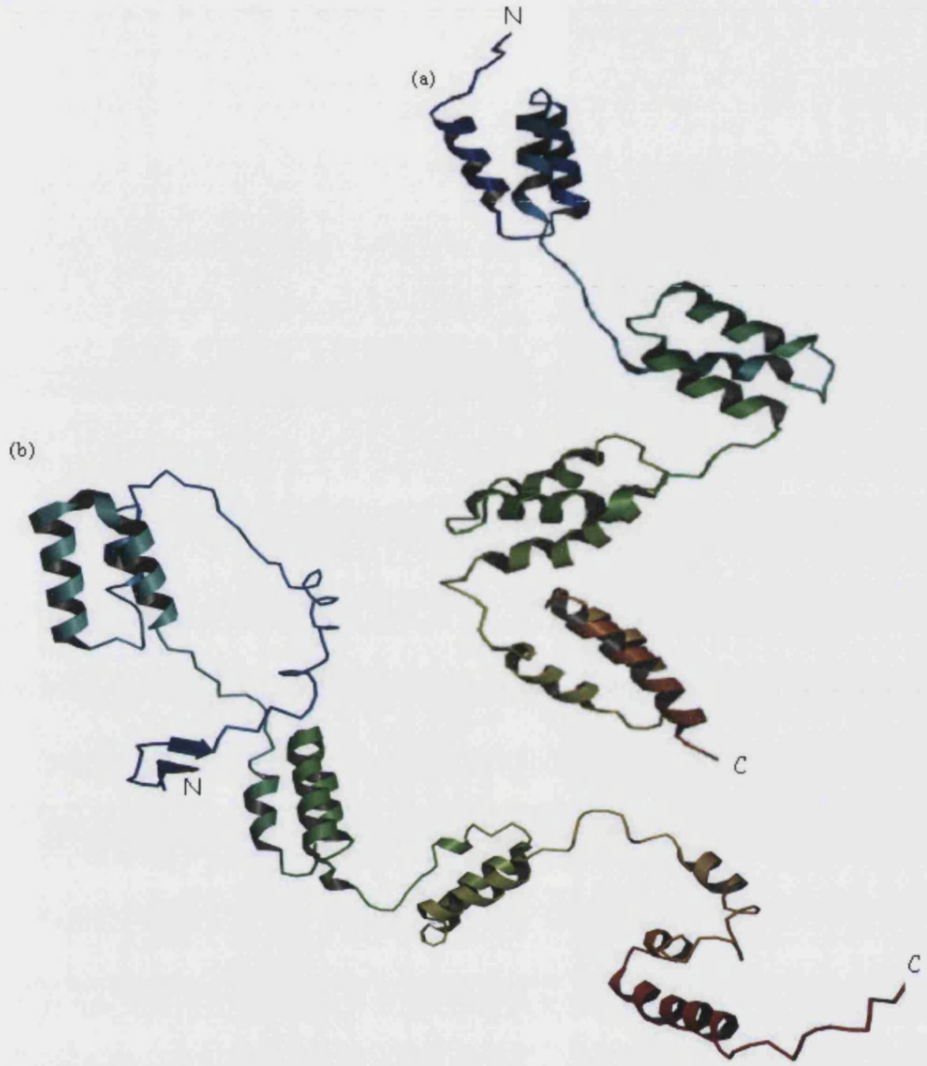


Figure 53 The 4-domain density structure of Sbi **(a)** by homology modelling and **(b)** the most probable structure predicted by SAXS. (b) contains longer N- and C-terminal peptides as DAMMIN is able to predict the structure of these long flexible peptides, such as the His-tag of Sbi-e, which is not possible to do by homology modelling without a suitable template to build the structure on. (a) therefore lacks the N-terminal His-tag and has a truncated C-terminal.

displayed in Figure 51a yields the fit with χ=2.0. Interestingly, the individual structures in the set appear rather similar, suggesting that only limited conformational space is sampled by Sbi in solution. Simultaneously, the results point to a larger conformational freedom in the C-terminal domain of Sbi compared with the rest of the protein. The ensemble of structures in Figure 51b has an elongated bend shape, which agrees well

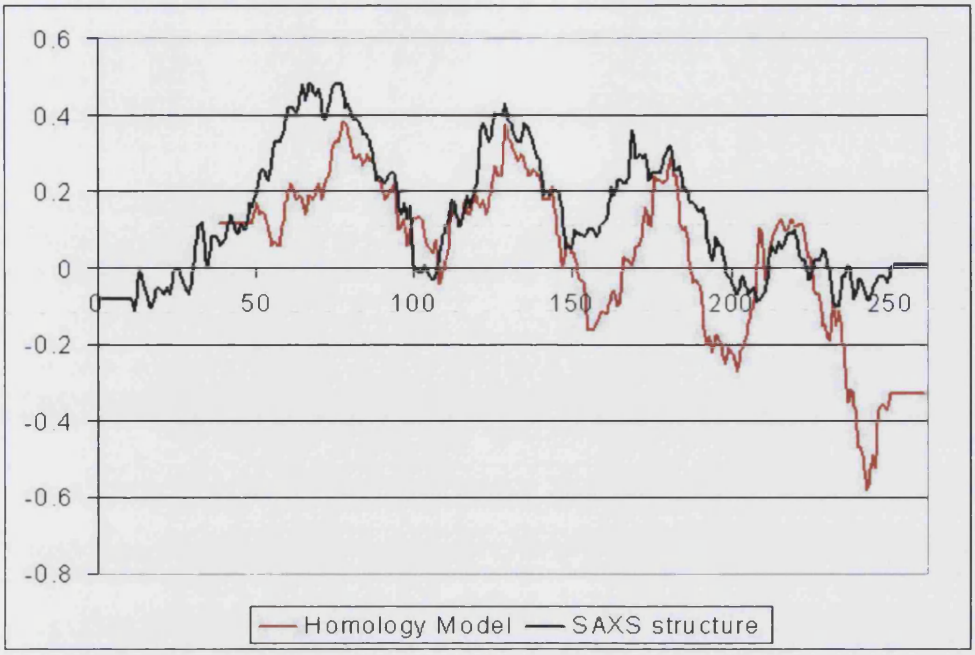


Figure 54 The Verify 3D plots for the structures in Figure 53. The regions for the Sbi-e models in the upper portion of the plot suggest that the N-terminal regions (excluding the His-tagged modelled in the SAXS structure) are correctly modelled. The C-terminal region is less accurate; with the Verify 3D result suggesting the model produced by SAXS is more accurate in this region.

with the *ab initio* and BUNCH reconstruction. Moreover, the N-terminal loops in the genetic algorithm models displays a folded back conformation, as already observed in the BUNCH solution (Figure 50b).

10.1.3.5 A model for the domain organisation of Sbi

Using the *ab initio* density model derived from the SAXS data, we manually fitted the Sbi models generated in Chapter 2 to construct a three-dimensional model of Sbi's predicted N-terminal region. The three-helix bundle motif units fit very well with the observed density of the first three domains of Sbi-e, suggesting that all four domains possibly adopt a similar structural fold. This homology model is shown in Figure 52.

The homology model of Sbi was compared to the SAXS structure. The homology model and SAXS structure are both shown in Figure 53. The Verify 3D plots for the two structures are overlaid in Figure 54; this shows the SAXS structure appears to be more favourable in terms of the side chain environment, however the backbone angles

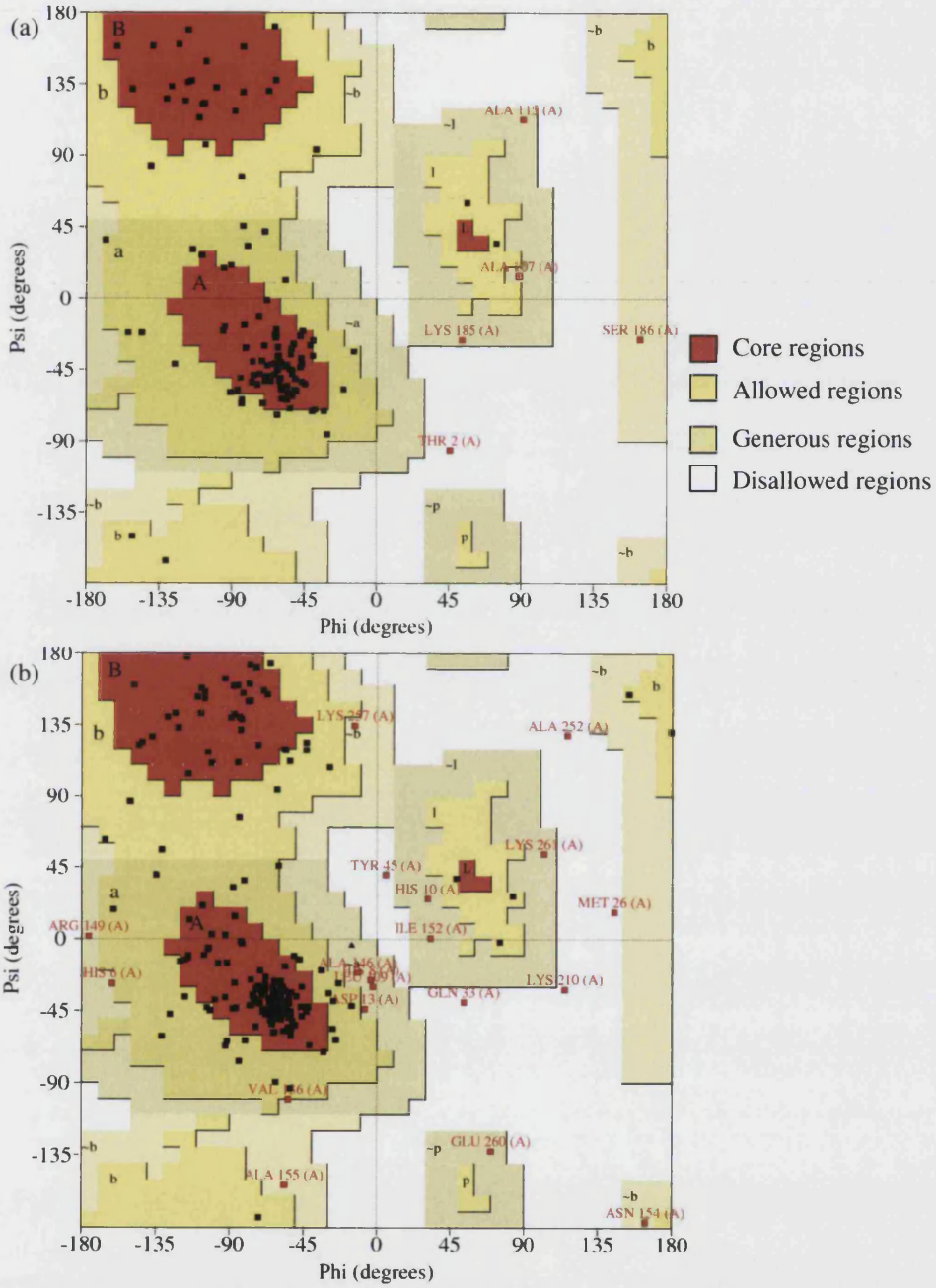


Figure 55 Ramachandran plots for the structures shown in Figure 53 determined by Procheck. **(a)** The homology model and **(b)** the model based on the SAXS structure

Table 25 Trypsin cleavage sites in the Sbi protein Sbi-e. Residues 1*and 263* are the N- and C-terminal residues in the molecule.

Sbi trypsin digestion sites	Point in Sbi-e sequence
1*	N-terminal residue
4	His-tag
37	Prior to helix 1
79	Between helix 1 and helix 2
97	Between 1st and 2nd Ig-binding domains
149	End of 2nd Ig-binding domain
186	β2GP1 binding domain
191	β2GP1 binding domain
200	β2GP1 binding domain
204	β2GP1 binding domain
263*	C-terminal residue

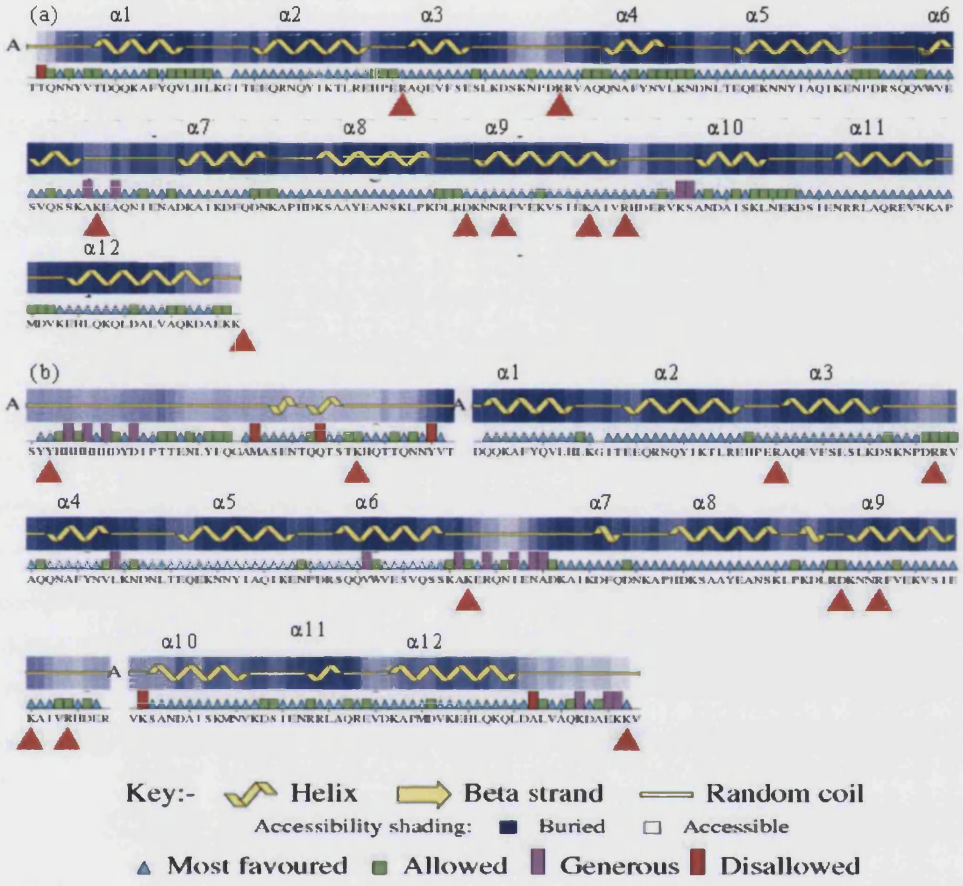
in the Ramachandran plots (Figure 55) shows that the SAXS structure contains more residues in disallowed regions than the homology model.

10.1.3.6 Mapping of trypsin sites on Sbi-e

Digestion of protein molecules by trypsin results in the cleavage of the peptide chain in regions not folded into tertiary structures, i.e. between domains, as trypsin can generally not cleave at its cleavage site if the protein backbone is folded into a globular structure. Therefore trypsin digestion of a protein can be used to reveal the domain boundaries of any domains in the protein's structure.

A trypsin digestion was carried out on the Sbi-e protein; the determined cleavage sites are shown in Table 25. When the cleavage sites were mapped onto both the Sbi models (Figure 56) the results generally support the proposed domain models. The first two sites detected do not appear on the homology model as it was not possible to model the N-terminal tag of Sbi-e (Y4 and K37), however as the N-terminal of the SAXS model is mainly unstructured, these two trypsin digestion sites do not cleave in a region that is expected to be structured. The first cleavage site mapped onto both Sbi structures (E79, Figure 56) is located between helices 2 and 3. The next two cleavage sites (R97 and R149) are located between domains 1 and 2, and domain 2 and 3

Figure 56 The secondary structure of the structural models shown in Figure 53 as determined by Procheck (Laskowski *et al* 1993), with the experimentally determined trypsin digestion sites marked (red arrow heads). **(a)** The Homology model and **(b)** the SAXS structure



respectively. This suggests that the predicted domain and helix boundaries in this part of the molecule are in good agreement, as trypsin is expected to cleave at sites that are accessible and not part of the core three-dimensional structure. The remaining restriction sites detected are R186, R191, K200 and R204. R191 cleaves the end of N-terminal helix 9 in both the homology model and SAXS structure, while K200 in the homology model cleaves the C-terminal end of this helix (see Figure 56 for the trypsin digestion sites on the secondary structure of the two models).

10.1.4 Discussion

SAXS revealed the N-terminal region of Sbi (Sbi-e) to be an elongated molecule composed of four domains all of similar size. For the first time, this provides insights into the three-dimensional domain arrangement of a bacterial immunoglobulin binding protein. The similarity with the domain structure of SpA may suggest that this region of the Sbi protein is composed of 4 three-helix bundles as predicted in Chapter 2. This would also support the prediction in Chapter 5 that Sbi and SpA have an evolutionary relationship, and that the β2GP1-binding domain and domain 3 of Sbi share an ancestral domain or gene with the Ig-binding domains of Sbi and SpA. Possible mechanisms by which the evolution of these two proteins occurred is discussed further in Chapter 7.

The sizes of the Sbi constructs were determined by three different methods. There are three sets of SE chromatography data for Sbi-e. The three elution profiles are very different (Figure 49), and suggest that an increase in the protein concentration results in an increase in the size of the Sbi-e particles, indicating the formation of large complexes. There are also two data sets for Sbi-S from the SAXS analysis; this also shows an increase in particle size with an increase in protein concentration (Table 24). This suggests that the Sbi constructs have the tendency to form large complexes composed of the Sbi monomers. This tendency could explain the ability of Sbi-1 to form insoluble complexes with Guinea pig serum IgG.

The SE chromatography results were carried out with the aim of increasing the purity of Sbi-e, while the DLS results were to analyse protein samples of Sbi-1, Sbi-2 and Sbi-S under different conditions in an attempt to reducing aggregation of the protein particle sizes, and improve the quality of the protein for the SAXS experiments and the possibility of NMR experiments. These experiments did show that the size of the protein particles does appear to be affected by the protein concentration, while the particle sizes for Sbi-1, Sbi-2 and Sbi-S also appear to be influenced by the presence or absence of the N-terminal His-tag. This is an area where further work is needed to gain a better understanding of the relationships between protein concentration, the presence of the His-tag on the protein and the particle size for the different Sbi constructs.

Trypsin digestion analysis of the Sbi-e construct revealed the presence of 9 cleavage sites (Table 25). Trypsin cleaves preferentially at lysine and arginine residues on exposed peptide chains (http://www.expasy.ch/tools/peptidecutter/peptidecutter_special_enzymes.html). Mapping these sites onto the homology model and the SAXS structure shows that most of the detected trypsin digestion sites in Sbi are on loops in the structure, indicating that the allocation of secondary structure in these regions is in good agreement with the predicted three-dimensional models (Figure 56). The only cleavage sites located on secondary structure elements are R191 and K200 are located on helix 9 for the homology model and R191 in the SAXS structure.

The trypsin digest results in combination with the other structure validation results suggests that the SAXS model of Sbi can be used to improve homology models for proteins whose structure cannot be solved by other structural techniques, such as x-ray crystallography. In addition, circular dichroism studies on the β_2GP1-binding domain (see Chapter 2) confirmed that this region of Sbi is α-helical, thereby also supporting the SAXS model. In light of the SAXS data showing that the Sbi protein is composed of four domains, we are confident that the model proposed is reliable.

11 Chapter 7

11.1 Reflections and perspectives

Sbi is an *S. aureus* cell surface protein. Zhang *et al* (1998, 1999) identified the presence of an N-terminal Ig-binding domain, followed by a second proposed Ig-binding domain, and a domain capable of binding the serum protein β2GP1. Comparison of the Sbi sequence to the SpA sequence revealed high homology in the part of Sbi located N-terminally from the proline rich region (see Figure 8 for the structure of the Sbi gene) with the Ig-binding domains of SpA. We therefore predicted that this region of Sbi is composed of domains with a structural fold similar to the SpA Ig-binding domains. We also predicted that Sbi contains four such domains.

Sbi also contains a proline-rich region between the N- and C-terminal regions. The length of the repeat unit matches the motif found in the proline pipe helix (Butcher *et al* 1996). Therefore a proline pipe helix was built for this region of the structure. We proposed that the proline pipe helix interacts with the bacterial cell wall in a similar fashion as the MSCRAMM proteins, which also contain proline rich sequences in the regions identified to be cell wall binding (see Chapter 2 for more details). The mechanism of cell wall anchoring by these sequence motifs still remains unclear.

It was not possible to identify any sequences or protein structures homologous to the C-terminal region of Sbi. This part of the protein therefore either has a novel function and fold or it shares its fold with a previously identified fold, but the sequence homology is too low for identification.

Homology modelling of the Sbi Ig-binding domains, using three-dimensional structures of SpA domains B and D, revealed that the amino acids on the face of the three-helix bundle structure of the Ig-binding domain, responsible for binding to IgG via the Fc fragment, are all conserved in both Ig-binding domains of Sbi (see Figure 14b & c). Residues on the face involved in Fab binding are not conserved in Sbi's Ig-binding domains (see Figure 14d). We predicted that Sbi could only bind IgG via the Fc binding site (se Chapter 2).

Binding studies with Sbi revealed that while both Sbi and SpA form insoluble complexes in the presence of intact serum IgGs, only Sbi precipitates in complex with IgG Fc fragments. This showed that although both Ig-binding proteins bind to the Fc fragment of IgG, SpA also requires Fab binding for precipitation. This observation supports the prediction in Chapter 2, based on sequence homology, that Sbi only binds to the Fc region of immunoglobulins.

We also examined a possible mechanism of complex formation by both SpA and Sbi. Hanson and Schumaker (1984) predicted that SpA in the presence of rabbit IgG forms a complex of IgG models stacked on top of each other with SpA Ig-binding molecules down each side (Figure 35). To investigate if such a model was possible we generated a model composed of two SpA domains, connected by a linker in complex with two Fc fragments. A similar model was created for both Sbi Ig-binding domains (see Figure 42 for the models of the SpA and Sbi complex with IgG). Interestingly, this revealed that while it is possible to fit two Fc molecules along in the SpA ladder model, it is not possible to do so for Sbi, because the shorter linker between the Ig-binding domains in Sbi causes steric clashes between the two Fc fragments. Our ladder model implies that, in the case of Sbi, Fc binding by both domains is more likely to cause cross-linking rather than stacking, seen in SpA. With this model we now can explain the formation of insoluble complexes between Sbi and Fc fragments, while the stacked complexes formed with SpA remain soluble in the absence of Fab.

