Structural and functional studies of intrinsically disordered fibronectin-binding proteins

Nicole Catherine Norris

PhD

University of York
Department of Biology

August 2009

For Ben

Abstract

Bacterial fibronectin-binding proteins (FnBPs) mediate adhesion of bacteria to host tissues through binding to the human protein fibronectin (Fn). FnBPs are predicted to contain a series of intrinsically disordered Fn-binding repeats (FnBRs), which undergo a disorder-to-order transition on binding up to five F1-modules ($^{1-5}$F1) from the N-terminal domain of Fn through a tandem β-zipper interaction. This work investigated these predictions for SfbI-5, an FnBR from the important human pathogen *Streptococcus pyogenes*.

SfbI-5 was shown to behave as an intrinsically disordered protein (IDP), with no stable secondary or tertiary structure. The C-terminal third of SfbI-5 was shown to have propensity for β-strand conformations and, thus, might contain a preformed contact site involved in the initial interaction between SfbI-5 and $^{1-5}$F1.

A high resolution structure of an SfbI-5 peptide in complex with ^{2}F1^{3}F1 demonstrated that peptide binding occurs through formation of a tandem β-zipper with the concomitant formation of a large, extended intermolecular interface. Conserved FnBR residues play roles similar to those played by residues in related complexes involving *Staphylococcus aureus* FnBRs. The effect of mutating these residues on SfbI-5 binding to $^{1-5}$F1 was shown to be modest, suggesting that large, extended interfaces, when formed by IDPs, might be tolerant to mutations in the IDP.

A putative FnBR, TickFnBR, was identified in a salivary protein from the blood-feeding tick *Ixodes scapularis*, which is a vector for human diseases. TickFnBR was shown to bind to $^{1-5}$F1 and also to Fn in human plasma. A high resolution structure demonstrated that a TickFnBR peptide binds to ^{2}F1^{3}F1 through a tandem β-zipper. TickFnBR is the first eukaryotic FnBR to be identified, suggesting the tandem β-zipper interaction is a useful mechanism for harmful organisms to target and exploit human Fn.

Table of Contents

List of Tables

List of Figures

17

List of Equations

Acknowledgements

Author's declaration

I declare that all the work contained herein is my own work unless explicitly acknowledged or referenced otherwise in the text.

1 Introduction

1.1 Overview

Streptococcus pyogenes is an important human pathogen that causes common mild infections of the throat and skin such as pharyngitis and impetigo, and may also cause more severe invasive infections such as streptococcal toxic shock syndrome (reviewed in Cunningham, 2000). A critical initial step in *S. pyogenes* infections is the adhesion of bacteria to host tissues. Adhesion can be mediated by bacterial cell-wall attached proteins that bind the human protein fibronectin (Fn). These Fn-binding proteins (FnBPs) also mediate invasion of epithelial cells. Fn acts as a bridging molecule between FnBPs and host cell-wall attached integrins, leading to internalization of *S. pyogenes* by the host cell. This probably allows *S. pyogenes* to avoid anti-microbial drugs and host-defence mechanisms (reviewed in Nitsche-Schmitz *et al.*, 2007; Schwarz-Linek *et al.*, 2006).

FnBPs are predicted to contain a series of intrinsically disordered Fn-binding repeats (FnBRs), which undergo a disorder-to-order transition and bind up to five of the most N-terminal F1-modules ($^{1-5}$F1) of Fn *via* a tandem β-zipper. In this interaction, the FnBR extends the triple-stranded anti-parallel β-sheet of sequential F1 modules by an additional anti-parallel strand (Schwarz-Linek *et al.*, 2003; Section 1.4). This thesis explores the proposed tandem β-zipper interaction between $^{1-5}$F1 from human Fn (Section 1.2) and SfbI-5, the fifth FnBR of streptococcal fibronectin-binding protein I (SfbI) from *S. pyogenes* (Section 1.3). Thus, this work helps to elucidate a molecular mechanism that *S. pyogenes* uses to adhere to and invade human cells.

To date, only prokaryotic FnBRs have been identified. This thesis reports the discovery, and structural and functional investigation, of the first eukaryotic FnBR, TickFnBR, from a salivary protein of *Ixodes scapularis*.

1.2 Fibronectin (Fn)

Fn is a glycoprotein important in many cellular processes, including cell adhesion and migration in, for example, early development (George *et al.*, 1993) and wound healing (reviewed in Grinnell, 1984; Magnusson & Mosher, 1998). The importance of Fn is demonstrated by its conservation in vertebrates (*e.g.,* Figure 1; DeSimone *et al.*, 1992),

and also because it has been shown that mice without the Fn gene die in early embryogenesis (George *et al.*, 1993). Fn exists as a soluble disulfide-linked dimer in plasma and as an insoluble network of fibrils in the extracellular matrix (ECM), and is particularly abundant in embryonic or injured tissue (reviewed in Magnusson & Mosher, 1998).

1.2.1 Overview of Fn structure

Each monomer of Fn is primarily composed of a series of repeating modules (Figure 2A). Sequence homology defines three types of Fn modules: F1, F2 and F3. The structure of more than one representative of each Fn module has been solved, revealing that each of the modules has a conserved tertiary fold (reviewed in Potts & Campbell, 1996). F1 modules have approximately 45 amino acid residues with a β-sandwich fold stabilized by two disulfide bonds (Figure 2B). They consist of five β-strands, labelled A through to E from the N- to C-terminus. Strands A and B form an anti-parallel β-sheet that folds on top of another anti-parallel β-sheet formed by strands C – E. Bacterial Fn-binding proteins (FnBPs) interact primarily with the relatively exposed E-strand (Section 1.4.1; Figure 4). F2 modules have approximately 60 residues and consist of two double-stranded anti-parallel sheets that fold on top of, and perpendicular to, each other (Figure 2C). They are also stabilized by two disulfide bonds. F3 modules have approximately 90 residues in a β-sandwich fold consisting of a three-strand and a four-strand anti-parallel β-sheet (Figure 2D). They are the only Fn modules not to contain any disulfide bonds. In addition to F1, F2 and F3 modules, Fn contains three alternatively spliced regions: extradomains A and B (EDA, EDB, respectively) and the variable region (V), which can be spliced in more than one place. Different splicing combinations of EDA, EDB and the V region of the single human Fn gene result in 20 Fn isoforms (reviewed in Pankov & Yamada, 2002).

There are plasma and cellular forms of Fn. Cellular Fn can be any of a variety of isoforms secreted by several different cell types, including fibroblasts, epithelial cells and macrophages, after which it is incorporated into the ECM (reviewed in Romberger, 1997). In contrast, plasma Fn does not contain EDA or EDB, although one subunit in the dimer contains the V region. Plasma Fn is secreted by hepatocytes into the blood, where it exists in a compact form. It is preferentially incorporated into blood clots during wound healing (reviewed in Magnusson & Mosher, 1998). Furthermore, it is an important part of the ECM, as demonstrated recently by Moretti *et al.* (2007) who used

a transgenic mouse model to show that up to 60% of the Fn in the ECM of some tissues is plasma-derived.

```
                          10        20        30        40        50
         ..         ..|....|....|....|....|....|....|....|....|....|
Human      QA-----QQMVQPQSPVAVSQSKPGCYDNGKHYQINQQWERTYLGNALVCTCYGGSRGFN
Horse      .D-----..II........G........................................
Dog        R.-----ARRLA.KFAAR..SER.....................................
Bovine     ..-----..I......LT....................................S.....
Mouse      ..-----..I...........F......................................
Chicken    GK--NRR.AQTASVPQATPA.G.QT.F...RY.........I....T.............
Frog       RRQAQQQ..QQVV.PQGTQDNHQK.......Y.............T.........G....
Zebrafish  KR--------.A.EHSVS.V.QD..VMD.QF.GAGEK...I...ST.L...H.-VSKIE
               :    *   :*:.*  .::*** ***.:*:***:*     ::

                          60        70        80        90       100       110
         ....|....|....|....|....|....|....|....|....|....|....|....|
Human      CESKPEAEETCFDKYTGNTYRVGDTYERPKDSMIWDCTCIGAGRGRISCTIANRCHEGGQ
Horse      .......P....................................................
Dog        .......P....................................................
Bovine     .......P....................................................
Mouse      .......P..............K.....................................
Chicken    .......P.........S.....E...................................K
Frog       .......S.........VS....E.........N..........................
Zebrafish  .K...D.....Y..VNSRS....E.......G.........SA..K.............H
           *:***:.****:**  .. :*:**:*******.*********:.**:*************:

                         120       130       140       150       160       170
         ....|....|....|....|....|....|....|....|....|....|....|....|
Human      SYKIGDTWRRPHETGGYMLECVCLGNGKGEWTCKPIAEKCFDHAAGTSYVVGETWEKPYQ
Horse      ............................................................
Dog        ............................................................
Bovine     ..................................................Q.........
Mouse      .......K...........L........................................
Chicken    ..................L.................L..R.Y.NT...............
Frog       ...................................V..R.Y.NT........Q.......
Zebrafish  ........Q...D..D..................V..R.Y.NT.....M...........
           ******.*:***:**.*:***:*************:**:*:*:*****:**.*******

                         180       190       200       210       220       230
         ....|....|....|....|....|..  .|....|....|....|....|....|....|...
Human      GWMMVDCTCLGEGSGRITCTSR--NRCNDQDTRTSYRIGDTWSKKDNRGNLLQCICTGNG
Horse      ...............N.......--...................................
Dog        .....................NTD....................................
Bovine     .......................--...................................
Mouse      ...............N.......--.................................V...
Chicken    .......................--.......K...........................
Frog       ...............N....S.K--.......K...........T.T.............
Zebrafish  ...I...................--.....................T.S.HV...L.....
           ***:*********.*****:*:  :*******:***********.*.**::***:*****

                         240       250
         .|....|....|....|....   |....
Human      RGEWKCERHTSVQTTSSGSGP--FTDVR
Horse      ........A.L....T....--....
Dog        ........A.L....T....--....
Bovine     ..........L....A...S--....
Mouse      ........A-L.SA.A...S--....
Chicken    ..........LH...T...SPS..N.Q
Frog       ........S.A.A.--.T.SNPI.NIQ
Zebrafish  ........A.LH...L.T.SRVV.N.Q
           ********: ::: *:*. .*:::
```

Figure 1. The N-terminal domain of Fn is conserved amongst vertebrates. Legend on p 24.

Figure 1. The N-terminal domain of Fn is conserved amongst vertebrates. An alignment between the Fn N-terminal domain (NTD) sequences from human, horse, dog, bovine, mouse, chicken, frog and zebrafish reveals this region is conserved amongst vertebrates. Residues are coloured based on chemical composition, and those residues identical to human Fn are shown as full stops (.) in the other vertebrate Fn sequences. Conserved residues are indicated below the alignment, where an asterisk (*), colon (:) or full stop (.) indicate that the sequences at those positions contain identical, conserved or semi-conserved residues, respectively. The residue numbering corresponds to mature human Fn (after cleavage of the signal sequence). Also indicated above the sequences are F1 modules (see text) in human Fn, with module boundaries defined by Potts *et al.* (1999). Residues predicted to form the five β-strands in each F1 module (based on the available F1 module structures) are indicated with arrows. Strands A – D are purple and the E strands are black.

The sequences were aligned using the program Clustal X v 2.0 (Larkin *et al.*, 2007). The species name and UniProt or UniParc accession number (UniProt ID) for the Fn sequences used in the alignment were: human, *Homo sapiens*, P02751; horse, *Equus caballus*, UPI00017961BB; dog, *Canis familiaris*, UPI0000EB0291; bovine, *Bos taurus*, P07589; mouse, *Mus musculus*, P11276; chicken, *Gallus gallus*, UPI0000ECB57B; frog, *Xenopus laevis*, Q91740; and zebrafish, *Danio rerio*, B0S602. The pdb codes and references for the F1 module structures used to determine the residues involved in a β-strand are: [1]F1[2]F1, 1qgb (Potts *et al.*, 1999); [2]F1[3]F1, 2cg7 (Rudiño-Piñera *et al.*, 2007) and [4]F1[5]F1, 1efb (Williams *et al.*, 1994).

In addition to the F1, F2 and F3 modules, EDA, EDB and the V region, Fn is also divided into larger functional domains, originally defined based on limited proteolysis of Fn (reviewed in Pankov & Yamada, 2002). The N-terminal domain (NTD) of Fn, comprising the first five F1 modules ([1−5]F1) and an additional 17 N-terminal residues, is the primary binding site of FnBPs (Figure 1; Figure 2A; Mosher & Proctor, 1980). It has also been shown to interact with other regions of Fn, and has been implicated in binding fibrin (Figure 2A; reviewed in Mao & Schwarzbauer, 2005; Wierzbicka-Patynowski & Schwarzbauer, 2003). Furthermore, [1−5]F1 is the only Fn domain essential for the cell-mediated assembly of Fn into fibrils in the ECM (Schwarzbauer, 1991). Other functional domains of Fn bind fibrin, heparin and gelatin (unfolded collagen) which are all components of the ECM, as well as integrins (Figure 2A; reviewed in Pankov & Yamada, 2002). Integrins are heterodimeric transmembrane cell-surface receptors that link the ECM with the actin cytoskeleton. Integrins, particularly α5β1 integrins, are important in Fn fibrillogenesis.
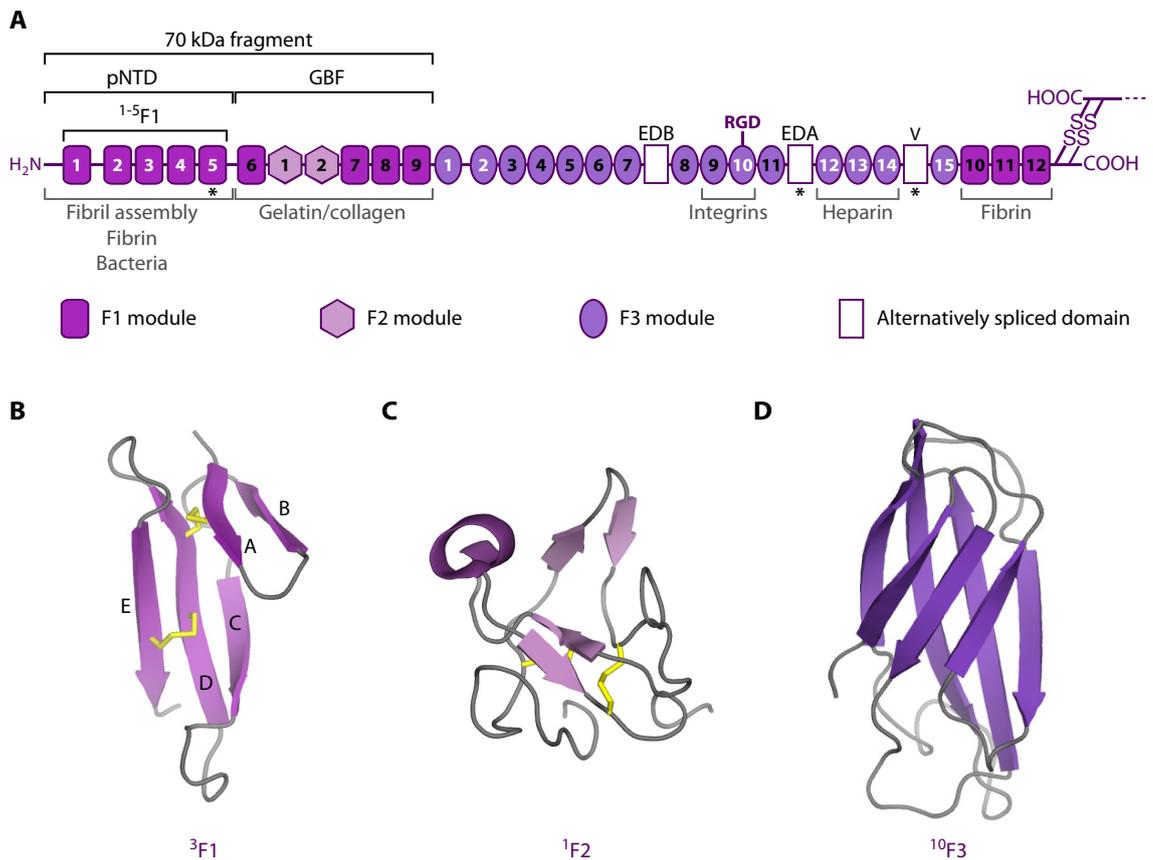
Figure 2. Fn modules, domain organization and structure. A. Schematic of Fn module and domain organization. Fn is a disulfide bond-linked dimer; the second subunit of the dimer is not shown in the schematic. Each Fn subunit is comprised of F1, F2, F3 modules and three alternatively spliced regions shown as rounded rectangles, hexagons, ellipsoids and rectangles, respectively. Each module is referred to as [n]FX where [2]F1 is the second F1 module from the N-terminus, and [10]F3 is the tenth F3 module from the N-terminus. Alternatively spliced regions EDA, EDB and the V region are indicated. The 70 kDa fragment, proteolytic NTD (pNTD) and the gelatin binding fragment (GBF) are also indicated. pNTD consists of [1-5]F1 with additional N- and C-terminal sequences that are 18 and 17 amino acid residues, respectively (*i.e.*, the human sequence in Figure 1). Binding sites for other molecules and bacteria are indicated below the schematic. The RGD sequence within the main integrin binding site ([9]F3[10]F3) is shown; other modules known to bind integrins are indicated with an asterisk (*). Regions of Fn that can interact with other regions of Fn are indicated by white numbering. This schematic is adapted from similar figures in Leiss *et al.*, 2008; Pankov & Yamada, 2002; Schwarz-Linek *et al.*, 2004a.

B – D. Representative structures of F1, F2 and F3 modules, respectively. Structures are shown in ribbon representation with disulfide bonds in yellow. Strands A – E are labelled in B. The pdb code and references used for these structures are: [3]F1, 2cg7 (Rudiño-Piñera *et al.*, 2007); [1]F2, 1e8b (Pickford *et al.*, 2001); [10]F3, 1fna (Dickinson *et al.*, 1994).

1.2.2 Fn fibrillogenesis

Fn fibrillogenesis is a cell-mediated process, rather than a spontaneous process, that can be summarized in five steps (reviewed in Geiger *et al.*, 2001; Mao & Schwarzbauer, 2005; Wierzbicka-Patynowski & Schwarzbauer, 2003). First, soluble, compact Fn binds to integrins on the cell surface. The main integrin binding site is an Arg-Gly-Asp (RGD) sequence in ^{10}F3 with an additional site that increases binding affinity, called the 'synergy site,' in ^{9}F3; additional sites in Fn can also bind integrins (Figure 2A). Second, binding of Fn dimers to integrins leads to integrin clustering, increasing the local Fn concentration outside the cell and also initiating signalling events and reorganization of the actin cytoskeleton inside the cell. Third, cell-tension generated by the intracellular actin cytoskeleton is transferred to extracellular Fn outside the cell *via* integrins, resulting in Fn lengthening from its compact form to expose previously cryptic (hidden) Fn-Fn binding sites. Fourth, short fibrils form as a result of inter-molecular Fn-Fn interactions; these are soluble in detergent. Fifth, a detergent-insoluble fibrillar matrix is formed irreversibly, possibly by inter-molecular disulfide bonds catalysed by the intrinsic disulfide isomerase ability of ^{12}F1 (reviewed in Mao & Schwarzbauer, 2005). Once formed, Fn fibrils play an important role in organizing cells and the ECM, because the Fn binds to integrins and other ECM proteins with its numerous functional domains (Figure 2A). However, it is also possible for bacteria to exploit the presence of Fn in both the ECM and the blood, to adhere to, avoid detection by, and invade the host.

1.3 Streptococcal Fn-binding protein I (SfbI)

1.3.1 *Streptococcus pyogenes*: an important human pathogen

Streptococcus pyogenes, also known as β-haemolytic Group A streptococcus (GAS), is a successful, prevalent, Gram-positive bacterial human pathogen. It causes common, mild infections of the throat and skin, including pharyngitis, impetigo and scarlet fever, and less-common, life-threatening invasive infections, such as streptococcal toxic shock syndrome, septicaemia and necrotizing fasciitis (reviewed in Cunningham, 2000). It also causes serious post-infection sequelae[1], including acute rheumatic fever, acute glomerulonephritis and rheumatic heart disease (reviewed in Cunningham, 2000). As well as causing a broad spectrum of diseases, *S. pyogenes* is prevalent, particularly in

---

[1] A sequela (plural sequelae) is a condition which is the consequence of a previous disease or injury.

less-developed countries (Carapetis *et al.*, 2005). For example, it is the most common cause of bacterial pharyngitis, causing a minimum of 616 million new cases worldwide each year (Carapetis *et al.*, 2005; Cunningham, 2000). The prevalence of mild *S. pyogenes* infections is significant for two reasons. First, these infections require treatment, which is costly both in time and money for health services. Second, mild infections can be the cause of subsequent invasive and post-infection diseases, which are estimated to cause over 1.78 million new cases each year globally and 517, 000 deaths per year (Carapetis *et al.*, 2005). The morbidity and mortality rates of *S. pyogenes* infections make it a major human pathogen.

1.3.2   *S. pyogenes* virulence factors, invasion and infection

S. pyogenes is a successful human pathogen because it has developed several virulence factors[2] for adhering to and colonizing the host while evading or overcoming the host's defence mechanisms (Table 1). Adhesion is a critical, initial step in infection during which *S. pyogenes* needs to overcome the host's innate immune defences at its two primary sites of infection: the mucosal epithelium of the upper respiratory tract, and the skin epidermis (reviewed in Courtney *et al.*, 2002). The innate immune defences the streptococci have to overcome are mechanical, chemical and biological. For example, the skin acts as a mechanical barrier and secretes fatty acids that chemically inhibit streptococcal growth. In the upper respiratory tract *S. pyogenes* can be trapped in mucus and removed mechanically by swallowing, and salivary mucins and immunoglobulins chemically reduce *S. pyogenes* adhesion. Furthermore, at both these sites, *S. pyogenes* is in competition with endogenous bacteria for adherence (reviewed in Courtney *et al.*, 2002).

Several virulence factors called adhesins are used by the bacteria to overcome these obstacles by either binding directly to cell-surface receptors (*e.g.,* the hyaluronic acid capsule, Table 1) or adhering to molecules that are in turn bound to cells (*e.g.,* MSCRAMMs, Table 1). There are many streptococcal adhesins, targeting numerous host molecules. Some adhesins are redundant, in that there is often more than one *S. pyogenes* adhesin capable of binding the same host molecule, and some adhesins are

---

[2] A virulence factor is a compound that contributes to an organism's ability to cause disease in the host. Virulence factors contribute to establishment and persistence of an infection in the host and dissemination of the infection to other areas of the host.

multi-functional and able to bind more than one type of host molecule (reviewed in Courtney *et al.*, 2002). For example, there are at least seven different streptococcal FnBPs (Section 1.3.3), and some types of M protein bind Fn in addition to components of the complement pathway (Table 1). Which adhesins are expressed depends on the strain, tissue-type and environmental cues, such as oxygen or carbon dioxide levels (reviewed in Kreikemeyer *et al.*, 2003). Concurrent with adhesion, the anti-phagocytic activity of M protein or the hyaluronic acid capsule protects *S. pyogenes* from phagocytosis (Table 1). With prolonged adhesion, the bacteria can replicate and colonize the host tissue, resulting in an *S. pyogenes* infection.

As a possible final step in the infection process, *S. pyogenes* may change from a local to a systemic infection, by invading deeper tissue and/or entering the blood. *S. pyogenes* potentially can do this by one of four mechanisms (reviewed in Nitsche-Schmitz *et al.*, 2007). First, it can be opportunistic, and invade damaged tissue after the host has been injured. Second, it can degrade surrounding tissue. For example, *S. pyogenes* can bind to and activate host plasminogen, which is a protease that degrades the ECM component fibrin. Third, hyaluronic acid-encapsulated bacteria can migrate between cells (Table 1). Fourth, *S. pyogenes* can be internalized by non-phagocytic host cells in either an M protein- or FnBP-mediated process (Table 1; Section 1.3.6). It is possible that subsequent host cell apoptosis releases *S. pyogenes* into surrounding tissue (reviewed in Kreikemeyer *et al.*, 2003). In summary, *S. pyogenes* is successful at adhering to, colonizing and infecting humans because of a large arsenal of virulence factors adept at taking advantage of host molecules.

Table 1. Description and function of some important *S. pyogenes* virulence factors. More detailed information on these and other *S. pyogenes* virulence factors is reviewed by Hynes (2004).

Virulence factor(s)	Description and Function
Hyaluronic acid capsule	Hyaluronic acid is a high molecular weight glycosaminoglycan that forms a capsule around *S. pyogenes* cells. The level of encapsulation varies by strain. The capsule confers resistance to phagocytosis and is a poor immunogen because it is similar to components in human connective tissue (reviewed in Bisno *et al.*, 2003). The capsule can also adhere to cells with the CD44 cell-surface receptor, which can lead to invasion of surrounding tissues by CD44-mediated translocation of the bacteria between cells by disruption of intracellular junctions (reviewed in Courtney *et al.*, 2002).

Virulence factor(s)	Description and Function
Lipoteichoic acid (LTA)	LTA is an amphipathic, cell-wall attached molecule that binds to epithelial cells and to Fn (reviewed in Courtney *et al.*, 2002). LTA binds Fn relatively weakly compared to the bacterial FnBP SfbI (Talay *et al.*, 1992). It is proposed by Courtney *et al.* (2002) that it forms the initial, weak, non-specific contact with host tissue by overcoming the electrostatic repulsion between the bacterial surface and the host tissue. Once this contact is formed, other adhesins make strong, tissue-specific interactions between *S. pyogenes* and the host tissue.
MSCRAMMs	MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) are a class of cell-surface localized adhesins that tightly and specifically bind to components of the ECM including Fn, fibrinogen and collagen (reviewed in Patti *et al.*, 1994). They are important in mediating adherence of *S. pyogenes* to host tissue. Fn-binding MSCRAMMS (*i.e.*, FnBPs) and their role in streptococcal internalization by non-phagocytic cells are discussed further in Sections 1.3.4 and 1.3.6.
M protein	M protein is a cell-wall attached α-helical coiled-coil protein. It has a variable N-terminal domain, including a hypervariable N-terminal tail of approximately 11 amino acid residues. The hypervariable region can be used to classify *S. pyogenes* strains into 'M types' by serotyping or by DNA sequencing of the M-protein gene, *emm*. Some M types are associated with particular *S. pyogenes* infections. For example, M1 is associated with invasive diseases (reviewed in Bisno *et al.*, 2003). M protein is a multi-functional virulence factor. It is anti-phagocytic, and it can adhere to host cells by both M type-dependent and -independent mechanisms. For some M types, M protein can adhere to host cells *via* Fn. M protein is also implicated in internalization of *S. pyogenes* by host cells (reviewed in Courtney *et al.*, 2002).
Pyrogenic exotoxins	A family of superantigens that non-specifically activate T-lymphocytes, resulting in the release of cytokines at toxic levels. The symptoms of streptococcal toxic shock syndrome, including high fever, headache, muscle and joint pain, hypotension and acute respiratory distress are attributed to the actions of exotoxins. It is unclear how exotoxins aid *S. pyogenes* infection, although their high conservation and prevalence in virulent strains indicates they are useful to the bacteria (reviewed in Fraser & Proft, 2008).
Streptolysins	Streptolysins O and S are both secreted toxins that can lyse a variety of cell types, including erythrocytes and platelets, by forming transmembrane pores (reviewed in Bisno *et al.*, 2003; Hynes, 2004).