Comparison of the binding patterns for the formation and size of soluble and insoluble complexes with both Sbi and SpA (in Chapter 3) reveals that each share their preferences in binding the different human IgG subclasses: both bind IgG1, IgG2, and IgG4 and show either weak or no binding to IgG3. Significant differences are observed in the presence of serum IgGs from different species: Sbi precipitates human, rabbit, mouse and Guinea pig serum IgGs while SpA precipitates human, pig and Guinea pig serum IgG, although Guinea pig IgG produces more precipitation in the presence of Sbi than with SpA. In the measurement of soluble complex sizes, SpA produces larger complexes in the presence of pig and cow serum IgGs than Sbi does. Sbi and SpA both appear to form 1:1 complexes in the presence of goat, horse, and sheep serum IgGs while Sbi also forms 1:1 complexes with pig and cow serum IgGs.

Interestingly, single Sbi Ig-binding domains are still capable of forming large complexes with human (soluble) and Guinea pig (insoluble) serum IgGs.

Determination of affinity constants for Sbi in the presence of the different serum IgGs, using SPR, reveals that Sbi binds mouse serum IgG most strongly, closely followed by human and Guinea pig. The weakest binding was found with sheep and pig IgGs. Both Sbi Ig-binding domains bind human IgG1 subclass strongly and also show similar affinity for IgG4. The binding of IgG2 revealed an order of magnitude difference in the affinity for IgG2 between the two domains (see Figure 33). Neither Sbi domains have any affinity for IgG3.

The differences found in complex formation between SpA and Sbi and the various serum IgGs was further analysed using homology modelling in Chapter 4. Models of SpA-Fc and Sbi-Fc complexes were generated using a crystal structure of SpA in complex with a human Fc fragment (see Chapter 4).

Analysis of the homology models of these complexes pointed to a number of substitutions in the Fc sequence that could affect the interaction and orientation between the C_H2 and C_H3 domains of the IgGs. The gain or loss of a large side chain at the C_H2-C_H3 domain interface in the Fc molecule could affect the elbow-angle between the two domains. Since the SpA binding site on the Fc molecule comprises residues on both C_H2 and C_H3, changes in the elbow-angle between the two domains are very likely to affect the binding of SpA (or Sbi) by changing the distance between the residues on the two C_H domains required to bind SpA (or Sbi). Thereby binding could be affected without changing the residues in the binding site itself.

In most cases it is not completely clear what effect, if any, the substitution may cause on binding. We were able to identify two substitutions that we believe may play an important role in the binding between the two Ig-binding proteins and the IgGs. In chicken and sheep IgG we found a substitution (H435R and D376Y, respectively) that could affect the elbow angle of the Fc molecule. The chicken substitution (H435R) and a neighbouring substitution (Y436F) also occur in the sequence for human IgG3, the only IgG subclass that does not bind SpA or Sbi. The H435R and Y436F substitutions appear to be able to eliminate Fc binding completely, as shown by

Jendeberg *et al* (1997). The Sheep IgG substitution introduces a large side chain at this position thereby possibly affecting the spacing and orientation between the two C_H domains.

Of the residues identified to be involved in the interaction between the C_H domains, most are well conserved. A few amino acids are less conserved; these include residues 315, 339, 375, and 378. Of these amino acids 315 is also involved in the interaction with the Ig-binding domain. For residues 339, 375 and 378, there is no apparent correlation between the substitution and the affinity of the IgG for SpA and Sbi.

Analysis of the interactions between the IgGs and the Sbi and SpA models identified a single site where a correlation between binding levels and the residue present could be determined. This is residue 315. Low binding species have a small uncharged or polar side chain at position 315, while the presence of large charged or polar side chains results in higher binding levels. Site directed mutagenesis of the IgGs at this position is needed to determine whether the substitutions really affect the ligand binding as proposed.

S. aureus infects a range of species; it is of economical importance in cattle where it causes mastitis, affecting milk production. However, our binding studies revealed that Sbi and SpA bind cow serum IgGs relativity poorly, as is the case for the IgGs from other farm animal species such as horse, sheep and goat. The Ig-binding proteins used in this study were both cloned from strains of *S. aureus* isolated from humans. To investigate whether the Ig-binding proteins of *S. aureus* strains isolated from human hosts have adapted to human proteins, we sequenced the SpA and Sbi genes from strains isolated from other species (see Chapter 5). The sequence analysis of the *spa* and *sbi* genes, however, showed no significant changes in the sequences of the Ig-binding regions of both genes that could account for the differences in IgG binding in the different host species; all sequences of the Ig-binding domains were essentially conserved in the different strains. This suggests that there is no tendency for acquiring host specific immunoglobulin binding in these strains. Since these animals are in regular contact with the human host future experiments should also include strains isolated from animals in the wild.

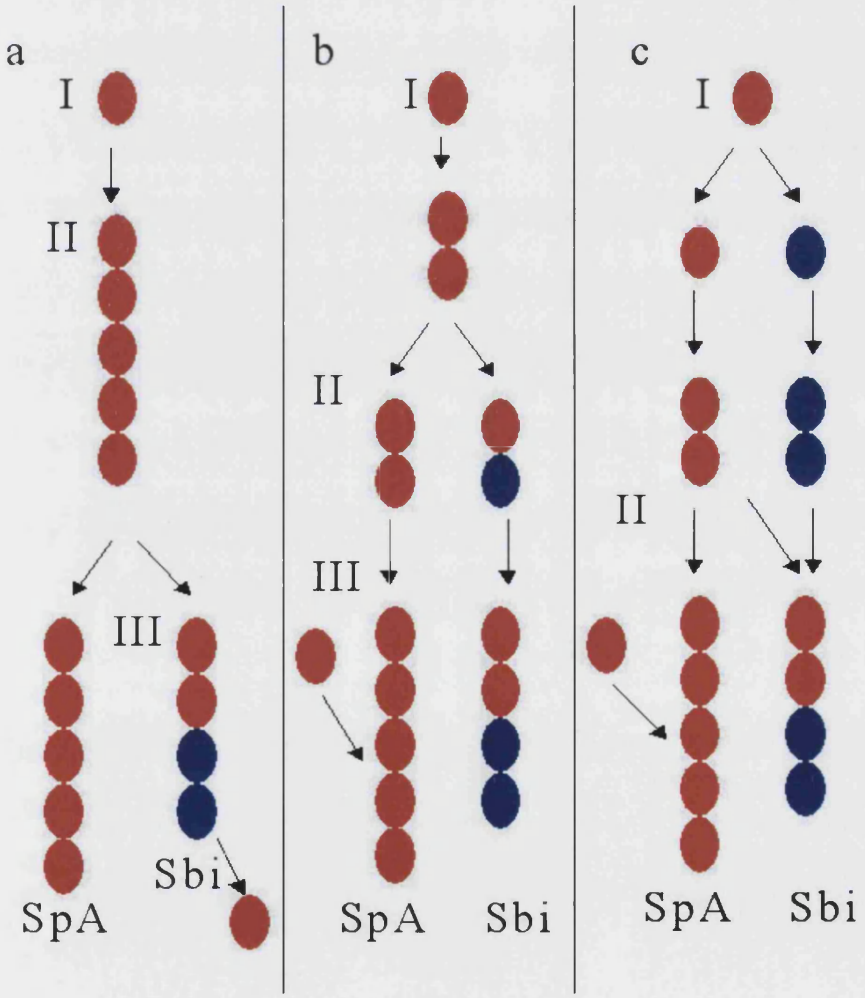
All attempts to crystallise the Sbi-e region of Sbi for structural studies have so far been unsuccessful. We believe this is inherent to the predicted multi-domain structure of this construct. We therefore attempted to determine the structure in solution. NMR is unsuitable because of the size-limitations of this technique, therefore we tried small-angle X-ray scattering. From X-ray scattering analysis it can be concluded that the Sbi-e has an elongated molecular shape and is monomeric in solution. An *ab initio* model derived from the scattering data reveals an elongated protein composed of 4 domains (see Figure 52), which neatly fits the models predicted in Chapter 2.

11.1.1 Implications for the evolutionary origin of Sbi and SpA

The structure determined by SAXS clearly shows Sbi contains 4 domains. The first three of these all appear to be the same fold, suggesting that domain 3 is similar to the two Ig-binding domains. The structure of the β2GP1-binding domain is harder to predict than the others, as the model is less well refined in this region, however we are not able to rule out the possibility of it also being a three-helix bundle, although in the most probable model, based on the SAXS *ab initio* structure shown in Figure 53b, the 3-helices are in a more open arrangement to the other domains.

We discussed in Chapter 5 the possibility that there is an evolutionary relationship between Sbi and SpA. The high similarity of the Ig-binding domains would suggest that they might originate from one ancestral domain, which has been duplicated. In Figure 57 I propose several possible routes of this evolution. The structural model for Sbi presented in Chapter 6 fits well with the scattering data and suggests a structural similarity between the Ig-binding domains of both Sbi and SpA and domain 3 and the β2GP1-binding domain of Sbi. This suggests that all five domains in SpA and all four domains in Sbi could be descended from a single Ig-binding domain type protein. We suggest that the ancestor had the function of Ig-binding, as it is more likely that the two C-terminal domains in Sbi have lost this function. This is more likely than the independent evolution of the Ig-binding domains in the two proteins. Therefore one of three possible routes for the evolution of the two proteins is the evolution of a protein containing either four or five similar Ig-binding type domains. This protein was than duplicated, one evolved into SpA with five domains, the other acquired mutations in its third and fourth domains resulting in a change of function and evolved into Sbi

Figure 57 Possible mechanisms for the evolution of the Ig-binding proteins SpA and Sbi. **(a)** The ancestor gene (I) consists of a single Ig-binding domain. This domain is duplicated (II) in a single gene until the gene consists of five domains. The gene is then duplicated (III) and one copy forms SpA. The other copy loses one domain while the other two undergo mutation to form Sbi. **(b)** The ancestor domain (I) is duplicated. The gene is than further duplicated (II). One copy undergoes further duplication of domains, while maintaining the Ig-binding function to form SpA (III). The other copy also duplicates its domains, but some of the domains undergo mutation, and this results in Sbi. **(c)** Duplication of the ancestral Ig-binding domain occurs at an early stage (I). One copy retains its function, and after undergoing duplication of the Ig-binding domain forms SpA (II). The other copy is only duplicated until it has two Ig-binding domains, but in this case they suffer mutations and develop novel functions. At some point two Ig-binding domains from SpA are transferred to this gene to form Sbi.



with four domains. In the second method a single or double Ig-binding domain ancestor split into two genes. Both underwent subsequent duplication of domains, with one acquiring mutations in some of its domains giving rise to Sbi, while the other formed SpA. A third method can be based on this. The ancestor consisted of a single Ig-binding domain. This was duplicated, and one took the route where it acquired mutations in its two Ig-binding domains, to later form domain 3 and β2GP1-binding domain of Sbi. Meanwhile the ancestor gene on the other route will have undergone little or no change in its double domain form. Then at some point in time a copy of this gene is inserted into the gene of the other protein containing the mutations to form the modern day Sbi. The original copy of the Ig-binding domains was duplicated in its gene to form today's SpA. One other piece of evidence to support the gain and/or loss of Ig-binding domains in the evolution of the Sbi and SpA genes is in our sequencing of the different SpA and Sbi genes (Chapter 5). We discovered several SpA genes containing only four Ig-binding domains rather than the more normal five. As the missing domains cover several different SpA domains it suggests that there is a mechanism that results in the loss of the Ig-binding domains under unknown conditions. An analysis of the linker sequences between the Ig-binding domains will provide evidence about the most likely evolutionary route. As the linkers between the Ig-binding domains in Sbi are different to those in SpA, it suggests that the molecules have evolved separately from an ancestral domain.

11.1.2 Remaining questions

11.1.2.1 Does Sbi bind other ligands?

SpA is known to bind several ligands other than Igs. These include gC1qR/p33, von Willebrand factor and the tumour necrosis factor-α receptor. The interactions with these ligands are not as well characterised as the interaction with IgG. However one of these appears to share the Ig-binding site. Romagnani *et al* (1982) showed that iodination of SpA eliminates IgG binding via its Fc region, while Nygren, Ghebrehiwet, and Peerschke (2000) showed that the same method also eliminates binding of gC1qR/p33. In Chapter 2 I have shown that the Ig-binding site is well conserved in Sbi as are the tyrosine residues targeted by the iodination of SpA. This suggests the possibility that Sbi, like SpA, binds gC1qR/p33. Perhaps Sbi also binds other SpA ligands?

Sbi binds the serum protein β2GP1 as well as IgG. β2GP1 is composed of short consensus repeat domains (Bouma *et al* 1999). Recent studies in our lab have shown that Sbi also binds to complement protein C3 (Burman *et al* submitted for publication). Does Sbi bind to any other host proteins? The C3 protein was detected by use of a pull down assay. While this method can be used to detect serum proteins bound by Sbi, it will not detect any ligands attached to host cells, and not present in the serum.

11.1.2.2 How is the expression of Sbi regulated?

Zhang *et al* (2000) discovered that when *S. aureus* is cultured in the presence of IgG, the expression of Sbi is up-regulated. However, how this occurs is unknown. The C-terminal region of Sbi, adjacent to the proline repeat region, is rich in tyrosine and threonine residues. Both residues are known to be involved in protein activation via the phosphorylation of these residues. However such activation sites are intracellular. Sbi does not contain any typical transmembrane region nor does it have any hydrophobic regions that could be expected to cross the membrane. Zhang *et al* (2000) showed that Sbi is located on the cell surface, but its mechanism of attachment is unknown, as Sbi lacks the LTXPG motif found in other adhesion proteins required for anchoring in the cell membrane. It is therefore unclear how it is attached to the cell and whether the tyrosine rich region is located intracellularly.

There are many unanswered questions remaining about Sbi. Future studies on this protein could reveal its role in *S. aureus* infections and whether it plays a greater role than the other Ig-binding protein SpA. Other areas of interest would include the function of individual ligands in infection and virulence for both Sbi and SpA, determined by the elimination of specific functional regions in addition to knocking out the entire gene. Also of interest is whether other *S. aureus* adhesion proteins show the same preference for ligands from a particular host, i.e. SpA appears to favour porcine IgGs. There is also further work needed to understand how the substitutions observed in Chapter 4 in the different IgG models affect binding and interactions with the Ig-binding proteins SpA and Sbi.

11.1.2.3 What is the effect of functional elimination of *S. aureus* Ig-binding proteins?

In vivo, proteins generally interact with several others operating in the same pathway. However knockout studies generally remove the entire protein, or render it inactive. However, as we discussed above, SpA and Sbi interact with a number of known ligands in vivo. Most of the binding sites in SpA are restricted to a single domain. What effect would the elimination of a single binding site have on the in vivo virulence of the bacteria?

The use of targeted mutagenesis in the gene to eliminate a function has the advantage that it only affects a single site and will only affect the interactions of ligands that bind to that site. This means all other interactions can still occur, and can tell us how important one function is compared to another in virulence. For example, could the elimination of the Fab binding site on SpA result in a lack of generation of B-cell superantigens and thereby decrease the virulence of *S. aureus*?

11.1.2.4 What is the role of the poly-proline repeat regions in SpA?

Unlike the number of polyproline repeats in Sbi, which is the same in all sequenced Sbi genes, the number of repeats in SpA varies from four to sixteen (see Chapter 5). Does this variation have any function? There is no apparent relationship between the number of repeats and the species from which the *S. aureus* strain was isolated. However, could there be a relationship between the number of repeats and the virulence of the strain?

11.1.2.5 Does *S. aureus* virulence vary from species to species?

The virulence studies discussed in Chapter 1 were all carried out on lab animals (mice, rats, rabbits). This could be important to consider as our binding studies (chapter 3) showed marked differences between the IgGs of mouse and rabbit (rat serum IgGs were not tested) with both Sbi and SpA. Sbi causes precipitation in the presence of these IgGs, while SpA forms large soluble complexes. This suggests these serum IgGs bind better to Sbi than SpA. Could this mean that Sbi plays a larger role in virulence in these species than does SpA?

11.1.2.6 Does the specificity of *S. aureus* adhesion proteins differ from species to species?

Regarding the relative binding of these Ig-binding proteins, SpA appears to bind best to human and porcine serum IgG while Sbi binds best to human and rodent serum IgGs (of those tested in Chapter 3). Other studies have characterised the binding of *S. aureus* proteins to ligands such as fibrinogen and fibronectin, but these adhesion proteins have not been characterised binding to the same ligand but produced by different mammalian species (Chapter 3, Kronvall *et al* 1970). Zhang *et al* (1999) identified β2GP1 as a ligand for Sbi, used newborn calf serum. However, recent studies in our lab used human serum and identify a novel Sbi ligand, complement component C3. No β2GP1 was identified in these affinity pull-down experiments. This suggests that Sbi (and perhaps also SpA) does indeed bind different ligands in different hosts.

We have determined that Sbi and SpA genes show very little variation between strains of *S. aureus* isolated in different species. Is this the same for other adhesion proteins, or do they show greater variability in protein sequence and how does this affect virulence of *S. aureus* in different species? If the *S. aureus* adhesion proteins including SpA and Sbi bind poorly to the serum and ECM proteins of an animal species, does it make *S. aureus* a poor pathogen in those species or is there compensation (for example, IgG could be bound weakly, while fibrinogen is bound strongly)? One way to investigate this would be comparing binding data with infection and virulence data for different species. Unfortunately, these data are not available: while the U.K. Department of Agriculture and Rural Development website (http://www.dardni.gov.uk/) lists infection data for Brucellosis and Tuberculosis, it does not have information published on *S. aureus* infection rates.

11.1.2.7 The role of SpA from earlier virulence studies

As well as virulence studies with the adhesion proteins such as ClfA and FnBPA, virulence studies have also been carried out with SpA (Callegan *et al* 1994, Gomez *et al* 2004, Jonsson *et al* 1985, Palmqvist *et al* 2002, Palmqvist *et al* 2005). This has also been recently classified as an adhesion protein as it has been discovered that it binds a component of the ECM: the von Willebrand factor (Foster and Hook 1998, Hartlieb *et*

al 2000) in addition to IgG. In the early virulence studies on SpA as well as other mutagenesis studies with SpA (for example: Jonsson *et al* 1985), SpA mutants were generated by chemical mutagenesis. Later studies used strains known to be defective in SpA. In all cases a small reduction in virulence was observed. The question is whether these strains had only the SpA gene eliminated or whether others were also affected such as Sbi. Testing for a reaction with IgG was used in determining the presence or activity of SpA. However we now know *S. aureus* contains two Ig-binding proteins SpA and Sbi. What we do not know is if they show different properties on the surface of *S. aureus* and whether all studies with *S. aureus* strains into the effect of SpA had functional Sbi or in effect they were studying the effect of a double mutant with the *S. aureus* strain also containing non-functional Sbi. Where detection of *S. aureus* mutants lacking IgG binding ability occurs, it would suggest the possible elimination of both genes, while when the presence of low-binding mutants is detected, it suggests that these strains could have only one or the other of the two Ig-binding genes active. The only way to answer this question would be to use *S. aureus* mutants of both SpA and Sbi generated by molecular biology, where it can be confirmed that only a single gene has been affected, and compare the action of the *S. aureus* mutants *in vitro* to both wild type *S. aureus* and a double knockout of both genes. Virulence studies with such *S. aureus* mutants would show whether one protein has a larger effect on virulence than the other.

11.1.2.8 Reflections on the role of *S. aureus* and its adhesion proteins in infection and future directions

Most virulence studies with *S. aureus* adhesion proteins (Moreillon *et al* 1995, Flock *et al* 1998, Jonsson *et al* 1985, Patel *et al* 1987, Callagen *et al* 1994, Palmqvist *et al* 2002) have shown that the loss of an adhesion protein results in some reduction in virulence (chapter 1). Interestingly, in studies where the loss of an adhesion protein is compared to the loss of a toxin such as α-haemolysin, the strain with the α-haemolysin mutation shows a much greater reduction in virulence (Callegan *et al* 1994, Jonsson *et al* 1985, Patel *et al* 1987). Unlike adhesion proteins α-haemolysin is a toxin secreted by *S. aureus* that is part of a group of channel forming toxins. These form channels in the membranes of the host cells, resulting in disruption of membrane potential and can cause cell death (Wilson, McNab, and Henderson 2002 Ch9). No

single adhesion protein knockout has the same level of effect as a knockout of α-haemolysin. This could be due to the differences in their individual roles in infections. What is unknown though, is how multiple knockouts of adhesion proteins affect virulence *in vivo*. We have not been able to find any studies with strains of *S. aureus* carrying more than a single gene knockout in an animal disease model. We do know from the work of O'Brien *et al* (2002) that multiple knockouts of adhesion proteins cause further increases in the lag time for platelet aggregation *in vitro*. In this work, using *S. aureus* knockout mutants of the ClfA, SdrE, and SpA genes, only the single ClfA knockout mutant had an increased lag time for aggregation compared to the use of wild type strains. However, double knockout mutants of SdrE or SpA and ClfA gave a further increase in lag time compared to the wild type, while the triple mutant recorded the longest lag time. This clearly shows that all 3 proteins appear to be required in an *in vitro* setting and the loss of the additional genes of SdrE and SpA increase the effect of the single ClfA mutation. The effect of these multiple knockout mutations on virulence *in vivo* is unknown. Perhaps the development of a drug therapy that targets multiple adhesion proteins would provide an effective way to treat and control *S. aureus* infections.

11.1.3 Conclusion

Sbi is an Ig-binding binding protein with high affinity for human, rabbit, Guinea pig and mouse serum IgGs, while SpA shows high affinity for human and pig serum IgGs. The N-terminal region of Sbi is predicted to be composed of 4 domains, all of the three-helix bundles typical of the SpA Ig-binding domains. Sbi contains two functional Ig-binding domains at its N-terminal, while the fourth predicted domain binds β2GP1 and domain three and the β_2GP1-binding domain together have recently been shown to be involved in the binding of complement component C3 (Zhang *et al* 1999, Burman *et al* submitted for publication). Comparison of the IgG sequences through homology modelling suggested that the IgG Fc residue 315 plays a role in determining affinity for Sbi and SpA, while the sequencing of the Sbi and SpA genes from strains isolated in different hosts revealed that the sequence of the Ig-binding domain is conserved across all the hosts.

Future directions for studies on Sbi should include virulence studies to determine its role in infection. Such a study would determine the role of Sbi by the generation of *sbi*⁻ strains, the introduction of Sbi into a non-virulent organism such as *L. lactis*, and the generation of Sbi mutants that lack the ability to bind one of its ligands (possibly more if it is determined that two or more different ligands share a binding site). The generation of such mutants would give a greater understanding of the role of Sbi in addition to an *sbi*⁻ mutant, as it would provide specific information about the role of the binding of individual ligands on the virulence of *S. aureus*. Another direction of research would be the generation of *S. aureus* mutants lacking more than one adhesion protein. *In vivo* virulence studies with strains carrying multiple knockouts would give a greater understanding of the roles the proteins play in virulence with respect to each other.

There is currently little known about the C-terminal regions of Sbi. We predict that Sbi is anchored in the cell wall by its proline rich region. More studies on this part of the protein are needed. In addition the function of the tyrosine-rich C-terminal region needs addressing. Structural studies on these parts of the protein could reveal if it has any structural homology to other known proteins, while the production of soluble protein of both the C-terminal region and the proline rich region could be used to determine if and how they interact with for instance *S. aureus* signalling proteins. This could be of particular interest as the C-terminal region has been suggested to be involved in regulation of the expression of Sbi.

12 Appendix 1

Table of *S. aureus* surface proteins, their ligands, and role in infectious diseases.

Protein	Ligand(s)	Mechanism of attachment to cells	Ligand located ...	Mammalian Receptor(s)	Roles/functions implicated in ...	Diseases implicated in ...	Known effects of KO	Other points of interest	References
ClfA	Fibrinogen	LPXTG-motif	Serum	>Yes - see Boden & Flock > GPIIb/IIIa	Platelet aggregation. Binding to platelets/ activation. Complement activation.		Increase in lag time for aggregation	*S. aureus* clumping has been known for 80years. Dominant cause of endocarditis	>Siboo *et al* (2001) >Boden & Flock (1989) >Loughman *et al* (2005) >O'Brien *et al*(2002) >Morellion *et al* (1995)
	P118/novel cell surface receptor		Platelets	Cell-surface	Direct platelet aggregation - fibrinogen independent				
	Fibrinogen-independent binding		? Complement	Cell-surface				>Does not occur in the absence of IgG/anti ClfA Ab. >Extended lag time for Complement-mediated activation due to assembly of complex	

Protein	Ligand(s)	Mechanism of attachment to cells	Ligand located ...	Mammalian Receptor(s)	Roles/functions implicated in ...	Diseases implicated in ...	Known effects of KO	Other points of interest	References
ClfA cont...	Anti-ClfA Ab	LPXTG-motif	Serum	FcγRIIa				>Fibrinogen-binding alone does not activate platelets >ClFA-Fibrinogen binding plus IgG cross linking	See above
	General					Endocarditis	50% reduction in cases in rats		
ClfB	Fibrinogen	LPXTG-motif	Serum	?					
	Cytokeratin 10		Squamous cells.	Cell surface	Nasal colonisation				Walsh *et al* (2004)
FnbpA	Fibrinogen	LPXTG-motif	Serum	?					
	Fibronectin		ECM	?	Adhesion to plasma clots			Expressed in exponential phase	Roche *et al* (2004)
	Elastin		ECM	?		Endocarditis/ Pneumonia		Lost of Fnbp eliminates elastin binding	Roche *et al* (2004)

Protein	Ligand(s)	Mechanism of attachment to cells	Ligand located ...	Mammalian Receptor(s)	Roles/functions implicated in ...	Diseases implicated in ...	Known effects of KO	Other points of interest	References
FnbpB	Fibrinogen	LPXTG-motif	Serum	?					
	Fibronectin		ECM	?	Adhesion to plasma clots			Expressed in exponential phase	Roche *et al* (2004)
	Elastin		ECM	?		Endocarditis/ Pneumonia		Lost of Fnbp eliminates elastin binding	Roche *et al* (2004)
FnbpA/B	Fibronectin	LPXTG-motif	ECM/ Epithelial cells		Internalisation of epithelial cells	Endocarditis	No difference	Most invasive strains bind Fn	>Flock *et al* (1996) >Dziewanowska *et al* (1999)
SdrE	Unknown	LPXTG-motif	N/a	N/a	Activation of platelets			Family includes ClfA/B, FnbpA/B and SdrC/D	Loughman *et al* (2005) Josefsson *et al* (1998)

Protein	Ligand(s)	Mechanism of attachment to cells	Ligand located ...	Mammalian Receptor(s)	Roles/functions implicated in ...	Diseases implicated in ...	Known effects of KO	Other points of interest	References
Cna	Collagen	LPXTG-motif	ECM	?	Attachment to cartilage	Septic arthritis	Reduced rate of arthritis	*S. aureus* causes 80% of cases	Patti *et al* (1994)
EpbS	Elastin		ECM	Has mammalian receptor				Elastin occurs in blood vessels. Different binding to mammalian receptor.	Park *et al* (1991)
Sbi	IgG Fc		Serum	Fc receptors					
	β_2GP1		Serum	Binds directly to platelets					
vWbp	Von Willebrand factor								

Protein	Ligand(s)	Mechanism of attachment to cells	Ligand located ...	Mammalian Receptor(s)	Roles/functions implicated in ...	Diseases implicated in ...	Known effects of KO	Other points of interest	References
SpA	IgG Fc	LPXTG-motif	Serum	Fc receptors	Binds most mammalian IgGs				Kronvall *et al* (1970)
	IgG Fab V_HIII		Serum	Fc receptors	>Supraclonal B-cell >All SpA domains bind Fab >Binds VH3 gene family only				>Palmqvist *et al* (2005) >Jansson, Uhlen, & Nygren (1998) >Robben, Salem, & Silvermann (1995)
	Von Willebrand factor		Serum	Platelets	Platelet adhesion	Endovascular disease			Hartlieb *et al* (2000)
	gC1qR/p33		Platelets/ endothelial cells	Cell surface		Endocarditis			Nguyen *et al* (2000)
	TNFR1			Cell surface		Pneumonia			Gomez *et al* (2004)
	General					Keratitis	Not a major virulence factor		Callegan *et al* (1994)
						Mastitis	No significant difference		Jonsson *et al* (1985)
						Arthritis/Septic Death	Reduced virulence		Palmqvist *et al* (2002)

Protein	Ligand(s)	Mechanism of attachment to cells	Ligand located ...	Mammalian Receptor(s)	Roles/functions implicated in ...	Diseases implicated in ...	Known effects of KO	Other points of interest	References
Alpha-toxin						Keratitis	Greatly reduced pathology	Pore-forming haemolytic toxin	Callegan *et al* (1994)
						Subcutaneous lesions	Major virulence factor		Jonsson *et al* (1985)
Alpha Haemolysin						Mastitis	Significantly reduced virulence	Double mutant is even less virulence	Jonsson *et al* (1985)
Coagulase						Mastitis	Significantly reduced virulence		Jonsson *et al* (1985)
						Endocarditis	None		Morellion *et al* (1995)

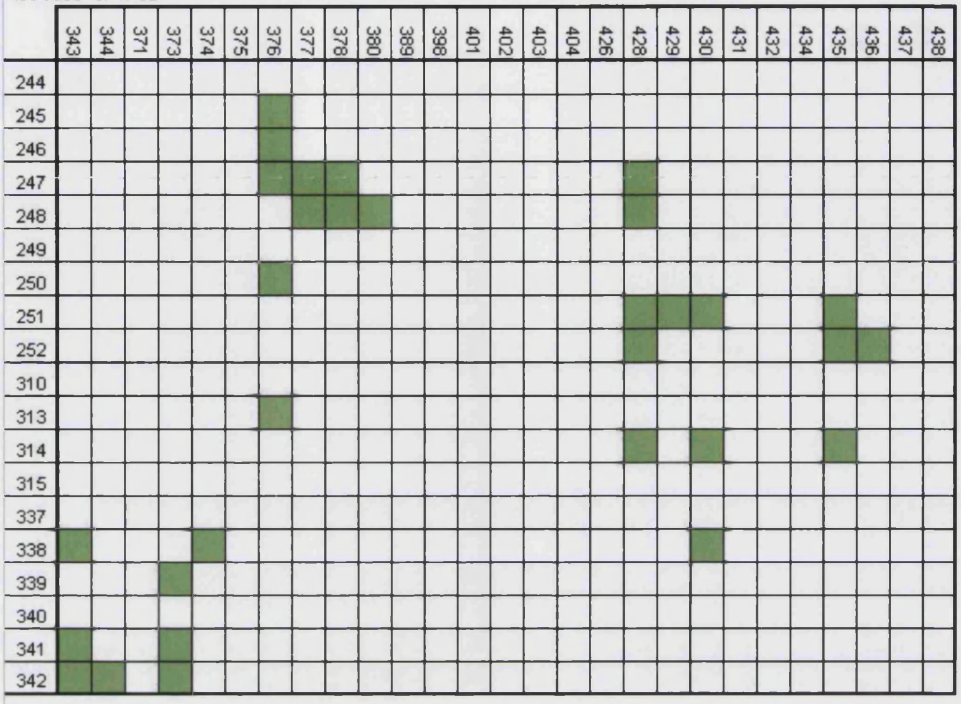
13 Appendix 2

Species of Serum IgG	Concentration (M)	k_s (s^{-1})	Chi squared (χ^2)
Chicken	6.37E-05	3.13E-03	645
	3.19E-05	2.01E-03	97
	2.12E-05	1.63E-03	11
Cow	1.26E-06	7.43E-03	9,240
	6.29E-07	4.31E-03	2,630
	3.15E-07	3.83E-03	11
Goat	1.07E-05	4.17E-03	68,200
	5.33E-06	2.71E-03	21,500
	2.66E-06	7.45E-04	3,230
Horse	9.84E-06	2.17E-03	772
	4.92E-06	1.42E-03	86
	2.46E-06	4.97E-04	117
Human	4.46E-07	1.07E-02	21,600
	2.23E-07	8.49E-03	6,080
	2.23E-07	6.43E-03	2,620
	1.49E-07	5.56E-03	974
	1.12E-07	4.03E-03	204
Guinea Pig	2.68E-07	1.55E-02	23,800
	1.34E-07	1.22E-02	10,200
	6.70E-08	8.92E-03	1,890
Mouse	3.51E-07	3.73E-02	65,800
	2.63E-07	3.19E-02	25,400
	1.76E-07	2.00E-02	4,580
Pig	7.83E-07	8.23E-03	68,100
	3.13E-07	6.77E-03	5,770
	1.57E-07	6.29E-03	14,100
Rabbit	8.67E-07	1.34E-02	10,800
	4.33E-07	7.53E-03	1,810
	1.73E-07	6.05E-03	1,280
Sheep	4.34E-05	3.75E-03	16,900
	2.17E-05	1.82E-03	4,050
	1.45E-05	1.36E-03	3,150

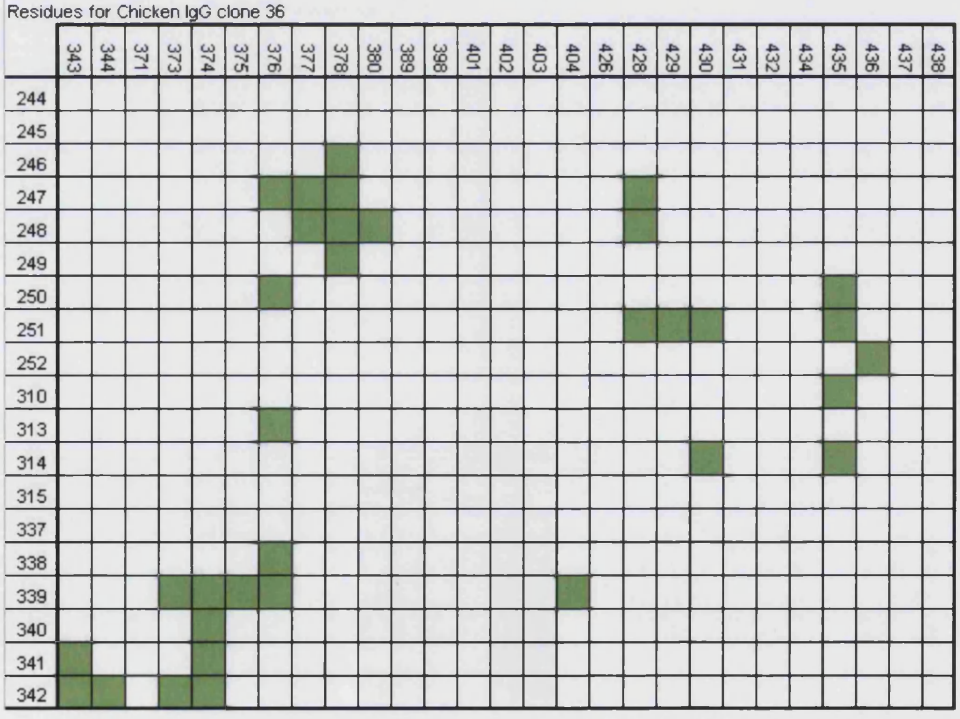
14 Appendix 3

Interaction patterns between the C_H2 and C_H3 domains of the IgG models. CONTACT was used to calculate residues within 5.2Å of a residue in the other C_H domain. Residues in the C_H2 domain (244 to 342) interacting with a residue in the C_H3 domain (343 to 438) are marked with a green square.