1.3.3 Domain organization of SfbI, a model FnBP

At least seven FnBPs are expressed by *S. pyogenes*: SfbI, SOF, FbaB, Pfbp, Sfbx, Fbp54 and FbaA (reviewed in Kreikemeyer *et al.*, 2003; Schwarz-Linek *et al.*, 2006). The most studied FnBP is SfbI (Streptococcus Fn-binding protein I). The domain organization of SfbI contains features typical of five of the seven *S. pyogenes* FnBPs and also other streptococcal and staphylococcal FnBPs (Figure 3). These features are: an N-terminal signal sequence that is involved in transport of the protein across the cell-membrane; a series of sequential Fn-binding repeats (FnBRs) at its C-terminus; and a cell wall- and membrane- spanning region that includes a cell-wall anchoring 'LPXTG' motif, where X is any amino acid (Figure 3; reviewed in Joh *et al.*, 1999; Schwarz-Linek *et al.*, 2006). SfbI also contains an upstream Fn-binding region (UR; Figure 3) that binds to the GBF of Fn (Figure 2A; Ozeri *et al.*, 1996), a proline-rich repeat region of unknown function, and an aromatic-rich N-terminal domain, which has been implicated in binding fibrinogen (Katerov *et al.*, 1998). The other two well-conserved FnBPs, Fbp54 and FbaA, do not appear to contain FnBRs, so will not be considered further (Courtney *et al.*, 1994; Ramachandran *et al.*, 2004; Terao *et al.*, 2001). FbaB and Pfbp are two FnBR-containing FnBPs that are mutually exclusive: either one or none of these two FnBPs occurs in any given *S. pyogenes* strain (Ramachandran *et al.*, 2004). Therefore, current knowledge suggests there can be at most four FnBR-containing FnBPs expressed by any given *S. pyogenes* strain.

The boundaries of FnBRs were originally defined based on sequence analysis; in SfbI these repeats were called R repeats (Figure 3; Section 1.3.5; Talay *et al.*, 1991; 1992). Subsequent structural studies by Schwarz-Linek *et al.* (2003) led to the FnBR boundaries that define the functional, repeating, $^{2-5}$F1-binding unit (Figure 3; Section 1.4.1). For the purpose of this thesis, an FnBP will henceforth be defined as a protein that contains at least one FnBR, but usually a series of adjacent FnBRs, that can each bind to $^{2-5}$F1 in Fn *via* a tandem β-zipper. The tandem β-zipper interaction will be discussed in Section 1.4.

1.3.4 FnBPs are important virulence factors

Three lines of evidence support the importance of FnBPs as virulence factors. First, the prevalence of the five FnBR-containing FnBPs in *S. pyogenes* indicates these proteins are important. For example, it has been shown that SfbI is present in 50–80% of

S. pyogenes isolates (Ma *et al.*, 2002; Natanson *et al.*, 1995; Ramachandran *et al.*, 2004). When the presence of all five known FnBR-containing FnBPs was investigated 97% of the 62 isolates had at least one of the five FnBR-containing FnBPs (Ramachandran *et al.*, 2004). *S. pyogenes* is under continual selective pressure to maintain infection; therefore the high prevalence of surface-bound FnBPs indicates FnBR-mediated binding to Fn is important for *S. pyogenes* infection.
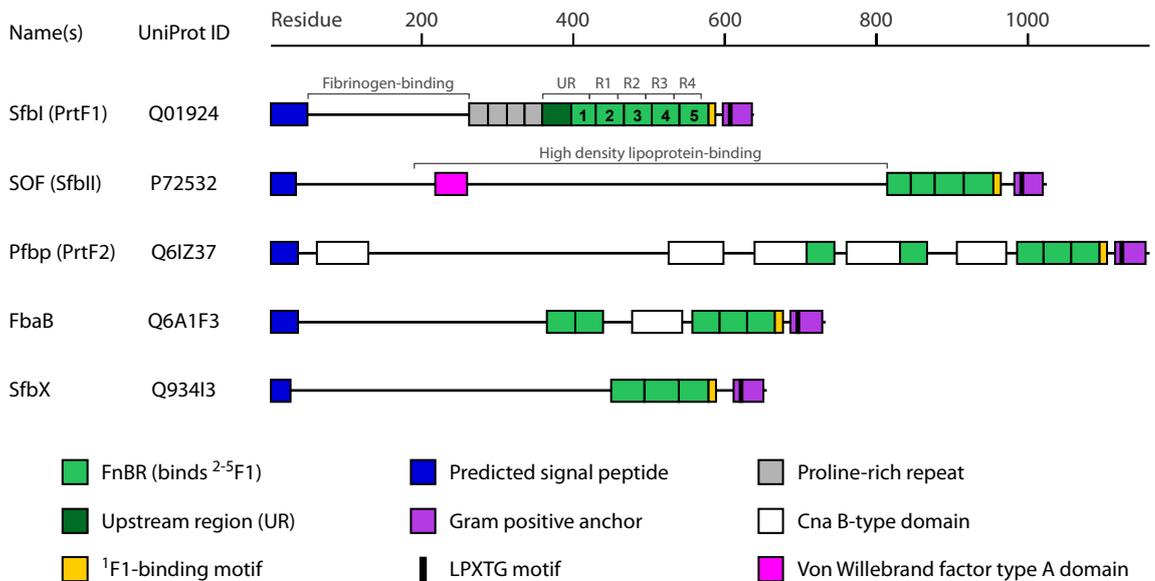
Figure 3. Domain organization of *S. pyogenes* FnBPs containing FnBRs. The UniProt ID and name of each FnBP is indicated, including names for allelic variants in brackets. The schematic of each FnBP shows the functional FnBRs (light green) as defined using the tandem β-zipper model (Schwarz-Linek *et al.*, 2003; Section 1.4.1). Each FnBR is predicted to bind $^{2-5}$F1 (Figure 2). Also shown are the ^{1}F1-binding region (yellow) at the C-terminus of the C-terminal FnBR in each FnBP, and the UR (dark green), which binds the GBF (Figure 2). The N-terminal signal sequence (blue) and a C-terminal Gram positive anchor (purple), which includes an LPXTG motif (black) are also shown. For SfbI, the proline rich repeats (grey) and R repeats (R1 – R4) identified by sequence analysis are indicated (Talay *et al.*, 1992; 1994), as is the original definition of the UR (Sela *et al.*, 1993). The latter two are shown above the SfbI schematic. Also shown (white and pink) are other domains predicted by the protein family database server Pfam, see Table 30 (p 157) for the potential function of these domains. Also indicated above the schematic for SfbI and SOF are regions implicated in fibrinogen-binding and high density lipoprotein-binding, respectively (Courtney *et al.*, 2006; Katerov *et al.*, 1998). The latter results in serum opacification, giving SOF its name (serum opacity factor). This figure is adapted from Figure 1 in Schwarz-Linek *et al.*, 2006.

Second, it has been demonstrated that intranasal vaccination with SfbI in an *in vivo* mouse model protects the mice from a lethal challenge with *S. pyogenes* (Guzman *et al.*, 1999). There was an 80% survival rate of mice vaccinated with SfbI against a challenge

with a heterologous *S. pyogenes* strain, compared to a 10% survival rate in the control mice. Subsequently, it was shown that the SfbI epitopes recognized by the host's B- and T-lymphocytes were all contained within the FnBR region (Schulze *et al.*, 2001; Schulze *et al.*, 2003). This also demonstrates that SfbI FnBRs are surface-exposed and accessible for binding by antibodies.

Third, FnBPs can mediate internalization of *S. pyogenes* into host cells, which correlates with recurring pharyngitis and asymptomatic carriage of the bacteria. Although considered an extracellular pathogen, in the mid-1990s *S. pyogenes* was shown *in vitro* to be internalized by non-phagocytic epithelial cells from a variety of cell-lines (reviewed in Molinari & Chhatwal, 1999) and subsequent *in vitro* studies demonstrated that FnBPs mediated internalization into HEp-2 cells[3] (discussed in Section 1.3.6). *S. pyogenes* internalization *in vivo* might allow the pathogen to avoid host immune defences and also anti-microbials, and it may also aid invasion of surrounding tissues. As discussed below, there is *in vivo* evidence that supports the former but not the latter.

S. pyogenes has been found internalized in both macrophages and epithelial cells in samples taken from the tonsils of individuals who either had recurring pharyngitis or were asymptomatic carriers of *S. pyogenes* (Österlund & Engstrand, 1997a; Österlund *et al.*, 1997b). This indicates that internalization might be partly responsible for the failure of antibiotics in treatment of recurring pharyngitis, and might also increase asymptomatic carriage of *S. pyogenes* through avoidance of the host immune defences. In support of this hypothesis, Sela *et al.* (2000) analysed the internalization of *S. pyogenes* isolates *in vitro* that were from asymptomatic carriers whose infection either did or did not respond to antibiotic treatment *in vivo*. They found that the antibiotic-resistant strains were internalized more efficiently than the antibiotic-sensitive strains, indicating internalization might contribute to antibiotic-resistance and persistence in the host and, thus, contribute to virulence. Furthermore, two studies have shown that SfbI was significantly more prevalent in antibiotic-resistant than in antibiotic-sensitive strains, where antibiotic-resistance was determined either by the failure of antibiotic treatment *in vivo* or by genotyping (Facinelli *et al.*, 2001; Neeman *et al.*, 1998). In these studies, SfbI was present in ~90% of resistant strains and only 20–

---

[3] HEp-2 cells are a human larynx epithelial cell-line originating from a carcinoma specimen taken from the larynx of a 57-year old man (Toolan, 1954).

30% of sensitive strains. In contrast, Brandt *et al.* (2001) found SfbI in only 24% of antibiotic-resistant strains, but they did not test antibiotic-sensitive strains. A possible explanation for this discrepancy is low sample size (17 strains), particularly compared to the study by Facinelli *et al.* (2001) where 74 antibiotic-resistant strains were tested. Taken together these studies indicate that internalization of *S. pyogenes* by host cells: occurs *in vivo*, can be mediated by SfbI, and contributes to virulence by increasing antibiotic resistance and asymptomatic carriage.

An *in vivo* mouse model of *S. pyogenes* infection, however, demonstrated that the expression of SfbI by an *S. pyogenes* strain associated with invasive disease attenuated virulence by reducing both mortality rates and bacterial dissemination (Nyberg *et al.*, 2004). The researchers proposed the attenuation might aid *S. pyogenes* persistence in its human host (Nyberg *et al.*, 2004). Given that repeated exposure to *S. pyogenes* increases the risk of acute rheumatic fever (Cunningham, 2000), it is possible that recurrent pharyngitis or asymptomatic carriage as a result of internalization may increase the likelihood of more serious sequelae. In conclusion, although FnBPs may attenuate the virulence of invasive infections, they are important virulence factors because they are prevalent, surface-exposed and appear to contribute to disease persistence and antibiotic-resistance.

1.3.5   Identification and characterization of SfbI as an adhesin

SfbI was discovered as a major *S. pyogenes* Fn adhesin by Talay *et al.* (1991; 1992) and also, independently, by Hanksi and Caparon (1992) who named it Protein F (also called prtF, Protein F1 or prtF1). Talay *et al.* (1992) used sequence analysis to identify a series of C-terminal 'R' repeats in SfbI that had sequence similarity to repeats in a *Staphylococcus aureus* FnBP (Figure 3). They cloned the first R repeat, R1, and showed that it bound Fn and could inhibit binding of either Fn to *S. pyogenes* or *S. pyogenes* to HEp-2 cells (Talay *et al.*, 1992). Furthermore, they demonstrated that the R repeats were surface exposed because antibodies raised against R1 bound to the cell surface of *S. pyogenes*. Sela *et al.* (1993) went on to show that the R repeats bound $^{1-5}$F1 in Fn and also identified an upstream Fn-binding region (UR) N-terminally adjacent to the repeats (Figure 3). The UR was shown to bind the GBF of Fn (Figure 2) and, along with the R repeats, mediate internalization of *S. pyogenes* into HEp-2 cells (Section 1.3.6; Ozeri *et al.*, 1996; Sela *et al.*, 1993; Talay *et al.*, 2000).

In addition to adherence to Fn in host tissue, SfbI might also help *S. pyogenes* colonize host tissues *via* mediating adhesion to collagen. Ten percent of 125 *S. pyogenes* isolates have been shown to bind collagen directly and an additional 61% bound collagen in the presence of Fn (Dinkla *et al.*, 2003). The 70 kDa Fn fragment and SfbI R repeats were necessary and sufficient for collagen recruitment. The 70 kDa Fn fragment contains $^{1-5}$F1 and the GBF of Fn (Figure 2). Therefore, it was proposed that SfbI binds to $^{1-5}$F1, resulting in recruitment of collagen *via* the GBF. It was shown, using scanning electron microscopy (SEM), that collagen recruitment resulted in the deposition of an Fn-collagen matrix on the bacterial cell-surface and aggregation of the streptococci, and that surface-bound Fn also mediated adherence to collagen fibres. Furthermore, collagen recruitment was implicated in blocking binding of *S. pyogenes* to phagocytic cells in the absence of serum, indicating that it might help immune evasion in a serum-free environment, such as the skin.

1.3.6 SfbI mediates internalization of *S. pyogenes* by host cells

In vitro studies have elucidated the internalization mechanisms used by *S. pyogenes*. The UR and R-repeats of SfbI were shown to be necessary and sufficient for internalization of *S. pyogenes* strain A40 by HEp-2 cells (Molinari *et al.*, 1997). Internalization of *S. pyogenes* was blocked when the *S. pyogenes* was incubated with recombinant SfbI, or with antibodies directed against either SfbI or the combined UR and R repeats. Furthermore, SEM revealed that latex beads coated with a peptide comprising the UR and R repeats were efficiently internalized by HEp-2 cells, indicating that this peptide was sufficient for internalization and did not require additional bacterial cell-surface components. Therefore, the Fn-binding region of SfbI mediates internalization of *S. pyogenes* by epithelial cells.

S. pyogenes SfbI-mediated internalization requires the presence of both integrins and full-length Fn and is enhanced by the UR (Ozeri *et al.*, 1998; Talay *et al.*, 2000). Ozeri *et al.* (1998) showed that both full-length Fn and integrins are required for internalization. Furthermore, internalization was inhibited by either the combined UR and R repeats or a peptide that contained the RGD sequence, the major integrin binding site in Fn (Figure 2). Therefore, Fn acts as a bridging molecule between the bacterial cell surface-attached SfbI and the host cell surface-attached integrins, resulting in integrin-mediated uptake of the bacteria. Also, SfbI R repeats recombinantly expressed by *Streptococcus gordonii* were found to be sufficient for bacterial adhesion to and

internalization by HEp-2 cells (Talay *et al.*, 2000). The presence of the UR in the same construct did not affect adhesion but increased internalization ten-fold. Therefore, the UR and R repeats in SfbI mediate internalization of *S. pyogenes via* Fn and integrins.

The host cells' integrin-mediated uptake of *S. pyogenes* involves integrin clustering and subsequent actin rearrangement, which leads to internalization of *S. pyogenes* in a type of membrane-bound organelle called a caveosome (reviewed in Nitsche-Schmitz *et al.*, 2007; Schwarz-Linek *et al.*, 2004a). SfbI-mediated integrin clustering was first demonstrated with immunostaining, where it was found that SfbI was unevenly distributed on the *S. pyogenes* cell-surface, and that recruited integrins had the same uneven distribution (Ozeri *et al.*, 2001). It was also shown that gold particles coated with SfbI, visualized with SEM, formed clusters on the surface of HEp-2 cells and were subsequently internalized (Rohde *et al.*, 2003). Given that integrin is recruited by SfbI-bound Fn, the observed clustering of gold particles was likely to be a result of integrin clustering. Inhibition of actin rearrangement also inhibited *S. pyogenes* internalization, demonstrating the involvement of actin rearrangement in internalization (Ozeri *et al.*, 2001). The formation of large invaginations by *S. pyogenes*, which appeared to form from the fusion of smaller invaginations called caveolae have been revealed by SEM studies (Rohde *et al.*, 2003). Internalized bacteria have been shown by transmission electron microscopy to be inside caveosomes, which meant the bacteria avoided fusion with lysosomes that could lead to bacterial cell-lysis (Rohde *et al.*, 2003). Therefore, SfbI appears to be an important mediator of internalization of *S. pyogenes* into an intracellular organelle protected from the host cell's defensive mechanisms.

1.3.7 SfbI FnBRs are probably intrinsically disordered

There is evidence to suggest that the FnBRs in SfbI are intrinsically disordered, that is, they contain no stable secondary or tertiary structure (intrinsic disorder is discussed further in Section 1.5). First, the R repeat region of SfbI was shown to contain no significant secondary structure by far-ultraviolet circular dichroism (CD; House-Pompeo *et al.*, 1996). Similar results were obtained for the repeat regions of FnBPs from *S. aureus* (FnBPA) and *Streptococcus dysgalactiae* (FnBA, FnBB). Furthermore, the difference between the CD spectra of pNTD in the presence and absence of the repeat region from FnBA indicated that complex formation between pNTD and the repeat region led to a significant increase in β-strand content. This led to the conclusion that FnBRs are intrinsically disordered and undergo a disorder-to-order transition on

binding Fn. Second, Penkett *et al.* performed a thorough nuclear magnetic resonance (NMR) analysis on the four Fn-binding 'D' repeats from *S. aureus* protein FnBPA (Figure 5A on p 40). Analysis of the chemical shifts, coupling constants, relaxation data and nuclear Overhauser effect data was used to demonstrate the D repeats were intrinsically disordered (Penkett *et al.*, 1997; Penkett *et al.*, 1998). Given the functional and sequence similarities between FnBRs in FnBPA and SfbI, it is likely that SfbI FnBRs are also disordered.

1.4   The tandem β-zipper model for an FnBR binding to Fn

1.4.1   The tandem β-zipper interaction

Studies attempting to find the minimal binding site in $^{1-5}$F1 for the FnBRs in SfbI or related streptococcal and staphylococcal FnBPs and *vice versa* gave confusing results (reviewed in Joh *et al.*, 1999; Schwarz-Linek *et al.*, 2004a), which were resolved by the structure of an *S. dysgalactiae* FnBP peptide, B3T, in complex with ^{1}F1^{2}F1 (Schwarz-Linek *et al.*, 2003). This complex showed that the FnBP peptide bound to the E-strand of sequential F1 modules, extending each triple-stranded anti-parallel β-sheet by an additional, anti-parallel strand (Figure 4A; Schwarz-Linek *et al.*, 2003). This novel interaction was named a tandem β-zipper.

1.4.2   The tandem β-zipper model

The structure (Figure 4A) and subsequent sequence analysis led to a new function-based definition of FnBRs (Figure 3; Schwarz-Linek *et al.*, 2003). Each FnBR is defined as a sequence that contains four short (~ eight residues), consecutive F1-binding regions that are in the correct order to bind to $^{2-5}$F1 in an anti-parallel fashion (Figure 4C). Therefore, each molecule of SfbI can theoretically bind to five molecules of Fn (Figure 4C). Additionally, some FnBPs have a ^{1}F1-binding motif C-terminal to the most C-terminal FnBR (Figure 4). This motif was identified because the B3T peptide consisted of the C-terminus of an FnBR and a ^{1}F1-binding motif (Figure 4). Combining the FnBR definition with the evidence for unbound FnBRs being intrinsically disordered (Section 1.3.7) led to the tandem β-zipper model for the interaction between an FnBR and $^{2-5}$F1. This model predicts that each FnBR is disordered and undergoes a disorder-to-order transition when binding to $^{2-5}$F1 *via* a tandem β-zipper.

1.4.3 Supporting evidence for the tandem β-zipper model

There are five studies that provide evidence in support of the tandem β-zipper model. First, before the inception of the model, Ozeri *et al.* (1996) used N- and C-terminal deletions of SfbI to identify the minimum upstream Fn-binding region in SfbI. As defined by the tandem β-zipper model, the region they identified consisted of the UR and the most N-terminal FnBR SfbI-1[4], which at the time was considered part of the UR (Figure 4C). The C-terminus of this construct included the seven most N-terminal residues of R1. The tandem β-zipper model predicts these residues bind ^{2}F1, and in support of this prediction it was found that deleting these seven residues greatly reduced binding of Fn (Ozeri *et al.*, 1996).

Second, Schwarz-Linek *et al.* provided evidence that supported the tandem β-zipper model using SfbI FnBRs (Schwarz-Linek *et al.*, 2004b). They dissected SfbI-5 into smaller peptides containing one or two F1-binding regions, and measured their binding affinity for F1 modules and F1 module pairs (*e.g.,* ^{2}F1^{3}F1, ^{4}F1^{5}F1). Using this approach, they showed that F1-binding regions bound their predicted F1 module partner. Further, they measured the binding of pNTD to SfbI-4, SfbI-5 and 'SfbI-4.5'. The latter consisted of the C-terminal half of SfbI-4 and the N-terminal half of SfbI-5, that is, it was similar to the previously defined R4 (Figure 4C). SfbI-4.5 contains F1-binding regions in the following order: ^{3}F1-^{2}F1-^{5}F1-^{4}F1. The tandem β-zipper model predicts that SfbI-4.5 will bind pNTD weakly because the F1-binding regions are not in the correct order for high affinity binding, that is, ^{5}F1-^{4}F1-^{3}F1-^{2}F1 (Figure 4C). This prediction was correct; using isothermal titration calorimetry (ITC) it was shown that pNTD bound to SfbI-4 with a dissociation constant (K_d) of 62 nM, but binding of pNTD to SfbI-5 and SfbI-4.5 was too tight and too weak, respectively, to measure the K_d using the chosen experimental parameters. SfbI-5 was predicted to bind pNTD more tightly than SfbI-4.5 because it contains an ^{1}F1-binding region (Figure 4). Thus, high affinity binding to $^{2-5}$F1 required F1-binding regions to be in the correct order.

---

[4] The first FnBR in SfbI, when numbering each FnBR from the N-terminus. The same nomenclature is used throughout. For example, FnBPA-10 is the tenth FnBR from the N-terminus in the protein FnBPA.

A

B3T

²F1   ¹F1

B

B3T

binds ²F1   binds ¹F1

FnBB 1042 GSTTEVEDSKPKLSIHFDNEWPKED 1066
SfbI   567 ETVTIVEDTRPKLVFHFDNNEPKVE 591

C

SfbI-5

UR   R1   R2   R3   R4

UR   1   2   3   4   5

9 8 7 2 1 6 5 4 3 2   5 4 3 2 1
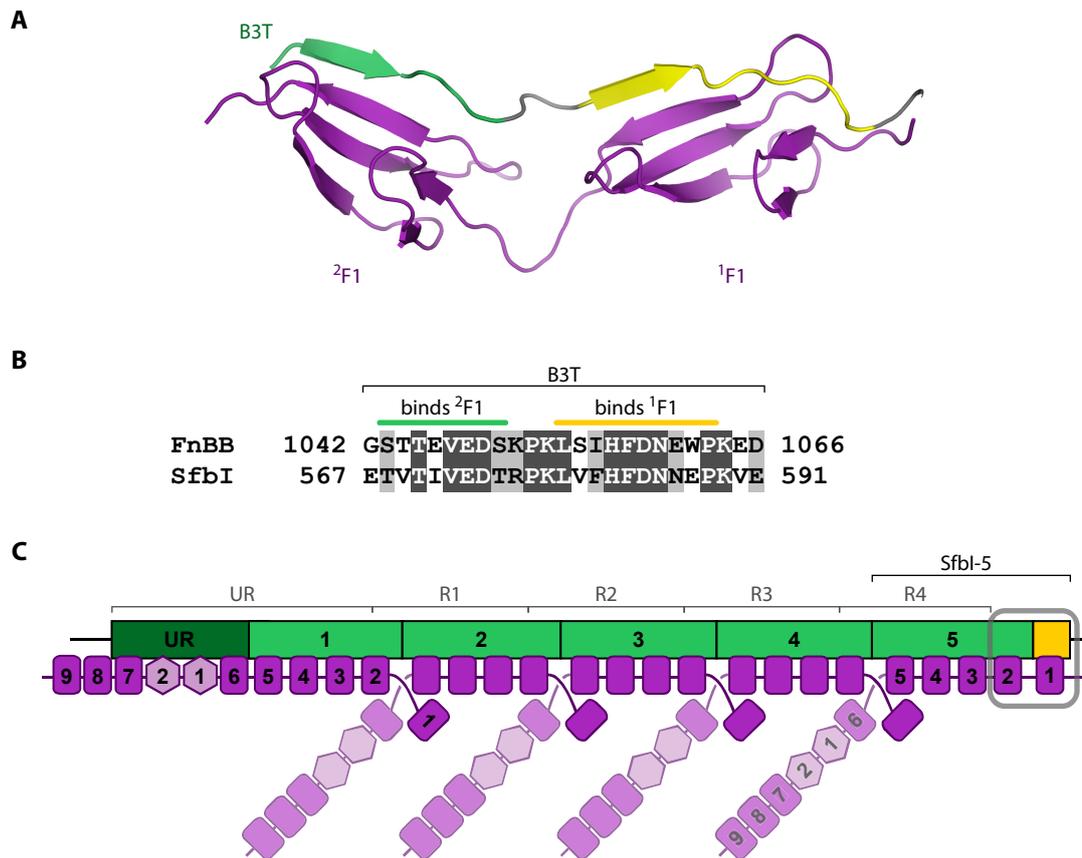
Figure 4. ¹F1²F1:B3T[5] structure led to the tandem β-zipper model, which predicts five Fn monomers can bind SfbI. A. Ribbon diagram of ¹F1²F1 (purple) in complex with B3T (light green and yellow), a peptide from *S. dysgalactiae* FnBB. Shown is the lowest energy structure from the family of NMR structures with PDB code 1o9a (Schwarz-Linek *et al.*, 2003). **B.** Alignment between B3T and a region of SfbI. Identical or conserved residues are shaded dark or light grey, respectively. Indicated above the alignment are the B3T residues involved in the interaction with ¹F1 (yellow) and ²F1 (light green) as identified by Schwarz-Linek *et al.* (2003) from the ¹F1²F1:B3T structure shown in A. This alignment shows that these sequences are highly similar, and hence, this region of SfbI is likely to form a complex with ¹F1²F1 similar to the ¹F1²F1:B3T complex. **C.** Schematic indicating how multiple Fn molecules could bind to a SfbI molecule. Only the NTD and GBF of Fn are shown, for clarity. Each FnBR (light green) is predicted to bind ²⁻⁵F1, except for SfbI-5 (indicated), which contains an additional ¹F1 binding region (yellow). Although the UR (dark green) is known to bind the GBF of Fn, nothing is known about the orientation of the GBF in relation to the UR, so the depiction here is entirely speculative. The module representation and colouring of Fn is as defined in Figure 2A. The grey boxed region designates the location of an interaction between ¹F1²F1 and SfbI homologous to the structure shown in A. Indicated above SfbI are the R repeats (R1–R4) and also the original definition of the UR (Sela *et al.*, 1993; Talay *et al.*, 1992). C is adapted from Figure 3 in Schwarz-Linek *et al.*, 2006.

---

[5] An intermolecular complex between molecules A and B is denoted A:B

Third, a putative FnBR has been characterized in the FnBP BBK32, from the Lyme-disease-causing spirochete *Borrelia burgdorferi*. The putative FnBR has a similar sequence to streptococcal and staphylococcal FnBRs (Raibaud *et al.*, 2005). *In silico* secondary structure predictors, CD, and size exclusion experiments all indicate that the putative FnBR is disordered (Kim *et al.*, 2004). Furthermore, the putative FnBR was shown to bind pNTD, with a concomitant increase in β-strand content, as judged by CD (Raibaud *et al.*, 2005). NMR binding studies with F1-module pairs and BBK32 peptides demonstrated that the tandem β-zipper model had correctly predicted the F1-module binding regions (Raibaud *et al.*, 2005). Further, the peptides appear to bind in an antiparallel fashion to the F1-module pairs, as predicted by the tandem β-zipper model.

Fourth, the tandem β-zipper model predicts *S. aureus* FnBPA contains eleven FnBRs, which greatly extends the region of FnBPA predicted to bind Fn (Figure 5A; Schwarz-Linek *et al.*, 2003). Five of the eleven FnBRs are incomplete, and therefore are predicted to bind Fn more weakly than the other six FnBRs. Meenan *et al.* (2007) confirmed this prediction with solid-phase binding assays and they also measured binding of the six high-affinity FnBRs and one low-affinity FnBR to pNTD using ITC. All high-affinity FnBRs bound pNTD tightly, with a K_d less than 50 nM, and the low-affinity FnBR bound to pNTD weakly, with a K_d of approximately 3 μM. Additionally, they showed 'D' repeats, which had previously been defined based on sequence analysis (Figure 5A), bound pNTD too weakly for the K_d to be measurable with the chosen ITC parameters. Finally, they showed that the sera of patients who had been infected by *S. aureus* contained antibodies to the high-affinity FnBRs in complex with Fn, demonstrating their findings are relevant *in vivo*.

Fifth, Bingham *et al.* (2008) recently solved the structures of peptides from FnBPA-1 and FnBPA-5 bound to F1 module pairs (Figure 5). Each FnBR was dissected into two peptides, predicted to bind either ^{2}F1^{3}F1 or ^{4}F1^{5}F1. In all four structures, each peptide bound to its predicted F1 module pair *via* a tandem β-zipper (Figure 5B). When taken together, therefore, these structures provide the first structural evidence for a tandem β-zipper interaction between an entire FnBR and $^{2-5}$F1.

Therefore, there is currently strong structural and biophysical evidence to support the new function-based definition of FnBRs predicted by the tandem β-zipper model, and the existence of FnBRs in streptococci, staphylococci and *B. burgdorferi*.
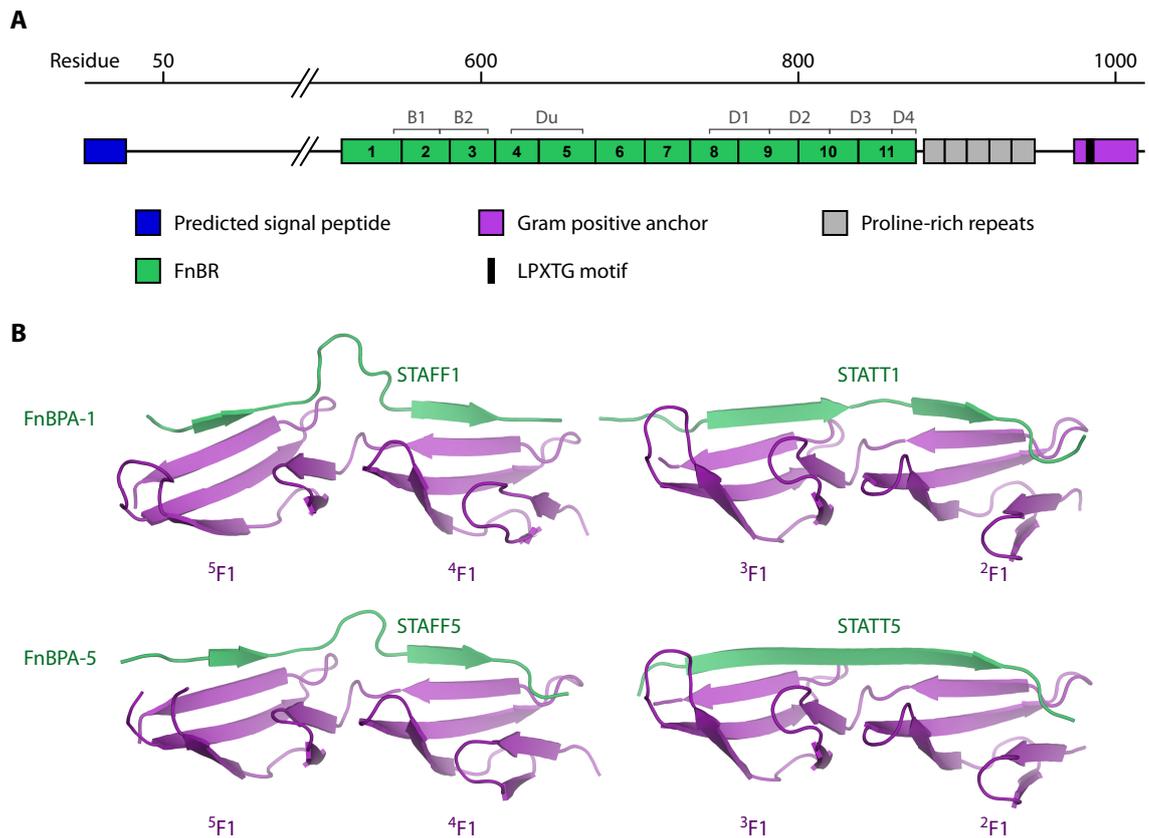
Figure 5. Predicted FnBRs in FnBPA from *S. aureus* bind to $^{2-5}$F1 *via* a tandem β-zipper.
A. Schematic showing the domain organization of FnBPA, with the functional FnBRs (green) as defined using the tandem β-zipper model indicated (Schwarz-Linek *et al.*, 2003). See the legend of Figure 3 for more detail of the other features shown. Sequence analysis led to identification of the proline rich repeats, and also the original definition of Fn-binding B and D repeats, which are indicated above the FnBPA schematic (Signäs *et al.*, 1989); subsequent Fn-binding studies identified the Du region, also indicated (Joh *et al.*, 1998). Schematic adapted from Figure 1 in Schwarz-Linek *et al.*, 2006. **B.** The high-resolution crystal structures of STAFF1 and STATT1 peptides from FnBPA-1 and STAFF5 and STATT5 peptides from FnBPA-5 (green) binding to F1 module pairs (purple) *via* a tandem β-zipper. The pdb codes of the of ^{2}F1^{3}F1:STATT1, ^{4}F1^{5}F1:STAFF1, ^{2}F1^{3}F1:STATT5 and ^{4}F1^{5}F1:STAFF5 structures are 2rkz, 2rky, 3cal and 2rl0, respectively (Bingham *et al.*, 2008).

1.5   Intrinsic disorder

SfbI is predicted to be intrinsically disordered (also known as intrinsically or natively unfolded; Section 1.3.7) and because this thesis investigates the disorder of SfbI-5 (Chapter 3), a brief introduction to intrinsic disorder follows.

A protein, or region of a protein, is intrinsically disordered when it lacks stable secondary and tertiary structure. Instead, it can be considered as an ensemble of conformations that rapidly interconvert. The scale between intrinsic disorder and order

has been divided into four states, although in reality the scale is probably continuous (Dyson & Wright, 2005). At one extreme is an intrinsically disordered protein (IDP) that exists as a random coil, in which the protein's dihedral angles are randomly distributed. Current knowledge suggests that a true random coil does not exist (reviewed by Uversky, 2002). Second, a protein may be mostly random coil but contain regions that transiently adopt some local secondary structure and/or contain hydrophobic clusters. Third, a protein could be an intrinsic 'molten globule', and have stable secondary structure and transient tertiary structure such that the protein is relatively compact. Fourth, an ordered, globular protein may contain disordered loops or flexible linkers.

Research into IDPs increased greatly in recent years as evidence mounted that intrinsic disorder is both prevalent and, because of the advantages of intrinsic disorder, associated with many important functions. The prevalence of IDPs has been suggested by *in silico* sequence analysis of representative genomes, which predicted that 4% of proteins in bacteria and 33% of proteins in eukaryotes contain a disordered region of at least 30 residues (Ward *et al.*, 2004). IDPs are associated with a variety of important functions mostly involving molecular recognition, such as regulation, signalling, chaperone activity, and involvement in macromolecular assemblies (reviewed in Dyson & Wright, 2005; Tompa, 2005). The involvement of IDPs in molecular recognition is probably the consequence of the 'benefits' of disorder. These include the extended conformation of IDP that facilitates a larger 'search area' for possible binding partners, leading to a faster association rate (Tompa, 2005). Also, some IDPs show binding diversity, and can bind to multiple partners by adopting different structures when bound (Tompa *et al.*, 2005). Furthermore, IDPs interact with their binding partners with large interfaces relative to their size, making them a more efficient use of a cell's resources at both the protein and DNA level (reviewed in Gunasekaran *et al.*, 2003).

Numerous physical and chemical properties can be used to identify and characterize IDPs; some examples follow (reviewed in Receveur-Brechot *et al.*, 2006). IDPs have a biased amino acid composition compared to globular proteins, with relatively few hydrophobic residues and more polar or charged residues. *In silico* sequence analysis can, therefore, be used to predict disordered regions (reviewed in Dosztanyi & Tompa, 2008). IDPs have a large hydrodynamic radius relative to globular proteins of the same molecular weight (MW), which can be detected using size exclusion chromatography

(SEC; *e.g.,* Longhi *et al.*, 2003). IDPs that are not molten globules will have minimal stable secondary structure, and this can be measured using CD (*e.g.,* House-Pompeo *et al.*, 1996). Fourth, NMR spectroscopy can investigate many properties of an IDP, including the presence of transient structure (Section 1.6; reviewed in Dyson & Wright, 2001; Mittag & Forman-Kay, 2007). In the work presented here, SfbI-5 was characterized with *in silico* sequence analysis, SEC and NMR spectroscopy.

1.6   Nuclear magnetic resonance (NMR) spectroscopy

NMR is the absorption of energy by atomic nuclei when in the presence of an applied magnetic field, $\boldsymbol{B_0}$ (Hore, 1995). All atomic nuclei have an intrinsic spin quantum number, I, determined by the number of neutrons and protons in the nucleus. For example, ^{12}C and ^{16}O have $I = 0$ and ^{1}H, ^{13}C and ^{15}N have $I = \frac{1}{2}$. Those nuclei with $I \neq 0$ have an angular momentum, called 'spin', which in combination with the positive charge of a nucleus, results in a magnetic dipole. When in the presence of $\boldsymbol{B_0}$, nuclei with $I = \frac{1}{2}$ may align either in the same or the opposite direction as $\boldsymbol{B_0}$. These two states have different energies, and nuclei can be excited from the lower energy state (in the same direction as $\boldsymbol{B_0}$) to the higher energy state by radio frequency electromagnetic radiation according to Equation 1, which results in resonance[6] (Hore, 1995).

$$\upsilon = \frac{\gamma B_0}{2\pi}$$

Equation 1. Frequency of electromagnetic radiation required for resonance when a nucleus is in the presence of an applied magnetic field. Abbreviations are υ, frequency; γ, gyromagnetic ratio, which is a constant that depends on the isotope (*e.g.,* ^{1}H or ^{13}C); B_0, magnitude (strength) of the applied magnetic field $\boldsymbol{B_0}$ (Hore, 1995).

Simply put, NMR spectroscopy is the detection of the resonance frequencies of nuclei in a sample. Which nuclei are detected depends both on the NMR experiment and the isotopes in the sample. Therefore, to detect carbon and nitrogen nuclei in an NMR

---

[6] A spectrometer is referred to by the resonance frequency of ^{1}H when placed in its magnetic field. For example, a 700 MHz spectrometer means that ^{1}H nuclei placed in that spectrometer will resonate at approximately 700 MHz. Given that the gyromagnetic ratio of ^{1}H is 2.675×10^8 rad s^{-1} T^{-1}, this means the magnetic field strength of the 700 MHz spectrometer is 16.4 T.

experiment it is necessary to use an isotope-labelled protein sample so that the carbon and nitrogen isotopes are not the natively abundant ^{12}C and ^{14}N, but ^{13}C and ^{15}N instead.

1.6.1   Chemical shifts

1.6.1.1   Definition of a chemical shift

NMR frequencies are converted to chemical shifts with respect to the resonance frequency of a reference compound according to Equation 2 (Hore, 1995). In biomolecular NMR spectroscopy, the resonance frequency of the methyl protons in reference compound DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid) is defined as zero by the International Union of Pure and Applied Chemistry. Unlike resonance frequencies, chemical shifts are dimensionless and independent of the magnetic field strength. Thus, using a chemical shift scale rather than a frequency scale facilitates comparison between NMR spectra recorded on spectrometers of different magnetic field strengths (Hore, 1995).

$$\delta = \frac{\upsilon - \upsilon_{\text{ref}}}{\upsilon_{\text{ref}}} \times 10^6$$

Equation 2. The chemical shift of a nucleus calculated from its resonance frequency. Abbreviations are δ, chemical shift in ppm; υ, frequency; υ_{ref}, frequency of a reference compound (Hore, 1995).

1.6.1.2   The chemical shift of a nucleus is affected by its chemical environment

Taken together, Equations 1 and 2 imply that the chemical shifts of all nuclei of any given isotope will be identical. Importantly this is not the case, because the nearby chemical environment of a nucleus affects the magnetic field it experiences. Therefore, because the effective magnetic field strength, B, experienced by a nucleus differs slightly from B_0, the nucleus' chemical shift also differs slightly from that calculated with B_0 (Equations 1, 2). Thus, the chemical shift of a nucleus is affected by its molecular environment and, hence, can be used to infer information about that chemical environment (Hore, 1995).

A simple way of understanding the effect of a nucleus' local environment on its chemical shift is to consider the electron density surrounding the nucleus (Hore, 1995). Electrons surrounding the nucleus are themselves affected by $\boldsymbol{B_0}$, and generate a small

magnetic field in the opposite direction of B_0 that shields the nucleus from B_0, reducing B relative to B_0. Therefore, any neighbouring electron-donating group that increases the electron density around the nucleus will increase shielding and, thus, lower the chemical shift of the nucleus. Conversely, neighbouring electron-withdrawing groups decrease shielding (*i.e.,* deshield) the nucleus, resulting in a higher chemical shift. Neighbouring groups do not have to be covalently bonded to affect the chemical shift of a nucleus; for example, a hydrogen bond increases the chemical shift of the involved hydrogen (Hore, 1995).

There are longer range effects on a nucleus' environment that affect the magnetic field strength experienced by the nucleus while not directly effecting its electron density. For example, the circulation of delocalized π-electrons in an aromatic ring, such as benzene, creates an electric current, resulting in a magnetic field that is perpendicular to the π-electrons, and shields a nucleus that is above or below the plane of the ring but deshields a nucleus that is adjacent to the ring (Figure 6; Hore, 1995). Therefore, this so-called 'ring current effect' is magnetically anisotropic. Other functional groups show magnetic anisotropy, including the peptide bond (reviewed in Wishart & Case, 2001). In addition to the direct effect caused by a group with magnetic anisotropy, neighbouring charges or dipoles can also indirectly affect the electron density of the nucleus by polarizing associated chemical bonds; this effect is called an electric field effect (reviewed in Wishart & Case, 2001).
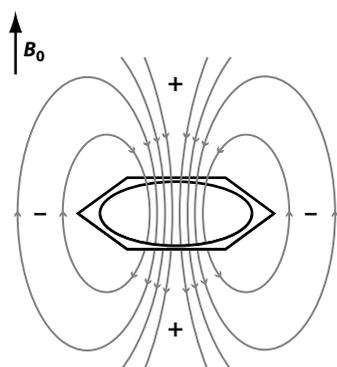
Figure 6. Schematic of magnetic flux lines in a benzene ring when in the presence of an applied magnetic field. When an applied magnetic field, B_0, is perpendicular to the plane of a benzene ring, it induces an electric current in the π-electrons, which results in a local magnetic field, depicted as grey magnetic flux lines. Nuclei above or below the plane of the ring are shielded, and nuclei adjacent to the ring are deshielded (Hore, 1995). Shielding and deshielding are indicated by a plus or minus sign, respectively.

Chemical shifts can be useful for obtaining information about the structure of a protein. For a nucleus in a protein the local and longer range effects of the molecular environment on the chemical shift can be considered as being contributed to by a range of effects: the local effects of backbone and sidechain functional groups and torsion

angles, ring current effects, hydrogen bonds and other close contact effects, electric field effects, and also other miscellaneous effects such as solvent, temperature and pH (reviewed in Wishart & Case, 2001). Therefore, the chemical shift of a nucleus in a protein is contributed to by the protein's primary, secondary and tertiary structure. This means that the chemical shifts of a folded protein are disperse, because its structure creates diverse chemical environments in which the nuclei reside (Figure 7A). In contrast, IDPs have low chemical shift dispersion, because IDPs lack stable secondary and tertiary structure (Figure 7B). However, as previously mentioned (Section 1.5), an IDP may contain transient secondary structure, with some regions of the IDP occupying φ and ψ angles typical of either α-helix or β-strand[7] more often than would occur randomly. Information about an IDP's transient secondary structure can be inferred from the chemical shifts of its nuclei as discussed in Section 1.6.2 because the observed chemical shift of a nucleus is a population-weighted average of the chemical shifts of that nucleus in all molecules in the sample (Dyson & Wright, 2001).
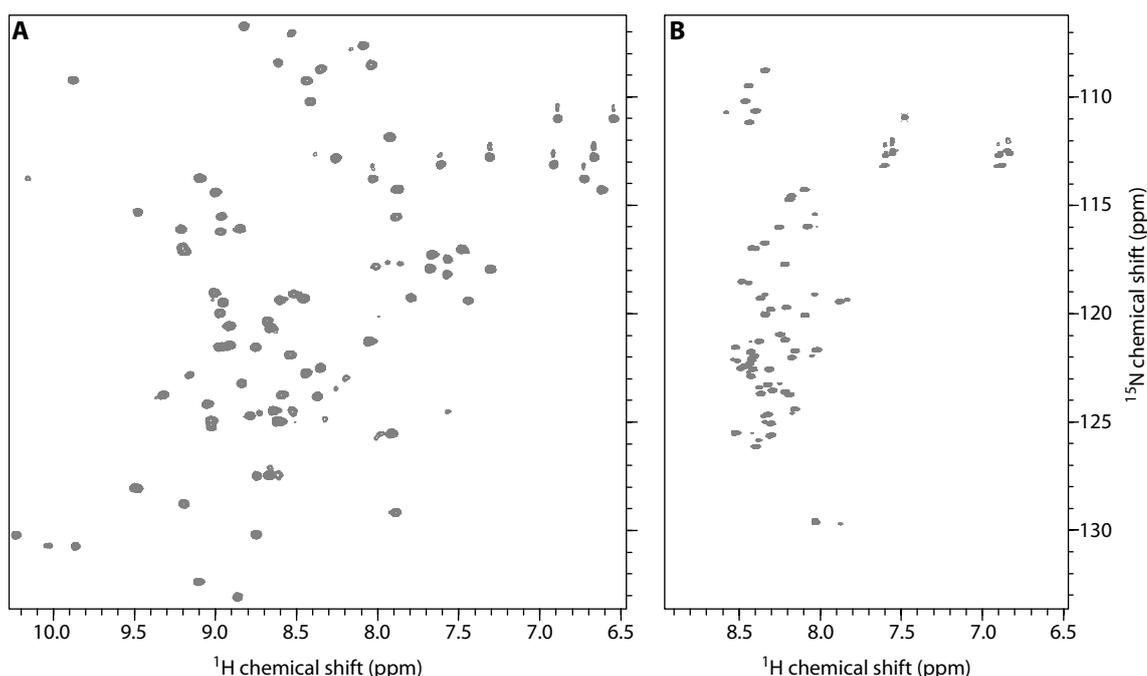
Figure 7. IDPs show low chemical shift dispersion compared to ordered proteins. A. The heteronuclear single-quantum correlation (HSQC) spectrum of a 10 kDa folded protein. **B.** The HSQC of a 6 kDa protein that is probably intrinsically disordered. The spectra were both acquired in our laboratory.

---

[7] The optimal φ and ψ angles for an α-helix are −57 ° and −47°, respectively. For a β-strand, the φ angle is usually between −60° and −180° and the ψ angle is usually between 30° and 180°, respectively (Sun *et al.*, 2004).

1.6.2   Chemical shift analysis to identify secondary structure propensity

1.6.2.1   Random coil chemical shifts

Secondary chemical shifts ($\Delta\delta$) are the differences between observed chemical shifts and sequence-corrected 'random coil' chemical shifts for a nucleus in a protein. $\Delta\delta$ values are a measure of the effect of secondary and tertiary structure on the nucleus' shifts, because it is assumed that by subtracting sequence-corrected random coil shifts, the effects of the primary structure on a nucleus' shifts are removed. Random coil shifts are the shifts of a nucleus in an amino acid residue when it has no stable secondary or tertiary structure, instead existing in a range of rapidly inter-converting conformations (Wishart & Sykes, 1994a). They are determined either experimentally from short peptides that are assumed to be random coil, or statistically-derived from chemical shift databases, where each residue in a structure is first classified as being either random coil, β-strand or α-helix (reviewed in Wishart & Sykes, 1994a; Wishart & Nip, 1998).

The random coil shifts of a residue's nuclei are affected by the chemical nature of a residue's sidechain and by neighbouring residues. In particular, the identity of the preceding residue primarily affects the shifts of amide proton and nitrogen ($^1H^N$ and $^{15}N^H$, respectively). Further, when proline is the following residue, it affects the shifts of $^1H^N$, the alpha proton, the alpha carbon and the carbonyl carbon ($^1H^\alpha$, $^{13}C^\alpha$, $^{13}C'$, respectively; Schwarzinger *et al.*, 2001; Wang & Jardetzky, 2002b). Sequence-dependent effects can be significant, with corrections as large as +4 ppm required for the $^{15}N^H$ shift of a residue where the preceding residue is an isoleucine (Wishart, 1995). Therefore, particularly for disordered proteins, where differences from random coil are likely to be smaller than for folded proteins, it is necessary to use sequence-corrected random coil shifts when calculating $\Delta\delta$ values to properly account for the effects of primary sequence (Schwarzinger *et al.*, 2001; Wishart & Nip, 1998).

Both experimentally and statistically determined random coil shifts and sequence correction factors have disadvantages and advantages, so these will now be discussed in turn. Experimentally determined random coil data are collected in defined conditions and at a known temperature and pH. This can be both an advantage, in that the data should be internally consistent, and a disadvantage, in that the conditions may vary significantly from those conditions under which the chemical shifts of the protein-of-interest were measured. In particular, different concentrations of urea are used to

minimize possible interactions, but urea might affect the random coil shifts (Wang & Jardetzky, 2002a; Wishart & Nip, 1998). Also, because assigning the chemical shifts of all possible combinations of three residue sequences would require experiments on 8000 peptides, the sequence correction factors are determined by measuring the neighbouring residue effects on a single residue type (either glycine or alanine) and assuming this sequence correction factor can be used for all residues, which might not be a valid assumption (Wang & Jardetzky, 2002a).

In contrast to experimentally-derived random coil shift datasets, statistically-derived chemical shift datasets are from proteins (rather than peptides), and provide estimates for not only random coil shifts but also α-helix and β-strand. However, in doing so, these databases classify all residues that are not α-helix or β-strand as random coil, not allowing for the possibility of a region with secondary structural propensity (SSP). Furthermore, to acquire enough data for each amino acid it is necessary to use experiments performed in a variety of conditions, pH and temperatures and with reference shifts other than that of DSS (Wang & Jardetzky, 2002b). This variability in experimental conditions and the natural lack of abundance for certain amino acid combinations could lead to sample bias (Wishart & Sykes, 1994a). However, the standard deviations for all chemical shifts and sequence correction factors is also known (Wang & Jardetzky, 2002b), which gives the researcher a measure of the uncertainty in the reported chemical shifts, which can be helpful in analysis of $\Delta\delta$ values.

1.6.2.2 Secondary structure propensity can be determined using secondary chemical shifts

The $\Delta\delta$ values are useful in determining secondary structure, because residues in either α-helix or β-strand deviate from random coil shifts in a known way. In particular, residues in a β-strand have negative $\Delta\delta$ values for $^{13}C^{\alpha}$ and $^{13}C'$ nuclei (that is, they are shifted to lower chemical shifts relative to random coil values), but have positive $\Delta\delta$ values for $^{1}H^{\alpha}$, $^{1}H^{N}$, $^{15}N^{H}$ and beta-carbon ($^{13}C^{\beta}$) nuclei. The opposite trends occur for residues in an α-helix. For folded proteins, this has allowed the development of a 'chemical shift index' (CSI) by Wishart *et al.* (1992; 1994b), which classifies each assigned nucleus as being either α-helix, β-strand or random coil based on a two-step process. The first step is to assign each residue a value of +1, 0 or −1 based on whether their $\Delta\delta$ values are above, within, or below a range around each random coil value. The second step is to assign a residue as β-strand when there is a local density of at least

three adjacent residues with the appropriate score, for example, −1 for $^{13}C^{\alpha}$. For α-helix, the requirement is a local density of four adjacent residues. Finally, when more than one nucleus type is assigned, a consensus CSI for each residue can be derived based on these residue assignments.

The quantification of SSP from Δδ values is not straight forward. This was demonstrated by Wang and Jardetzky's (2002a) analysis of their first chemical shift database, which contained α-helix, β-strand, and random coil shifts and accompanying standard deviations for the $^{1}H^{\alpha}$, $^{1}H^{N}$, $^{15}N^{H}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ nuclei for each of the 20 residues. In their analysis two key observations were made. First, that the secondary structure of folded proteins can be more accurately predicted by using shifts from all assigned nuclei and not just those used most commonly for CSI calculations (*i.e.,* $^{1}H^{\alpha}$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei). In particular, they found that $^{1}H^{N}$ and $^{15}N^{H}$ were useful for distinguishing β-strand from random coil. Second, they found that for each nucleus type, the Δδ value between β-strand and random coil or α-helix and random coil could differ significantly by residue type. For example, the average Δδ value for $^{13}C^{\alpha}$ is −1.1 and +2.9 for β-strand and α-helix, respectively. However, the range of Δδ values for $^{13}C^{\alpha}$ is −1.8 − −0.4 and +1.6 − +4.5 for β-strand and α-helix, respectively (Wang & Jardetzky, 2002b). Also, the standard deviation for each chemical shift value can vary significantly. This indicates that the nuclei of some residue types are more sensitive determinates of, for example, β-strand, than other residue types. These observations were used to devise the SSP score (Equation 3) for analysis of any SSP in IDPs (Marsh *et al.*, 2006).

$$SSP_i = \frac{\sum_{j=i-2}^{i+2} \sum_{X} \left\{ \text{or} \begin{array}{l} \frac{|\Delta\delta X_{j\,obs}|}{\sigma_j\alpha} \text{if}\left(\Delta\delta X_{j\,obs}\right)\left(\Delta\delta X_j\alpha\right)>0 \\ \frac{|\Delta\delta X_{j\,obs}|}{\sigma_j\beta} \text{if}\left(\Delta\delta X_{j\,obs}\right)\left(\Delta\delta X_j\beta\right)>0 \end{array} \right.}{\sum_{j=i-2}^{i+2} \sum_{X} \left\{ \text{or} \begin{array}{l} \frac{|\Delta\delta X_j\alpha|}{\sigma_j\alpha} \text{if}\left(\Delta\delta X_{j\,obs}\right)\left(\Delta\delta X_j\alpha\right)>0 \\ \frac{|\Delta\delta X_j\beta|}{\sigma_j\beta} \text{if}\left(\Delta\delta X_{j\,obs}\right)\left(\Delta\delta X_j\beta\right)>0 \end{array} \right.}$$

Equation 3. SSP score for residue *i* with weighted averaging over a five-residue window centred on *i*. Annotation: $\Delta\delta X_{j\,obs}$ is the observed Δδ value for nucleus X; $\Delta\delta X_j\alpha$ and $\Delta\delta X_j\beta$ are the average Δδ values for fully formed α-helix and β-strand, respectively; $\sigma_j\alpha$ and $\sigma_j\beta$ are the standard deviations of $\Delta\delta X_j\alpha$ and $\Delta\delta X_j\beta$, respectively. The summation over X is for all chemical shifts for the assigned $^{1}H^{\alpha}$, $^{1}H^{N}$, $^{15}N^{H}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ nuclei for each residue. Equation defined by Marsh *et al.* (2006).

For each residue i, its SSP score is a weighted average of the chemical shifts for all nuclei for a five-residue window centred on i (Equation 3). The output SSP score is a measure of the fraction of protein molecules that are exhibiting a particular secondary structure at that position, where a score of −1 or +1 indicates all molecules have either fully-formed β-strand or α-helix, respectively, at that position. The SSP score weights each observed $\Delta\delta$ value by first predicting the secondary structure type of the nucleus based on the sign of the observed $\Delta\delta$ value, and then weighting the observed $\Delta\delta$ value by the standard deviation of the corresponding $\Delta\delta$ value from the random coil dataset (Equation 3).

This approach gives the SSP score three key advantages over the $\Delta\delta$ values or the CSI. First, it combines all available chemical shift information for all nuclei for each residue. This is likely to improve the accuracy of the prediction, as shown by Wang and Jardetzky (2002a) for their probability-based secondary structure identification (PSSI) method for folded proteins. Second, the weighting method gives greater weight to residues that have greater resolving ability, as determined by their average $\Delta\delta$ values and associated standard deviations for each secondary structure type. Third, unlike the CSI or PSSI, the SSP score is not ternary; it does not require each residue to be either α-helix, β-strand or random coil. Therefore, the SSP score was used to analyse the chemical shift data presented in this thesis.