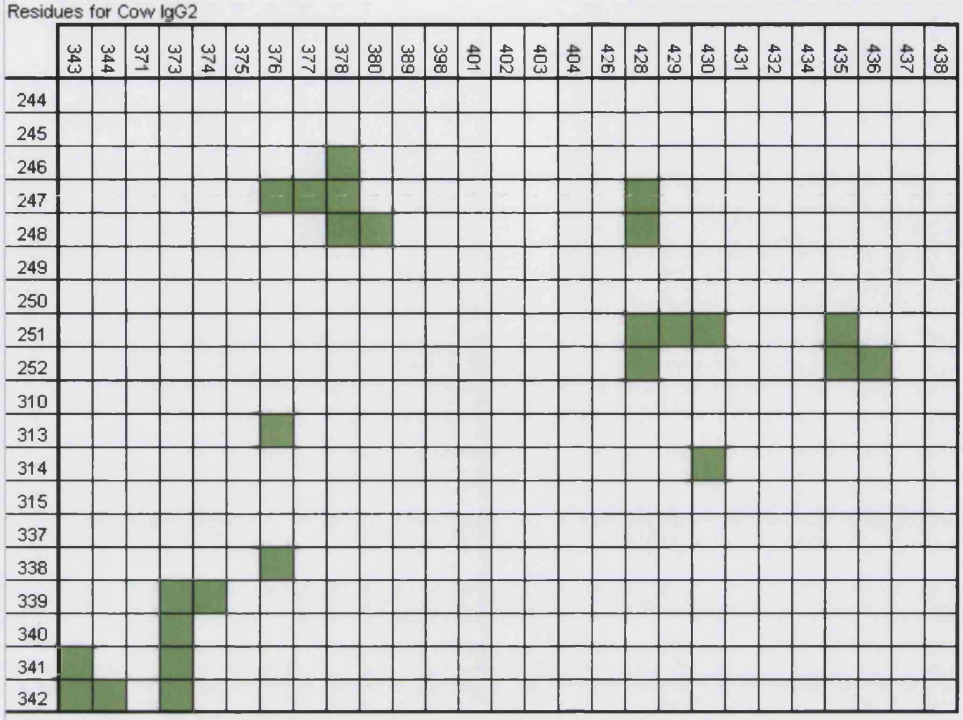
Residues for 1FC2



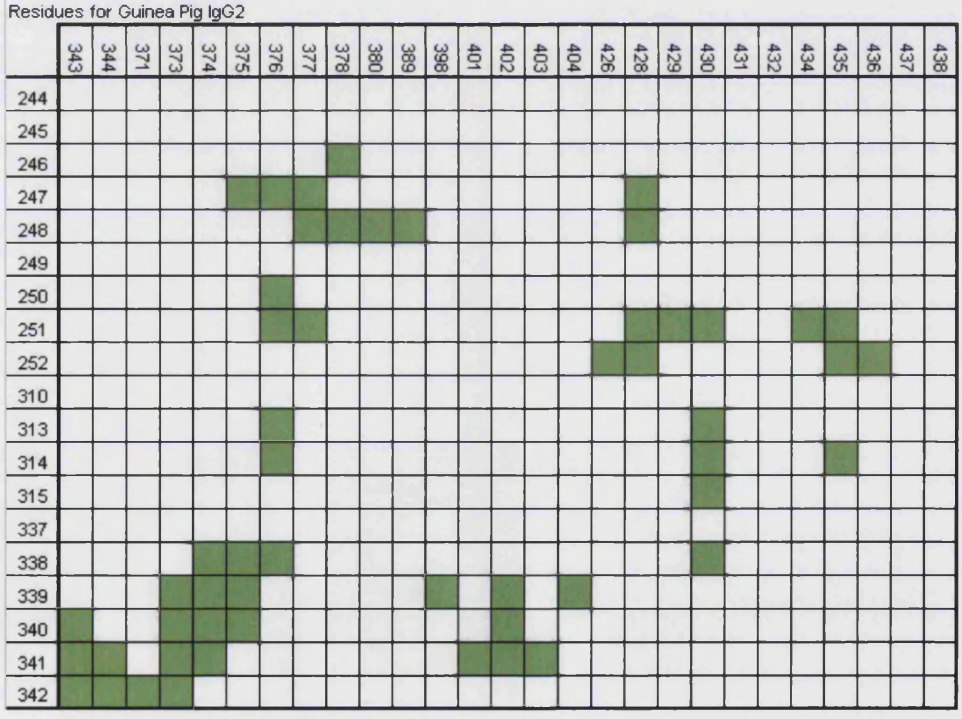
Residues for Chicken IgG clone 36



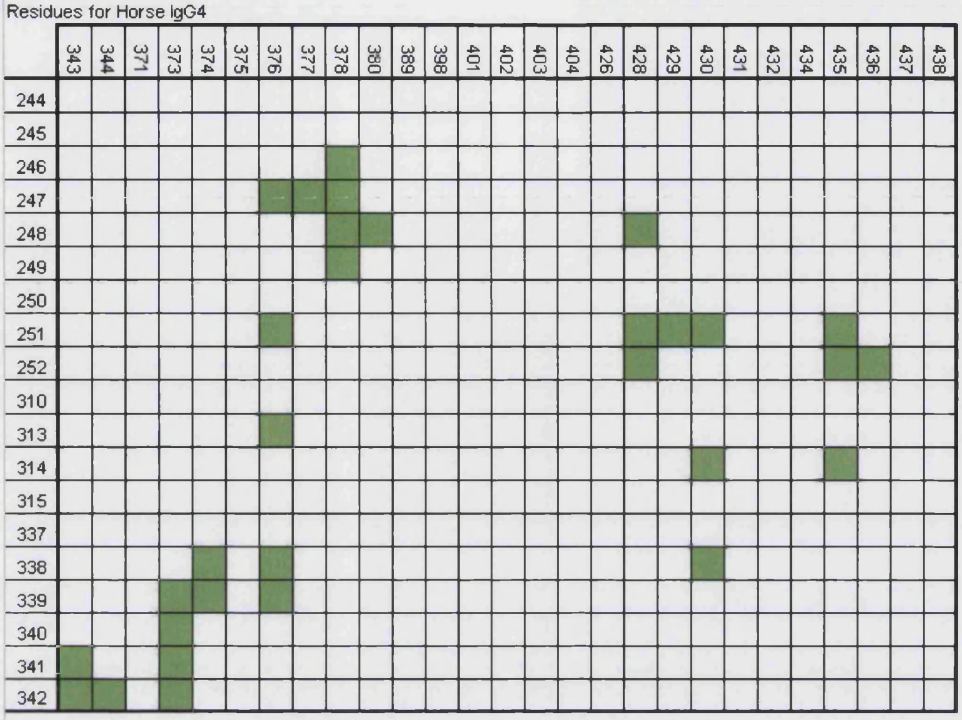
Residues for Cow IgG2



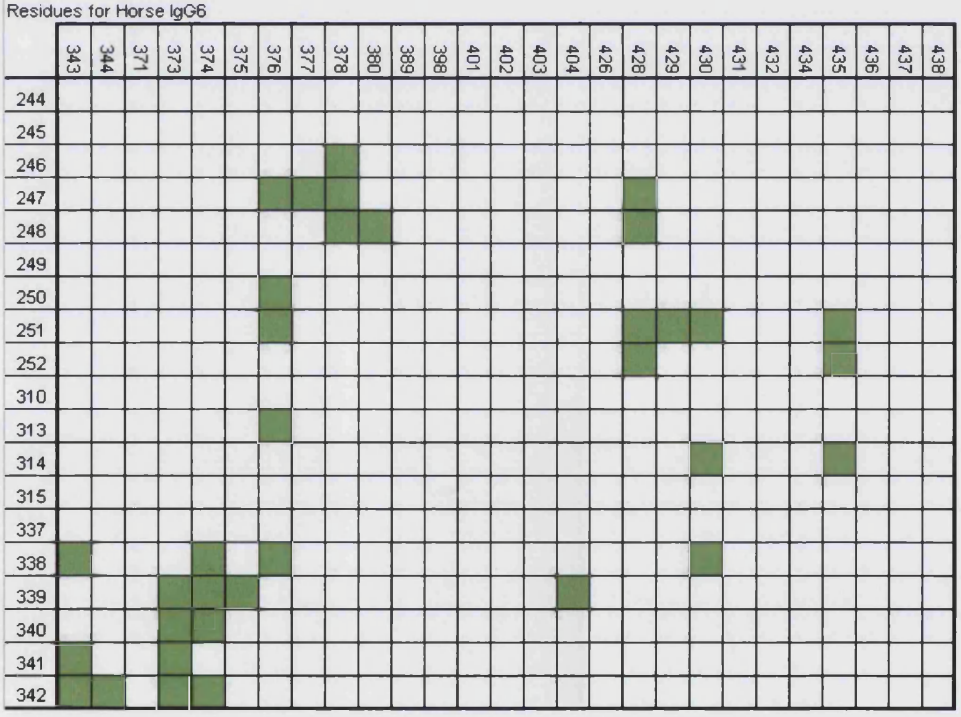
Residues for Guinea Pig IgG2



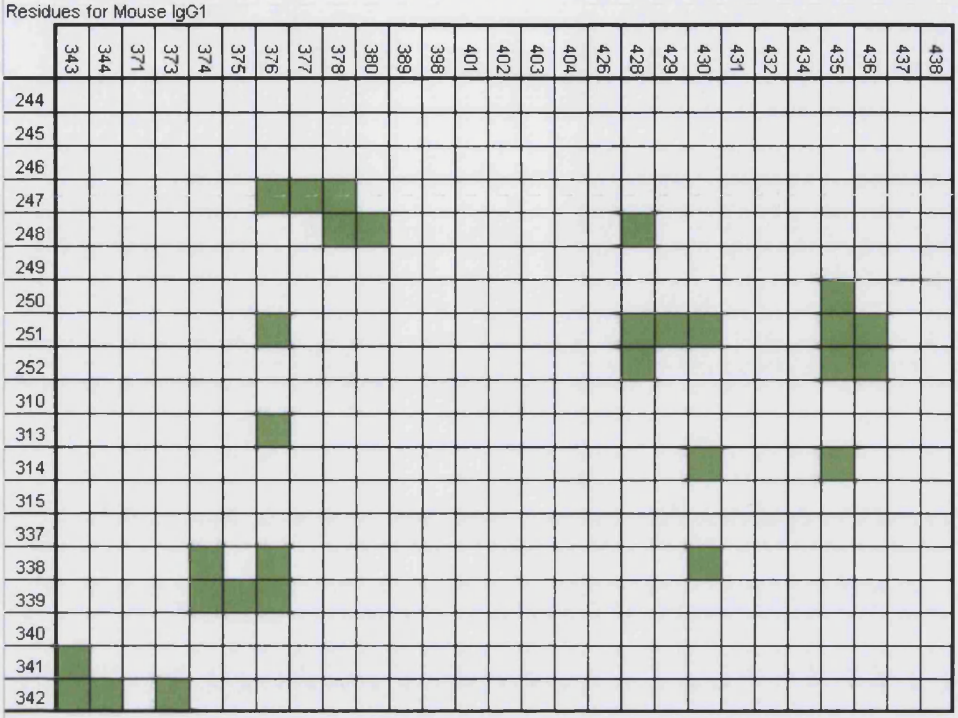
Residues for Horse IgG4



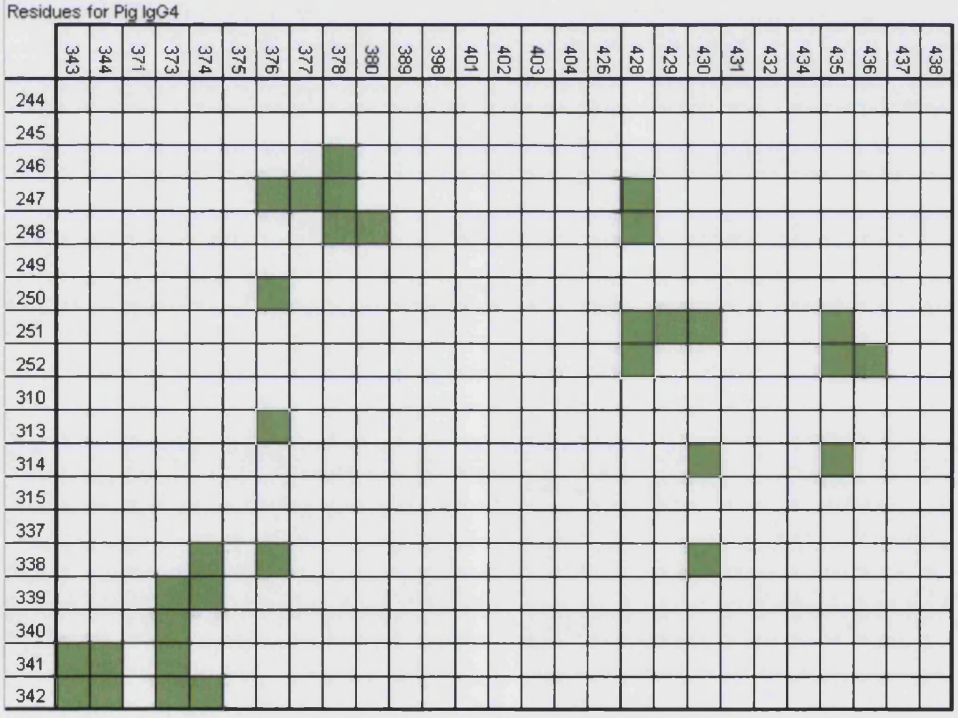
Residues for Horse IgG6



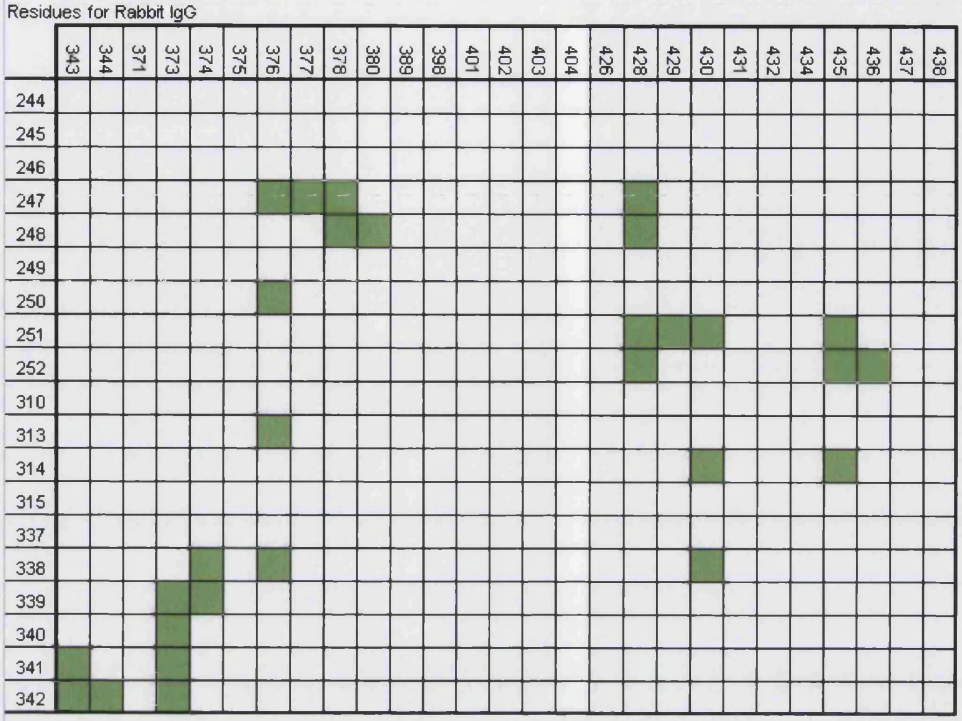
Residues for Mouse IgG1



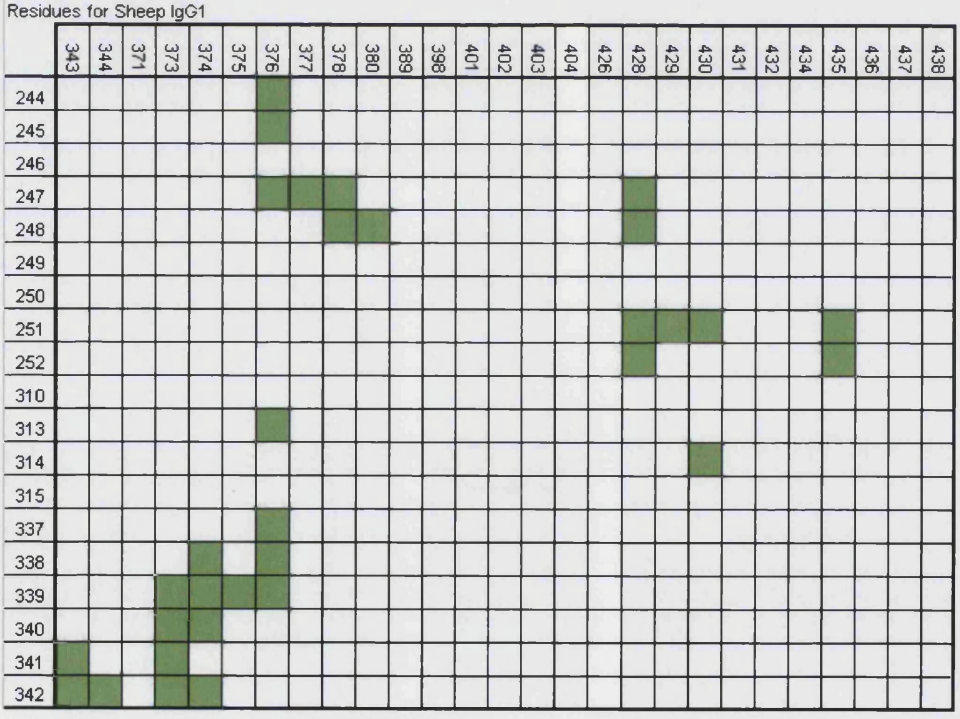
Residues for Pig IgG4



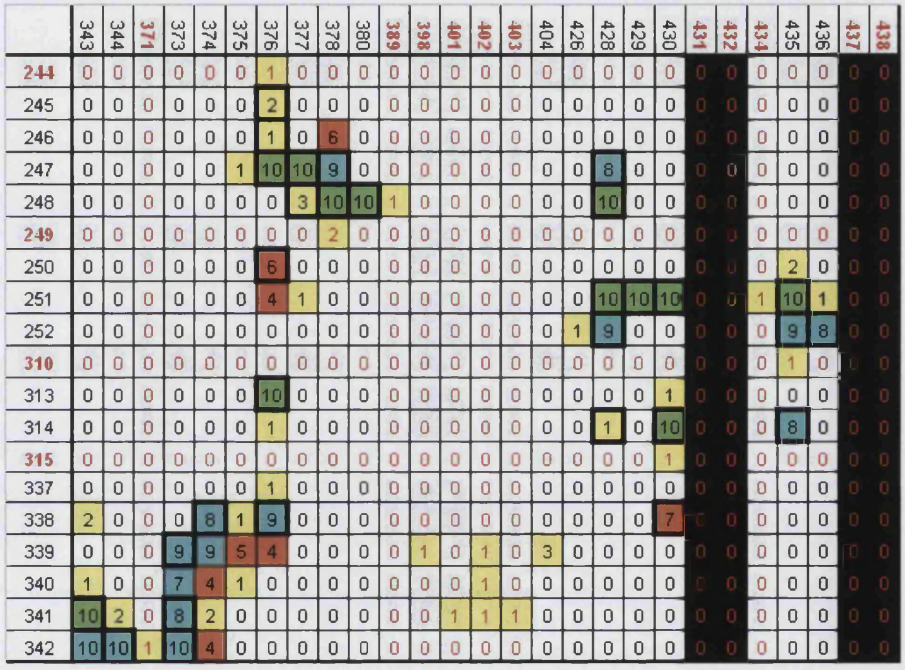
Residues for Rabbit IgG



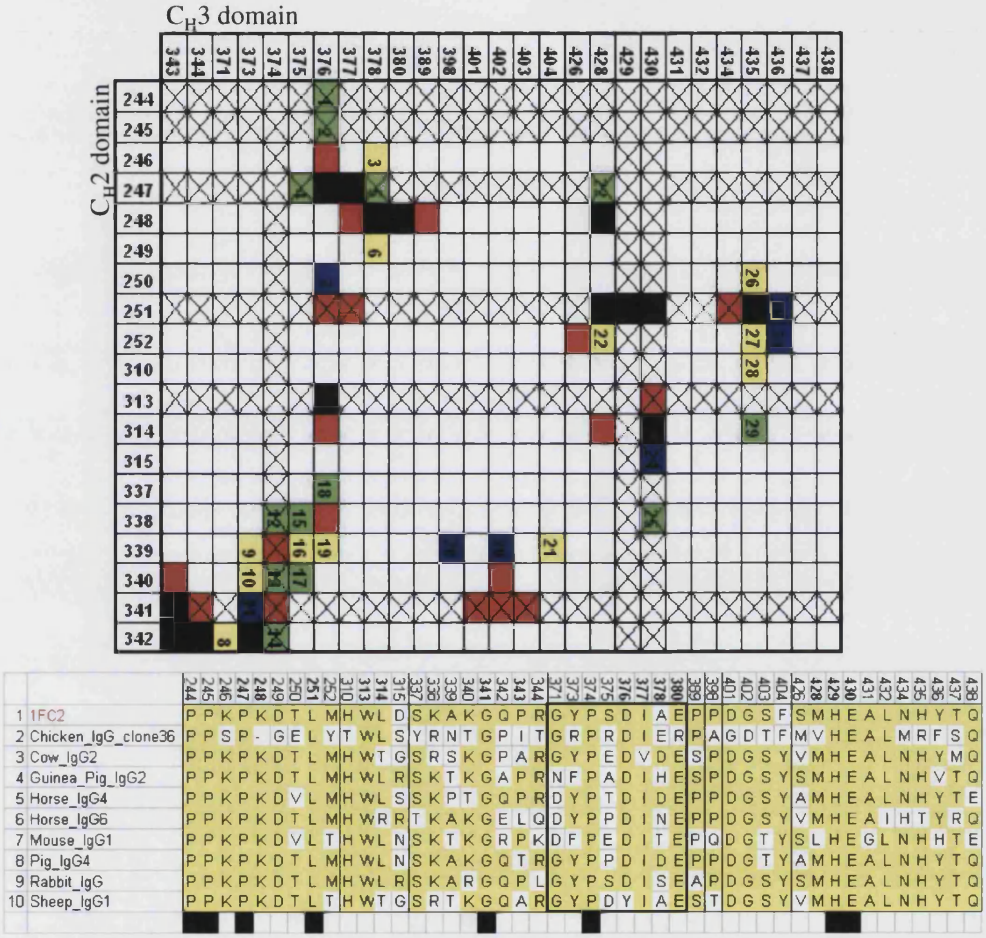
Residues for Sheep IgG1



Summary of the interactions between the C_H2 and C_H3. The number indicates the number of times an interaction between two residues is detected. Residue numbers in red are residues that are conversed in all models



Summary of interactions observed above between the IgG C_H2 and C_H3 domains. The numbered squared refer to the interactions listed in Table 9. Interactions detected in all models are marked in black. Changed interaction pattern: marked in red do not correspond to changes in sequence; blue corresponds to changes in the C_H2 domain; green to changes in the C_H3 domain; yellow to changes in both domains.



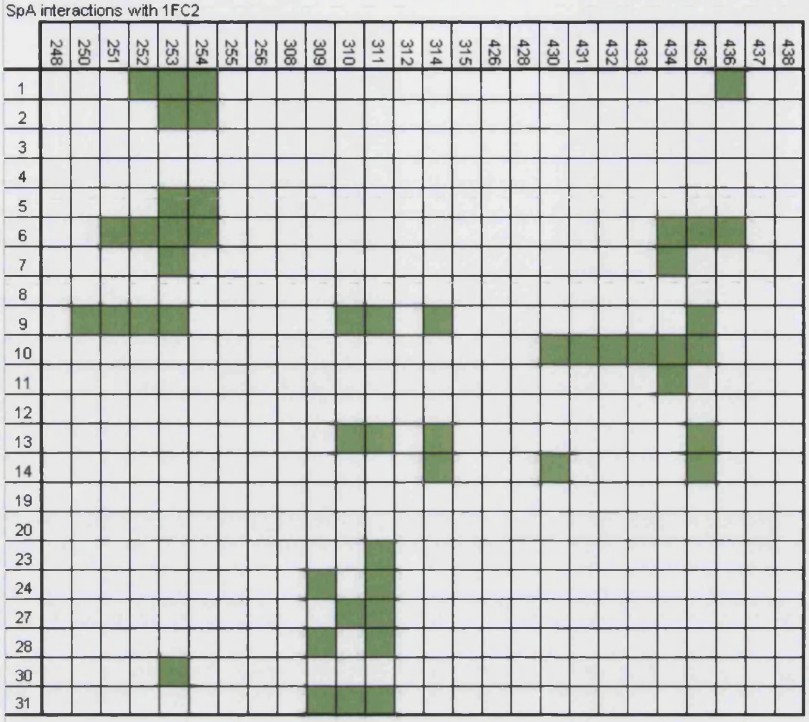
		244	245	246	247	248	249	250	251	252	310	313	314	315	337	339	340	341	342	343	344	371	373	374	375	376	377	378	380	389	398	401	402	403	404	426	428	429	430	431	432	434	435	436	437	438
1	1FC2	P	P	K	P	K	D	T	L	M	H	W	L	D	S	K	A	K	G	Q	P	R	G	Y	P	S	D	I	A	E	P	P	D	G	S	F	S	M	H	E	A	L	N	H	Y	T	Q
2	Chicken_IgG_clone36	P	P	S	P	-	G	E	L	Y	T	W	L	S	Y	R	N	T	G	P	I	T	G	R	P	R	D	I	E	R	P	A	G	D	T	F	M	V	H	E	A	L	M	R	F	S	Q
3	Cow_IgG2	P	P	K	P	K	D	T	L	M	H	W	T	G	S	R	S	K	G	P	A	R	G	Y	P	E	D	V	D	E	S	P	D	G	S	Y	V	M	H	E	A	L	N	H	Y	M	Q
4	Guinea_Pig_IgG2	P	P	K	P	K	D	T	L	M	H	W	L	R	S	K	T	K	G	A	P	R	N	F	P	A	D	I	H	E	S	P	D	G	S	Y	S	M	H	E	A	L	N	H	V	T	Q
5	Horse_IgG4	P	P	K	P	K	D	V	L	M	H	W	L	S	S	K	P	T	G	Q	P	R	D	Y	P	T	D	I	D	E	P	P	D	G	S	Y	A	M	H	E	A	L	N	H	Y	T	E
6	Horse_IgG6	P	P	K	P	K	D	T	L	M	H	W	R	R	T	K	A	K	G	E	L	Q	D	Y	P	P	D	I	N	E	P	P	D	G	S	Y	V	M	H	E	A	I	H	T	Y	R	Q
7	Mouse_IgG1	P	P	K	P	K	D	V	L	T	H	W	L	N	S	K	T	K	G	R	P	K	D	F	P	E	D	I	T	E	P	Q	D	G	T	Y	S	L	H	E	G	L	N	H	H	T	E
8	Pig_IgG4	P	P	K	P	K	D	T	L	M	H	W	L	N	S	K	A	K	G	Q	T	R	G	Y	P	P	D	I	D	E	P	P	D	G	T	Y	A	M	H	E	A	L	N	H	Y	T	Q
9	Rabbit_IgG	P	P	K	P	K	D	T	L	M	H	W	L	R	S	K	A	R	G	Q	P	L	G	Y	P	S	D	I	S	E	A	P	D	G	S	Y	S	M	H	E	A	L	N	H	Y	T	Q
10	Sheep_IgG1	P	P	K	P	K	D	T	L	T	H	W	T	G	S	R	T	K	G	Q	A	R	G	Y	P	D	Y	I	A	E	S	T	D	G	S	Y	V	M	H	E	A	L	N	H	Y	T	Q

Sequence of the IgG models in the regions identified to be involved in interacting with the other domain. Conserved residues are highlighted in yellow. Residues conserved in all sequences have a black mark under them.

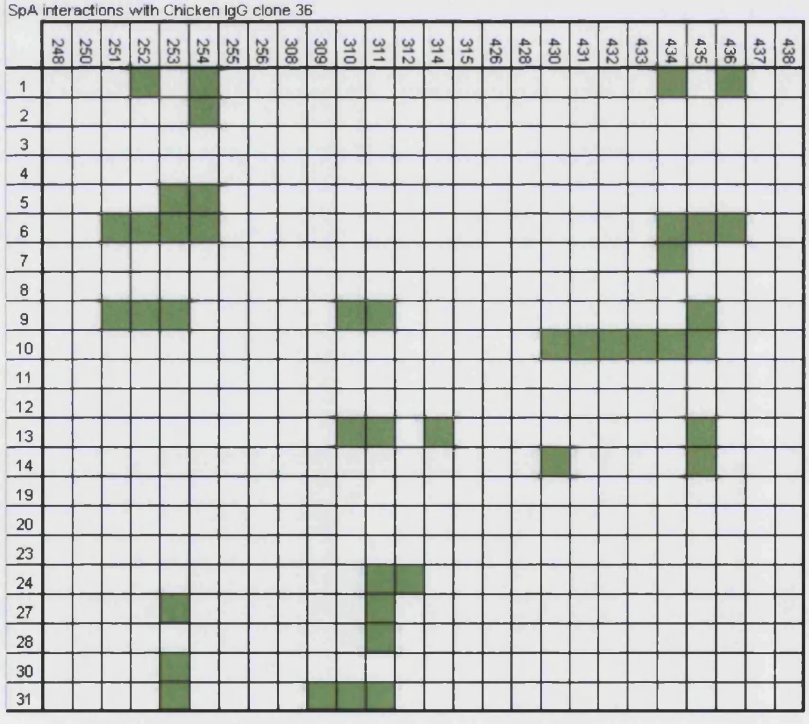
15 Appendix 4

Interaction patterns for SpA with the IgG models. CONTACT was used to calculate residues between 1.8 Å and 5.2Å of a residue in the other half of the complex. Residues in the SpA domain (1 to 31) interacting with a residue in the IgGmodel (248 to 438) are marked with a green square.