1.6.3 Chemical exchange in NMR spectroscopy

Chemical exchange is important in understanding some chemical shift perturbation assays performed in this work. Chemical exchange in NMR occurs when there is an equilibrium between two or more states in an NMR sample, and the chemical environment of a nucleus or group of nuclei differs between the states (reviewed in Hore, 1995). For example, in a mixture of a protein with its ligand, the protein can be in either a bound or unbound state, and the chemical environment of at least some of its nuclei will change accordingly. Consider a nucleus which is in equilibrium between state A and B, with a resonance frequency of υ_A and υ_B, respectively. The NMR timescale is determined by the resonance frequency difference between υ_A and υ_B. Slow exchange occurs when the exchange rate between states A and B is significantly slower than the NMR timescale. In this case, two peaks are detected for the nucleus, one each at υ_A and υ_B. Fast exchange occurs when the exchange rate between states A and B is significantly faster than the NMR timescale. In this case, a single peak is observed for

the nucleus, with a resonance frequency that is the weighted average of the population of states A and B. Therefore, if half of the population was in state A and half in state B, the observed chemical shift would be $\frac{1}{2}(\upsilon_A + \upsilon_B)$. Intermediate exchange occurs when the exchange rate between the bound and unbound state is approximately equal to the NMR timescale. In this scenario, the linewidths of the peaks broaden and start to merge, and are thus often undetectable.

1.7 Hidden Markov models (HMMs)

HMMs can be used to model any data taken at regular time intervals or a linear sequence of symbols or numbers. They have been commonly used in speech recognition for decades (Rabiner, 1989), were first applied to modelling multiple sequence alignments (MSAs) by Krogh *et al.* (1994), and, among a range of other uses they are now used to create and maintain widely-used databases of protein families (*e.g.,* Pfam and SMART databases, Finn *et al.*, 2008; Letunic *et al.*, 2009). The databases are created by searching protein sequence databases with libraries of 'profile' HMMs built from manually curated MSAs. In this thesis, a profile HMM was used to identify FnBRs in the UniProt database.

A profile HMM is a formal probabilistic model of an MSA, where each column in the MSA is modelled as a series of match, insertion and deletion states (Figure 8; reviewed in Eddy, 1996; 1998; 2004). There is a probability associated with 'transition' from one state to another, and also emission probabilities for all 20 amino acids at each match and insertion state (Figure 8B). For example, the HMM in Figure 8B was 'built' from a four-column MSA (Figure 8A) containing 70 sequences using the HMMER2 package. Columns 1, 3 and 4 are represented by match states M_1, M_2 and M_3 respectively. Column 2 contains residues in only five of the 70 sequences, so it is modelled as an insertion. That is, there is an increase in the state-transition probability between M_1 and insertion state I_1 (Figure 8). If some of the sequences had contained more than one inserted residue at this position, then the state-transition probability from I_1 back to itself would have increased, and the state-transition probability from I_1 to M_2 would have correspondingly decreased. 25 sequences did not have a residue in column 3, so the state-transition probability from M_1 to deletion state D_2 is correspondingly higher. Column 1 contains primarily bulky hydrophobic residues, column 3 is enriched in lysine and column 4 is primarily glutamate, and these distributions are apparent from the amino acid emission probabilities associated with each match state (Figure 8). Thus, not

only does a profile HMM use position-specific probabilities for residues, but also for initiating and extending an insertion or deletion.
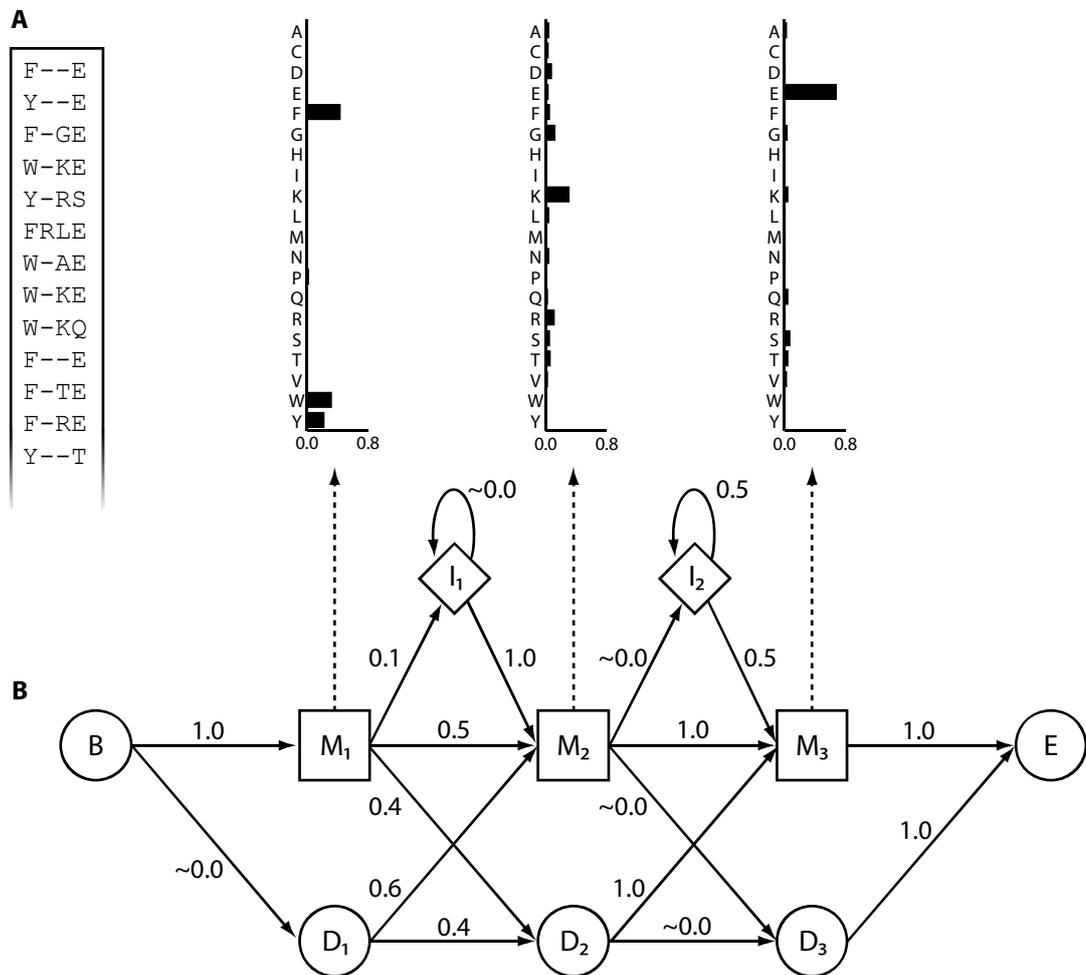
Figure 8. A schematic of an HMM of a MSA. A 70 sequence MSA (part of which is shown in **A.**), was used to build a profile HMM with the HMMER2 package (Eddy, 2003a). This HMM (represented in **B.**) has several features common to all profile HMMs. That is, it consists of a series of match (M), insertion (I) and deletion (D) states represented as squares, diamonds and circles, respectively. A begin (B) and an end (E) state are both shown as circles. State transitions are shown as solid arrows, with their approximate associated probability for transmission from one state to the next indicated next to the arrow. Both match and insertion states are emitting states: each of these states has emission probabilities for all 20 amino acids associated with it. The emission probabilities for each of the match states are represented by a dashed arrow and a bar graph. The insertion state emission probabilities are not shown, but in HMMER2 the emission probabilities are mostly evenly distributed between all amino acids with a slight bias towards hydrophilic amino acids. In HMMER2 this distribution is based on insertions in protein families in an early version of the Pfam database (Eddy, 2003b). Deletion, begin (B) and end (E) states are non-emitting states. This figure is based on Figure 2 in Eddy, 1998 . The HMM file created from the 70 sequence MSA by HMMER2 contains log odds scores that were converted to transmission and emission probabilities with the help of the HMMER user manual (Eddy, 2003b).

Position-specific probabilities for residues, insertions and deletions in HMMs contrasts with the commonly used pairwise alignment algorithms (*e.g.,* BLAST), which use position-independent scoring matrices (*e.g.,* PAM, Gonnet) and gap penalties. Position-specific probabilities are useful because some regions of a protein will be more likely to have insertions or deletions and also have a different amino acid distribution than other regions. For example, the surface of a globular protein is more likely to have hydrophilic loops that can be of variable length, whereas the core of a protein is less likely to have insertions and deletions and more likely to be hydrophobic. Applying this logic to FnBRs, insertions and deletions are more likely to be in the linker regions between F1-binding regions. Therefore, when building an FnBR MSA to act as a seed for an HMM, the emphasis should be on aligning conserved regions likely to be forming interactions, and not on minimizing gaps in the MSA.

Once an HMM has been built, it can then be used to search protein databases for new domains that fit the HMM; the identified sequences are called 'hits'. Whether a sequence is identified as a hit is determined by the sequence's expectation value (E-value), which is the expected number of false positives with a raw score at least as high as the hit within the searched sequence database (Eddy, 2003b). Identical sequences in two databases of different sizes searched by the same HMM will have different E-values; the sequence in the smaller database will have a better (lower) E-value. For example, if a hit has a score of 20 and an E-value of 1.0 in an 1,000 sequence database, then it is expected that one sequence out of the 1,000 could get a score of 20 or better by chance, without actually containing the domain modelled by the HMM. The same sequence searched with the same HMM in a 10 sequence database will still have a score of 20, but its E-value will be 0.01. The default E-value cut-off used by HMMER is 10, because hits with larger E-values are considered too likely to be false positives.

All identified domains can then be aligned to the HMM, creating a new MSA. This MSA can then be used to build a new HMM, which can then be used to search the protein database again. This iterative process is repeated until no new sequences are identified. Thus, besides generating an alignment of all FnBRs from which conserved residues can be identified, this procedure has the added advantage of potentially identifying new FnBPs.

1.8 *Ixodes scapularis* is an important vector for human disease

The HMM searches presented in this thesis (Chapter 4) identified four putative FnBPs from the saliva of the deer tick *Ixodes scapularis*. This thesis investigates whether one of the putative FnBRs is functional (Chapter 5). Hence, a brief introduction to this species and its importance as a vector for human disease follows.

1.8.1 Life cycle of *I. scapularis*

Ticks are bloodsucking, eight-legged arachnids from the subclass Acari, which also includes mites (Sonenshine, 1991). Tick species are subdivided into two major families: the 'hard ticks' (Ixodidae) and 'soft ticks' (Argasdiae). *I. scapularis*, formerly also known as *I. dammini* (Oliver *et al.*, 1993), is an ixodid tick species found in the central and eastern United States. It has a lifespan of about two years, during which it undergoes three developmental stages: larva, nymph and adult. To survive, it feeds at each development stage on a host, which can be any of a wide variety of mammals, birds and reptiles (summarized in Table 2). White-footed mice (*Peromyscus leucopus*) are the primary hosts of *I. scapularis* larvae and nymphs, and white-tailed deer (*Odocoileus virginianus*) are a major host of adults (Sonenshine, 1993).

1.8.2 The feeding process of *I. scapularis*

I. scapularis feeds by first piercing and then ripping and tearing the epidermis with horizontal cutting motions of its chelicerae, which bear denticles (backward projecting teeth; Figures 9, 10; Sonenshine, 1991). Next, the hypostome is inserted (Figures 9, 10), which is also covered in denticles, and provides initial attachment. The palps stay splayed on the skin (Figures 9, 10). Next, attachment 'cement' is secreted within 5–30 minutes of attachment, which spreads into and around the wound and hardens into a latex-like material that strengthens the attachment of the tick to the host (Figure 10). Additional cement secretions are made over the next 48–72 hours, adding layers to the cement. The tick feeds by alternating between ingesting fluids from the feeding lesion and salivating into the lesion. The tick feeds slowly, only increasing its feeding rate in the last 12–24 hours, during which its weight increases rapidly until it reaches engorgement and detaches. It is not known how the tick detaches from the cement. During the feeding process the female adult can increase its weight 100-fold.

Table 2. The life cycle of *I. scapularis*. This information is compiled from Figure 3.1 in Sonenshine, 1991, and from Anderson, 2002.

Steps in *I. scapularis* life cycle

1. Larva hatches from egg

2. Larva uses questing behaviour to search for the host by climbing vegetation or surfaces

3. Having engaged a host, while on the host, the larva:

   a. Explores the host for a suitable site, penetrates the dermis with its mouthparts and subsequently secretes a cement to aid attachment

   b. Ingests the blood pool created by lacerating the dermis with its mouthparts

   c. Reaches engorgement after feeding for 2.5–5 days

   d. Detaches its mouthparts from the host

4. The larva disengages (drops off) from the host

5. The larva moults into a nymph

6. The nymph repeats steps 1–4 with a second host, feeding for 3–5 days

7. The nymph moults into an adult

8. The adult repeats steps 1–4 with a third host, females feeding for 7–10 days

9. Additionally, the adults mate on the host. The male stays on the host to mate again, the female disengages from the host

10. The female lays 2000–3000 eggs
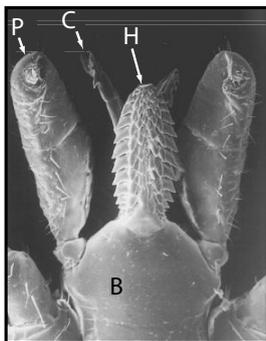
Figure 9. Ventral view of the capitulum of an *I. scapularis* adult female. The capitulum is the 'head', consisting of palps and mouthparts. Abbreviations are: P, palp; C, chelicerae; H, hypostome; B, basis capituli. Taken from Figure 1 in Anderson, 2002.
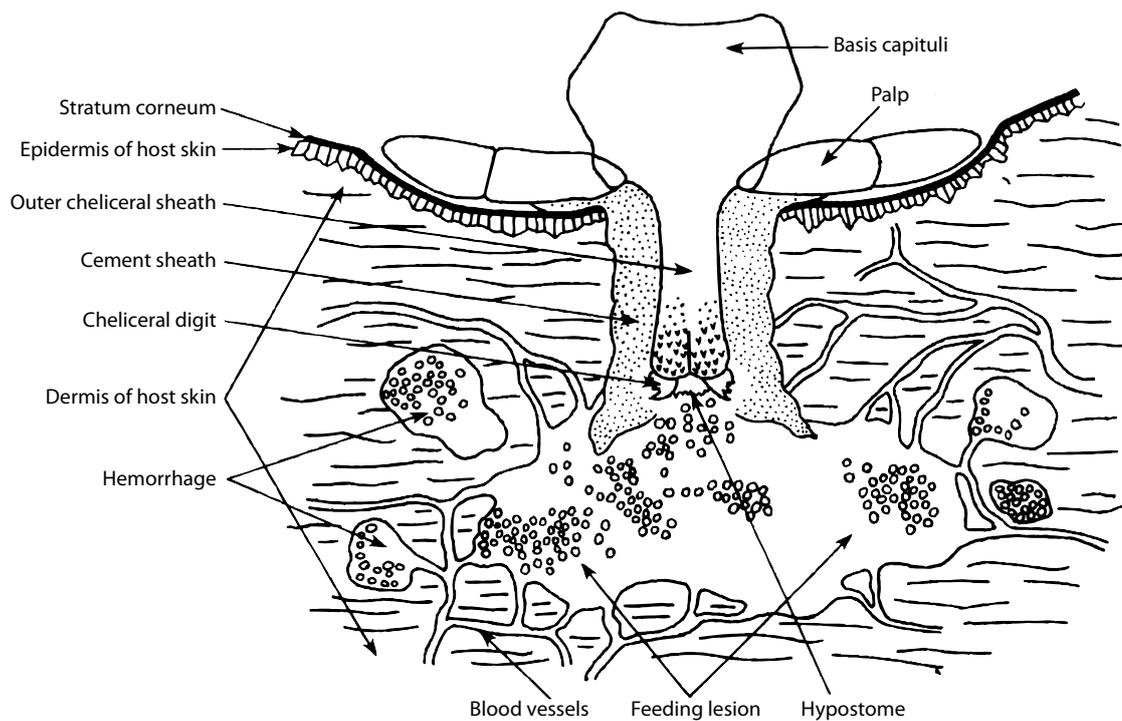
Figure 10. Schematic of the tick's feeding lesion and surrounds. The feeding lesion is filled with blood and inflammatory infiltrates. The cheliceral digits have denticles on their tips, which are used for initially penetrating the skin. The hypostome also has denticles, which help the tick attach while the cement sheath is secreted and solidifies. Taken from Figure 7.19 in Sonenshine, 1991.

1.8.3   *I. scapularis* saliva

1.8.3.1   *I. scapularis* saliva is anti-haemostatic, anti-inflammatory and immunomodulatory

Saliva is important in the tick's days-long feeding process. Saliva allows the tick to secrete excess water and ions, which concentrates the tick's blood meal and maintains the tick's water balance. Saliva also helps the tick maintain a blood pool because it contains a large array of compounds to combat the host's haemostatic, inflammatory and cell-mediated immune responses (reviewed in Valenzuela, 2004). For example, *I. scapularis* combats the host's haemostatic response with prostaglandin E2 and ixolaris. The former inhibits both platelet aggregation and vasoconstriction and the latter inhibits clotting by binding to Factor X, an important protease in the tissue factor pathway. Furthermore, *I. scapularis* saliva inhibits components of the host's inflammatory response to tissue injury, including inhibition of phagocytosis of bacteria by neutrophils and also proteolysis of bradykinin, which causes oedema (tissue swelling) and pain (reviewed in Valenzuela, 2004). Finally *I. scapularis* saliva inhibits the cell-mediated immune response. For example, it contains a protein that binds

interleukin-2, affecting cells with interleukin-2 receptors such as activated T- and B-cells (reviewed in Brossard & Wikel, 2004). *I. scapularis* is so successful at targeting the immune response of one of its primary hosts, *P. leucopus*, that the mouse does not exhibit an immune response even after being fed on repeatedly by *I. scapularis* (reviewed in Anderson, 2002). Therefore, the plethora of biologically-active compounds in tick saliva are critical to its days-long feeding process, and hence, to its success as a parasite.

1.8.3.2 *I. scapularis* saliva mediates disease transmission

I. scapularis is known as the primary vector of Lyme disease in the United States, but can also transmit other diseases, such as rickettsiosis, babesiosis and encephalitis (Sonenshine, 1993). There are several features of the tick's feeding process that contribute to its usefulness as a vector. First, the tick's mouthparts penetrate the epidermis and inject the pathogen from the tick's salivary glands into the host's dermal layers. Second, repeated salivation by the tick into the host during a prolonged feeding time of days gives the pathogen ample opportunity to infect. Third, the pathogen is protected from the host's immune system by anti-inflammatory and immunomodulatory compounds in the saliva, which have been shown to enhance pathogen transmission.

This 'salivary-activated transmission' has been demonstrated for *I. scapularis* transmission of *B. burgdorferi* (reviewed in Brossard & Wikel, 2004; Nuttall & Labuda, 2004). Mice were injected with *B. burgdorferi* in the presence or absence of *I. scapularis* salivary gland lysate, and the bacterial load was determined eight weeks later. It was shown that the presence of saliva significantly increased bacterial dissemination (Zeidner *et al.*, 2002). Therefore, not only is tick saliva critical to the success of the tick as a parasite, but it also contributes to the success of other pathogens for which the tick is a vector.

1.8.3.3 Research into *I. scapularis* salivary components is important

As a result of the role of the saliva of *I. scapularis* in its success as both a parasite and a disease vector to both human and animals, understanding the saliva's biologically active components is important. Knowledge of tick saliva has increased in recent years with high throughput research such as the *I. scapularis* genome project and expressed sequence tag analysis of tick salivary glands. This knowledge is sought for two key

reasons. First, biologically active components could be novel drug targets. Second, because acquired host immunity to saliva reduces pathogen transmission, there is research into the development of vaccines using components of saliva (reviewed in Brossard & Wikel, 2004). These vaccines could either inhibit salivary-activated pathogen transmission by targeting compounds that help transmission, or the vaccines could elicit a high immune response at the feeding lesion against salivary proteins, which would indirectly kill transmitted pathogens (reviewed in Brossard & Wikel, 2004; Valenzuela, 2004). The advantage of a tick vaccine is that it could prevent transmission of all the pathogens for which it is a vector.

1.9 Aims

The overall aim was to investigate the proposed tandem β-zipper interaction between SfbI-5, the fifth FnBR of SfbI from *S. pyogenes*, and $^{1-5}$F1 from Fn.

Therefore, the specific aims of the research presented here were:

1. To confirm and characterize the predicted disordered state of SfbI-5 (Chapter 3);

2. To investigate the secondary structure propensity of SfbI-5 using chemical shift analysis of the backbone resonances of SfbI-5, and so, to determine if the regions of SfbI-5 predicted to bind F1-modules have propensity for β-strand (Chapter 3);

3. To determine the feasibility of characterizing the $^{1-5}$F1:SfbI-5 complex with NMR spectroscopy (Chapter 3);

4. To demonstrate that SfbI-5 can interact with F1-modules *via* a tandem β-zipper by solving the structure of an SfbI-5 peptide in complex with an F1-module pair (Chapter 4);

5. To investigate the role of conserved FnBR residues in SfbI-5 by:

   a. Identifying conserved FnBR residues by using HMMs to rigorously search the UniProt database for all FnBRs (Chapter 4);

   b. Measuring the effect of mutating conserved FnBR residues in SfbI-5 on SfbI-5 binding to $^{1-5}$F1 or Fn (Chapter 4);

   c. Analysing the results with respect to available F1-module:FnBR complex structures (Chapter 4).

During the course of the research, putative FnBRs from the deer tick *I. scapularis* (Chapter 4) were identified, including TickFnBR.

Hence, the additional following aims were pursued:

6. To determine if TickFnBR could bind $^{1-5}$F1 both in the absence and presence of another FnBR, and hence, if it was likely TickFnBR bound to $^{1-5}$F1 *via* a tandem β-zipper (Chapter 5);

7. To determine if TickFnBR could bind to Fn in human plasma (Chapter 5);

8. To demonstrate that a TickFnBR peptide interacts with F1-module pairs *via* a tandem β-zipper by solving the structure of the complex (Chapter 5).

2 Methods

2.1 Materials

2.1.1 Chemicals, reagents and enzymes

All standard laboratory chemicals and reagents were of analytical grade and were purchased as specified in Table 3.

Table 3. Chemicals or reagents used in this work. Standard laboratory chemicals not listed here were purchased from Fisher Scientific.

Chemical or reagent	Abbreviation	Company
Agar		Oxoid
Agarose (Electrophoresis grade)		Melford
4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride	AEBSF	Melford
Ampicillin, sodium salt	Amp	Melford
1,3-Bis (tris (hydroxymethyl) methylamino) propane	Bis-tris propane	Sigma Aldrich
Deoxynucleotide triphosphates	dNTPs	Novagen
Deuterium oxide	D_2O	GOSS scientific.
2,2-Dimethyl-2-silapentane-5-sulfonate	DSS	GOSS Scientific
Dithiothreitol	DTT	Melford
100 basepair DNA ladder		New England Biosciences
Isopropyl β-D-thiogalactopyranoside	IPTG	Melford
Mark 12 protein markers		Invitrogen
Sodium azide	NaN_3	Sigma Aldrich
SYBR safe DNA gel stain		Invitrogen
Tris (hydroxymethyl) aminomethane (Tris)	Tris	Invitrogen
Tryptone		Oxoid
Yeast extract		Melford
Yeast nitrogen base without amino acids or ammonium sulfate	YNB	Sigma Aldrich

2.1.2 Buffers and solutions

All buffers and solutions were prepared using high purity water (HPW), which had been distilled, de-ionized and then purified with a PURELAB Ultra system (Elga). The

compositions of all buffers used in this work are summarized in Table 4, except those used in protein electrophoresis, which are summarized in Table 10.

Table 4. Composition of buffer and media solutions. Bacterial growth broth and plates were prepared as described in Sambrook *et al*. (2001). Unless specifically stated, the pH of these solutions was not adjusted.

Buffer or media	Composition
LB-Amp broth	1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% NaCl (w/v), 100 µg/mL Amp
LB-Amp plates	LB-Amp broth with 1.5% (w/v) agar
Glutathione elution buffer	50 mM Tris, 10 mM glutathione, 150 mM NaCl, pH ~8 at temperature of experiment (either room temperature or 4°C).
Phosphate buffered saline (PBS)	140 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4
M9 broth	0.2% (w/v) glucose or ^{13}C glucose, 15 mM $^{15}NH_4Cl$, 48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 2 mM $MgSO_4$, 0.17% YNB, 100 µg/mL Amp
Tris buffered saline (TBS)	50 mM Tris, 150 mM NaCl, pH 7.5

2.1.3 Enzymes

The protease Factor Xa and the type II restriction endonucleases *Bam* H1 and *Dpn* I were purchased from New England Biosciences. *PfuUltra* high-fidelity DNA polymerase was purchased from Stratagene. The hydrolases lysozyme and deoxyribonuclease (DNase) were purchased from Sigma Aldrich. These enzymes were used with their supplied buffers where applicable.

2.1.4 *Escherichia coli* strains

All *E coli* strains were purchased as competent cells from Stratagene. Solopack Gold or XL1-Blue cells were used for molecular biology and plasmid maintenance and BL21 or BL21 Gold cells were used for protein overexpression. The genotypes of these strains are shown in Table 5.

Table 5. Genotypes of the *E. coli* strains used for molecular biology, plasmid maintenance and protein expression. All strains were purchased as competent cells from Stratagene.

Strain	Genotype
Solopack Gold	Tet^r Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F′ *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r) Amy Cam^r]
Xl1-Blue	Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F′ *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r)]
BL21	B F$^-$ *ompT hsdS*(r$_B^-$ m$_B^-$) *dcm$^+$ gal*
BL21 Gold	B F$^-$ *ompT hsdS*(r$_B^-$ m$_B^-$) *dcm$^+$* Tet^r *gal endA* Hte

2.1.5 Peptides and proteins

Peptides and proteins purchased for this work are summarized in Table 6. All full-length FnBRs were produced by overexpressing and purifying a recombinant construct, as described in Section 2.2. The F1 module pairs ^{2}F1^{3}F1 and ^{4}F1^{5}F1 were generously provided by both Miss Gemma Harris and Dr Richard Bingham (Potts Group, Department of Biology, University of York), who expressed and purified the F1 module pairs using a procedure similar to that described previously (Bright *et al.*, 2000). Recombinant $^{1-5}$F1 was generously provided by either Dr Caroline Milner (Department of Biochemistry, University of Oxford) or the Höök laboratory (Centre for Extracellular Matrix Biology, Institute of Biosciences and Technology, Texas A & M University System Health Science Centre, Houston). This construct differed from pNTD because it did not contain the 17 residues N-terminal to $^{1-5}$F1. Human plasma was purchased from Sera Laboratories.

Table 6. Peptides and proteins purchased for this work. Both synthetic peptides had an acetyl group on their C-terminus and an amide group on their N-terminus. See Footnote 8 on p 62 for information on residue numbering.

Name	Description	Company
pNTD	30 kDa proteolytic fragment from human plasma Fn	Sigma Aldrich
IxTT1	Synthetic peptide, corresponds to residues 36–53 of UniProt ID Q4PN68	Alta bioscience
PyTT5	Synthetic peptide, corresponds to residues 560–577 of UniProt ID Q01924	Alta bioscience

2.2 Molecular biology of pGEX FnBR constructs

2.2.1 Introduction

Genes encoding SfbI-5 (residues 541–591[8] in UniProt ID Q01924) and TickFnBR (residues 18–54[8] of a putative FnBP from *I. scapularis*, UniProt ID Q4PN68) were cloned into the pGEX-5X-2 vector (GE Lifesciences) creating, respectively, pGEX-SfbI-5 and pGEX-TickFnBR. The pGEX-SfbI-5 construct was created by Sophie Raibaud (Schwarz-Linek *et al.*, 2003). Mutants of SfbI-5 were created from this construct as described in Section 2.2.3. The pGEX-TickFnBR construct was created as described in Section 2.2.4.

All these pGEX constructs express the FnBR fused to the C-terminus of glutathione *S*-transferase (GST). Expression of the GST-FnBR fusion from the pGEX construct is under the control of a *tac* promoter, which allows strong expression of the GST-fusion by *E. coli* RNA polymerase when the culture is induced with IPTG (Figure 11; Smith & Johnson, 1988). The GST facilitates purification of the expressed GST-FnBR fusion from other soluble *E. coli* protein by glutathione affinity chromatography. Also, there is a specific Factor Xa cleavage site between the GST and the FnBR sequences that, when cleaved, leaves a non-native Gly-Ile di-peptide sequence at the N-terminus of the FnBR.

Note that unless otherwise specified, all molecular biology techniques were carried out according to Sambrook *et al.* (2001), or according to manufacturer's instructions in the case of commercial kits, with the exception of the transformations of *E. coli* with plasmid DNA, which were carried out as described in Section 2.2.2.

---

[8] Residue numbering used throughout this thesis for FnBPs always includes the predicted N-terminal signal sequence.

A

B
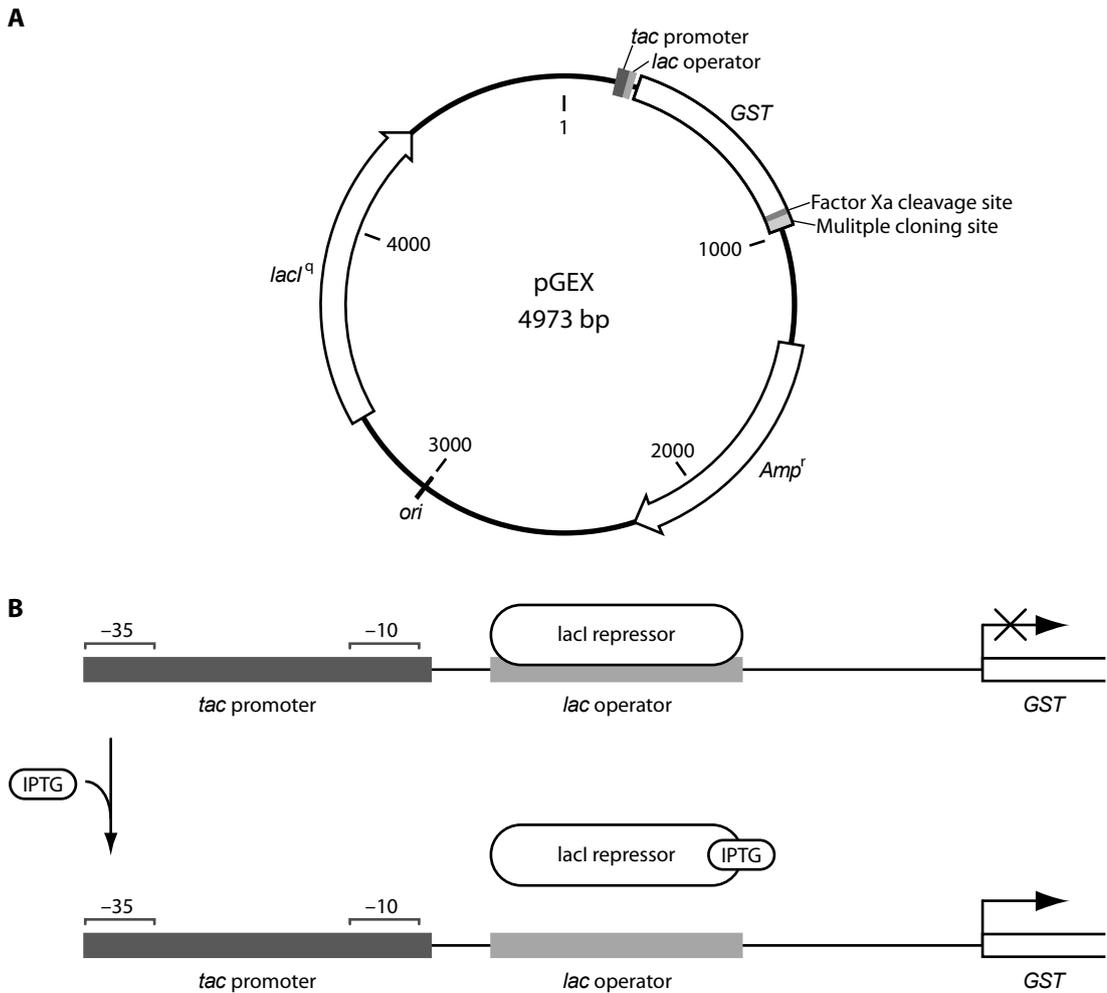
Figure 11: Several properties of the pGEX vector allow strong, inducible expression of a GST-fusion by *E. coli*. A. A schematic of the pGEX vector. The pGEX vector contains the following components that allow it to confer strong, inducible expression of a GST-fusion by *E. coli* (Smith & Johnson, 1988). (i) The origin of replication (*ori*), which facilitates replication of the pGEX vector in *E. coli*. (ii) The β-lactamase gene (*Amp*r), which confers resistance to ampicillin. This selects for *E. coli* cells successfully transformed with the pGEX vector when the cells are grown in the presence of ampicillin. (iii) A multiple cloning site at the C-terminus of the *GST* gene, which facilitates cloning of the desired protein into the pGEX vector. (iv) The *tac* promoter, which facilitates strong expression of the GST-fusion by *E. coli* RNA polymerase because it is a hybrid of the *E. coli trp* and *lacUV5* promoters. Promoters from *E. coli* contain two conserved regions, the -10 and -35 regions. The *tac* promoter contains the *E. coli* consensus sequences for both these regions, the -35 region is from the *trp* promoter and the -10 region is from the *lacUV5* promoter (Amann *et al.*, 1983). These changes resulted in the *tac* promoter expressing approximately five times more of a test protein compared with the *lacUV5* promoter (Amann *et al.*, 1983). (v) The *lac* operator, which, when bound by the lacI repressor, represses expression of the GST-fusion. On addition of IPTG, the IPTG binds to the lacI repressor, releasing it from the *lac* operator and allowing expression of the GST-fusion. Therefore, expression from the pGEX vector is inducible. (vi) The lacI repressor gene (*lacI*q), which ensures there is enough lacI repressor in the cell to repress expression of the GST-fusion. **B.** Expression of the GST-fusion from the pGEX vector is inducible with IPTG. See (v) above for more explanation.

2.2.2 Transformation of *E. coli* cells with plasmid DNA

The transformation protocol followed was developed by Pope and Kent (1996). Briefly, fresh LB-Amp plates (Table 4) were pre-warmed at 37°C and competent *E. coli* cells that had been stored at −80°C and had undergone no more than one freeze–thaw cycle were thawed on ice. Plasmid DNA (0.5 µL, 10 ng/µL) or a reaction mixture containing newly created construct (0.5–1 µL) was added to the thawed cells (30–50 µL) and the tubes were flicked gently to mix and returned to the ice. As a negative control, an aliquot of cells without DNA was also prepared. After at least 5 minutes, the cells were spread onto the pre-warmed plates. For small volumes of cells, spreading was made easier by adding 100 µL of pre-warmed LB-Amp broth (Table 4) to the plates directly prior to adding the cells, this idea was adapted from Sambrook *et al.* (2001). The plates were incubated at 37°C overnight, and then stored at 4°C.

2.2.3 pGEX-SfbI-5 mutant constructs

The pGEX-SfbI-5 mutant constructs were created using the Quikchange II site-directed mutagenesis kit (Stratagene). First, the pGEX-SfbI-5 construct was sequenced (Section 2.2.5). Then, mutagenic primers were designed for this sequence using the Quikchange Primer Design Program (www.stratagene.com/sdmdesigner/default.aspx). This program designs two complementary primers for each mutation, such that each primer is between 25–45 basepair (bp) long, with 10–15 bp of DNA on either side of the mutation, with a predicted melting temperature of at least 78°C, and a GC nucleotide base content of at least 40%. The mutated SfbI residue, the name of the associated pGEX-SfbI-5 mutant construct, and primers used to create the mutation are shown in Table 7.

Table 7. Primers used to create pGEX-SfbI-5 mutant constructs. The mutation with respect to full length SfbI, the name of the associated pGEX-SfbI-5 construct and the forward (a) and reverse (b) primers are shown. A mutation of V552A indicates residue 552 in full length SfbI, which is a valine, is mutated to alanine. Primers were purchased from MWG-Biotech.

Mutation	Construct		Primer (5′ → 3′)
V552A	pGEX-Py501	a	gggaggtcaaagtgagtctgctgaatttactaaagacactc
		b	gagtgtctttagtaaattcagcagactcactttgacctccc
E553A	pGEX-Py502	a	gaggtcaaagtgagtctgttgcatttactaaagacactcaaac
		b	gtttgagtgtctttagtaaatgcaacagactcactttgacctc

Mutation	Construct	Primer (5′ → 3′)	
F554A	pGEX-Py503	a	ggaggtcaaagtgagtctgttgaagctactaaagacactcaaacagg
		b	cctgtttgagtgtctttagtagcttcaacagactcactttgacctcc
K556A	pGEX-Py504	a	gtgagtctgttgaatttactgcagacactcaaacaggcatgagc
		b	gctcatgcctgtttgagtgtctgcagtaaattcaacagactcac
K556E	pGEX-Py505	a	aagtgagtctgttgaatttactgaggacactcaaacaggcatgagc
		b	gctcatgcctgtttgagtgtcctcagtaaattcaacagactcactt
D557A	pGEX-Py506	a	gtctgttgaatttactaaagccactcaaacaggcatgagcg
		b	cgctcatgcctgtttgagtggctttagtaaattcaacagac
T558A	pGEX-Py507	a	ctgttgaatttactaaagacgctcaaacaggcatgagcgg
		b	ccgctcatgcctgtttgagcgtctttagtaaattcaacag
E573A	pGEX-Py508	a	aacagtgaccattgttgcagatacgcgtccgaagt
		b	acttcggacgcgtatctgcaacaatggtcactgtt
D574A	pGEX-Py509	a	cagtgaccattgttgaagctacgcgtccgaagttag
		b	ctaacttcggacgcgtagcttcaacaatggtcactg
T575A	pGEX-Py510	a	gtgaccattgttgaagatgcgcgtccgaagttagtg
		b	cactaacttcggacgcgcatcttcaacaatggtcac
Q548A	pGEX-Py511	a	ccggagtattgatgggaggtgcaagtgagtctgttgaattta
		b	taaattcaacagactcacttgcacctcccatcaatactccgg
K556E, T558A	pGEX-Py512	a	gttgaatttactgaggacgctcaaacaggcatgagcg
		b	cgctcatgcctgtttgagcgtcctcagtaaattcaac
M562A	pGEX-Py513	a	ctaaagacactcaaacaggcgcgagcggtttcagtgaaacag
		b	ctgtttcactgaaaccgctcgcgcctgtttgagtgtctttag
G564A	pGEX-Py514	a	tcaaacaggcatgagcgctttcagtgaaacagtga
		b	tcactgtttcactgaaagcgctcatgcctgtttga
G564Q	pGEX-Py515	a	gacactcaaacaggcatgagccagttcagtgaaacagtgaccatt
		b	aatggtcactgtttcactgaactggctcatgcctgtttgagtgtc
I571A	pGEX-Py516	a	tttcagtgaaacagtgaccgctgttgaagatacgcgtccg
		b	cggacgcgtatcttcaacagcggtcactgtttcactgaaa

The Quikchange method takes advantage of the methylated state of plasmid DNA isolated from *dam*[+] strains of *E. coli* (most commonly used *E. coli* strains are *dam*[+], including Solopack Gold). The site-directed mutagenesis protocol was performed as per manufacturer's instructions, and is outlined in Figure 12. For each mutant, two *E. coli*

colonies transformed with the mutated plasmid were used to make a stock solution of each mutated plasmid with a QIAprep Spin Miniprep Kit (Qiagen). Plasmid DNA was sequenced (Section 2.2.5) to confirm the presence of the mutation.
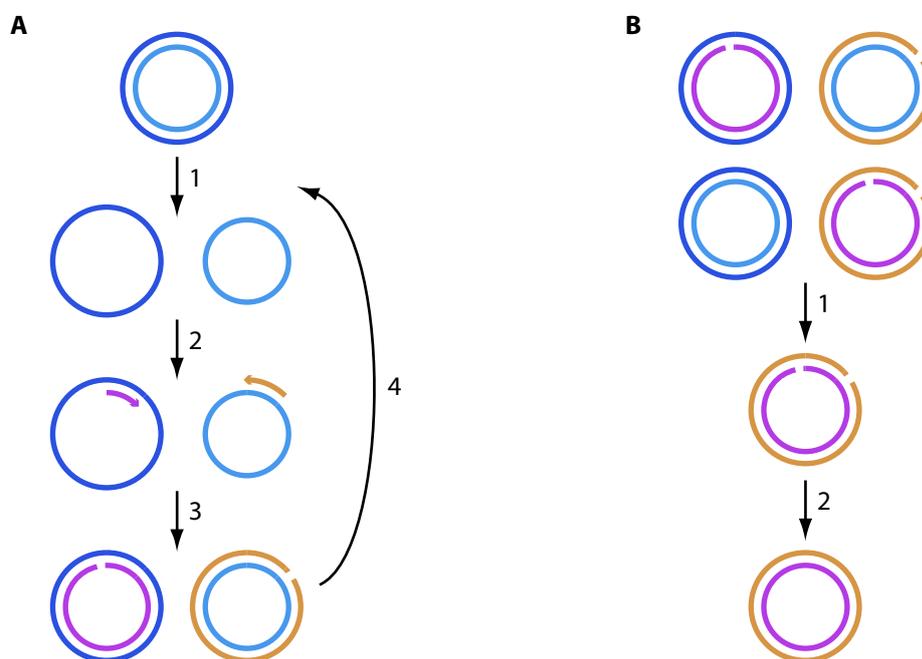
Figure 12. Schematic of the Quikchange II site-directed mutagenesis method. A. Mutated plasmid was created using temperature cycling. 1) Plasmid DNA isolated from a *dam*$^+$ strain of *E. coli* (light and dark blue circles) was denatured at 95°C. This DNA has been methylated by the dam methylase present in *E. coli*. 2) Complementary primers containing the mutation (purple and orange arrows) were then annealed to the denatured plasmid at 55°C. 3) The primers were extended by the high fidelity DNA polymerase *PfuUltra* at 68°C, which created nicked circular strands that contained the mutation (purple and orange circles). The strands are not methylated because the dNTPs used in the reaction were not methylated. 4) Steps 1–3 were repeated 16 times. **B.** Mutated plasmid was selected for on the basis of its methylation state. After temperature cycling, the reaction contained a mixture of methylated, non-methylated and hemi-methylated plasmids. Only non-methylated strands contained the mutation. 1) The endonuclease *Dpn* I was used to specifically digest methylated and hemi-methylated plasmid DNA for 1 h at 37°C. This left undigested double-stranded non-methylated plasmid DNA, which contained the mutation on both strands and had two staggered nicks. 2) The nicked plasmid DNA was transformed into XL1-Blue Supercompetent *E. coli* cells (Section 2.2.2), where *E. coli* repair enzymes ligated the nicks. The large overhang between the two nicked strands meant that a separate ligation reaction was not necessary.

2.2.4   pGEX TickFnBR construct

Four primers, TickFwdL, TickFwdS, TickRevL and TickRevS were used in a polymerase chain reaction (PCR) to create the *TickFnBR* gene construct (Figure 13A).

TickFwdL and TickRevL were the PCR template; these long primers overlap at their 3′ ends and together encode the entire *tickFnBR* construct (Figure 13A). *tickFnBR* encodes residues 18–54 of the putative *I. scapularis* FnBP, UniProt ID Q4PN68, using codons optimized for *E. coli*. Also, the 5′ and 3′ ends of the construct contain, respectively, 15 bp and 14 bp with homology to either side of the *Bam* HI cleavage site in pGEX-5X-2 (Figure 13).

A

TickFwdS

TickFwdL

```
1   GAAGGTCGTGGGATCATGGTGGAAGTTGTCGGTGGCAATGTGGAGTTCGATATTCACTCC 60
1    E  G  R  G  I  M  V  E  V  V  G  G  N  V  E  F  D  I  H  S  20
```

TickFwdL

TickRevL

```
61  ACCGGCGGTGGCATCTCGGGTCACAATCAAGGTCCTCAAACTGAAGTTGAGGACACAGGT 120
21   T  G  G  G  I  S  G  H  N  Q  G  P  Q  T  E  V  E  D  T  G  40
```

TickRevL

TickRevS

```
121 GAAAGTTGATGATAAGATCCCCGGAATTC 149
41   E  S  *  *  *              42
```

B

Bam HI

```
921  ATCGAAGGTCGTGGGATCCCCGGAATTCCC 948
      I  E  G  R  G  I  P  G  I  P
```
Factor Xa

Figure 13: Sequences used to create pGEX-TickFnBR. A. *tickFnBR* gene sequence with corresponding protein translation shown underneath. The four primers used to create the construct are shown above the DNA sequence, with forward primers in red and reverse primers in blue. **B.** pGEX-5X-2 sequence immediately surrounding the *Bam* HI cleavage site, with corresponding protein translation shown underneath. Factor Xa and *Bam* HI cleavage sites are indicated. In both A and B identical regions of DNA that are used in the In-Fusion reaction to clone *tickFnBR* into pGEX-5X-2 are green.

To create the pGEX-TickFnBR construct, an In-Fusion Dry-Down PCR Cloning Kit (Clontech) was used to clone the purified *tickFnBR* PCR product into pGEX-5X-2 that had previously been linearized with *Bam* H1. This In-Fusion reaction was used to transform competent Solopack Gold *E. coli* cells (Section 2.2.2). The In-Fusion Kit was kindly supplied by Dr Jared Cartwright (Technology Facility, Department of Biology,

University of York) and advice on implementing the accompanying manufacturer's protocol was kindly provided by Dr Mark Saw (Technology Facility, Department of Biology, University of York). Plasmid stock solutions were created from two successfully transformed *E. coli* colonies with a QIAprep Spin Miniprep Kit (Qiagen) and sequenced in both the forward and reverse directions (Section 2.2.5).

2.2.5 DNA sequencing

All DNA sequencing was performed by the Genomics Laboratory (Technology Facility, Department of Biology, University of York). Plasmid DNA that had been prepared with either a QIAprep Spin Miniprep Kit or QIAGEN Midi Kit (Qiagen) was supplied for sequencing within a concentration range of 100–150 ng/μL as determined by absorbance at 260 nm. Sequencing was always performed in both the forward and reverse directions, using forward and reverse primers designed to sequence DNA cloned into a pGEX vector. These primers were kindly supplied by the Genomics Laboratory, and their sequences are shown in Table 8. Results were analysed using the freeware programs FinchTV (Geospiza, www.geospiza.com/Products/finchtv.shtml) and CLC Sequence Viewer (CLC bio, www.clcbio.com).

Table 8. Primers used for sequencing the multiple cloning site of the pGEX cloning vector in both the forward and reverse directions (pGEXfor and pGEXrev, respectively).

Name	Sequence (5′ → 3′)
pGEXfor	gggctggcaagccacgtttggtg
pGEXrev	ccgggagctgcatgtgtcagagg

2.3 Producing and purifying FnBRs

2.3.1 Overexpression of ^{15}N- or ^{15}N, ^{13}C-labelled GST-SfbI-5

Two *E. coli* BL21 Gold colonies, which had been transformed with the appropriate pGEX construct, were each used to inoculate isotope-labelled M9 broth (Table 4) and incubated for 18 h with shaking at 37°C. This culture and all others in this section were shaken at approximately 200 rotations per minute (rpm). Each of the cultures was then used to inoculate 50 mL isotope-labelled M9 broth in a 250 mL flask at a ratio of 1:25 (v/v), and incubated with shaking at 37°C for 3.5 h until the cultures reached an optical density at 600 nm (OD$_{600}$) of ~0.8 (Section 2.3.6). The cultures were then

transferred to sterile tubes and centrifuged (3,000×g, 8 min, 4°C) and the supernatant removed. Each of the cell pellets was gently resuspended in 5 mL isotope-labelled M9 broth and transferred to 500 mL isotope-labelled M9 broth in a 2.5 L baffled flask. The cultures were shaken at 37°C for ~3.5 h until the OD$_{600}$ reached ~0.45, when they were cooled briefly under running water and then transferred to shaking at 30°C. Protein expression was induced with IPTG (final concentration 0.3 mM) when the cultures' OD$_{600}$ was ~0.9, after which the cultures were grown for a further 14 h. Cultures were then centrifuged (5,000×g, 12 min, 4°C) and the broth removed. The cell pellets were resuspended in cold PBS (Table 4) and pooled to give a final volume of 35 mL and immediately lysed and purified (Sections 2.3.3, 2.3.4).

2.3.2 Overexpression of non-labelled GST-SfbI-5 or GST-TickFnBR

A single *E. coli* BL21 Gold colony, which had been transformed with the appropriate pGEX construct, was used to inoculate LB-Amp broth (Table 4), which was incubated for ~18 h with shaking at 37°C. This culture and all others in this section were shaken at ~200 rpm. This culture was then inoculated into 500 mL LB-Amp broth in a 2 L baffled flask at a ratio of 1:100 (v/v), and incubated with shaking at 37°C. Protein overexpression was induced at either 30°C or 37°C. For induction at 30°C, cultures were grown at 37°C until they reached an OD$_{600}$ of ~0.5 (Section 2.3.6), when they were cooled briefly under running water and then transferred to shaking at 30°C. At both temperatures, cultures were induced with IPTG (final concentration 0.3 mM) when the OD$_{600}$ of the culture was between 0.8 and 1.2. Cultures induced at 30°C and 37°C were grown for a further 16–18 h or 3–4 h after induction, respectively. Cultures were then centrifuged (5,000×g, 12 min, 4°C), and the supernatant removed. Cell pellets were resuspended in cold PBS (Table 4) and either lysed (Section 2.3.3) or transferred to storage tubes, and centrifuged a second time (5,000×g, 10 min, 4°C) to remove the PBS. Cell pellets were stored at either −20°C or −80°C for short and long term storage, respectively.

2.3.3 *E. coli* cell lysis

Pellets of cells expressing GST-SfbI-5, a GST-SfbI-5 mutant or GST-TickFnBR were thawed, if necessary, and then resuspended in cold PBS to give a theoretical OD$_{600}$ of ~100, which was estimated from the final OD$_{600}$ reached by the culture directly before harvesting the cells by centrifugation. The cell suspension was lysed using a French

Press (1,500 psi). To improve protein solubility, Triton X-100 (final concentration 1% v/v) was added to the lysed cells which were then mixed thoroughly (Sambrook & Russell, 2001) and centrifuged (48,000×g, 30 min, 4°C). The supernatant containing the soluble proteins was decanted from the insoluble cellular protein, and then the supernatant was immediately processed by affinity chromatography.

2.3.4 Affinity chromatography

2.3.4.1 Affinity chromatography of ^{15}N-labelled GST-SfbI-5, ^{15}N, ^{13}C-labelled GST-SfbI-5 and GST-TickFnBR

[15]N-labelled GST-SfbI-5, ^{15}N, ^{13}C-labelled GST-SfbI-5 and GST-TickFnBR were all purified at 4°C on a series of 5 mL GSTrap HP columns (GE Lifesciences) connected in a tandem array, attached to an ÄKTApurifier 100 chromatography system (GE Lifesciences). After pre-equilibration of the column with PBS, soluble cellular protein was loaded onto the column using a different but related method for each GST-fusion. Soluble cellular protein from 1 L of culture containing either ^{15}N-labelled GST-SfbI-5 or ^{15}N, ^{13}C-labelled GST-SfbI-5 was loaded at a flow rate of 1 mL/min onto two GSTrap columns in tandem. The ^{15}N, ^{13}C-labelled GST-SfbI-5 sample was then reloaded onto the column by looping the flow-through, which contained soluble *E. coli* proteins and any unbound GST-fusion, back into the sample vessel until the entire sample had passed over the column approximately three times. Soluble cellular protein from a 2 L culture containing GST-TickFnBR was loaded onto five GSTrap columns in tandem at a flow rate of 0.3 mL/min. For all GST-fusions, after the loading step the column was washed with PBS (10 column volumes (CV), 2–3 mL/min) and then the GST-fusion was eluted with freshly prepared glutathione elution buffer (5 CV, 2–3 mL/min; Table 4). However, the glutathione elution buffer used to elute either ^{15}N-labelled or ^{15}N, ^{13}C-labelled GST-SfbI-5 did not contain 150 mM NaCl. During elution 3 mL fractions were automatically collected. Fractions containing GST-fusion were identified by monitoring the absorbance at 280 nm, and these fractions were pooled. The pooled eluates and the flow-through were stored at −20°C.

2.3.4.2 Affinity chromatography of GST-SfbI-5 and GST-SfbI-5 mutants

GST-SfbI-5 and its mutants were purified in groups of five or six at room temperature using a combination of batch and column affinity chromatography. Glutathione Sepharose 4 Fastflow agarose beads (GE Lifesciences) were pre-equilibrated in PBS

and soluble cellular protein containing the GST-fusion was added to the beads at a ratio of ~10 mL of soluble cellular protein per 1 mL beads. The bead-protein slurry was incubated at 4°C for at least 2 h with gentle rocking on a gel shaker and then centrifuged ($500 \times g$, 5 min, 4°C) to pellet the beads. Most of the unbound soluble cellular protein was then carefully removed so as not to disturb the beads, this was the 'flow-through'. The beads were then washed by adding ~30 mL of chilled PBS and gently inverting, and were then centrifuged ($500 \times g$, 5 min, 4°C) and most of the supernatant was carefully removed. The wash step followed by centrifugation was repeated a further three times, once with chilled PBS and twice with chilled TBS (Table 4). The beads were then resuspended in TBS to give a slurry that was ~50% (v/v) beads, which was then transferred to a Poly-Prep gravity flow column (Bio-Rad Laboratories). A further three 5 mL washes with TBS were used to both transfer any remaining beads to the column and to help settle the beads in the column. The GST-fusion was eluted from the column under gravity flow with glutathione elution buffer (Table 4) in five 0.5 CV fractions. Those fractions containing GST-fusion were identified by measuring their absorbance at 280 nm relative to glutathione elution buffer (Section 2.3.6) and these samples were pooled and stored along with the flow-through at −20°C.

2.3.4.3   Further affinity chromatography steps

Samples from the overexpression and purification of GST-fusions were examined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE; Section 2.3.5). Typical gel samples are summarized in Table 9. When gel analysis indicated there was a significant amount of GST-fusion that had not bound to the affinity column and remained in the flow-through, the flow-through was repurified with either one or two additional affinity chromatography steps. The same protocol was used for these additional affinity chromatography steps as for the initial affinity chromatography of the GST-fusion from the soluble cellular protein (Sections 2.3.4.1, 2.3.4.2).

2.3.4.4   Affinity chromatography to remove GST from cleaved GST-TickFnBR

After cleavage of GST-TickFnBR by the protease Factor Xa (Section 2.3.7), the cleaved sample was purified by GST affinity chromatography to remove the GST and any uncleaved GST-fusion. The same protocol was used to remove GST from TickFnBR as for the initial GST-TickFnBR affinity chromatography, with the exception that the flow-through was collected in fractions. Fractions that contained TickFnBR were

identified by monitoring absorbance at 220 nm, and these fractions were pooled and stored at +4°C until purification, which was always performed within a few days.

Table 9. Abbreviations and descriptions of samples that were collected during protein overexpression and purification and that were then analysed by SDS PAGE.

Sample	Abbreviation	Description
Pre-induction	−I	Cell sample was taken just before induction with IPTG. Cells were pelleted and then resuspended in PBS to give a theoretical OD_{600} of 10, and then diluted 1:1 (v/v) with 3× sample buffer stock solution (Table 10). Samples were vortexed briefly before heating and again before loading the gel. The increased concentration of sample buffer and vortexing were necessary for breaking up DNA and reducing sample viscosity.
Post-induction	+I	Cell sample was taken after induction, directly before harvesting cells by centrifugation. Sample was prepared using the same procedure as for −I.
Total cellular protein	T	Sample was taken after lysis, but before centrifugation, and contained all *E. coli* cellular protein.
Soluble cellular protein	S	Sample was taken after lysis and centrifugation, just before affinity chromatography. It contained all soluble *E. coli* cellular protein.
Insoluble cellular protein	I	Sample was taken at the same time as T, but was then centrifuged and the supernatant removed, before being resuspended in the same volume of PBS as the original sample. Thus, this sample contained all insoluble protein and could be compared directly to the T and S samples, without needing to resuspend all the insoluble protein from the culture.
Flow-through	F	Sample was taken of the flow-through from the loading step during affinity chromatography, and contained all unbound GST-fusion and soluble cellular protein. When the flow-through was repurified, the flow-through from this second affinity chromatography step was called F2, and so on.
Eluate fractions (individual or pooled)	E	Sample was taken of either the individual eluate fractions or pooled eluate. To avoid overloading the gels, samples were diluted 1 in 10 before adding sample buffer stock solution.