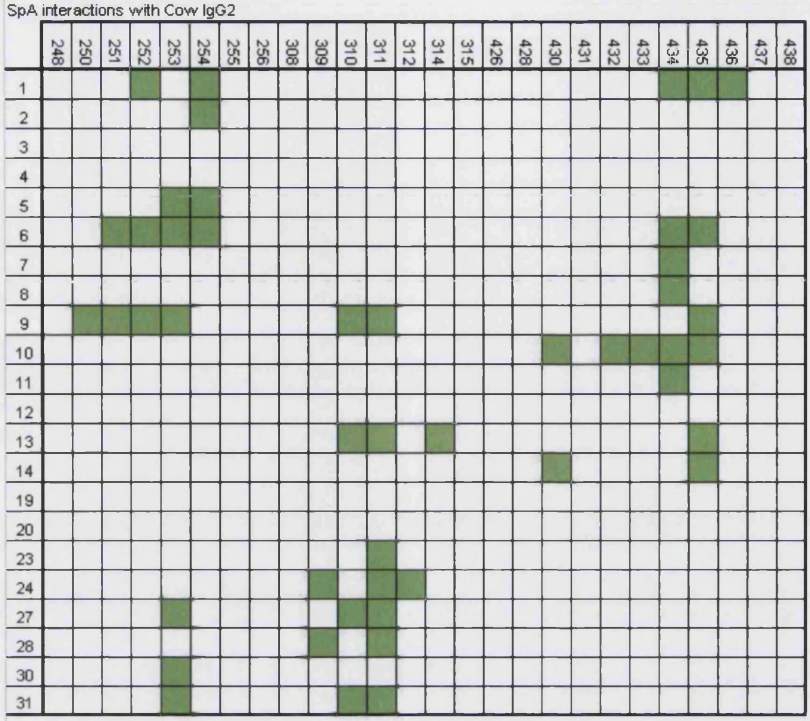
SpA interactions with 1FC2



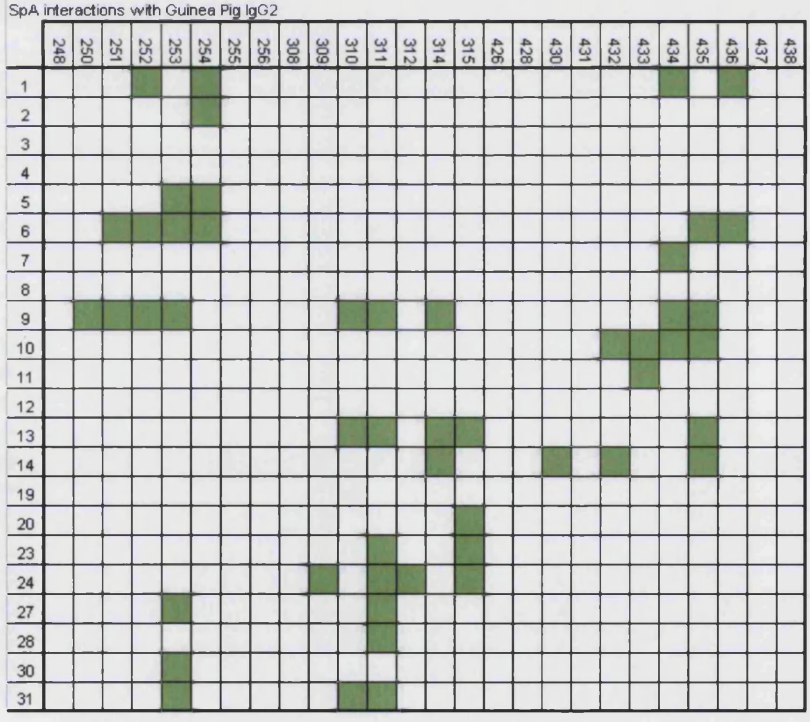
SpA interactions with Chicken IgG clone 36



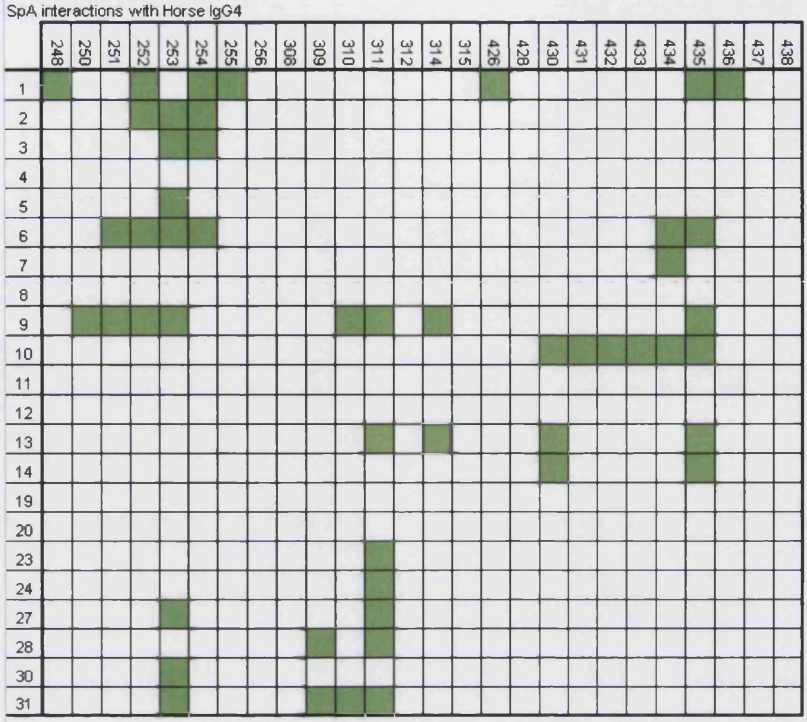
SpA interactions with Cow IgG2



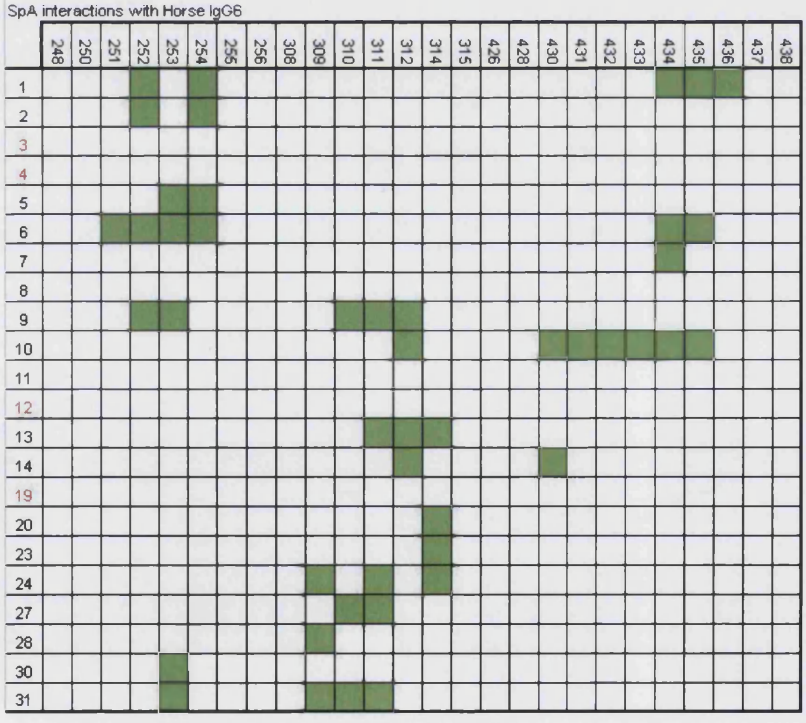
SpA interactions with Guinea Pig IgG2



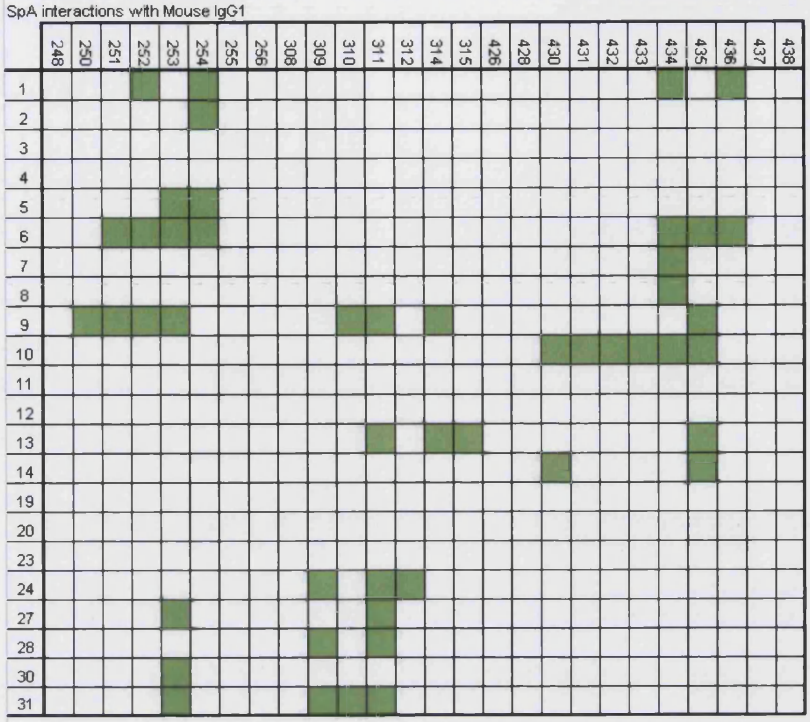
SpA interactions with Horse IgG4



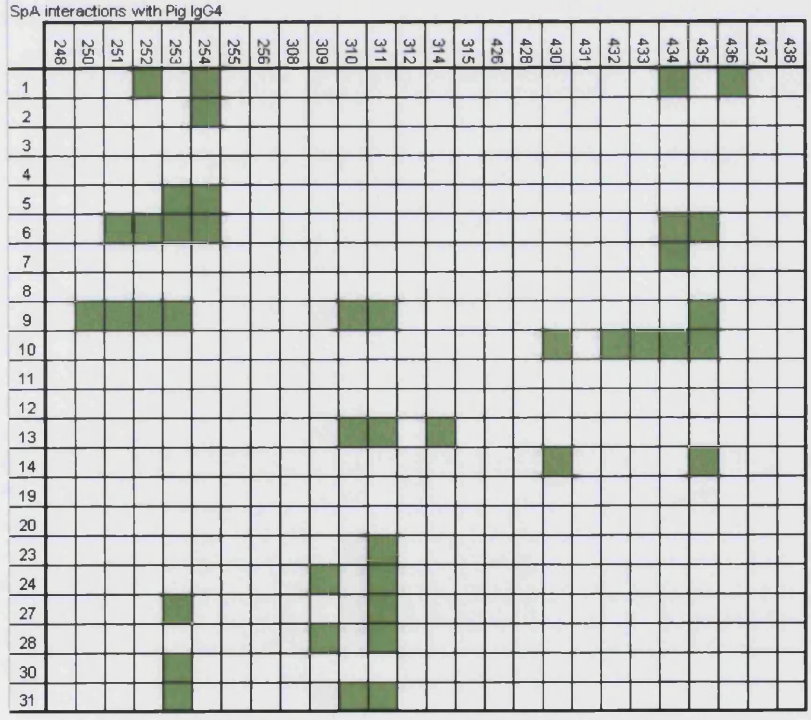
SpA interactions with Horse IgG6



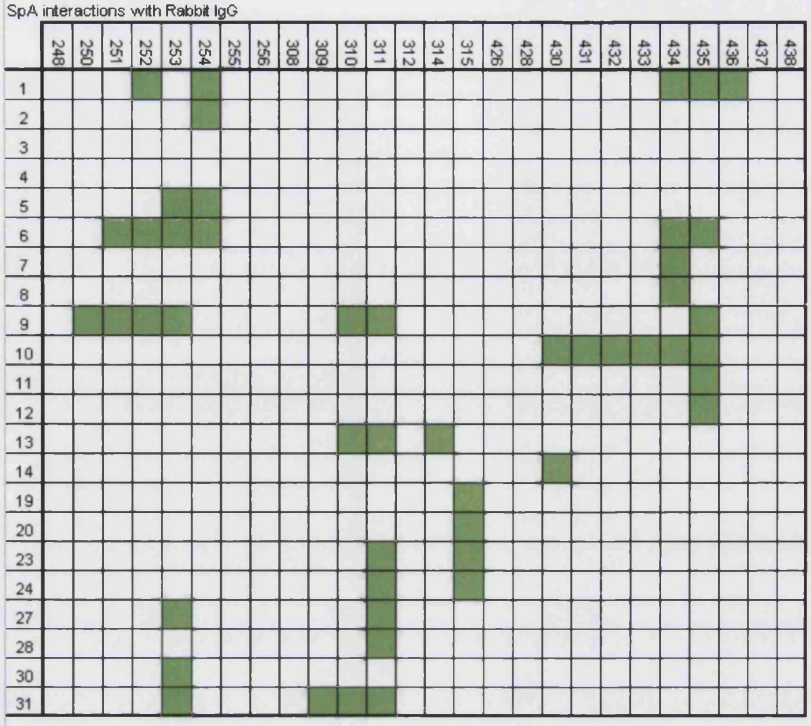
SpA interactions with Mouse IgG1



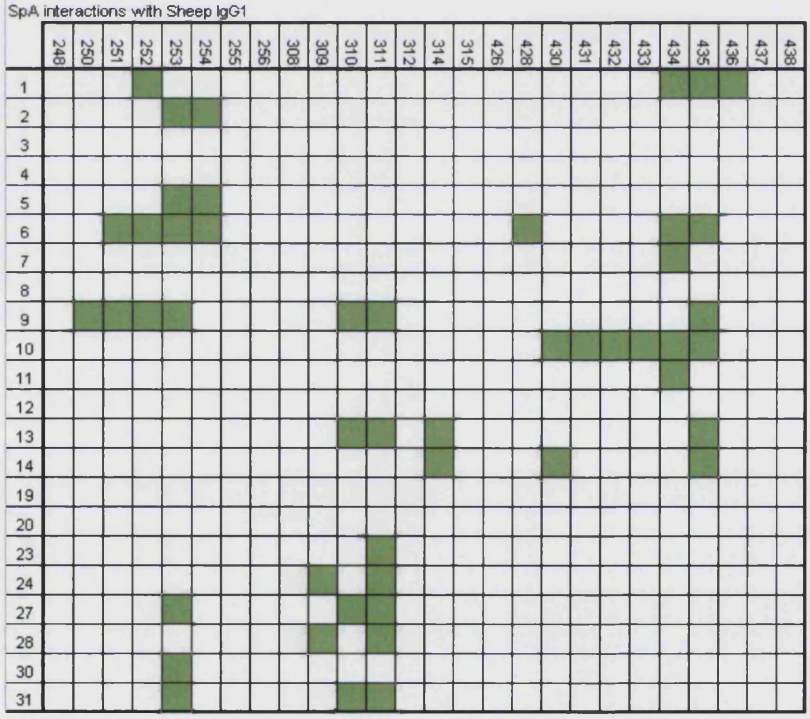
SpA interactions with Pig IgG4



SpA interactions with Rabbit IgG



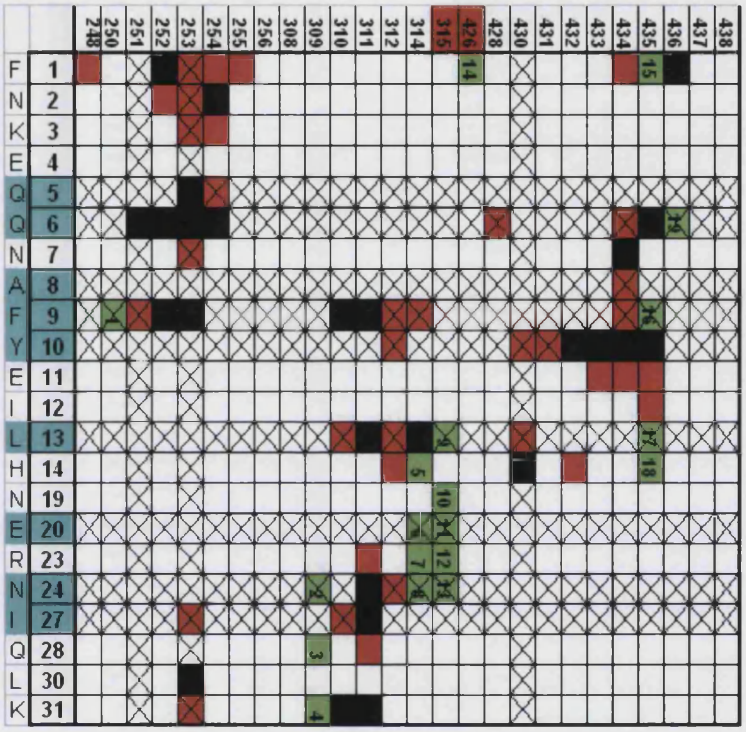
SpA interactions with Sheep IgG1



Summary of the interactions between SpA and the IgG models. The number indicates the number of times an interaction between two residues is detected. Residue numbers in red are residues that are conversed in all models

	248	250	251	252	253	254	255	256	308	309	310	311	312	314	315	426	428	430	431	432	433	434	435	436	437	438
1	1	0	0	10	1	9	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	8	5	10	0	0
2	0	0	0	2	3	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	10	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	10	10	10	10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	9	10	4	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
9	0	8	9	10	10	0	0	0	0	0	10	10	1	4	0	0	0	0	0	0	0	1	9	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	9	7	10	10	10	10	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
13	0	0	0	0	0	0	0	0	0	0	7	10	1	10	2	0	0	1	0	0	0	0	7	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0	10	0	1	0	0	8	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	7	0	1	2	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	7	0	10	4	1	2	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	8	0	0	0	0	0	4	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	7	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	9	0	0	0	0	6	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Summary of interactions observed above between SpA and the IgG models. The numbered squared refer to the interactions listed in Table 10. Interactions detected in all models are marked in black. Changed interaction pattern: marked in red do not correspond to changes in sequence; green indicates where a substitution occurs in the IgG sequences. The sequence for SpA is given.



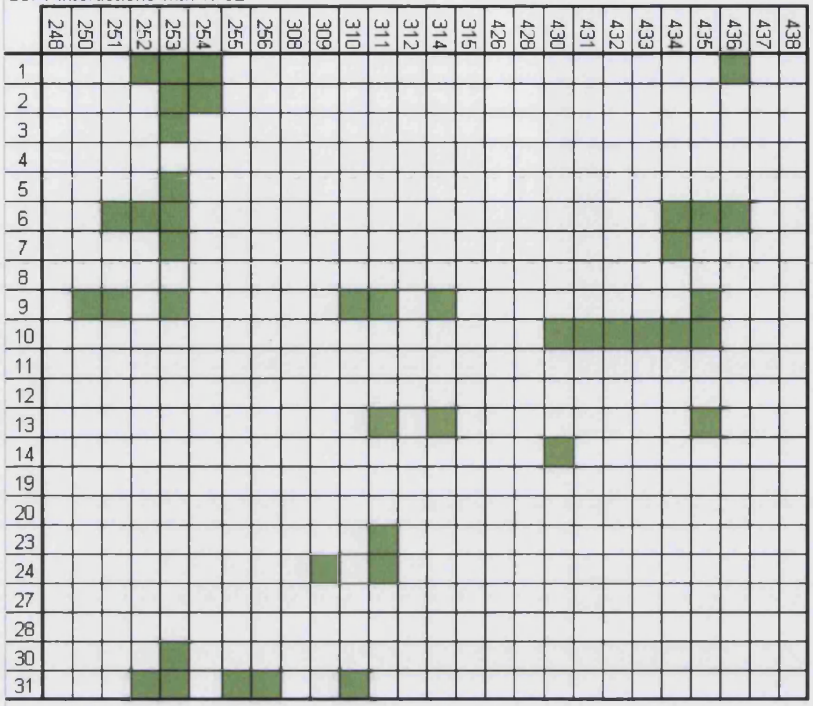
Sequence of the IgG models in the regions identified to be involved in interacting with SpA. Conserved residues are highlighted in yellow. Residues conserved in all sequences have a black mark under them.

		248	250	251	252	253	254	255	256	308	309	310	311	312	314	315	426	428	430	431	432	433	434	435	436	437	438
1	1FC2	K	T	L	M	I	S	R	T	V	L	H	Q	N	L	D	S	M	E	A	L	H	N	H	Y	T	Q
2	Chicken_IgG_clone36	-	E	L	Y	I	S	L	D	V	S	T	Q	D	L	S	M	V	E	A	L	P	M	R	F	S	Q
3	Cow_IgG2	K	T	L	M	I	T	G	T	I	Q	H	Q	D	T	G	V	M	E	A	L	H	N	H	Y	M	Q
4	Guinea_Pig_IgG2	K	T	L	M	I	S	L	T	I	Q	H	Q	D	L	R	S	M	E	A	L	H	N	H	V	T	Q
5	Horse_IgG4	K	V	L	M	I	S	R	T	I	Q	H	K	D	L	S	A	M	E	A	L	H	N	H	Y	T	E
6	Horse_IgG6	K	T	L	M	I	S	R	T	I	Q	H	Q	D	R	R	V	M	E	A	I	R	H	T	Y	R	Q
7	Mouse_IgG1	K	V	L	T	I	T	L	T	I	M	H	Q	D	L	N	S	L	E	G	L	H	N	H	H	T	E
8	Pig_IgG4	K	T	L	M	I	S	R	T	I	Q	H	Q	D	L	N	A	M	E	A	L	H	N	H	Y	T	Q
9	Rabbit_IgG	K	T	L	M	I	S	R	T	I	T	H	Q	D	L	R	S	M	E	A	L	H	N	H	Y	T	Q
10	Sheep_IgG1	K	T	L	T	I	S	G	T	I	Q	H	Q	D	T	G	V	M	E	A	L	H	N	H	Y	T	Q

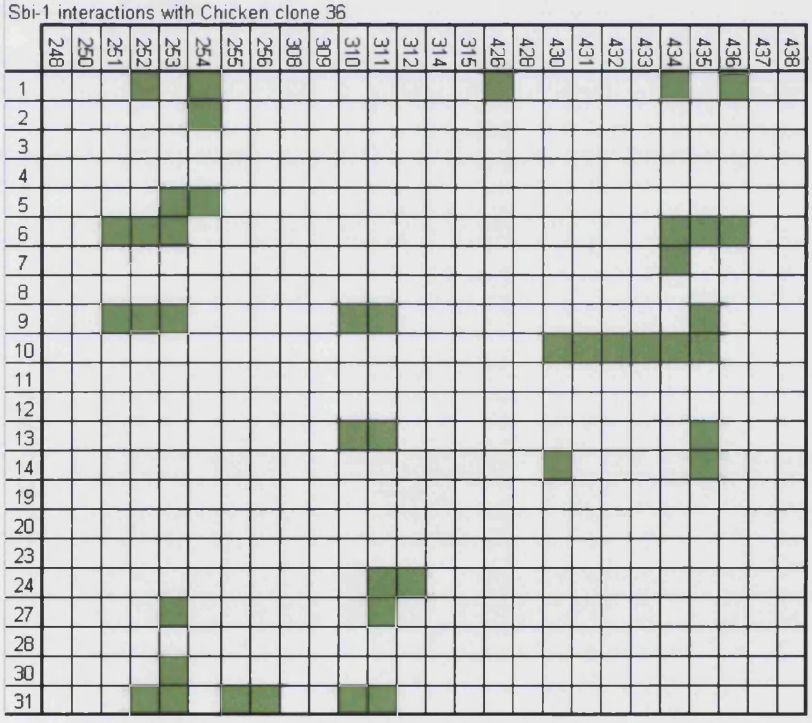
16 Appendix 5

Interaction patterns for Sbi-1 with the IgG models. CONTACT was used to calculate residues between 1.8 Å and 5.2Å of a residue in the other half of the complex. Residues in the SpA domain (1 to 31) interacting with a residue in the IgGmodel (248 to 438) are marked with a green square.

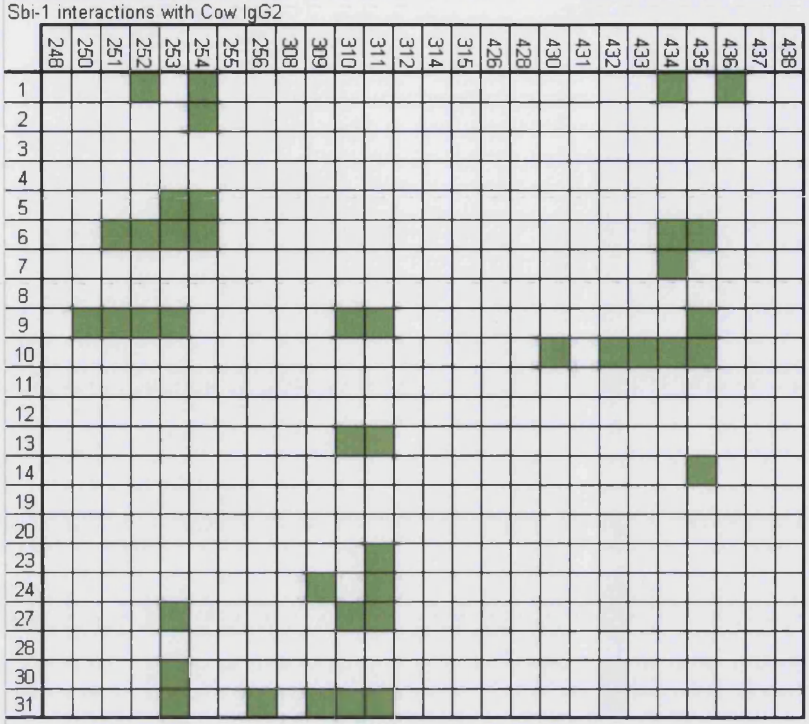
Sbi-1 interactions with 1FC2



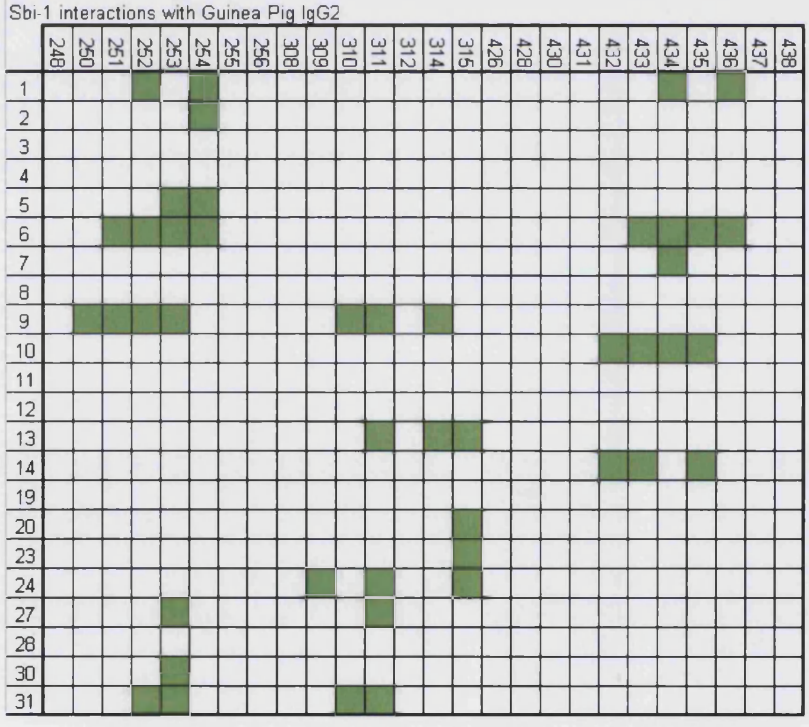
Sbi-1 interactions with Chicken clone 36



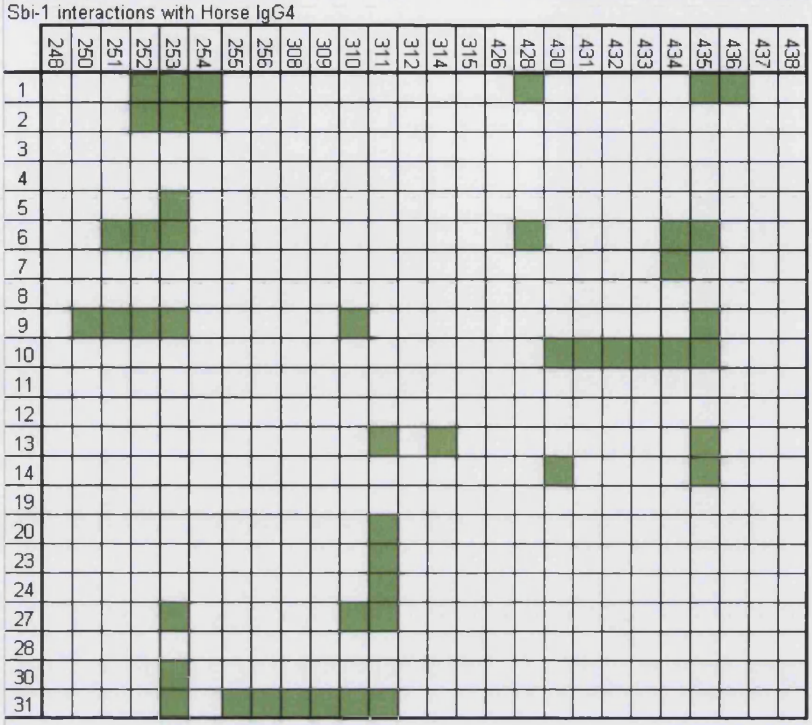
Sbi-1 interactions with Cow IgG2



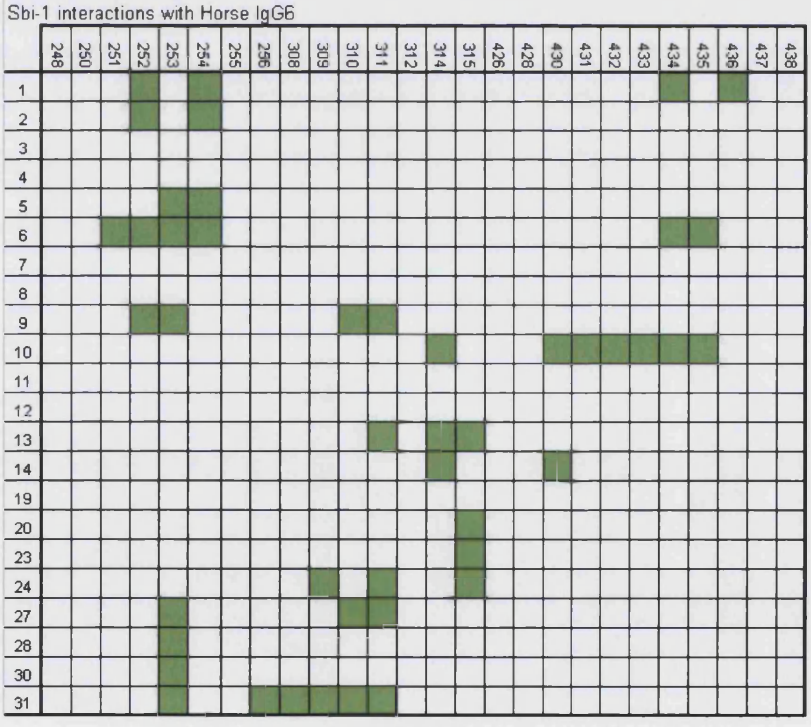
Sbi-1 interactions with Guinea Pig IgG2



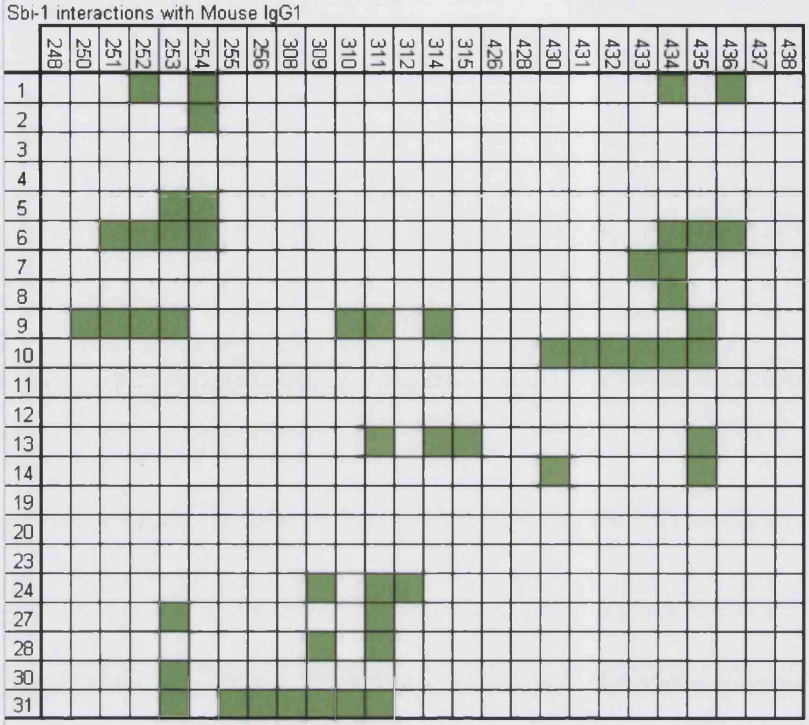
Sbi-1 interactions with Horse IgG4



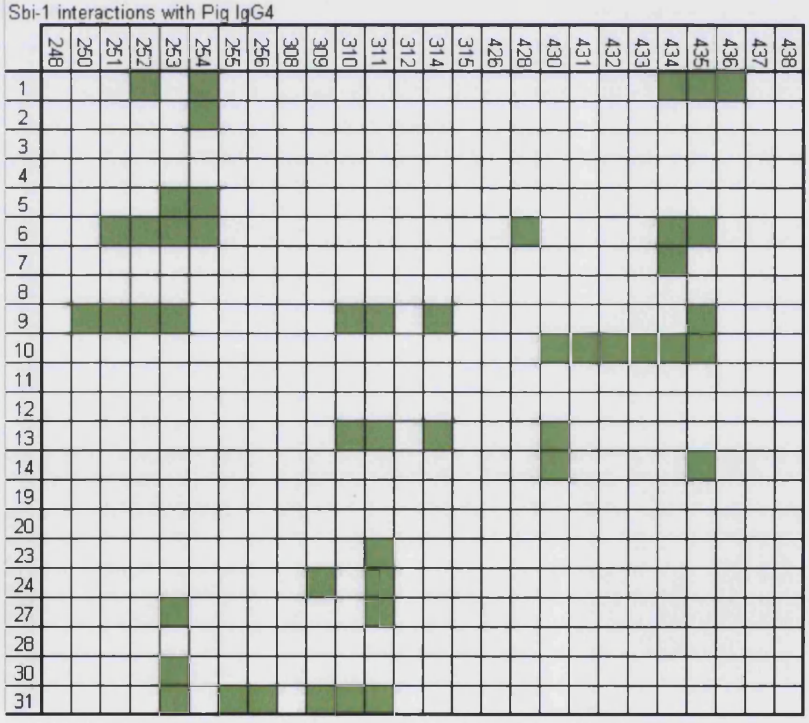
Sbi-1 interactions with Horse IgG6



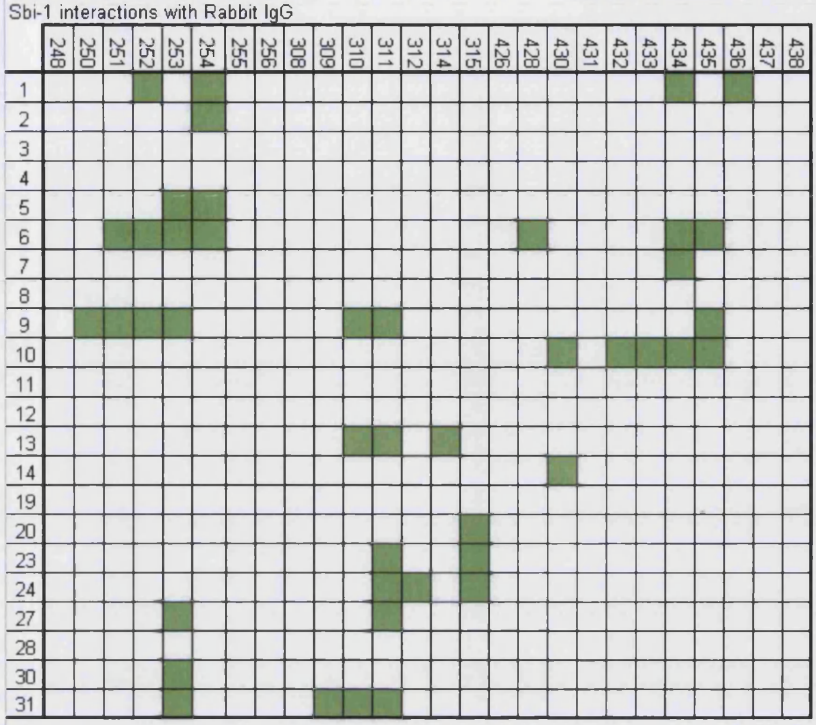
Sbi-1 interactions with Mouse IgG1



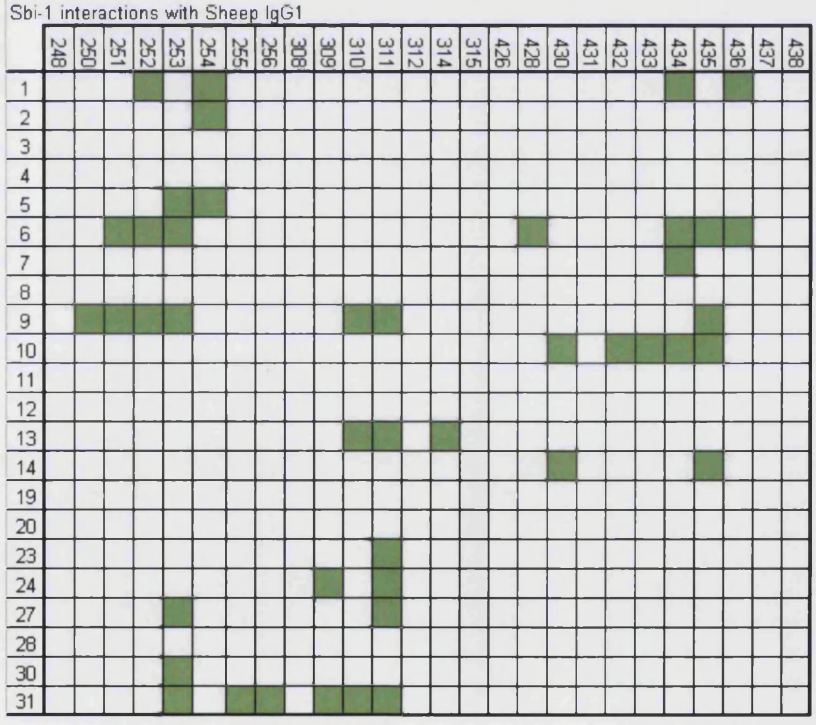
Sbi-1 interactions with Pig IgG4



Sbi-1 interactions with Rabbit IgG



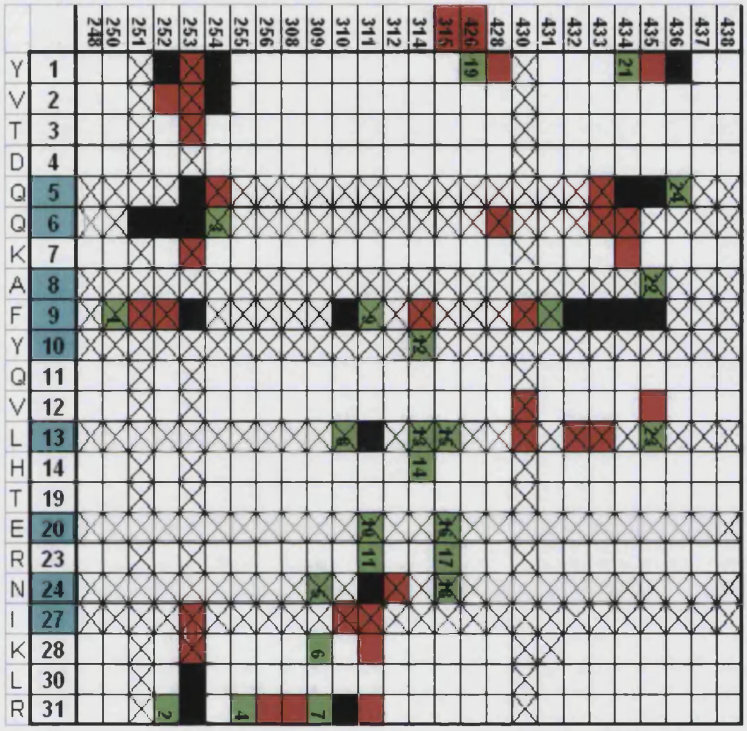
Sbi-1 interactions with Sheep IgG1



Summary of the interactions between Sbi-1 and IgG models. The number indicates the number of times an interaction between two residues is detected. Residue numbers in red are residues that are conversed in all models

	248	250	251	252	253	254	255	256	308	309	310	311	312	314	315	426	428	430	431	432	433	434	435	436	437	438
1	0	0	0	10	2	10	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	8	2	10	0	0
2	0	0	0	2	2	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	10	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	10	10	10	6	0	0	0	0	0	0	0	0	0	0	4	0	0	0	1	10	10	5	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
9	0	8	9	9	10	0	0	0	0	0	10	9	0	4	0	0	0	0	0	0	0	0	8	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	9	6	10	10	10	10	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	5	10	0	8	3	0	0	1	0	0	0	0	4	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	8	0	1	1	0	7	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	6	0	0	3	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	7	0	10	3	0	3	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	9	0	0	0	0	0	3	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	3	10	0	6	8	3	7	10	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Summary of interactions observed above between Sbi-1 and the IgG models. The numbered squared refer to the interactions listed in Table 11. Interactions detected in all models are marked in black. Changed interaction pattern: marked in red do not correspond to changes in sequence; green indicates where a substitution occurs in the IgG sequences. The sequence for Sbi-1 is given.



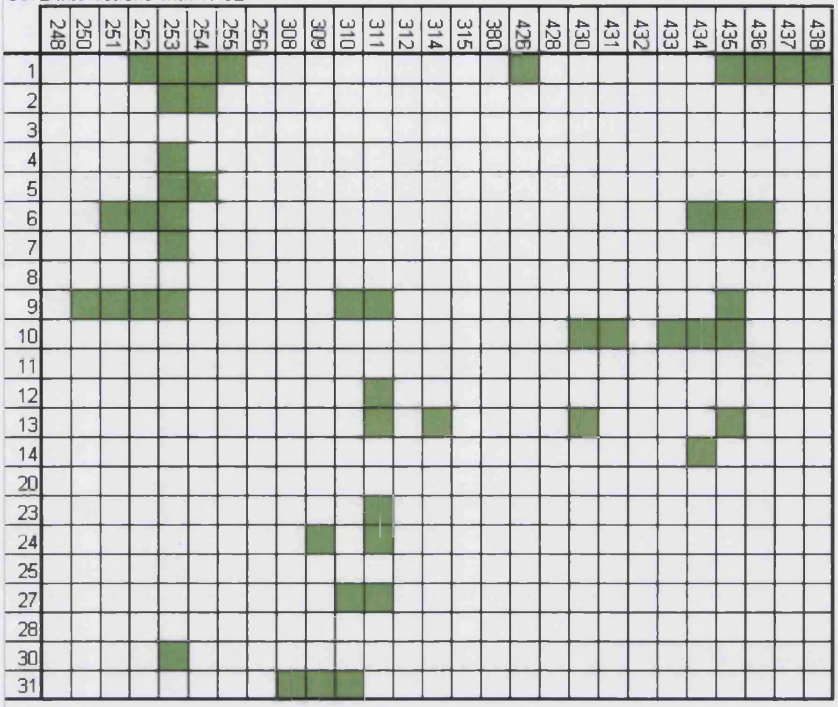
Sequence of the IgG models in the regions identified to be involved in interacting with the other domain. Conserved residues are highlighted in yellow. Residues conserved in all sequences have a black mark under them.