2.3.5 SDS PAGE analysis of protein samples

The Novex protein gel electrophoresis system (Invitrogen) was used for all SDS PAGE. Specifically, the gels used were NuPage Novex Bis-Tris pre-cast gels (8 cm × 8 cm × 1 mm, 4% stacking gel and either 10% or 4–12% gradient separating gel). A working sample buffer stock solution (3×) was made up from 100 µL 1 M DTT and 250 µL LDS (lithium dodecyl sulfate) sample buffer stock solution (Table 10). Samples were heated at 70°C for 10 minutes to help denature the sample and reduce disulfide bonds. Typically 10 µL of sample was loaded on to the gel per lane, and where possible samples were made up such that no more than 15 µL and no less than 5 µL of sample was needed to get clear bands with good resolution. If samples were thought to be too concentrated, they were diluted first with 1× sample buffer.

Table 10. Composition of buffers used in SDS PAGE. Sample and running buffers were dilutions of stock solutions (Invitrogen), except for DTT which was purchased as a solid powder (Melford). Abbreviations: EDTA, ethylenediaminetetraacetic acid; MES, 2-(*N*-morpholino) ethanesulfonic acid; MOPS, 3-(*N*-morpholino) propanesulfonic acid.

Buffer	Composition
Sample buffer (1x)	95 mM DTT, 134 mM Tris, 1.9% LDS, 9.5% glycerol, 0.46 mM EDTA, 0.21 mM SERVA blue G250, 0.17 mM Phenol Red, pH 8.5
MES SDS Running Buffer	50 mM MES, 50 mM Tris, 0.1% SDS, 1 mM EDTA, pH 7.3
MOPS SDS Running Buffer	50 mM MOPS, 50 mM Tris, 0.1% SDS, 1 mM EDTA, pH 7.7
Coomassie staining solution	10% (v/v) acetic acid, 50% (v/v) methanol, 0.2% (w/v) Coomassie blue
Destaining solution	10% acetic acid, 10% ethanol

The gels were run in either MES or MOPS SDS Running Buffer (Invitrogen; Table 10), depending on the MW of the protein(s) of interest: MOPS is better for medium to large proteins, MES is better for smaller proteins. Gels were run in an XCell SureLock Mini-Cell gel tank (Invitrogen) at a constant voltage (200 V) with a Basic PowerPac (Bio-rad) until the dye front was near the bottom of the gel (typically 40 min). Primarily, a 4–12% gradient gel was run with MES SDS Running Buffer unless specifically stated otherwise.

When the run was finished, each gel was rinsed in HPW, covered in Coomassie staining solution (Table 10), and stained with gentle rocking until the gel was dark purple (10–

60 min). The stain was removed and the gel was rinsed a few times with HPW and then covered in destaining solution (Table 10) and destained with gentle rocking until the background was reduced enough that faint bands could be detected (3–18 h). Finally the gels were photographed with a GeneGenius bioimaging system (Syngene) using GeneSnap software (Syngene).

2.3.6 Ultraviolet-visible absorption spectroscopy

All UV-visible data were recorded on a Biophotometer spectrophotometer (Eppendorf) using either a 0.2 or 1 cm path length cuvette against an equivalent buffer-only blank. Samples were diluted and/or the pathlength was changed so that absorbance readings were less than 2 and ideally between 0.1 and 1. Culture densities were measured at 600 nm using plastic cuvettes. Protein samples were measured at 280 nm and 320 nm simultaneously using plastic Uvettes (Eppendorf), which are transparent at these wavelengths and have a 0.2 cm pathlength in one direction and a 1 cm pathlength in the other. The absorbance at 280 minus the absorbance at 320 nm ($A_{280-320}$) is calculated by the Biophotometer. Protein concentrations were estimated from $A_{280-320}$ using the Beer-Lambert Law (Equation 4) and the molar extinction coefficients were predicted from primary sequence data by the web tool ProtParam (www.expasy.ch/tools/protparam.html; Equation 5; Table 11). The $A_{280-320}$ was usually used instead of the absorbance at 280 nm (A_{280}) to estimate protein concentration, because any absorbance at 320 nm is not likely to be caused by soluble proteins, but rather is a measure of sample turbidity.

$$A = \varepsilon c l$$

Equation 4. The Beer-Lambert Law. Annotation: A, absorbance; ε, molar extinction coefficient; c, concentration; l, pathlength.

$$\varepsilon = 1490 \times N_{Tyr} + 5500 \times N_{Trp} + 125 \times N_{Cystine}$$

Equation 5. Predicted molar extinction co-efficient of a protein in water based on its primary sequence. Annotation: ε, molar extinction coefficient; N_{Tyr}, N_{Trp} and $N_{Cystine}$, number of tyrosines, tryptophans and cystines in the protein of interest, respectively (Gasteiger *et al.*, 2005).

Table 11. Molar extinction coefficients calculated for Fn fragments and GST-FnBR fusions. None of the FnBRs studied in this work had any tryptophans, tyrosines or cystines, so they do not absorb at 280 nm. Therefore, the molar extinction coefficient for all GST-FnBR fusions was that predicted for GST. All cysteines were assumed to be cystines in Fn fragments, none were predicted to be cystines for GST.

Protein	ε (M^{-1} cm^{-1})
pNTD or $^{1-5}$F1	63130
^{2}F1^{3}F1	24450
^{4}F1^{5}F1	26970
GST or GST-FnBR fusion	42860

2.3.7 Factor Xa cleavage of GST constructs

GST was cleaved from SfbI-5, SfbI-5 mutants and TickFnBR with the protease Factor Xa, which specifically recognises the sequence Ile-Glu-Gly-Arg, cleaving C-terminal to the arginine residue (Figure 13B). First, the pooled GST-fusion samples were exchanged from glutathione elution buffer into TBS (Table 4) by either using Vivaspin 20 centrifugal concentrators (10 kDa nominal MW cut off, MWCO; Sartorius) or by dialysis with Spectra/Por dialysis membrane (6–8 kDa MWCO, Fisher Scientific). Then, $CaCl_2$ was added to a final concentration of 2 mM. Second, when cleaving a GST-fusion for the first time, a trial cleavage was performed to optimize Factor Xa concentration and cleavage time at room temperature. Three samples of purified GST-fusion (each 50 µL) were cleaved at three different protease concentrations. The cleavage was monitored over time by taking samples at regular intervals and stopping the protease by adding SDS-PAGE sample buffer stock solution (Table 10). These samples were analysed by SDS-PAGE to ascertain optimal cleavage time (Section 2.3.5). Third, the remaining GST-fusion sample was cleaved with Factor Xa at the optimal concentration and time at room temperature.

2.3.8 Reversed-phase high performance liquid chromatography

Reversed-phase high performance liquid chromatography under acidic conditions (acidic rpHPLC) was performed to repurify pNTD and also on cleaved samples containing SfbI-5 or its mutants, GST and any intact GST-fusion. Samples of cleaved TickFnBR were first repurified by affinity chromatography to remove GST and any

intact GST-fusion (Section 2.3.4.4), and then purified by rpHPLC under basic conditions (basic rpHPLC).

All rpHPLC was performed on a Jupiter 250×10 mm column (Phenomenex) connected to an ÄKTApurifier 10 system (GE Lifesciences) at a flow rate of 3 mL/min. pNTD, SfbI-5 or its mutants were purified on a C4[9] column and TickFnBR was purified on a C12 column. Solvents for acidic rpHPLC were HPW and acetonitrile, with 0.15% and 0.1% (v/v) trifluoroacetic acid (TFA), respectively. For basic rpHPLC, solvents were acetonitrile and a solution of ammonium carbonate (10 mM in HPW) that was filtered through a 0.22 µM filter before use. All solvents were degassed by an inline degasser (Cambridge Scientific Instruments).

Before performing a series of rpHPLC purification 'runs', a Pre-run method was performed to clean the column of any proteins from previous runs that had eluted during storage (storage solution 65% acetonitrile, 35% HPW). This was followed by a blank run of the method to be used in the subsequent purification to check that the baseline was reasonable. Then injected samples were eluted using a gradient of 10–90% acetonitrile, with a shallow gradient over either 2.5 or 2.0 CV for acidic and basic rpHPLC, respectively. The shallow gradient was optimized for different purifications, but typical values were 33–40% acetonitrile for SfbI-5, 22−42% acetonitrile for pNTD and 15–20% acetonitrile for TickFnBR. Eluate was monitored using the absorbance at 280 nm and 220 nm. Fractions (3 mL) were collected automatically, pooled where appropriate, and lyophilized. Pooled fractions from basic rpHPLC were resuspended in HPW and re-lyophilized at least two more times to help remove the trace amounts of ammonium carbonate.

2.3.9 Yield determination of lyophilized recombinant FnBRs

An appropriate number of 1.5 mL tubes were weighed on a four decimal-place balance at least three times each. FnBRs purified by rpHPLC and then lyophilized were

---

[9]The name 'C4' describes the number of carbons in the hydrocarbon ligand attached to the solid support of the rpHPLC column. That is, a C4 column has a butyl hydrocarbon ligand: $-CH_2-CH_2-CH_2-CH_3$. Similarly, a C12 column has a dodecyl hydrocarbon ligand. The longer the hydrocarbon, the more hydrophobic it is, so the better it is at purifying more hydrophilic substances. Therefore, C4 columns are usually used for purifying proteins and C12 columns are used for purifying peptides.

resuspended in HPW (650 µL) and each transferred to a pre-weighed 1.5 mL tube. The initial tube was then washed with HPW (650 µL), which was then transferred to the same pre-weighed tube. An aliquot of this solution (50–100 µL) was taken and lyophilized separately, then analysed by electrospray mass spectrometry (Section 2.3.10). The remaining solution was re-lyophilized and then the tube, which now contained lyophilized FnBR, was re-weighed on the same balance, in triplicate. Reproducibility between subsequent measurements of the same tube was improved by performing all weighing with the tube on a small cork stand, which served to minimize electrostatic effects from the plastic tube and lyophilized FnBR powder. The mass of FnBR peptide in each tube was calculated by taking the average of the initial weights for each tube and subtracting the average of the weights of each tube with lyophilized FnBR. This mass was then corrected assuming that 20% of the FnBR mass was water. This correction was based on past experience in our laboratory of comparison between the molar amounts of FnBRs when calculated by weight and when determined by quantitative amino acid analysis.

2.3.10 Electrospray mass spectroscopy (ESMS)

All ESMS was carried out by Ms Berni Strongitharm (Molecular Interactions Laboratory, Technology Facility, Department of Biology, University of York) on an ABI Qstar tandem mass spectrometer.

2.4   Nuclear magnetic resonance spectroscopy

2.4.1   Sample preparation

Lyophilized protein or peptide was resuspended in 400–600 µL of 10% D_2O in HPW, to a final concentration of ~100 µM, except the ^{15}N, ^{13}C-labelled SfbI-5 sample which was prepared at a concentration of ~1 mM. The concentrations of F1-module pairs or pNTD were determined by $A_{280-320}$ and the concentrations of FnBR peptides were determined by the weight of the lyophilized peptide. All samples were within a pH range of 5–7 and were slowly adjusted to the desired pH with small aliquots of 0.1 M NaOH or HCl. All samples kept long-term also contained 0.03% NaN_3 to inhibit growth of microbes. NaN_3 was added prior to adjusting the pH.

For chemical shift perturbation assays, aliquots of unlabelled peptide were prepared from a stock solution made up in HPW at the same pH as the corresponding NMR sample. These aliquots were lyophilized, and then resuspended with the NMR sample during the chemical shift perturbation assay. The method minimized the dilution of the NMR sample throughout the assay.

2.4.2 Experiments

All experiments were acquired using standard Bruker pulse sequences on a Bruker AVII 700 MHz spectrometer, equipped with a triple resonance (HCN) probe and three-axis pulse-field gradients. The experiments were kindly performed by either Dr Jennifer Potts (Departments of Biology and Chemistry, University of York) or Mr Andy Brentnall (Potts group, Department of Biology, University of York). One-dimensional ^{1}H spectra were acquired with a spectral window of 10,000–12,500 Hz, 64–256 scans and 8192 complex points. Parameters for the two and three-dimensional experiments are summarized in Table 12.

Table 12. The pulse programs and parameters for the two- and three-dimensional NMR experiments acquired for this work. All experiments were performed with ^{15}N, ^{13}C-labelled SfbI-5, except for HSQC 2 and HSQC 3, which are the parameters for chemical shift perturbation assays with either TickFnBR titrated into ^{15}N-labelled ^{4}F1^{5}F1 or IxTT1 titrated into ^{15}N-labelled ^{2}F1^{3}F1, respectively. All experiments were performed at 298 K.

Experiment	Pulse program	# of scans	# of complex points			Spectral window (Hz)		
			^{1}H	^{15}N	^{13}C	^{1}H	^{15}N	^{13}C
HSQC	hsqcetf3gpsi	16	2048	512	–	12500	1845	–
HNCO	hncogp3d	16	1024	64	128	12500	1845	3522
HN(CA)CO	hncacogp3d	16	2048	40	128	12500	1845	3522
CBCA(CO)NH	cbcaconhgp3d	16	2048	48	128	10000	1845	13324
CBCANH	cbcanhgp3d	16	2048	48	128	10000	1845	13224
HSQC 2	hsqcetf3gpsi	8	2048	256	–	12500	2242	–
HSQC 3	hsqcetf3gpsi	8	2048	256	–	12500	2242	–

2.4.3 Data processing

Experimental data was processed with NMRPipe (Delaglio *et al.*, 1995) using scripts written by Mr Andy Brentnall. In each dimension the resolution and signal to noise ratio

were enhanced prior to Fourier transformation by zero-filling in all dimensions, application of a Gaussian window function in the ^{1}H dimension, and application of both linear prediction and a squared sine-bell window function in both indirect dimensions (^{15}N and ^{13}C). Additionally, a solvent filter which helps to remove the water signal from the spectrum was applied to the three-dimensional spectra.

The spectra were referenced to DSS directly in the ^{1}H dimension and indirectly in the ^{15}N and ^{13}C dimensions. Indirect referencing was achieved using Equation 6 with a frequency ratio with respect to DSS of 0.10132912 and 0.25144953 for the ^{15}N and ^{13}C dimensions, respectively (Harris *et al.*, 2008). Small changes were made to the referencing of the three-dimensional triple-resonance spectra of SfbI-5 so that the differences between both the ^{1}H^N and ^{15}N^H chemical shifts in these spectra were minimized with respect to the corresponding chemical shifts in the HSQC spectrum of SfbI-5.

$$\upsilon_X = \Xi_X \times \upsilon_H$$

Equation 6. Indirect referencing of a frequency for nuclei other than hydrogen. The reference frequency for nucleus X (υ_X), that is, 0 ppm in the X-dimension of a spectrum, can be determined indirectly from the reference frequency of DSS in a one-dimensional ^{1}H spectrum (υ_H) and the frequency ratio of nucleus X with respect to DSS (Ξ_X).

2.4.4 Resonance assignment of ^{15}N, ^{13}C-labelled SfbI-5

All spectra were both visualized and analysed with the CCPN analysis software (Vranken *et al.*, 2005). Sequence-specific resonance assignment of SfbI-5 ^{1}H^N, ^{15}N^H, ^{13}C$^\alpha$, ^{13}C$^\beta$ and ^{13}C$'$ nuclei was achieved using a procedure reviewed by Cavanagh *et al.* (1995) and Kanelis *et al.* (2001). This assignment process is described in more detail in Section 3.3.2.2.

2.4.5 Secondary chemical shift analysis of ^{15}N, ^{13}C-labelled SfbI-5

2.4.5.1 Chemical shifts of ^{15}N, ^{13}C-labelled SfbI-5

The chemical shifts of SfbI-5 used in the secondary chemical shift analysis (Section 2.4.5.2) were measured from the acquired spectra of ^{15}N, ^{13}C-labelled SfbI-5 (Table 12) using the following rationalization. The ^{1}H^N and ^{15}N^H chemical shifts were

those measured from the HSQC spectrum because out of the spectra recorded this spectrum has the highest resolution in these dimensions. In contrast, the chemical shift used for each carbon nucleus was the average of all chemical shifts measured for that nucleus in the three-dimensional triple-resonance spectra. This approach was used for the following reasons. Both the HN(CA)CO and HNCO experiments detected the chemical shifts of the SfbI-5 ^{13}C′ nuclei. However, only the HN(CA)CO experiment detected all the ^{13}C′ chemical shifts, because the HNCO experiment did not detect the ^{13}C′ nuclei of residues that directly preceded proline or the ^{13}C′ nucleus of the C-terminal residue. However, the HN(CA)CO spectrum had a lower resolution, a lower signal to noise ratio and was more overlapped than the HNCO spectrum. Therefore, using the average chemical shift for each ^{13}C′ nuclei was deemed the best compromise. The same rationale was applied to the ^{13}C$^{\alpha}$ and ^{13}C$^{\beta}$ chemical shifts obtained from the CBCA(CO)NH and CBCANH experiments.

2.4.5.2   Calculation of secondary chemical shifts

Secondary chemical shifts of the SfbI-5 ^{1}H^{N}, ^{15}N^{H}, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$ and ^{13}C′ nuclei were calculated as follows. The random coil dataset and corresponding sequence correction values were used to calculate the sequence-corrected random coil chemical shift values for the SfbI-5 nuclei according to Equation 7. Next, for each nucleus the sequence-corrected random coil chemical shifts were subtracted from the corresponding experimentally determined chemical shift to give the Δδ value. This procedure was performed with two random coil datasets and corresponding sequence correction factors yielding two datasets of Δδ values. The random coil datasets used were an experimentally-derived dataset from Wishart *et al*. (1995) and a statistically-derived dataset from Wang and Jardetzky (2002b), which will be referred to as the WS and WJ2 datasets, respectively. The WJ2 dataset contains random coil values and sequence correction factors for the ^{1}H^{N}, ^{15}N^{H}, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$ and ^{13}C′ nuclei of each residue. The WS dataset contains random coil values for the ^{1}H^{N}, ^{15}N^{H}, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$ and ^{13}C′ nuclei of each residue and sequence correction factors for the ^{1}H^{N}, ^{15}N^{H}, ^{13}C$^{\alpha}$ and ^{13}C′ nuclei in the preceding residue (*i.e.,* CX$_{i-1}$ in Equation 7). The WS dataset also contains random coil values the ^{1}H^{N}, ^{15}N^{H}, ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$ and ^{13}C′ nuclei for each residue when followed by a proline, which were used in the calculation of sequence-corrected random coil chemical shifts where appropriate. All Δδ calculations were performed in Excel (Microsoft) with the aid of a macro kindly written by Mr Stephen Austin (University of York).

$$\delta X_{i \text{ corrected}} = \delta X_{i \text{ coil}} + BX_{i+1} + CX_{i-1}$$

Equation 7. The sequence-corrected random coil chemical shift for a nucleus of type X in residue *i*. Annotation: $\delta X_{i \text{ corrected}}$, the sequence-corrected random coil chemical shift for a nucleus of type X in residue *i*; $\delta X_{i \text{ coil}}$, the random coil chemical shift for a nucleus of type X in residue *i*; BX_{i+1} and CX_{i-1}, sequence correction factor for a nucleus of type X in the residue C-terminal or N-terminal to residue *i*, respectively.

2.4.5.3   Analysis of secondary chemical shifts

The $\Delta\delta$ values were used to calculate both the CSI and also the SSP score for the nuclei of residues in SfbI-5. The CSI of each assigned nucleus was calculated from WS-derived $\Delta\delta$ values according to the procedure proposed by Wishart and colleagues (1992; 1994b) and summarized in Section 1.6.2.2. The threshold values within which $\Delta\delta$ values were considered to be a random coil were ± 0.2 ppm for $^{1}H^{N}$ $\Delta\delta$ values and ± 0.5 ppm for $^{15}N^{H}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ $\Delta\delta$ values. These threshold values are those used in CSI calculations implemented in NMRView for the WS dataset (Johnson & Blevins, 1994; Schwarzinger *et al.*, 2000; Schwarzinger *et al.*, 2001). The consensus CSI for each residue was then calculated by a simple 'majority rules' approach for all the assigned nuclei in the residue (Wishart & Sykes, 1994a).

The SSP score for SfbI-5 residues was calculated using the SSP Perl script kindly made freely available by Marsh *et al.* (2006). This script calculates the SSP score according to Equation 3 (p 48) from an input random coil shift dataset and experimental chemical shifts. However, the script does not perform sequence-correction to the random coil shifts. Therefore, the experimentally-determined SfbI-5 chemical shifts input into the SSP were first adjusted by adding the corresponding sequence correction values (BX_{i+1} and CX_{i-1} in Equation 7), and then the SSP scores were calculated from the WJ2 dataset and the adjusted SfbI-5 chemical shifts.

2.5 Size exclusion chromatography

2.5.1 SEC for apparent MW determination

SEC for determining the apparent MW of proteins was performed on a Superdex 200 10/300 GL column (GE Lifesciences) attached to an AKTApurifier 10 system (GE Lifesciences), which was fitted with a 500 µL sample loop. Filtered and degassed PBS (Table 4) was used to pre-equilibrate the column and for all subsequent experiments. Injected samples (250 µL) were eluted at a flow rate of 3 mL/min for 2 CV, while monitoring absorbance at 215 and 280 nm.

2.5.2 Apparent MW and stokes radius of pNTD, SfbI-5 and pNTD bound to SfbI-5

RpHPLC purified and lyophilized pNTD and SfbI-5 were made up in PBS to concentrations of 10 µM and 150 µM, respectively. The concentration of pNTD was calculated using A_{280} (Section 2.3.6) and SfbI-5 concentration was calculated by using the weight of the lyophilized sample (Section 2.3.9). Samples of SfbI-5, pNTD, SfbI-5 with pNTD, and globular protein standards were made up in PBS (Table 4) and analysed by SEC (Section 2.5.1). Linear regression analysis in Excel (Microsoft) was performed on the relationship between the logarithm of the MW and retention for the globular protein standards. The resulting line-of-best fit (Equation 8) was used to calculate the apparent MWs of the species in the other samples.

Table 13. Name and composition of samples used to determine apparent MW of pNTD, SfbI-5 and pNTD bound to SfbI-5.

Sample name	Sample composition in PBS
pNTD only	6 µM pNTD
SfbI-5 only	10 µM SfbI-5
pNTD and SfbI-5 (1:0.8)	6 µM pNTD, 5 µM SfbI-5
pNTD and SfbI-5 (1:1.6)	6 µM pNTD, 10 µM SfbI-5
Standards 1	500 µg ribonuclease, 500 µg ovalbumin, 500 µg aldolase, 100 µg ferritin
Standards 2	500 µg chymotrypsinogen A, 500 µg albumin, 500 µg catalase

$$Log(MW) = -0.248(R) + 8.65$$

Equation 8. The linear correlation between the logarithm of the MW (Da) of globular protein standards and their retention in SEC. Where R is the retention volume (mL) of globular proteins standards when separated by SEC on a Superdex 200 column in PBS. Linear regression was performed using the Log(MW) as the independent variable and R as the dependent variable. The resulting line-of-best fit was then rearranged to make Log(MW) the dependent variable.

Uversky (1993) determined the relationship between the MW of proteins and their Stokes radii for both folded proteins and proteins denatured in 6M guanidine hydrochloride using data derived from the literature (Equations 9 and 10, respectively). Equation 9 was used to estimate the experimental Stokes radii of pNTD, SfbI-5 or pNTD:SfbI-5 from their apparent MW. This was possible because the apparent MW of proteins determined by SEC is related to their Stokes radii. By the same rationale, the theoretical MWs of pNTD, SfbI-5 and a 1:1 complex of pNTD with SfbI-5 and Equations 9 and 10 were used to estimate the theoretical Stokes radii of these species if they behaved like globular or denatured proteins, respectively. A similar analysis of an IDP has been performed by Longhi *et al.* (2003).

$$Log(R_S) = 0.369 Log(MW) - 0.254$$

Equation 9. The relationship between the MW and Stokes radii (R_S) of folded globular proteins (Uversky, 1993).

$$Log(R_S) = 0.533 Log(MW) - 0.682$$

Equation 10. The relationship between the MW and the Stokes radii (R_S) of denatured globular proteins (Uversky, 1993).

2.5.3   Size exclusion chromatography combined with multi-angle laser light scattering

Size exclusion chromatography (SEC) combined with multi-angle-laser light scattering (MALLS) was performed on either a Superdex 200 or a Superdex 75 10/300 GL column (GE Lifesciences) that was attached to a Shimadzu HPLC system linked to a Dawn HELEOS-II 18-angle light scattering detector (Wyatt) and Optilab rEX refractive index monitor (Wyatt). The column was pre-equilibrated with filtered and degassed PBS at a flow rate of 0.5 mL/min. Injected protein solutions (100 µL) at a concentration

of at least 1 mg/mL in PBS were eluted at a flow rate of 0.5 mL/min for 1.25 CV, while monitoring absorbance at 280 nm, the refractive index and the light scattering at multiple angles. The accompanying ASTRA V software (Wyatt) was used to calculate the MW of the species present in each sample from the refractive index and MALLS data collected during the SEC run.

2.6 HMM building and searching

2.6.1 Starting multiple sequence alignments (MSAs)

Two separate HMM building and searching 'rounds' were performed, one before any Fn:FnBP crystal structures were solved, and one after six crystal structures were solved. The first, relatively 'naïve' round started with a manual MSA of FnBRs based on that of Schwarz-Linek *et al.* (2004a). Added to this MSA were the remaining FnBRs from *S. aureus* FnBPA, also as defined by Schwarz-Linek *et al.* (2003), with the exception of FnBPA-7 as it is predicted not to be functional. Also, the FnBR from the *B. burgdorferi* protein BBK32 was included (Raibaud *et al.*, 2005; UniProt accession O50835, residues 146–190). Sequences with greater than either 70% or 90% pairwise identity to any other sequence in these starting MSAs were removed using the 'Remove Redundancy' feature in the Jalview program (Clamp *et al.*, 2004; Waterhouse *et al.*, 2009), creating two new MSAs called, respectively, nr70 and nr90 (Table 14).

In the second round, the same FnBRs were used as in the first round, except the FnBR from BBK32 was omitted. For the three remaining repeats for which there was structural information (FnBPA-1, FnBPA-5, Bingham *et al.*, 2008; ^{2}F1- and ^{3}F1-binding regions for SfbI-5, Section 4.2), an alignment was made by aligning structurally-equivalent residues. Then the remaining FnBRs were manually aligned to this three-sequence alignment, creating the starting MSA for the second round. The structures were used to inform the position of insertions in this MSA, which led to changes in the MSA compared to the starting MSA in the first round (Table 14). For example, since neither of the two ^{4}F1^{5}F1:FnBP structures had insertions in the linker region between the ^{4}F1- and ^{5}F1-binding regions, this was avoided in the alignment. All of the β-sheet interactions making up the tandem β-zipper were standard anti-parallel β-sheets, with no β-bulges. Therefore, no insertions or deletions were allowed in the F1-binding regions identified from the structures. Sequences which were identical to other sequences in the starting MSA were removed to give a new MSA called 's100' (Table 14).

2.6.2 Iterative HMM building and searching method and FnBR selection criteria

In both rounds, the nr70, nr90 and s100 MSAs (Table 14) were used to start an iterative cycle of building and then searching with HMMs as described in Table 15. Putative FnBRs identified by searching had to meet the following two selection criteria to be included in the MSA for the next cycle. First, if they had been experimentally found to be functional, then they were always included. If an FnBP contained at least one FnBR with a better (smaller) E-value than a functional FnBR, then all FnBRs from that FnBP were included. Second, previously unidentified putative FnBPs had to make biological sense. That is, they had to be from an organism that could potentially use an FnBP because the organism comes into contact with Fn. Second, the putative FnBP had to have access to Fn by being predicted to be either secreted (*e.g.,* by having a predicted signal sequence) or an outer-membrane protein. Effectively, because both FnBPA-2 and FnBPA-6 had poor E-value scores, but both have been shown to bind pNTD (Meenan *et al.*, 2007), this meant that the 'biological sense' selection criterion was more restrictive than the E-value selection criterion.