		248	250	251	252	253	254	255	256	308	309	310	311	312	314	315	426	428	430	431	432	433	434	435	436	437	438
1	1FC2	K	T	L	M	I	S	R	T	V	L	H	Q	N	L	D	S	M	E	A	L	H	N	H	Y	T	Q
2	Chicken_IgG_clone36	-	E	L	Y	I	S	L	D	V	S	T	Q	D	L	S	M	V	E	A	L	P	M	R	F	S	Q
3	Cow_IgG2	K	T	L	M	I	T	G	T	I	Q	H	Q	D	T	G	V	M	E	A	L	H	N	H	Y	M	Q
4	Guinea_Pig_IgG2	K	T	L	M	I	S	L	T	I	Q	H	Q	D	L	R	S	M	E	A	L	H	N	H	V	T	Q
5	Horse_IgG4	K	V	L	M	I	S	R	T	I	Q	H	K	D	L	S	A	M	E	A	L	H	N	H	Y	T	E
6	Horse_IgG6	K	T	L	M	I	S	R	T	I	Q	H	Q	D	R	R	V	M	E	A	I	R	H	T	Y	R	Q
7	Mouse_IgG1	K	V	L	T	I	T	L	T	I	M	H	Q	D	L	N	S	L	E	G	L	H	N	H	H	T	E
8	Pig_IgG4	K	T	L	M	I	S	R	T	I	Q	H	Q	D	L	N	A	M	E	A	L	H	N	H	Y	T	Q
9	Rabbit_IgG	K	T	L	M	I	S	R	T	I	T	H	Q	D	L	R	S	M	E	A	L	H	N	H	Y	T	Q
10	Sheep_IgG1	K	T	L	T	I	S	G	T	I	Q	H	Q	D	T	G	V	M	E	A	L	H	N	H	Y	T	Q

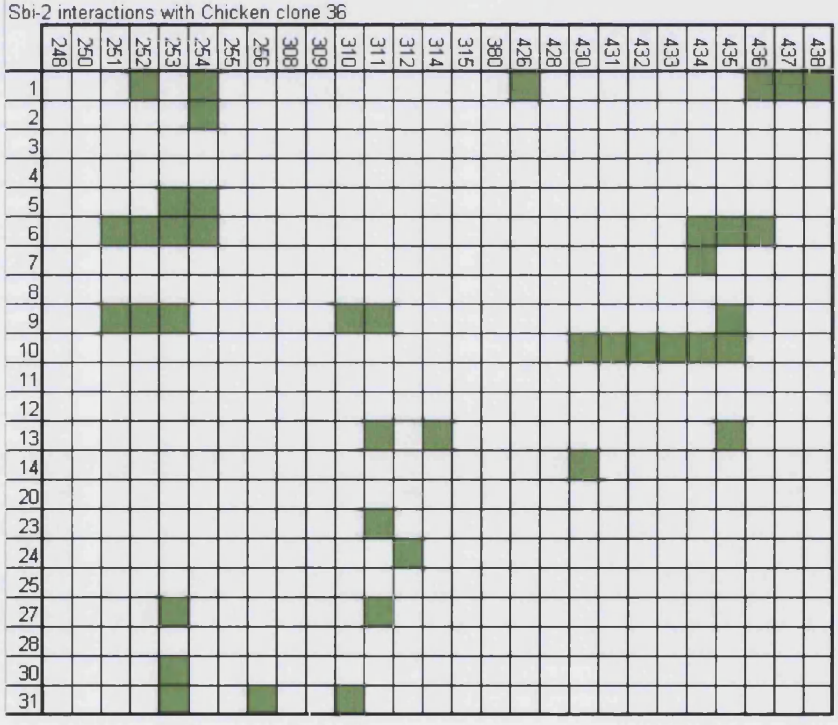
17 Appendix 6

Interaction patterns for Sbi-2 with the IgG models. CONTACT was used to calculate residues between 1.8 Å and 5.2Å of a residue in the other half of the complex. Residues in the Sbi-2 domain (1 to 31) interacting with a residue in the IgGmodel (248 to 438) are marked with a green square.

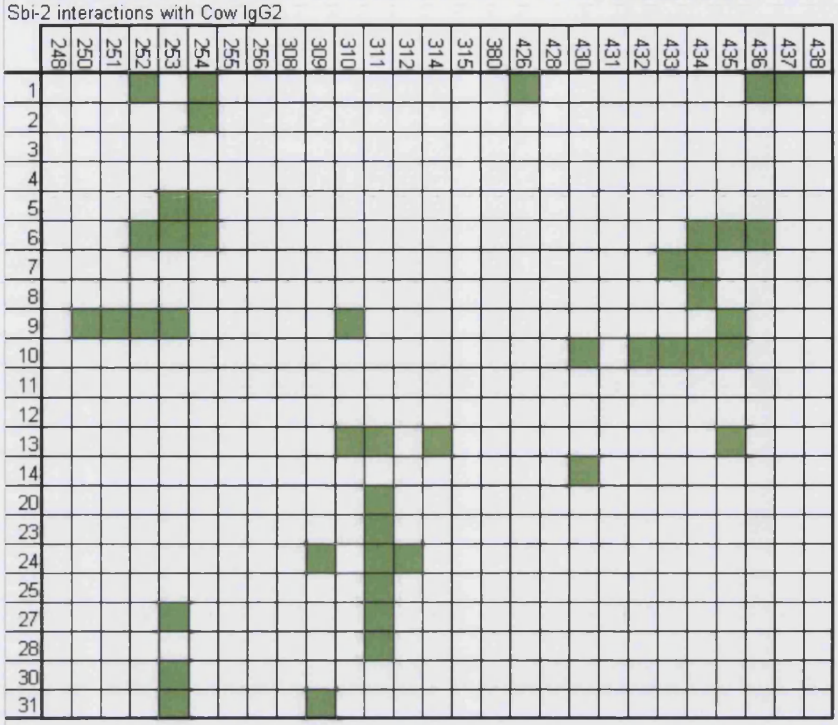
Sbi-2 interactions with 1FC2



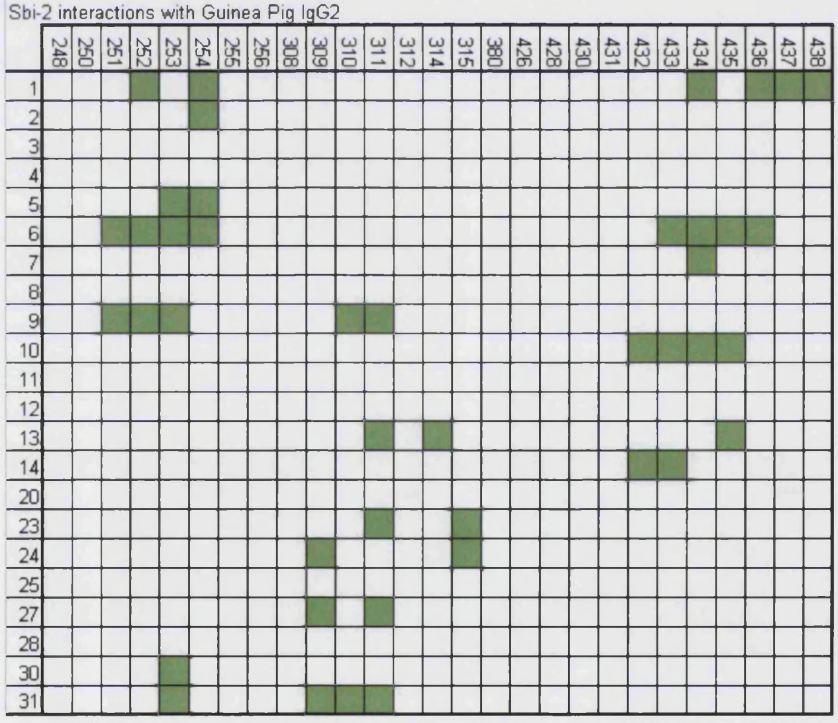
Sbi-2 interactions with Chicken clone 36



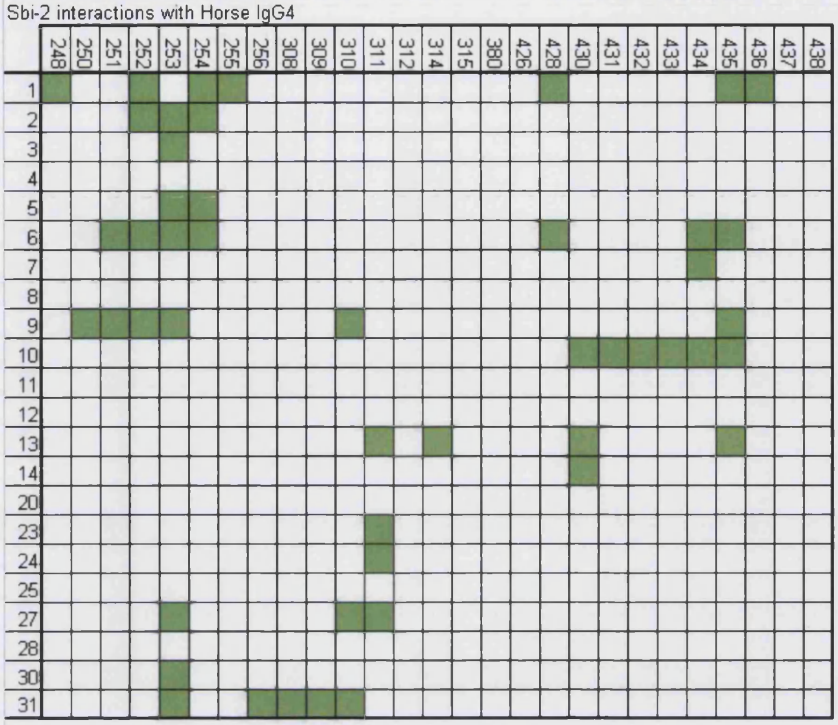
Sbi-2 interactions with Cow IgG2



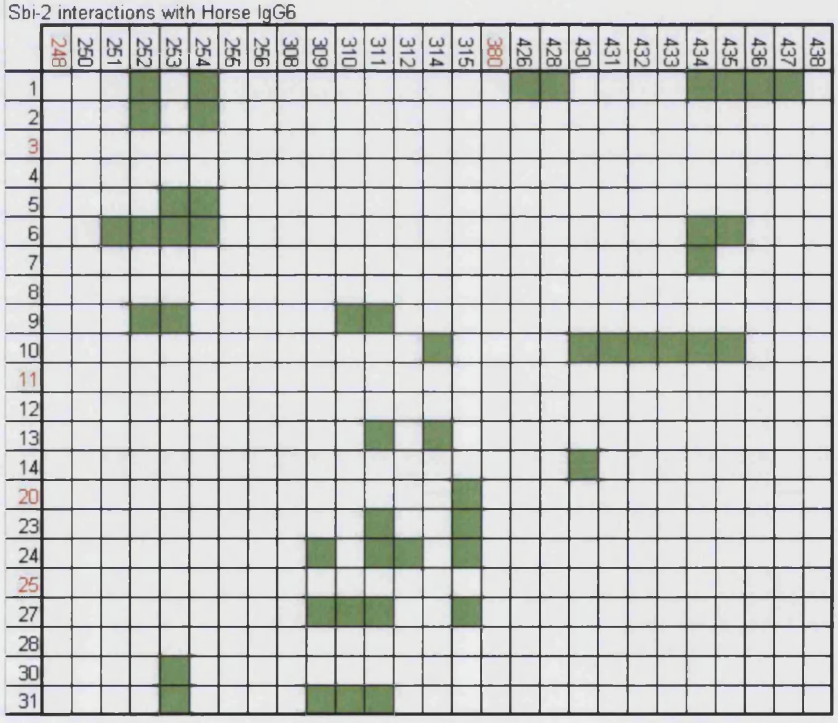
Sbi-2 interactions with Guinea Pig IgG2



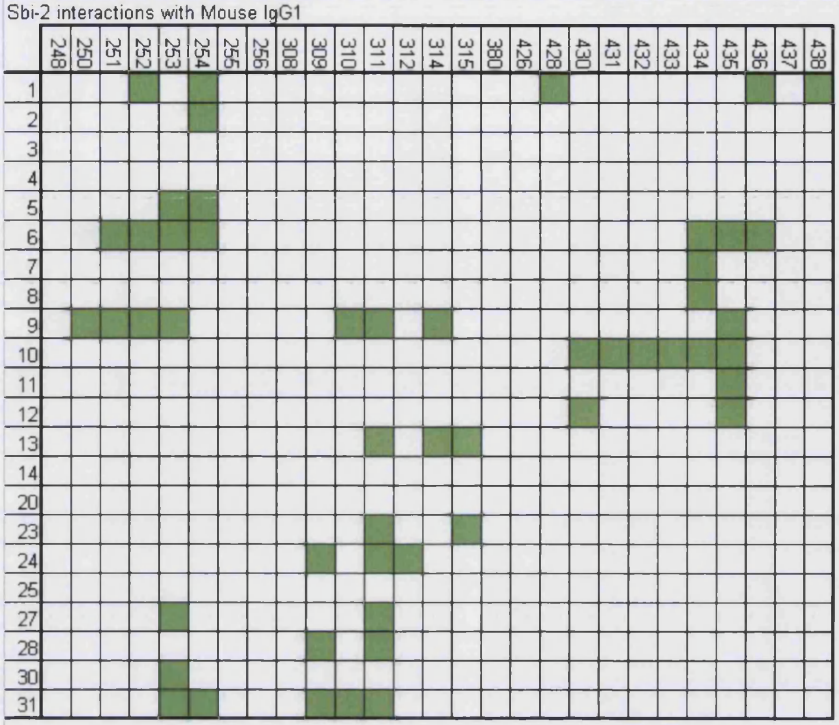
Sbi-2 interactions with Horse IgG4



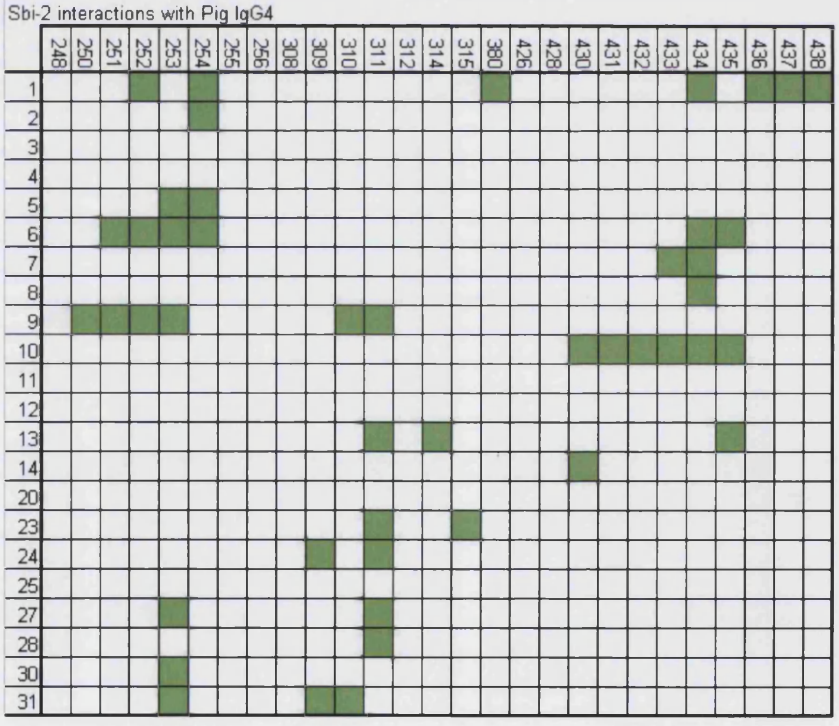
Sbi-2 interactions with Horse IgG6



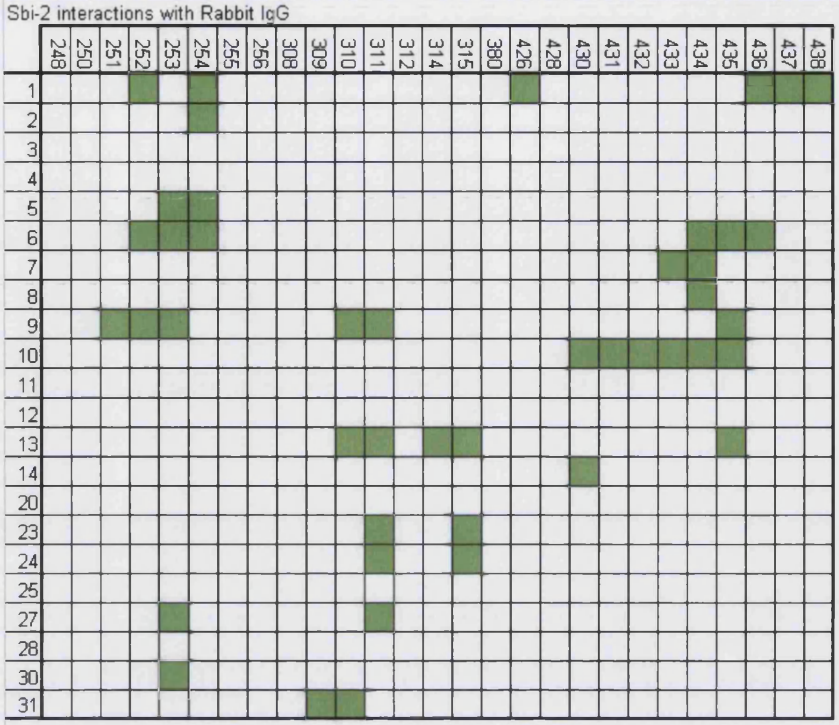
Sbi-2 interactions with Mouse IgG1



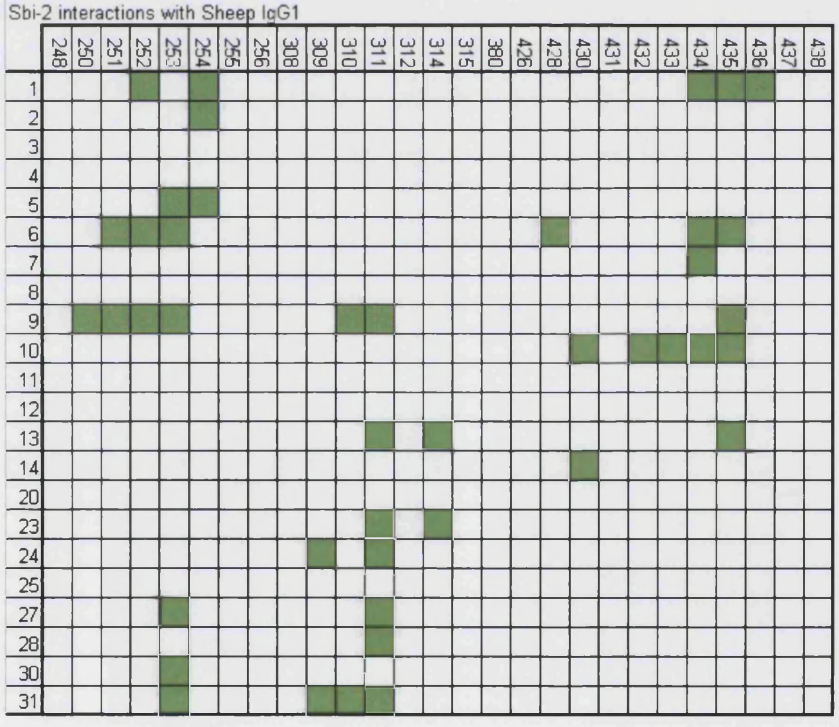
Sbi-2 interactions with Pig IgG4



Sbi-2 interactions with Rabbit IgG



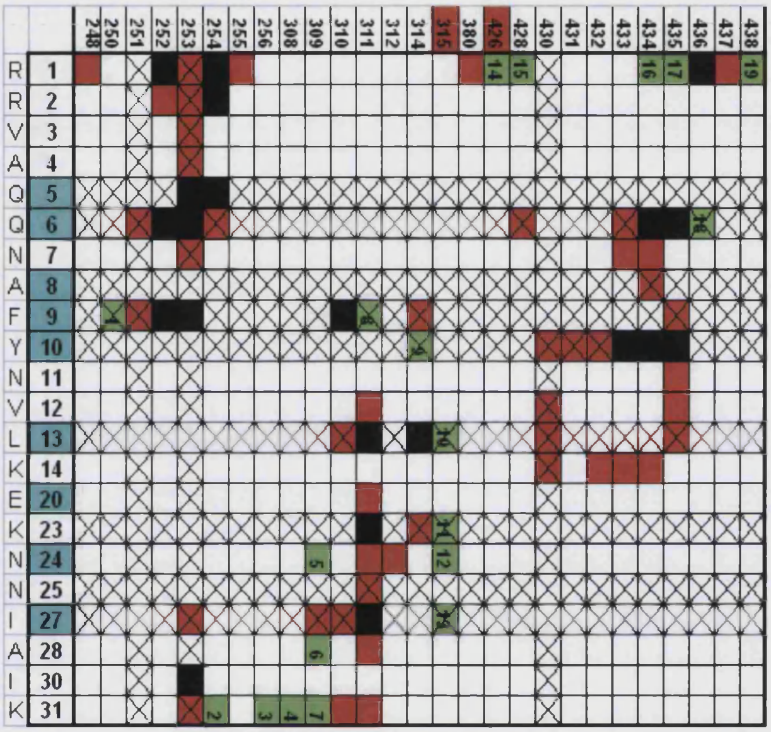
Sbi-2 interactions with Sheep IgG1



Summary of the interactions between Sbi-2 and the IgG models. The number indicates the number of times an interaction between two residues is detected. Residue numbers in red are residues that are conversed in all models

Sbi2	248	250	251	252	253	254	255	256	308	309	310	311	312	314	315	380	426	428	430	431	432	433	434	435	436	437	438
1	1	0	0	10	1	10	2	0	0	0	0	0	0	0	0	1	5	3	0	0	0	0	4	4	10	7	6
2	0	0	0	2	2	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	8	10	10	8	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	10	10	6	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	9	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
9	0	6	9	10	10	0	0	0	0	0	10	8	0	1	0	0	0	0	0	0	0	0	0	7	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	9	7	9	10	10	10	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
13	0	0	0	0	0	0	0	0	0	0	2	10	0	10	2	0	0	0	2	0	0	0	0	8	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	1	1	1	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	10	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	7	0	8	4	0	3	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	7	0	0	0	0	2	3	10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	8	1	0	2	2	9	9	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Summary of interactions observed above between Sbi-2 and the IgG models. The numbered squared refer to the interactions listed in Table 12. Interactions detected in all models are marked in black. Changed interaction pattern: marked in red do not correspond to changes in sequence; green indicates where a substitution occurs in the IgG sequences. The sequence for Sbi-2 is given.



Sequence of the IgG models in the regions identified to be involved in interacting with the other domain. Conserved residues are highlighted in yellow. Residues conserved in all sequences have a black mark under them.

		248	250	251	252	253	254	255	256	308	309	310	311	312	314	315	426	428	430	431	432	433	434	435	436	437	438
1	1FC2	K	T	L	M	I	S	R	T	V	L	H	Q	N	L	D	S	M	E	A	L	H	N	H	Y	T	Q
2	Chicken_IgG_clone36	-	E	L	Y	I	S	L	D	V	S	T	Q	D	L	S	M	V	E	A	L	P	M	R	F	S	Q
3	Cow_IgG2	K	T	L	M	I	T	G	T	I	Q	H	Q	D	T	G	V	M	E	A	L	H	N	H	Y	M	Q
4	Guinea_Pig_IgG2	K	T	L	M	I	S	L	T	I	Q	H	Q	D	L	R	S	M	E	A	L	H	N	H	V	T	Q
5	Horse_IgG4	K	V	L	M	I	S	R	T	I	Q	H	K	D	L	S	A	M	E	A	L	H	N	H	Y	T	E
6	Horse_IgG6	K	T	L	M	I	S	R	T	I	Q	H	Q	D	R	R	V	M	E	A	I	R	H	T	Y	R	Q
7	Mouse_IgG1	K	V	L	T	I	T	L	T	I	M	H	Q	D	L	N	S	L	E	G	L	H	N	H	H	T	E
8	Pig_IgG4	K	T	L	M	I	S	R	T	I	Q	H	Q	D	L	N	A	M	E	A	L	H	N	H	Y	T	Q
9	Rabbit_IgG	K	T	L	M	I	S	R	T	I	T	H	Q	D	L	R	S	M	E	A	L	H	N	H	Y	T	Q
10	Sheep_IgG1	K	T	L	T	I	S	G	T	I	Q	H	Q	D	T	G	V	M	E	A	L	H	N	H	Y	T	Q

18 References

Aybay C. Differential binding characteristics of protein G and protein A for Fc fragments of papain-digested mouse IgG. *Immunol Lett.* 2003 Feb 3;85(3):231-5.

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 1997 25:3389-3402.

Bas de Laat H, Derksen RH, de Groot PG. beta2-glycoprotein I, the playmaker of the antiphospholipid syndrome. *Clin Immunol.* 2004 Aug;112(2):161-8. Review.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch Biochem Biophys.* 1978 Jan 30;185(2):584-91.

Boden MK, Flock JI. Fibrinogen-binding protein/clumping factor from Staphylococcus aureus. *Infect Immun.* 1989 Aug;57(8):2358-63.

Boulin, C.J., Kempf, R., Gabriel, A. and Koch, M.H.J. Data acquisition systems for linear and area X-ray detectors using delay line readout. *Nucl. Instrum. Meth. A,* 1988 269: 312-320.

Bouma B, de Groot PG, van den Elsen JM, Ravelli RB, Schouten A, Simmelink MJ, Derksen RH, Kroon J, Gros P. Adhesion mechanism of human beta(2)-glycoprotein I to phospholipids based on its crystal structure. *EMBO J.* 1999 Oct 1;18(19):5166-74.

Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 1991 Jul 12;253(5016):164-70.

Burman J, Atkins K, Lango L, de Groot P.G, Bagby S, Svergun D, Jenkins A.T.A, Isenman D.E.I, van den Elsen J.M.H, *Staphylococcus aureus* protein Sbi, a double-edged sword in immune evasion. in prep.

Butcher, D. J., Nedved, M. L., Neiss, T. G., Moe, G. R. Proline pipe helix: structure of the tus proline repeat determined by 1H NMR. *Biochemistry* 1996 35 pp. 698

Callegan MC, Engel LS, Hill JM, O'Callaghan RJ. Corneal virulence of Staphylococcus aureus: roles of alpha-toxin and protein A in pathogenesis. *Infect Immun.* 1994 Jun;62(6):2478-82.

Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 1994 Sep 1;50(Pt 5):760-3.

Deisenhofer J. 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.* 1981 Apr 28;20(9):2361-70.

Dziewanowska K, Patti JM, Deobald CF, Bayles KW, Trumble WR, Bohach GA. Fibronectin binding protein and host cell tyrosine kinase are required for internalization of Staphylococcus aureus by epithelial cells. *Infect Immun.* 1999 Sep;67(9):4673-8.

Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG. The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). *Proc Natl Acad Sci U S A.* 2002 May 28;99(11):7687-92.

Fischer D. Modeling three-dimensional protein structures for amino acid sequences of the CASP3 experiment using sequence-derived predictions. *Proteins.* 1999 Suppl 3:61-5.

Flock JI, Hienz SA, Heimdahl A, Schennings T. Reconsideration of the role of fibronectin binding in endocarditis caused by Staphylococcus aureus. *Infect Immun.* 1996 May;64(5):1876-8.

Foster TJ, Hook M. Surface protein adhesins of Staphylococcus aureus. *Trends Microbiol.* 1998 Dec;6(12):484-8.

Foster TJ. The Staphylococcus aureus "superbug". *J Clin Invest.* 2004 Dec;114(12):1693-6.

Friguet B, Chaffotte AF, Djavadi-Ohaniance L, Goldberg ME. Under proper experimental conditions the solid-phase antigen does not disrupt the liquid phase equilibrium when measuring dissociation constants by competition ELISA. *J Immunol Methods.* 1995 May 11;182(1):145-50.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.;Protein Identification and Analysis Tools on the ExPASy Server; *(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press* 2005 pp. 571-607

Goetz H, Kuschel M, Wulff T, Sauber C, Miller C, Fisher S, Woodward C. Comparison of selected analytical techniques for protein sizing, quantitation and

molecular weight determination. *J Biochem Biophys Methods.* 2004 Sep 30;60(3):281-93.

Gomez MI, Lee A, Reddy B, Muir A, Soong G, Pitt A, Cheung A, Prince A. Staphylococcus aureus protein A induces airway epithelial inflammatory responses by activating TNFR1. *Nat Med.* 2004 Aug;10(8):842-8.

Gouda H, Shiraishi M, Takahashi H, Kato K, Torigoe H, Arata Y, Shimada I. NMR study of the interaction between the B domain of staphylococcal protein A and the Fc portion of immunoglobulin G. *Biochemistry.* 1998 Jan 6;37(1):129-36.

Gouda H, Torigoe H, Saito A, Sato M, Arata Y, Shimada I. Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures. *Biochemistry.* 1992 Oct 13;31(40):9665-72.

Graille M, Stura EA, Corper AL, Sutton BJ, Taussig MJ, Charbonnier JB, Silverman GJ. Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. *Proc Natl Acad Sci U S A.* 2000 May 9;97(10):5399-404.

Green RJ, Frazier RA, Shakesheff KM, Davies MC, Roberts CJ, Tendler SJ. Surface plasmon resonance analysis of dynamic biological interactions with biomaterials. *Biomaterials.* 2000 Sep;21(18):1823-35.

Greene C, McDevitt D, Francois P, Vaudaux PE, Lew DP, Foster TJ. Adhesion properties of mutants of Staphylococcus aureus defective in fibronectin-binding proteins and studies on the expression of fnb genes. *Mol Microbiol.* 1995 Sep;17(6):1143-52.

Greenwood D, Slack RCB, Peutherer JF (edited by). Medical microbiology: a guide to microbial infections, pathogensis, immunity, laboratory diagnosis and control *15th ed. New York: Edinburgh: Churchill Livingstone,* 1997

Guinier, A. *La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. Ann. Phys. (Paris),* 1939 12: 161-237.

Hammel M, Kriechbaum M, Gries A, Kostner GM, Laggner P, Prassl R. Solution structure of human and bovine beta(2)-glycoprotein I revealed by small-angle X-ray scattering. *J Mol Biol.* 2002 Aug 2;321(1):85-97.

Hanson DC, Phillips ML, Schumaker VN. Electron microscopic and hydrodynamic studies of protein A-immunoglobulin G soluble complexes. *J Immunol.* 1984 Mar;132(3):1386-96.

Hanson DC, Schumaker VN. A model for the formation and interconversion of protein A-immunoglobulin G soluble complexes. *J Immunol.* 1984 Mar;132(3):1397-409.

Hartleib J, Kohler N, Dickinson RB, Chhatwal GS, Sixma JJ, Hartford OM, Foster TJ, Peters G, Kehrel BE, Herrmann M. Protein A is the von Willebrand factor binding protein on Staphylococcus aureus. *Blood.* 2000 Sep 15;96(6):2149-56.

Jansson B, Uhlen M, Nygren PA. All individual domains of staphylococcal protein A show Fab binding. *FEMS Immunol Med Microbiol.* 1998 Jan;20(1):69-78.

Jendeberg L, Tashiro M, Tejero R, Lyons BA, Uhlen M, Montelione GT, Nilsson B. The mechanism of binding staphylococcal protein A to immunoglobin G does not involve helix unwinding. *Biochemistry.* 1996 Jan 9;35(1):22-31

Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol.* 1999 Apr 9;287(4):797-815.

Jones TA, Zou JY, Cowan SW, Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A.* 1991 Mar 1;47 (Pt 2):110-9.

Jones TA, Zou JY, Cowan SW, Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A.* 1991 Mar 1;47 (Pt 2):110-9.

Jonsson P, Lindberg M, Haraldsson I, Wadstrom T. Virulence of Staphylococcus aureus in a mouse mastitis model: studies of alpha hemolysin, coagulase, and protein A as possible virulence determinants with protoplast fusion and gene cloning. *Infect Immun.* 1985 Sep;49(3):765-9.

Josefsson E, McCrea KW, Ni Eidhin D, O'Connell D, Cox J, Hook M, Foster TJ. Three new members of the serine-aspartate repeat protein multigene family of Staphylococcus aureus. *Microbiology.* 1998 Dec;144 (Pt 12):3387-95.

Karplus K, Barrett C, and Hughey R, Hidden Markov Models for Detecting Remote Protein Homologies, *Bioinformatics,* 1998 vol. 14, no. 10, pp. 846-856

Karplus K, Barrett C, Hughey R. Hidden Markov models for detecting remote protein homologies. *Bioinformatics.* 1998 ;14(10):846-56.

Katayama Y, Ito T, Hiramatsu K. A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. *Antimicrob Agents Chemother.* 2000 Jun;44(6):1549-55.

Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol.* 2000 Jun 2;299(2):499-520.

Kirton SB, Baxter CA, Sutcliffe MJ. Comparative modelling of cytochromes P450. *Adv Drug Deliv Rev.* 2002 Mar 31;54(3):385-406.

Kishore U, Reid KB. C1q: structure, function, and receptors. *Immunopharmacology.* 2000 Aug;49(1-2):159-70.

Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr D Biol Crystallogr.* 1996 Jul 1;52(Pt 4):842-57.

Koch, M.H.J. and Bordas, J. *X-ray diffraction and scattering on disordered systems using synchrotron radiation. Nucl. Instrum. Methods,* 1983 208: 461-469.

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

Koreen L, Ramaswamy SV, Graviss EA, Naidich S, Musser JM, Kreiswirth BN. spa typing method for discriminating among Staphylococcus aureus isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *J Clin Microbiol.* 2004 Feb;42(2):792-9.

Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005 May;22(5):1208-22. Epub 2005 Feb 9.

Kreiswirth B, Kornblum J, Arbeit RD, Eisner W, Maslow JN, McGeer A, Low DE, Novick RP. Evidence for a clonal origin of methicillin resistance in Staphylococcus aureus. *Science.* 1993 Jan 8;259(5092):227-30.

Kronvall G, Frommel D. Definition of staphylococcal protein A reactivity for human immunoglobulin G fragments. *Immunochemistry.* 1970 7:124-127.

Kronvall G, Seal US, Finstad J, Williams RC Jr. Phylogenetic insight into evolution of mammalian Fc fragment of gamma G globulin using staphylococcal protein A. *J Immunol.* 1970 Jan;104(1):140-7.

Kumar S, Tamura K & Nei M MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment *Briefings in Bioinformatics* 2004 5:150-163.

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

Lin TH, Quinn TP, Grandgenett D, Walsh MT. Secondary structural analysis of retrovirus integrase: characterization by circular dichroism and empirical prediction methods. *Proteins.* 1989 ;5(2):156-65.

Linh BK, Thuy DT, My LN, Sasaki O, Yoshihara S. Application of Agar Gel Diffusion Test to the Diagnosis of Fasciolosis in Cattle and buffaloes in the Red River Delta of Vietnam. *JARQ.* 2003; 37(3):201-205

Liu Y, Ames B, Gorovits E, Prater BD, Syribeys P, Vernachio JH, Patti JM. SdrX, a serine-aspartate repeat protein expressed by Staphylococcus capitis with collagen VI binding activity. *Infect Immun.* 2004 Nov;72(11):6237-44.

Loughman A, Fitzgerald JR, Brennan MP, Higgins J, Downer R, Cox D, Foster TJ. Roles for fibrinogen, immunoglobulin and complement in platelet activation promoted by Staphylococcus aureus clumping factor A. *Mol Microbiol.* 2005 Aug;57(3):804-18.

Lowy FD. Antimicrobial resistance: the example of Staphylococcus aureus. *J Clin Invest.* 2003 May;111(9):1265-73.

Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature.* 1992 Mar 5;356(6364):83-5.

McDonnell JM. Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Curr Opin Chem Biol.* 2001 Oct;5(5):572-7.

McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000 16, 404-405.

Meininger DP, Rance M, Starovasnik MA, Fairbrother WJ, Skelton NJ. Characterization of the binding interface between the E-domain of staphylococcal protein A and an antibody Fv-fragment *Biochemistry.* 2000 Feb 15;39(6):1541.

Moreillon P, Entenza JM, Francioli P, McDevitt D, Foster TJ, Francois P, Vaudaux P. Role of Staphylococcus aureus coagulase and clumping factor in pathogenesis of experimental endocarditis. *Infect Immun.* 1995 Dec;63(12):4738-43.

Morris A L, MacArthur M W, Hutchinson E G & Thornton J M S tereochemical quality of protein structure coordinates. P *roteins,* 1992 12, 345-364.

Murzin A. G., Brenner S. E., Hubbard T., Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 1995 247, 536-540

Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986 Sep;3(5):418-26.

Nguyen T, Ghebrehiwet B, Peerschke EI. Staphylococcus aureus protein A recognizes platelet gC1qR/p33: a novel mechanism for staphylococcal interactions with platelets. *Infect Immun.* 2000 Apr;68(4):2061-8.

Ni Eidhin D, Perkins S, Francois P, Vaudaux P, Hook M, Foster TJ. Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of Staphylococcus aureus. *Mol Microbiol.* 1998 Oct;30(2):245-57.

Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins.* 1991 11(4):281-96.

Nieba L, Nieba-Axmann SE, Persson A, Hamalainen M, Edebratt F, Hansson A, Lidholm J, Magnusson K, Karlsson AF, Pluckthun A. BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Anal Biochem.* 1997 Oct 15;252(2):217-28.

Nilsson M, Bjerketorp J, Guss B, Frykberg L. A fibrinogen-binding protein of Staphylococcus lugdunensis. *FEMS Microbiol Lett.* 2004 Dec 1;241(1):87-93.

O'Brien L, Kerrigan SW, Kaw G, Hogan M, Penades J, Litt D, Fitzgerald DJ, Foster TJ, Cox D. Multiple mechanisms for the activation of human platelet aggregation by Staphylococcus aureus: roles for the clumping factors ClfA and ClfB, the serine-aspartate repeat protein SdrE and protein A. *Mol Microbiol.* 2002 May;44(4):1033-44.

Oda M, Kozono H, Morii H, Azuma T. Evidence of allosteric conformational changes in the antibody constant region upon antigen binding. *Int Immunol.* 2003 Mar;15(3):417-26.

Palmqvist N, Foster T, Tarkowski A, Josefsson E. Protein A is a virulence factor in Staphylococcus aureus arthritis and septic death. *Microb Pathog.* 2002 Nov;33(5):239-49.

Palmqvist N, Silverman GJ, Josefsson E, Tarkowski A. Bacterial cell wall-expressed protein A triggers supraclonal B-cell responses upon in vivo infection with Staphylococcus aureus *Microbes Infect.* 2005 Jul 20; [Epub ahead of print]

Parham P. The immune system *Garland Publishing/Elsevier Science* 2000

Park PW, Roberts DD, Grosso LE, Parks WC, Rosenbloom J, Abrams WR, Mecham RP. Binding of elastin to Staphylococcus aureus. *J Biol Chem.* 1991 Dec 5;266(34):23399-406.

Park PW, Rosenbloom J, Abrams WR, Rosenbloom J, Mecham RP. Molecular cloning and expression of the gene for elastin-binding protein (ebpS) in Staphylococcus aureus. *J Biol Chem.* 1996 Jun 28;271(26):15803-9

Patel AH, Nowlan P, Weavers ED, Foster T. Virulence of protein A-deficient and alpha-toxin-deficient mutants of Staphylococcus aureus isolated by allele replacement. *Infect Immun.* 1987 Dec;55(12):3103-10.

Patti JM, Bremell T, Krajewska-Pietrasik D, Abdelnour A, Tarkowski A, Ryden C, Hook M. Staphylococcus aureus collagen adhesin is a virulence determinant in experimental septic arthritis. *Infect Immun.* 1994 Jan;62(1):152-61.

Peerschke EI, Ghebrehiwet B. Human blood platelet gC1qR/p33. *Immunol Rev.* 2001 Apr;180:56-64.

Petoukhov, M.V. and Svergun, D.I. *Global rigid body modelling of macromolecular complexes against small-angle scattering data. Biophys J,* 2005 89: 1237-1250.

Petoukhov, M.V., Eady, N.A., Brown, K.A. and Svergun, D.I. *Addition of missing loops and domains to protein models by x-ray solution scattering. Biophys J,* 2002 83: 3113-3125.

Porod, G., *General theory. In: O. Glatter and O. Kratky (Editors), Small-angle X-ray scattering. Academic Press, London,* 1982 pp. 17-51.

Prescott LM, Harley JP, Klein DA Microbiology *McGraw-Hill* 2002

Raju, T.S., Mulkerrin, M.G., Parker, M., De Vos, A.M., Gazzano-Santoro, H., Totpal, K., Ultsch, M.H. Impact of Fc Glycans on The Effector Functions Vary with the Antibody mechanism of Action 2003 To be Published

Roben PW, Salem AN, Silverman GJ. VH3 family antibodies bind domain D of staphylococcal protein A. *J Immunol.* 1995 Jun 15;154(12):6437-45.

Roche FM, Downer R, Keane F, Speziale P, Park PW, Foster TJ. N-terminal A domain of fibronectin-binding proteins A and B promotes adhesion of Staphylococcus aureus to elastin. *J Biol Chem.* 2004 Sep 10;279(37):38433-40.

Romagnani S, Giudizi MG, del Prete G, Maggi E, Biagiotti R, Almerigogna F, Ricci M. Demonstration on protein A of two distinct immunoglobulin-binding sites and their role in the mitogenic activity of Staphylococcus aureus Cowan I on human B cells. *J Immunol.* 1982 Aug;129(2):596-602.

Saphire EO, Parren PW, Pantophlet R, Zwick MB, Morris GM, Rudd PM, Dwek RA, Stanfield RL, Burton DR, Wilson IA. Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science.* 2001 Aug 10;293(5532):1155-9.

Sasso EH, Silverman GJ, Mannik M. Human IgM molecules that bind staphylococcal protein A contain VHIII H chains. *J Immunol.* 1989 Apr 15;142(8):2778-83.

Schneewind O, Fowler A, Faull KF. Structure of the cell wall anchor of surface proteins in Staphylococcus aureus. *Science.* 1995 Apr 7;268(5207):103-6.

Sheffield P, Garrard S, Derewenda Z. Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. *Protein Expr Purif.* 1999 Feb;15(1):34-9.

Siboo IR, Cheung AL, Bayer AS, Sullam PM. Clumping factor A mediates binding of Staphylococcus aureus to human platelets. *Infect Immun.* 2001 May;69(5):3120-7.

Silverman GJ, Goodyear CS. A model B-cell superantigen and the immunobiology of B lymphocytes. *Clin Immunol.* 2002 Feb;102(2):117-34.

Silverman GJ, Goodyear CS. Confounding B-cell defences: lessons from a *staphylococcal* superantigen. *Nature Reviews: Immunology.* 2006 Jun;102(6):465-75.

Stulik K, Pacakova V, Ticha M. Some potentialities and drawbacks of contemporary size-exclusion chromatography. *J Biochem Biophys Methods.* 2003 Jun 30;56(1-3):1-13.

Surolia A, Pain D, Khan MI Protein A: nature's universal anti-antibody *TIBS* 1982 Feb; 74-76

Svergun DI, Koch MH. Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol.* 2002 Oct;12(5):654-60.

Svergun, D.I. *Determination of the regularization parameter in indirect transform methods using perceptual criteria. J. Appl. Crystallogr.,* 1992 25: 495 - 503.

Svergun, D.I. *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J,* 1999 76: 2879-86.

Svergun, D.I. and Koch, M.H.J. *Small angle scattering studies of biological macromolecules in solution. Rep. Progr. Phys.,* 2003 66: 1735-1782.

Svergun, D.I., Barberato, C. and Koch, M.H.J. *CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr.,* 1995 28: 768-773.

Swofford D. L. PAUP and Other Methods. Phylogenetic Analysis using Parismony, 2000 4th Edition. Sinauer Associates, Sunderland, Mass.

T. Huber, G. Faulkner and P. Hugenholtz. Bellerophon; a program to detect chimeric sequences in multiple sequence alignments, *Bioinformatics* 2004 20 2317-2319.

Tashiro M, Montelione GT. Structures of bacterial immunoglobulin-binding domains and their complexes with immunoglobulins. *Curr Opin Struct Biol.* 1995 Aug;5(4):471-81.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997 Dec 15;25(24):4876-82.

Todar K, Todar's Online Testbook of Bacteriology *http://textbookofbacteriology.net/staphy.html*

Walsh EJ, O'Brien LM, Liang X, Hook M, Foster TJ. Clumping factor B, a fibrinogen-binding MSCRAMM (microbial surface components recognizing adhesive matrix molecules) adhesin of Staphylococcus aureus, also binds to the tail region of type I cytokeratin 10. *J Biol Chem.* 2004 Dec 3;279(49):50691-9.

Wann ER, Gurusiddappa S, Hook M. The fibronectin-binding MSCRAMM FnbpA of Staphylococcus aureus is a bifunctional protein that also binds to fibrinogen. *J Biol Chem.* 2000 May 5;275(18):13863-71.

Wilson M, McNab, Henderson B, Bacterial disease mechanisms: an introduction to cellular microbiology *Cambridge University Press* 2002

Wilson WW Light scattering as a diagnostic for protein crystal growth - A practical approach *Journal of Structural Biology* 2003 142:56-65

Wisconsin Package Version 10.2 Genetics Computor Group (GCG) *Madison, Wisc.*

Wood P. Understanding immunology *Pearson Education Limited* 2001

Yuan, X., Werner, J.M., Lack, J., Knott, V., Handford, P.A., Campbell, I.D. and Downing, A.K. *Effects of the N2144S mutation on backbone dynamics of a TB-cbEGF domain pair from human fibrillin-1. J Mol Biol,* 2002 316: 113-25.

Zhang L, Jacobsson K, Strom K, Lindberg M, Frykberg L. Staphylococcus aureus expresses a cell surface protein that binds both IgG and beta2-glycoprotein I. *Microbiology.* 1999 Jan;145 (Pt 1):177-83.

Zhang L, Jacobsson K, Vasi J, Lindberg M, Frykberg L. A second IgG-binding protein in Staphylococcus aureus. *Microbiology.* 1998 Apr;144 (Pt 4):985-91.

Zhang L, Rosander A, Jacobsson K, Lindberg M, Frykberg L. Expression of staphylococcal protein Sbi is induced by human IgG. *FEMS Immunol Med Microbiol.* 2000 Jul;28(3):211-8.