In both rounds, the iterative HMM building and searching cycle was repeated for a total of three iterations, since in the third iteration no new FnBPs were identified with any of the three HMMs. However, in the second round, after the third iteration, a new cycle was performed using the third iterations' search results as a starting point, but with an E-value cut-off of 1000 instead of the default of 10. The first search in this cycle led to the identification of BBK32 protein (UniProt ID Q9L890). Therefore, the first selection criterion was changed: FnBRs were included using the BBK32 FnBR as the cut-off above which an FnBP needed at least one FnBR with a better E-value to be included. This new cycle was repeated a further two times until no new FnBPs were identified.

Table 14. Starting FnBR MSAs used in the first and second rounds of HMM building and searching. When building an HMM from an MSA, HMMER2 interprets each column of an alignment as either a match state, which may also contain deletions; or an insertion state, which is then linked to the previous match state (Figure 8). Whether a column was modelled as a match or an insertion in the HMM is indicated in the MSAs shown below in two ways. First, match state columns contain capital letters, and gaps are indicated by a hyphen, whereas insertion state columns contain lower case letters and gaps are indicated by a full stop. Second, the bottom line in the MSA, named #=GC RF, shows match states as an x and insertion states as a full stop. These representations of the MSAs as interpreted by HMMER2 were generated when building the HMM.

Name	MSA	
nr70	SfbI-1	`.eVYGNQQNPVDI-.......DKKLPNETGFSGNM-.--VETEDTKEP..`
	SfbI-2	`gvLMGGQSESVEFT.......KdTQTGMSGQTTPQ-.--VETEDTKEP..`
	SfbII-2	`vsTQENKdPIVDIT.......EDTQPGMSGSNDAT-.--VVEEDTTPQrp`
	F2-1	`.eVITQQGPNLEI-.......EETLPLESGASGGT-.--TTVEDSRPV..`
	F2-3	`dvIIGGQGEVVDTT.......EDTQSGMTGHSGST-.--TEIEDSKSS..`
	F2-2	`ykPTKGSGQVIDI-.......EEKLPDEQGHSGST-.--TEIEDSKPS..`
	FnBA-1	`enNLGGQSEEITIT.......EDSQSGMSGQNPGSG.NETVVEDTQTSqe`
	FnBA-2	`diVLGGPGQVIDFT.......EDSQPGMSGNNSH--.--TITEDSKPSqe`
	FnBA-4	`evIIGGQGQVIDFT.......EDTQTGMSGAGQVES.-PTITEETHKP..`
	FnBPA-1	`..GPIIQNNKFEYK.......EDTIKETLTGQYDKNlVTTVEEEYDSS..`
	FnBPA-2	`..-----TLDIDY-.......HTAIDGGGGYVDGY-.-IETIEETDSS..`
	FnBPA-4	`..--------IDF-.......EESTHENSKHHADV-.-VEYEEDTNPG..`
	FnBPA-5	`ggQVTTESNLVEFD.......EESTKGIVTGAVSD-.-HTTVEDTK--..`
	FnBPA-6	`..EYTTESNLIELV.......DE-LPEEHGQAQGP-.---VEEITK--..`
	bbk32-1	`..EP-IESNEIDLTidsdlrpKSSLQGIAGSNSISYtDEIEEEDYDQY..`
	#=GC RF	`..xxxxxxxxxxxx.......xxxxxxxxxxxxxxx.xxxxxxxxxx..`
nr90	SfbI-1	`.eVYGNQQNPVDI-.......DKKLPNETGFSGNM...-VETEDTKEP..`
	SfbI-2	`gvLMGGQSESVEFT.......KdTQTGMSGQTTPQ...-VETEDTKEP..`
	SfbI-5	`gvLMGGQSESVEFT.......KdTQTGMSGFSETV...-TIVEDTRPKlv`
	SfbII-2	`vsTQENKdPIVDIT.......EDTQPGMSGSNDAT...-VVEEDTTPQrp`
	SfbII-3	`dvLVGGQSDPIDIT.......EDTQPGMSGSNDAT...-VVEEDTVPKrp`
	F2-1	`.eVITQQGPNLEI-.......EETLPLESGASGGT...-TTVEDSRPV..`
	F2-3	`dvIIGGQGEVVDTT.......EDTQSGMTGHSGST...-TEIEDSKSS..`
	F2-2	`ykPTKGSGQVIDI-.......EEKLPDEQGHSGST...-TEIEDSKPS..`
	F2-4	`dvIIGGQGQVVETT.......EDTQTGMHGDSGCK...-TEVEDTKLVql`
	FnBA-1	`enNLGGQSEEITIT.......EDSQSGMSGQNPGSg.nETVVEDTQTSqe`
	FnBA-2	`diVLGGPGQVIDFT.......EDSQPGMSGNNSH-...-TITEDSKPSqe`
	FnBA-3	`evIIGGQGQVIDFT.......EDTQSGMSGDNSHTd.gTVLEEDSKPSqe`
	FnBA-4	`evIIGGQGQVIDFT.......EDTQTGMSGAGQVE..sPTITEETHKP..`
	FnBA-5	`eiIMGGQSDPIDMV.......EDTLPGMSGSNEAT...-VVEEDTRPKlq`
	FnBB-1	`.eVIIQQGPILGL-.......EETLPTEEHQSGDT...-TTIEDTRP-..`
	FnBB-2	`dkPITEASQSIDF-.......EETLPTEQGQSGST...-TEVEDTKGP..`
	FnBB-3	`evIIGGQGEIVDI-.......EENLPTEQGQSGST...-TEVEDTKGP..`
	FnZ-3	`.gQTGGQGPVIETT.......EDTQKGMSGQSGGT...-IESENTKKP..`
	FnZ-4	`evMIGGQGQTIETT.......EDTQKGMSGQSGGT...-IESEDTKKP..`
	FnZ-5	`evMIGGQGQIIDFS.......ENTQSGMSGQSGDT...-TVIEDTKKS..`
	FnZ-6	`eiIIGGQGQIIDFS.......EDTQPGMSGQSGGT...-TIVEDTKKP..`
	FnBPA-1	`..GPIIQNNKFEYK.......EDTIKETLTGQYDKNlvTTVEEEYDSS..`
	FnBPA-2	`..-----TLDIDY-.......HTAIDGGGGYVDGY...IETIEETDSS..`
	FnBPA-3	`..-----AIDIDY-.......HTAVDSEAGHVGG-...--YTESSEESnp`
	FnBPA-4	`..--------IDF-.......EESTHENSKHHADV...VEYEEDTNPG..`
	FnBPA-5	`ggQVTTESNLVEFD.......EESTKGIVTGAVSD...HTTVEDTK--..`
	FnBPA-6	`..EYTTESNLIELV.......DE-LPEEHGQAQGP...--VEEITK--..`
	FnBPA-8	`..-------HVDI-.......KSELGYEGGQNSGN...QSFEEDTEEDkp`
	FnBPA-9	`kyEQGGNIVDIDF-.......-DSVPQIHGQNKGN...QSFEEDTEKdkp`
	FnBPA-10	`kyEHGGNIIDIDF-.......-DSVPHIHGFNKHT...EIIEEDTNKdkp`
	FnBPA-11	`syQFGGH-NSVDFE.......EDTLPKVSGQNEGQ...QTIEEDTTP-..`
	bbk32-1	`..EP-IESNEIDLTidsdlrpKSSLQGIAGSNSISytdEIEEEDYDQY..`
	#=GC RF	`..xxxxxxxxxxxx.......xxxxxxxxxxxxxxx...xxxxxxxxxx..`

Name	MSA	
s100	FnBPA-1	..GPIIQNNKFEYKEDTIKETLTGQYDKNLvTTVEEEYDSS..
	FnBPA-5	ggQVTTESNLVEFDEESTKGIVTGAVSD--.HTTVEDTK--..
	SfbI-5	gvLMGGQSESVEFTKdTQ-TGMSGFSET--.VTIVEDTRPKlv
	SfbI-1	.eVYGNQQNPVDIDKKLP--NETGFSGN--.MVETEDTKEP..
	SfbI-2	gvLMGGQSESVEFTKdTQ-TGMSGQTTP--.QVETEDTKEP..
	SfbI-3	gvLMGGQSESVEFTKdTQ-TGMSGQTAS--.QVETEDTKEP..
	SfbII-2	vsTQENKdPIVDITEDTQ-PGMSGSNDA--.TVVEEDTTPQrp
	SfbII-3	dvLVGGQSDPIDITEDTQ-PGMSGSNDA--.TVVEEDTVPKrp
	SfbII-4	diLVGGQSDPIDITEDTQ-PGMSGSNDA--.TVIEEDTKPKrf
	F2-1	.eVITQQGPNLEIEETLP--LESGASGG--.TTTVEDSRPV..
	F2-3	dvIIGGQGEVVDTTEDTQ-SGMTGHSGS--.TTEIEDSKSS..
	F2-2	ykPTKGSGQVIDIEEKLP--DEQGHSGS--.TTEIEDSKPS..
	F2-4	dvIIGGQGQVVETTEDTQ-TGMHGDSGC--.KTEVEDTKLVql
	FnBA-1	enNLGGQSEEITITEDSQ-SGMSGQNPGSGnETVVEDTQTSqe
	FnBA-2	diVLGGPGQVIDFTEDSQ-PGMSGNNS---.HTITEDSKPSqe
	FnBA-3	evIIGGQGQVIDFTEDTQ-SGMSGDNSHTDgTVLEEDSKPSqe
	FnBA-4	evIIGGQGQVIDFTEDTQ-TGMSGAGQVES.PTITEETHKP..
	FnBA-5	eiIMGGQSDPIDMVEDTL-PGMSGSNEA--.TVVEEDTRPKlq
	FnBB-1	.eVIIQQGPILGLEETLP--TEEHQSGD--.TTTIEDTRP-..
	FnBB-2	dkPITEASQSIDFEETLP--TEQGQSGS--.TTEVEDTKGP..
	FnBB-3	evIIGGQGEIVDIEENLP--TEQGQSGS--.TTEVEDTKGP..
	FnBB-4	evIIGGQGEVVDIEESLP--TEQGQSGS--.TTEVEDSKPKls
	FnZ-3	.gQTGGQGPVIETTEDTQ-KGMSGQSGG--.TIESENTKKP..
	FnZ-4	evMIGGQGQTIETTEDTQ-KGMSGQSGG--.TIESEDTKKP..
	FnZ-5	evMIGGQGQIIDFSENTQ-SGMSGQSGD--.TTVIEDTKKS..
	FnZ-6	eiIIGGQGQIIDFSEDTQ-PGMSGQSGG--.TTIVEDTKKP..
	FnBPA-2	..-----TLDIDYHTAID--GGGGYVDGY-.IETIEETDSS..
	FnBPA-3	..-----AIDIDYHTAVD--SEAGHVGGY-.TESSEESNP-..
	FnBPA-4	..--------IDFEESTH-ENSKHHADV--.VEYEEDTNPG..
	FnBPA-6	..EYTTESNLIELVDELP--EEHGQAQGPVeEITK------..
	FnBPA-8	..-------HVDIKSELG--YEGGQNSGN-.QSFEEDTEEDkp
	FnBPA-9	kyEQGGNIVDIDFDSVPQ---IHGQNKGN-.QSFEEDTEKdkp
	FnBPA-10	kyEHGGNIIDIDFDSVPH---IHGFNKHT-.EIIEEDTNKdkp
	FnBPA-11	.sYQFGGHNSVDFEEDTL-PKVSGQNEGQ-.QTIEEDTTP-..
	#=GC RF	..xxxxxxxxxxxxxxxxxxxxxxxxxxxx.xxxxxxxxxx..

Table 15. The steps involved in building an HMM and searching with it. All steps described were performed, in order, on either a Linux PC (first round) or a Windows PC (second round). The programs used were HMMER2 (Eddy, 2003a), Jalview (Clamp *et al.*, 2004; Waterhouse *et al.*, 2009) and a Perl script 'hmmdomains' kindly written by Dr Peter Ashton (Technology Facility, Department of Biology, University of York). The UniRef90 database was downloaded on 27/02/2006 and 10/03/2008 from www.uniprot.org/downloads for the first and second rounds, respectively. The UniRef90 database was also downloaded on 22/01/2009 and searched with the final hmm from nr70 in the first round and s100 in the second. An example of the commands used in each step is shown for the first nr70 MSA (Table 14).

Step	Description	Command for first nr70 MSA
1.	An HMM was built from an MSA, and a file generated showing how HMMER2 interpreted the input MSA.	`hmmbuild -o nr70-1in.txt nr70-1.hmm nr70.fa`
2.	The HMM was calibrated to improve how the E-values are calculated during searching.	`hmmcalibrate nr70-1.hmm`
3.	This HMM was used to locally search the UniRef90 database. The default E-value cut-off of 10 was used except for the second cycle of the second round, for which 1000 was used.	`hmmsearch nr70-1.hmm uniref90.fasta >nr70-1results.txt`
4.	The FnBRs were extracted from the search results using the 'hmmdomains' Perl script. Those FnBRs which did not meet the selection criteria were not included in the next step (see Section 2.6.2 for details)	`hmmdomains <nr70-1results.txt >nr70-1domains.txt`
5.	The FnBR domains were realigned to the HMM so that the sequences would be correctly aligned to each other.	`hmmalign -q nr70-1.hmm nr70-1domains.txt >nr70-1domains.aln`
6.	The MSA was visually inspected and occasionally improved manually, by ensuring related proteins had insertions aligned with respect to each other.	
7.	Similar sequences were removed from the new MSA, using the same pairwise identity cut-off that was used to create the starting MSA (either 70%, 90% or 100% for nr70, nr90 and s100, respectively).	Jalview > Edit > Remove redundancy > select '70' Save as nr70-1-70.fa

2.6.3 Construction of the final HMM used in analysis of conserved FnBR residues

Comparison between the results from the final round of searching with nr70 and s100 HMMs revealed that some sequences were missing from the two UniRef90 databases

used in these searches (Section 4.1.2.3, p 159). Therefore, a search with both the final nr70 and s100 HMMs was performed on the UniRef90 database downloaded on 22/01/2009 using an E-value cut-off of 1000. Identified proteins were considered a 'true positive' and therefore a putative FnBR-containing FnBP if they met the 'biological sense' selection criteria (Section 2.6.2) and had at least one FnBR above an FnBR that was considered highly likely to be a true positive. For the search with the final nr70 HMM from the first round, the FnBR used as the cut-off was thrombostasin (UniProt ID Q95P73), which was identified in the second round as highly likely to be a true positive by both the 'biological sense' criterion and because it had an E-value below BBK32 Q9L890 (Section 4.1.2.2, p 157). For the search with the final s100 HMM the FnBR used as a cut-off was a 'BBK32-like protein' (UniProt ID Q1CNX8) from *Borrelia hermsii*, which was considered highly likely to be a true positive by the 'biological sense' criterion and because the full-length protein was homologous to the known FnBP BBK32. This sequence had also been identified in the second round but had not been included because it had an E-value worse than BBK32 Q9L890. A final HMM, called 'FnBR-final' was built from all non-identical FnBRs identified with this last s100 HMM search. The FnBR-final HMM was then represented as an HMM 'logo' as a means of visualising the conserved residues in FnBRs (www.sanger.ac.uk/Software/analysis/logomat-m/; Schuster-Bockler *et al.*, 2004).

2.6.4 FnBP sequence annotation for family identification

The FnBP sequences identified in the final nr70 search were retrieved using the European Bioinformatics Institute's sequence retrieval service (www.ebi.ac.uk/srs/srsc), and were then searched with two new HMMs, one to identify URs and one to identify [1]F1-binding regions. Finally, these FnBPs were annotated with FnBRs as identified in the final nr70 search, as well as the URs, [1]F1-binding regions, signal sequences predicted by the Signal P server (www.cbs.dtu.dk/services/SignalP; Bendtsen *et al.*, 2004) and Pfam domains (pfam.sanger.ac.uk; Finn *et al.*, 2008) using CLC Protein Workbench (www.clcbio.com).

2.6.5 Calculation of FnBR residue conservation

FnBR residue conservation was calculated from the MSA used to create the FnBR-final HMM (Section 2.6.3). The conservation of each position in the alignment was calculated with the Scorecons server (Valdar, 2002; www.ebi.ac.uk/thornton-

srv/databases/cgi-bin/valdar/scorecons_server.pl) using the default settings and the 'Valdar01' method (Valdar & Thornton, 2001).

2.7 Crystallography

2.7.1 Sample preparation

Solutions of $^2F1^3F1$ with PyTT5 or IxTT1 were prepared from lyophilized protein and peptide resuspended in HPW and then mixed together to give a final concentration of 300–400 µM $^2F1^3F1$ and either ~3.9 mM PyTT5 peptide or 4.7 mM IxTT1. The pH of the unbuffered solution was raised from an initially acidic pH to a pH of ~7 by gradual addition of 0.1 M NaOH. All protein solutions and stock solutions were stored at –20°C.

2.7.2 Crystallization screens

Screening for conditions in which a solution of $^2F1^3F1$ and an FnBR peptide crystallized was achieved using three commercially available crystallization screens. These were the Index screen (Hampton Research; 96 conditions), Crystal screens 1 and 2 (Hampton Research, 96 conditions total), and Clear Strategy Screens I and II (CSS; Molecular Dimensions, 48 conditions total). Each screen was prepared using a 96-well crystal tray (MRC Wilden) in which each well contained 54 µL of crystallization screen solution and the 300 nL sitting drop contained 150 nL protein solution and 150 nL of well solution. The wells in the CSS screen also contained 9 µL of a 1 M Tris buffer (either pH 7 or 7.5). The drops were aliquoted using an automated Mosquito robot (TTP Labtech), and the tray was then sealed with a clear film and stored at 18°C.

Initial conditions in which crystals were obtained were optimized by designing a 96-well screen around the initial condition, in which the pH and either the salt or the precipitant were systematically varied. When setting-up the optimization screen, 1 mL of each well solution in the optimization screen was prepared to minimize pipetting error, and then a crystal tray was set-up using the same method as for the commercial screens.

2.7.3 Cryo-cooling of crystals for data collection

Cryoprotecting solutions were prepared for each crystallization condition that contained crystal(s) of interest; these solutions consisted of the same components at the same

concentration as the well solution, but with an anti-freeze agent added in place of the equivalent volume of water. The anti-freeze agent was either glycerol or polyethylene glycol (PEG) 400. A series of cryoprotecting solutions were prepared with a range of anti-freeze agent concentrations. Next, the minimum amount of glycerol or PEG 400 in the cryoprotecting solution that was sufficient to stop the formation of ice crystals was found. A loop containing a cryoprotecting solution was flash-cooled in liquid nitrogen and then an image of the diffraction pattern was acquired to test for diffraction from ice crystals, which appeared as characteristic rings. Cryoprotecting solutions with a minimum amount of anti-freeze agent that still offered protection from ice crystals were then used during cryo-cooling of the crystals.

Crystals were retrieved with a cryoloop, transferred to a cryoprotectant solution and then flash-cooled in liquid nitrogen using one of two methods. In the first method, cryoprotecting solution (~0.5 µL) was added directly to the drop containing the crystal(s), a crystal was retrieved using a cryoloop and then directly flash-cooled in liquid nitrogen. When this method was used, all crystals in a drop were retrieved and cryo-cooled in quick succession. Alternatively, a crystal was retrieved from the drop with the cryoloop and transferred to a ~1 µL drop of the cryoprotecting solution and then retrieved and flash-cooled in liquid nitrogen. The original drop was resealed. In both methods crystals were soaked in the cryoprotecting solution for 30–60 s before cryo-cooling.

2.7.4 Data collection

All data was collected either 'in-house' at the York Structural Biology Laboratory (YSBL; Department of Chemistry, University of York) or at the European Synchrotron Radiation Facility (ESRF). 'In-house' diffraction data were acquired at one of two stations. Both stations used a MicroMax 007HF rotating anode X-ray generator (Rigaku) and were equipped with Osmic multilayer optics and an Oxford Cryosystems cryostream that maintained the crystal at 120 K. Station 1 was equipped with a mar345 imaging plate detector (MarResearch) and Station 2 was equipped with an R-axis IV++ imaging plate detector (Rigaku).

All cryo-cooled crystals were 'test-shot' by acquiring two images of the diffraction pattern using in-house Station 1 at φ angles of 0° and 90°. Typically, an image was acquired for an oscillation angle of 0.5°, an exposure time of ~120 s and a detector

distance of between 175 mm and 300 mm. The two images for each crystal were then visually inspected, indexed and integrated in MOSFLM (Leslie, 2006) to determine the cell dimensions, space group, mosaicity and the resolution that gave a mean signal-to-noise ratio, $I/\sigma(I)$, of ~2. The mosaicity is a measure of the spot overlap in the diffraction pattern. High mosaicity can result in spot intensities that cannot be accurately measured. The mean $I/\sigma(I)$ is the ratio between the mean spot intensity, I, and the mean standard deviation of the measured intensities, $\sigma(I)$, within a given resolution range. The cell dimension, space group, mosaicity and mean $I/\sigma(I)$ can be used to calculate the φ angle, detector distance, oscillation angle and number of images required to obtain a complete dataset with the minimum number of images (Blow, 2002).

In most cases, the four crystals of each complex with the highest quality diffraction pattern, as judged by a combination of resolution and mosaicity, were sent to the ESRF. There, the crystals were again test-shot and a dataset was collected for the crystal with the highest quality diffraction pattern. The parameters used for the collection of datasets, which were subsequently used to solve three ^{2}F1^{3}F1:PyTT5 structures (Structures I, II and III), are summarized in Table 32 (p 170). The parameters used for collection of the ^{2}F1^{3}F1:IxTT1 dataset are summarized in Table 41 (p 252).

Most in-house data collection was performed with the generous help and advice of Dr Richard Bingham (formerly of our laboratory, now at School of Applied Sciences, University of Huddersfield). Additionally, data collection from the ^{2}F1^{3}F1:IxTT1 crystal on in-house Station 2 was kindly performed by Mr Sam Hart (YSBL, Department of Chemistry, University of York). Data for ^{2}F1^{3}F1:PyTT5 Structures I and III were collected on beamlines ID14-4 and ID23-1, respectively, at 100 K. Data collection from the ESRF was kindly performed by both Dr Johan Turkenburg and Mr Sam Hart (YSBL, Department of Chemistry, University of York).

2.7.5 Data processing

Data was indexed and integrated in MOSFLM (Leslie, 2006) using the procedure detailed in MOSFLM's documentation. Briefly, data was auto-indexed in MOSFLM using two images separated by ~90° and the space group was chosen based on the penalty score output by MOSFLM. Next, the mosaicity of the two images used in indexing was determined, and the highest mosaicity was input into the program before the cell parameters were refined using at least two sets of five images separated by

~90°. Next, the images were integrated using profile fitting. Finally, the integrated data was sorted, scaled and converted to structure factor amplitudes using SCALA (Evans, 2006) from the CCP4 program suite (CCP4, 1994; Potterton *et al.*, 2003). SCALA also randomly selected a 5% sample of reflections that were subsequently used to calculate the free R factor, R_{free}, during refinement (Section 2.7.7).

2.7.6 Molecular replacement

Molecular replacement with F1 molecules from known ^{2}F1^{3}F1 structures was used to obtain an initial model of each ^{2}F1^{3}F1:FnBR peptide structure that was correctly orientated and positioned in the unit cell (Blow, 2002). A molecular replacement solution for subsequent refinement was achieved in two steps for all structures except for ^{2}F1^{3}F1:PyTT5 Structure III. In the first step, a molecular replacement solution was found by searching with a model of ^{2}F1 and a model of ^{3}F1. The F1 modules were searched for separately because the inter-module orientation of ^{2}F1 with respect to ^{3}F1 can change (Rudiño-Piñera *et al.*, 2007). For ^{2}F1^{3}F1:PyTT5 Structure I, the molecular replacement was performed by the BALBES server (www.ysbl.york.ac.uk/~fei/balbes/index.html, Long *et al.*, 2008) with search models of ^{2}F1 and ^{3}F1 corresponding to residues 64–106 and 107–151, respectively, from the ^{2}F1^{3}F1 structure with pdb code 2cg6 (Rudiño-Piñera *et al.*, 2007). For the other structures, molecular replacement was performed by Phaser from the CCP4 program suite (McCoy *et al.*, 2007). The search models were ^{2}F1 (residues 64–106) and ^{3}F1 without the loop between strands C and D (residues 108–125, 134–151); they were from the ^{2}F1^{3}F1:FnBPA-1 peptide structure (pdb code 2rkz; Bingham *et al.*, 2008).

In a correct molecular replacement solution, the C-terminus of each ^{2}F1 molecule was near the N-terminus of a corresponding ^{3}F1 molecule, and each F1 molecule in a module pair was in the same asymmetric unit. However, not all ^{2}F1 and ^{3}F1 molecules in the initial molecular replacement solution fulfilled these criteria because the ^{2}F1 and ^{3}F1 molecules were searched for separately. Therefore, each initial molecular replacement solution was inspected for a ^{2}F1 molecule with a C-terminus near the N-terminus of a ^{3}F1 molecule, and these two F1 molecules were combined into a new ^{2}F1^{3}F1 model with the correct inter-module orientation. This model was used in a second molecular replacement run in Phaser. The molecular replacement solution from this second step was used as input for a rigid body refinement followed by a restrained refinement in Refmac 5 (Murshudov *et al.*, 1997).

Note that molecular replacement was not necessary for the third ^{2}F1^{3}F1:PyTT5 dataset, which had the same space group and cell dimensions as the second ^{2}F1^{3}F1:PyTT5 dataset. Therefore, the structure factor amplitudes from the third ^{2}F1^{3}F1:PyTT5 dataset and the model of ^{2}F1^{3}F1 from ^{2}F1^{3}F1:PyTT5 Structure II were used as the input for a rigid-body refinement followed by a restrained refinement in Refmac 5.

2.7.7 Structure refinement and validation

For each structure, the model and structure factors from the initial restrained refinement (Section 2.7.6) were used in iterative rounds of model building in Coot (Emsley & Cowtan, 2004) followed by further restrained refinement in Refmac 5. Initially, as much of the F1 molecules as possible was modelled and then the FnBR peptide was modelled, first as poly-alanine anti-parallel to the E-strand of the F1 module, and then, after refinement, as the FnBR sequence. Waters were modelled using the water finding functions in both Coot and Refmac 5. During model building, the quality of the model was judged both by using Coot's validation tools (particularly density fit analysis and rotamer analysis), and also by a decrease in the R factor (R_{work}) and R_{free} after refinement. Both R_{work} and R_{free} are a measure of the agreement between the model and the observed data. Finally, both ^{2}F1^{3}F1:PyTT5 Structure III and the ^{2}F1^{3}F1:IxTT1 structure were processed with the MolProbity server (molprobity.biochem.duke.edu/, Davis *et al.*, 2007), and problem regions identified by MolProbity were re-modelled before re-analysing the structures with the MolProbity server.

2.8 Isothermal titration calorimetry (ITC) and associated analysis

2.8.1 ITC data collection for pNTD binding to either SfbI-5, an SfbI-5 mutant or TickFnBR

The binding of pNTD to SfbI-5, SfbI-5 mutants[10] and TickFnBR were measured by ITC on a VP-ITC microcalorimeter (MicroCal) in PBS (Table 4). Experiments with either SfbI-5 constructs or TickFnBR were performed at 37°C or 25°C, respectively. The heats evolved during injections in these experiments were close to the limit-of-detection for the VP-ITC microcalorimeter. Therefore, much effort was made to maximize the reproducibility and consistency of these experiments. To determine reproducibility of the ITC experiments, the experiments with the following SfbI-5 mutants were

---

[10] Henceforth, the term 'SfbI-5 construct' will be used to encompass both SfbI-5 and SfbI-5 mutants.

performed twice: SfbI-5, F554A, K556A, K556E, G564Q, D574A and T575A; and those with T558A, three times. Additionally, two control experiments were performed in which PBS was injected into either PBS or into pNTD in PBS.

2.8.1.1 Solution preparation

PBS for dialysing pNTD for use in ITC was made up as follows. Components for 1 L of 10× PBS stock were weighed out to two decimal places, dissolved, and made up to 1 L with a volumetric flask. The stock PBS was filtered and degassed and stored at room temperature. 2 L of 1× PBS was made up from this stock on the day of dialysis using 1 L and 100 mL volumetric flasks. At no stage was the pH adjusted.

pNTD was purchased as vials of powder lyophilized from PBS with sucrose (0.5 mg/vial, Sigma Aldrich) and stored at −20°C. All pNTD used in the ITC experiments was from the same batch number (047K4086). The pNTD present in 2 vials (*i.e.,* 1 mg including PBS and sucrose) were each dissolved in 750 µL HPW. Both vials were then rinsed with 500 µL HPW by washing one vial with gentle inversion, transferring this solution to the next vial, and washing again. The dissolved pNTD solution and wash solution were pooled to give ~2 mL, then dialysed in a Slide-A-lyzer dialysis cassette (0.5–3 mL capacity, 10 kDa MWCO, Pierce) against 2 L of PBS with gentle stirring (16–18 h, room temperature), during which time the volume increased to ~2.5 mL.

The dialysed pNTD was retrieved, and the concentration of this pNTD stock solution calculated from an $A_{280–320}$ reading (Section 2.3.6). An aliquot of pNTD stock solution was removed that was ~100 µL more than the volume required to give a final volume of 2.2 mL when diluted to the desired concentration for the ITC experiment. The remaining dialysed pNTD stock solution was stored at 4°C and used within 5 days[11]. The aliquot of pNTD stock solution was centrifuged (16,000×*g*, 5 min, room temperature) to remove any particulate matter, and then the supernatant was transferred to a new tube. The concentration of this pNTD stock solution aliquot was calculated from the average of at least five separate $A_{280–320}$ readings (Section 2.3.6). The dialysis

---

[11] There was an exception made for one experiment, when the dialysed pNTD solution was used after storage for 10 days at 4°C. However, the integrity of the pNTD sample was verified by SDS PAGE analysis the day before use (*i.e.,* the ninth day of storage).

buffer was filtered and degassed to minimize particulate matter and then used to dilute the pNTD stock to create a pNTD solution for the ITC cell (2.2 mL) at the desired concentration (usually 2 μM). The concentration of the cell sample was determined from the average of at least five separate $A_{280-320}$ readings (Section 2.3.6).

The syringe solution (1 mL) was prepared from lyophilized SfbI-5 construct or TickFnBR using one of two methods. In the first method, the peptide was diluted with PBS from the dialysis of pNTD (dialysis PBS) to give a concentration of 300 μM based on the yield calculated from its mass assuming 20% water (Section 2.3.9). This was then sonicated in a sonicator bath (5 min), centrifuged (16,000×g, 1 min, room temperature) and then an aliquot was added to dialysis PBS to give the desired syringe concentration (usually 30 μM). For these dilutions, to ensure pipetting was as accurate and precise as possible, every pipetting step used in preparation of the stock and syringe solution was 'calibrated' as follows. Assuming x μL of dialysis PBS or syringe stock solution needed to be pipetted, the pipette's volume was changed until it pipetted x μL of HPW as determined by weighing pipetted volumes on a 4 decimal place balance (Mettler Toledo), and assuming 1 mL water weighs 1 g. When $x \pm 2$ μL HPW had been pipetted at least two times consecutively, x μL of dialysis PBS or syringe stock solution was pipetted.

It was subsequently realised that the first method had the following two disadvantages: some of the SfbI-5 mutants were not soluble at 300 μM, and had a very fine, light precipitate. Also, when repeating an ITC experiment at a later date the syringe solutions had to be made up from a stock solution of SfbI-5 mutant made up in dialysis PBS from a different dialysis. To avoid both these problems, the final four mutants (M562A, G564A, G564Q and I571A) and TickFnBR were dissolved in a known volume of HPW until no precipitate was visible, aliquoted, and then re-lyophilized. On the day of the ITC experiment, an aliquot was resuspended in 1 mL dialysis PBS to give 30 μM. Again, pipetting was 'calibrated' as described above.

The pH of the pNTD cell solution and of the SfbI-5 construct or TickFnBR syringe solution was measured. The pH of the syringe solution was increased with dilute sodium hydroxide (~2 μL) until it was within ± 0.01 pH units of the pNTD solution. Although the pH of the pNTD solution was not adjusted, for all ITC experiments the pH of both cell and syringe solutions were within the range 7.40–7.50.

2.8.1.2 Experimental setup

Each ITC experiment consisted of two titrations: the binding titration, with pNTD in the cell and an SfbI-5 construct or TickFnBR in the syringe; and a control titration, with dialysis PBS in the cell, and the same SfbI-5 construct or TickFnBR solution in the syringe. Before each experiment, the cell and syringe solutions and dialysis PBS for the final wash in the cleaning step (5 mL) were degassed with stirring in a ThermoVac degasser (10 min; MicroCal). Before the control titration, a further 5 and 2.5 mL dialysis PBS were degassed with stirring (10 min); the former was used for the cleaning steps (Table 16) and the latter was used as the cell solution for the control titration. The degassing steps were performed at a temperature 2°C lower than the temperature of the corresponding ITC experiment.

Table 16. Cleaning methods used before an ITC titration.

Name	Method
A	The cell was set to the correct experimental temperature (37°C), to allow the system to warm up while the washing was taking place. The cell was washed with 5% Decon 90 (100 mL; Fisher Scientific) and HPW (300 mL) using the ThermoVac to pull the solutions into and out of the cell. It was then washed with PBS (5 × 2 mL) and finally with degassed dialysis PBS (2.5 mL) by filling the cell with PBS using the filling syringe and then removing. The cell was then immediately filled with the cell solution.
	The injection syringe was washed with 5% Decon 90 (2 × 4 mL), HPW (3 × 4 mL) and dialysis PBS (3 × 4 mL). For each 4 mL wash, ~2 mL was pulled through the syringe, the injection port was closed while the loading syringe was emptied, and then the port was opened and a final 2 mL was pulled through, without pulling through air. The solution that remained in the syringe was allowed to drain before the next wash, and the outside of the syringe was wiped gently with a tissue. Finally, degassed dialysis PBS (2.5 mL) was pulled through, the injection port was closed and the syringe was purged and refilled twice using the ITC software. The port was opened and left to drain, and then filled with the syringe solution.

Name	Method
B	As in method A, the cell was set to the correct experimental temperature (37°C or 25°C). The cell was washed with dialysis PBS (2 × 2 mL) and 5% Decon 90 (3 × 3 mL). The wash solutions were agitated in the cell by gently pipetting the solution up and down with the filling syringe three times before removing from the wash solution from the cell. The final Decon 90 wash was left in the cell for at least 10 minutes, and was agitated six times before removing. Then, the cell was washed with HPW (6 × 2 mL) and dialysis PBS (4 × 2 mL), and agitated three times for each wash. Finally, the cell was washed with degassed dialysis PBS (2.5 mL) by gently filling and then removing, and then immediately filled with the cell solution.
	The injection syringe was washed with dialysis PBS (2 × 4 mL) and 5% Decon 90 (2 ×°4 mL), and was left with the final Decon 90 wash for at least ten minutes, during which time the outside of the injection syringe was washed with 5% Decon 90 and HPW and gently dried with a tissue. Then, the syringe was washed with HPW (4 × 4 mL) and dialysis PBS (4 × 4 mL). Each 4 mL wash was carried out as described in method A. Finally, degassed PBS (2.5 mL) was pulled through as described in method A.
Rinse	This consisted of the dialysis PBS and degassed dialysis PBS wash steps described in method A for both the cell and the syringe.

The cell and syringe were cleaned before each binding titration, and rinsed or cleaned before each control titration (Table 16). Throughout these ITC experiments, there were problems with consistency between the binding and control titration (Section 4.3.3.1), so two changes were made to the cleaning procedure in an effort to improve consistency. First, after the initial four experiments (with SfbI-5, E573A, T575A and a repeat of T575A), the rinsing step in-between the binding and control titrations was changed to a cleaning step. Second, the cleaning method was changed from method A to method B at the recommendation of a MicroCal technician (Table 16) for the final six experiments, those with M562A, G564A, G564Q, I571A and TickFnBR (G564Q was repeated).

After washing the ITC cell and syringe, and degassing the cell and syringe solutions, the ITC cell and syringe were filled with their appropriate solutions following the manufacturer's instructions. Briefly, the pNTD solution or degassed dialysis PBS was loaded into the cell with a filling syringe (a 2.5 mL polypropylene disposable syringe equipped with a long, blunt needle). During loading, the tip of the filling syringe's needle was held just above the bottom of the cell. When the solution could be seen at the top of the cell, the remaining solution in the syringe was tapped gently to expel

~200 µL of the remaining cell solution into the cell to help remove any trapped air from within the cell. Care was taken not to add air in this process, by leaving a small amount of liquid in the syringe. Then the excess solution from the cell was retrieved by resting the tip of the needle on a ledge at the top of the cell. The excess pNTD solution was frozen at −20°C. The syringe solution was pulled up gently through the ITC injection syringe with a loading syringe (a 2.5 mL disposable syringe attached to a clear plastic tube that in turn was attached to injection syringe's port) until the solution had just entered the loading syringes' tube, then the injection syringe's port was closed. The injection syringe was then purged and refilled two times with syringe solution, and placed carefully in the cell.

The parameters for the ITC experiments were based on those reported for the pNTD and SfbI-5 titration performed by Schwarz-Linek *et al*. (2004b). The parameters for individual ITC experiments (summarized in Table 17) were varied based on the predicted K_d given the mutation and also, as more ITC experiments were performed, on the results of previous experiments too. The binding and control titrations always had the same parameters.

An ITC experiment goes through two equilibration steps, the 'pre-stirring' and 'final baseline equilibration' steps for equilibration between the reference and the cell before and after stirring starts. All automatic equilibration options were turned off to allow manual selection of the start time for each of these two equilibration steps, since the software did not allow the baseline sufficient time to equilibrate given the small size of the heats being measured. Best results were achieved when the system had been allowed to equilibrate for 45–60 minutes, whereas the software judged the system ready after ~10 minutes.

2.8.2   ITC data collection for ^{2}F1^{3}F1 binding to IxTT1

The binding of ^{2}F1^{3}F1 to IxTT1 (Section 2.1.5) was measured by ITC in 10 mM sodium/potassium phosphate buffer, pH 7.4 at 25°C. Solution preparation and experimental setup were carried out as described for pNTD and SfbI-5 constructs (Section 2.8.1) with the following exceptions. 10 mM sodium/phosphate buffer, pH 7.4 (2 L) was prepared by diluting 100 mM sodium/potassium phosphate 10-fold using volumetric flasks. ^{2}F1^{3}F1 was dialysed against this buffer in a Slide-A-lyzer dialysis cassette (0.5–3 mL capacity, 2 kDa MWCO, Pierce) for ~18 h. An aliquot of IxTT1 was

lyophilized and resuspended in dialysis buffer after the dialysis of ^{2}F1^{3}F1 to give a 1 mL syringe solution at a concentration of approximately 1.1 mM. This concentration was based on the weight of the lyophilized peptide assuming that 10% of the weight was water. 10% rather than 20% water was assumed based on the experience of our laboratory when working with synthetic FnBR peptides. The experiment parameters used for the titration differed from those in Table 17 as follows: the reference power was 10 µcal/s and the total number of injections was 33, with injections 2–33 having a volume of 8 µL, an injection speed of 0.5 µL/s and an injection spacing of 360 s. A control titration was not performed for this titration. Instead, the last five injections were used to control for heat of dilution effects as described in Section 2.8.3.

Table 17. Summary of parameters used in ITC experiments of pNTD with SfbI-5 mutant peptides in PBS. The injection syringe of the VP-ITC calorimeter was upgraded partway through this series of experiments, and this changed the stirring speed and injection speed, the values of these parameters before and after the upgrade are shown separated by a '/'. The injection speed is not set directly; rather the duration for the injection is calculated by the microcalorimeter software based on the injection volume and default injection speed. However, in two experiments the duration was accidentally changed to that of the first injection, to give an injection speed of 1 µL/s.

Parameter	Setting
Total number of injections	35–50
Cell temperature (°C)	37
Reference power (µCal/s)	4
Initial delay (s)	60
Stirring speed (rpm)	300 / 307
Feedback mode	Low
ITC equilibration options	All off
Volume of first injection (µL)	2
Injection volume (µL)	5–7
Injection speed (µL/s)	0.42 / 0.5
Injection spacing (s)	240–330
Filter period (s)	2
Cell volume (mL)	1.4193

2.8.3   Raw data processing

The raw ITC data (μcal/s versus s) from the binding and control titrations was analysed with Origin 7.0 software with the ITC module (MicroCal). With this software, the baseline for the titration was automatically calculated then manually adjusted such that the baseline went through the middle of the noise at the end of each titration, with care taken not to overfit the noise. The width of each injection peak calculated by Origin was only altered if an unusual spike in the data was present at the end of an injection, which happened rarely. The integrated peak area (μcal/injection) for each injection was then calculated by Origin.

Next, an appropriate control was subtracted from the binding titration. This was either the corresponding injection from the control titration from the same ITC experiment, or the average peak area of the last five or ten injections from the binding titration. The aim of this step is to compensate for any heat effects that are not caused by binding of the protein in the cell to the peptide in the syringe. Examples include the 'heat of dilution' from diluting the same solution into itself, and also heat from any mismatch between components in the cell and syringe solutions besides, for example, pNTD and the SfbI-5 construct. For tight binding, such as that seen for pNTD and SfbI-5 constructs, saturation of all pNTD binding sites was reached within the binding titration, so the injection heats at the end of the titration can also act as a control for other heat effects. In many experiments the integrated heats in the control titration were not similar in size to the integrated heats at the end of the titration. Therefore, for these experiments the average area of the last five or ten injections in the binding experiment were used as the control. The number of injections averaged (five or ten) was a trade-off between compensating for noise within the data and by which injection the binding titration had reached saturation. For tight binding mutants (K_d of 15 nM or smaller) the average of the last ten injections was used, for relatively weak binding mutants (K_d of 25 nM or greater) the average of the last five injections was used.

After subtraction of the appropriate control from the integrated peak area data, this data was fitted to a single binding site model in Origin (Equations 11, 12) using non-linear least squares regression analysis, in which Origin aims to maximize the agreement between the model and the data by reducing the χ^2 value. The data was processed with both the control titration and the end of the binding titration as a control, and the fit that gave the lower χ^2 value was used as the control for that ITC experiment (Table 18).

$$\Delta Q(i) = Q(i) + \frac{\Delta V_i}{V_0}\left[\frac{Q(i) + Q(i-1)}{2}\right] - Q(i-1)$$

Equation 11. The change in heat content for the ith injection. Annotation: V_0, volume of the cell detected calorimetrically ('active' cell volume); ΔV_i, volume of the ith injection; Q, total heat content of the solution contained in V_0 after the ith injection (Equation 12); ΔQ, change in Q from injection i compared to injection $(i-1)$ (MicroCal, 2004).

$$Q = \frac{nM_t \Delta H V_0}{2}\left[1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} - \sqrt{\left(1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t}\right)^2 - \frac{4X_t}{nM_t}}\right]$$

Equation 12. The total heat content in the active cell volume after the ith injection. Annotation is as defined in Equation 11 and: $\Delta H°$, molar change in enthalpy (heat) on ligand binding; M_t, bulk concentration of macromolecule (pNTD) in V_0 after the ith injection; X_t, bulk concentration of ligand (FnBR peptide) in V_0 after the ith injection; K, association constant; n, molar ratio of binding sites (MicroCal, 2004).

Input 'constants' for the single binding site model are the concentration of protein in the cell and peptide in the syringe, output variables are the association constant (K), the change in binding enthalpy ($\Delta H°$) and the molar ratio of binding (n), with associated errors (Table 18). The concentration of pNTD or ^{2}F1^{3}F1 used was that calculated for the cell solution after dilution from pNTD or ^{2}F1^{3}F1 stock solution (Section 2.8.1.1). The concentration used for the peptide in the syringe was that required to give a molar ratio of 1 after non-linear regression analysis by Origin. That is, the peptide concentration of the SfbI-5 construct or IxTT1 was manually varied until the molar ratio for the binding curve was 1.0, at which point the molar ratio was fixed, and the data was fitted again. So the peptide concentration was treated as a variable and the molar ratio was treated as a constant. For some experiments with SfbI-5 constructs, the pooled syringe sample, left over from the control and binding experiments, was analysed by quantitative amino acid analysis (Section 2.8.4). This was then used as the SfbI-5 construct concentration for that experiment, and the molar ratio was left to vary.

Table 18. Thermodynamic parameters for all ITC experiments performed with pNTD and SfbI-5 constructs. The standard errors from the fit are shown. The method of control for the heats of dilution is indicated in the 'Control' column. Abbreviations are: C, control titration; 5 or 10, average of the last five or ten injections of the binding titration, respectively; n, molar ratio of binding, which was only varied when amino acid analysis was performed on the SfbI-5 construct syringe sample, otherwise it was fixed to 1; K, association constant; $\Delta H°$, change in enthalpy on binding.

Construct	[pNTD] (µM)	[construct] (µM)	Control	n	K (×10⁶ M⁻¹)	$\Delta H°$ (kcal/mol)
SfbI-5	2.02	25.4	10	1	277 ± 27	−44.74 ± 0.20
	1.97	24.3	C	0.973 ± 0.002	298 ± 31	−46.09 ± 0.21
V552A	1.95	31.0	C	1	94.1 ± 7.2	−44.25 ± 0.21
E553A	1.98	22.9	C	1	119 ± 12	−45.19 ± 0.27
F554A	1.99	28.1	C	0.877 ± 0.005	36.6 ± 3.0	−43.52 ± 0.34
	2.07	29.7	5	1	26.8 ± 2.1	−42.18 ± 0.29
K556A	1.44	21.1	10	0.941 ± 0.003	700 ± 120	−47.06 ± 0.31
	1.72	23.8	10	1	538 ± 59	−44.36 ± 0.19
K556E	1.47	26.2	10	1	1610 ± 200	−46.66 ± 0.17
	1.52	24.1	10	0.878 ± 0.002	890 ± 120	−51.36 ± 0.24
D557A	2.03	28.1	C	1	72.3 ± 5.3	−43.15 ± 0.21
T558A	2.01	24.2	C	1	17.2 ± 1.1	−36.30 ± 0.26
	2.01	23.5	5	1	18.8 ± 0.84	−39.12 ± 0.18
	3.09	41.9	5	1.010 ± 0.004	18.2 ± 0.96	−38.17 ± 0.21
M562A	2.08	30.5	C	1	122 ± 6.3	−40.04 ± 0.12
G564A	1.98	28.0	5	1	34.9 ± 2.5	−38.46 ± 0.23
G564Q	1.99	25.9	5	1	19.2 ± 1.8	−34.43 ± 0.32
	2.02	26.6	C	1	21.1 ± 1.0	−34.96 ± 0.17
I571A	1.99	29.3	C	1	130 ± 5.4	−44.47 ± 0.10
E573A	1.98	26.7	10	1	71.8 ± 4.0	−43.60 ± 0.16
D574A	2.00	25.0	5	1	27.9 ± 1.3	−45.08 ± 0.18
	4.00	49.4	5	0.922 ± 0.002	20.9 ± 0.9	−47.49 ± 0.18
T575A	2.01	24.0	10	1	90.0 ± 6.5	−44.48 ± 0.20
	1.92	24.0	10	1	89.9 ± 3.6	−45.22 ± 0.12

2.8.4 Correction to SfbI-5 construct concentration based on amino acid analysis

The following samples underwent quantitative amino acid analysis performed by Alta Biosciences (University of Birmingham): F554A, K556A, K556E, T558A, D574A and SfbI-5. The latter four samples were from the final titrations for that sample. The results from this analysis comprised molar concentrations for each amino acid, with the following caveats. First, asparagine and glutamine are completely converted to aspartic acid and glutamic acid, respectively. Second, there is a 5% and 10% loss of threonine and serine, respectively. Therefore, these results were used to calculate the concentration of each sample as follows. First, the amino acid composition of the SfbI-5 construct was used to normalise the molar concentrations of each amino acid by dividing the reported values by the number of that amino acid in the mutant. The reported value for glutamic acid was normalized by the total number of glutamines and glutamic acids, and the value reported for aspartic acid was normalized by the total number of asparagines and aspartic acids. Second, the SfbI-5 mutant's mean concentration and standard deviation was calculated from the normalized concentrations of all amino acids except threonine and serine. Third, the number of standard deviations from the mean was calculated for each normalized amino acid concentration. Any value that was more than two standard deviations away from the mean was considered an outlier. Fourth, a new mean and standard deviation were calculated by not including the greatest outlier. Fifth, the third and fourth steps were repeated until there were no more outliers contributing to the mean. There were no more than three outliers per sample. This mean value was used as the SfbI-5 construct concentration for that sample in ITC data analysis (Section 2.8.3; Table 18).

2.8.5 Calculations of thermodynamic parameters from ITC data

The processed data consisted of an association constant (K) and change in binding enthalpy ($\Delta H°$) for each ITC experiment (Table 18). For those SfbI-5 mutants whose ITC experiments were repeated, the mean value of each parameter was used for subsequent analysis. Then, the error for the mean was calculated using two different methods, giving two different values for the error. The larger of these two values was used as the error for all subsequent calculations. The first method was to calculate each value's difference from the mean, and take the largest difference as the error. The second method was to combine the parameters' standard errors using Equation 13. The

standard errors were calculated by Origin from non-linear least squares regression analysis of the experiment's data. For mutants whose ITC experiments were not repeated, the error of a mutant's parameters was the standard error from Origin. The association constants and change in enthalpies were then used to calculate the K_d, the molar change in entropy on binding ($\Delta S°$), the Gibbs free energy ($\Delta G°$) and the change in Gibbs free energy of the SfbI-5 mutants relative to SfbI-5 ($\Delta\Delta G°$; Table 19). Finally, whether energy–enthalpy compensation had occurred was determined using the method described by Sharp (2001).

$$E = \frac{\sqrt{s_1^2 + s_2^2 + \ldots + s_n^2}}{n}$$

Equation 13. Method 2 for calculating the uncertainty associated with a mean. Annotation: s_n, standard error for the nth value for which the mean is being calculated, n, total number of values used to calculate the mean; E, the error associated with the mean.

Table 19. Equations for calculating thermodynamic parameters and their associated errors. Annotation: K, association constant; K_d, dissociation constant; $\Delta G°$, Gibbs free energy (subscripts M and S represent SfbI-5 mutant and SfbI-5, respectively); R, gas constant; T, temperature; $\Delta S°$, change in entropy; $\Delta H°$, change in enthalpy; E_x, error in thermodynamic parameter x. The equations for associated errors assume the variables are independent. Therefore, there is not a simple equation for the calculation of error for $\Delta S°$, since it is calculated from two dependent variables, $\Delta H°$ and $\Delta G°$.

Equation for thermodynamic parameter	Equation for associated error
$K_d = \dfrac{1}{K}$	$E_{K_d} = \dfrac{E_K}{K} K_d$
$\Delta G° = -RT \ln K$	$E_{\Delta G°} = \dfrac{\lvert -RT \rvert}{K} E_K$
$\Delta S° = \dfrac{\Delta H° - \Delta G°}{T}$	–
$\Delta\Delta G°_M = \Delta G°_M - \Delta G°_S$	$E_{\Delta\Delta G°_M} = \sqrt{E_{\Delta G°_M}^2 + E_{\Delta G°_S}^2}$

2.9 Solid phase binding assays

A series of solid phase binding assays were performed with GST-fusions of FnBRs and either human plasma or $^{1-5}$F1. The general concept of these 'GST pull-down' experiments is depicted in Figure 14.
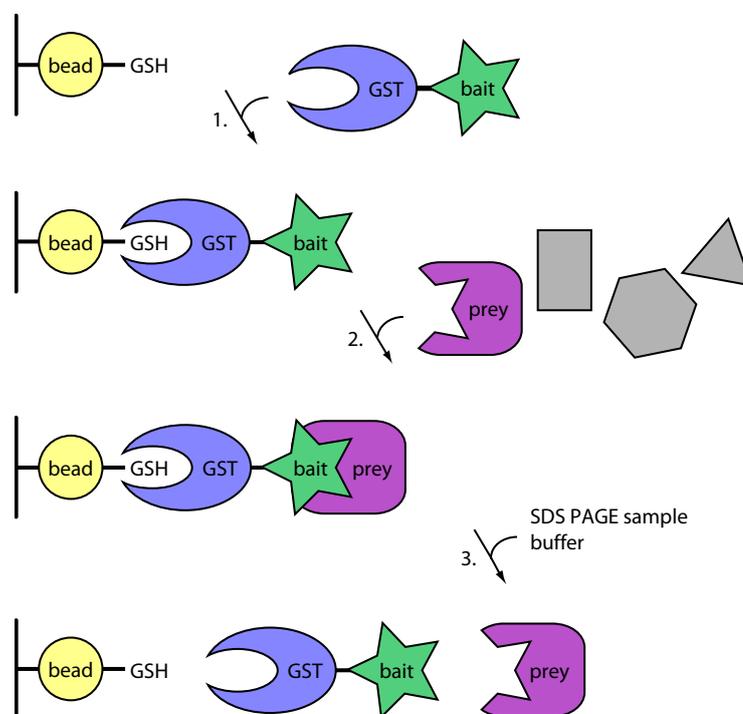
Figure 14. Schematic of a GST pull-down experiment. The general concept of a GST pull-down experiment can be summarized in the following four steps. 1) GST fused to the 'bait' protein is added to glutathione (GSH), which is covalently attached to a solid support such as agarose or a magnetic particle (bead). For pull-down experiments in this work, the bait protein was always an FnBR. The GST-fusion is added in excess and any unbound GST-fusion is washed away with a series of wash steps. 2) Either pure 'prey' protein or a complex mixture containing the 'prey' protein is added to the bead-GSH:GST-bait complex. Therefore, a bead-GSH:GST-bait:prey complex may be formed. In this work, the prey protein was Fn in plasma (Section 2.9.1), or pure $^{1-5}$F1 in the presence or absence of another FnBR that might compete for binding with the GST-FnBR (Section 2.9.2). After a binding step in which prey protein is added in an amount that is predicted to be in excess of the bound GST-bait protein, other proteins and unbound prey protein are washed away. 3) SDS PAGE sample buffer (Table 10) is added, and the sample containing the glutathione beads, GST-bait and any bound prey protein is heated. This step denatures the proteins present, which results in separation of the GST-bait and prey proteins from each other and from the glutathione beads. This sample can then be analysed by SDS PAGE to detect the MW of any prey proteins.

2.9.1 GST pull-downs of Fn from human plasma

All GST pull-downs of Fn from human plasma (Sera Laboratories) were performed with magnetic MagneGST Glutathione Particles (Promega), and a MagnaRack stand for 1.5 mL tubes (Invitrogen). Magnetic glutathione particles were used instead of glutathione agarose because it reduced the risk of working with human plasma by avoiding centrifugation of human plasma in mini-spin columns. The human plasma was supplied as 7.1 mL aliquots to avoid multiple freeze-thaw cycles and stored at −80°C until use.

2.9.1.1 Production of GST-fusions used in pull-down experiments

GST-SfbI-5, six GST-SfbI-5 mutants, GST-TickFnBR and GST were overexpressed in *E. coli* at 30°C as described in Section 2.3.2, except that for each construct a 250 mL culture was grown in a 2 L flask. The six GST-SfbI-5 mutants expressed were E553A, F554A, K556A, T558A, G564A and E573A (Table 7). GST was overexpressed by transforming *E. coli* with the 'empty' pGEX-5X-2 plasmid. This expressed GST with the Factor Xa cleavage site (4 residues) and the multiple-cloning site (14 residues) at the C-terminus of the GST. After harvesting the cells from each culture by centrifugation (5,000×g, 20 min, 4°C), the cell pellets containing the overexpressed GST-fusion were each resuspended, aliquoted to give a theoretical OD_{600} of ~60 when resuspended in 500 µL, and then centrifuged again (4,600×g, 20 min, 4°C). The aliquoted cell pellets were stored at −80°C until use.

2.9.1.2 Preparation of GST-fusion bound glutathione particles

For each GST-fusion, a slurry of glutathione particles was prepared as follows. Aliquots of cell pellets of each mutant were thawed and resuspended in ice-cold 450 µL PBS to give a final volume of ~500 µL. The protease inhibitor AEBSF and lysozyme were added to each aliquot to give final concentrations of 1 mM and ~1 mg/mL, respectively. Each aliquot was mixed and then lysed by three freeze/thaw cycles, which consisted of freezing each aliquot with liquid nitrogen and thawing in cold water. Then DNase (Sigma Aldrich) and Triton X-100 were added to a final concentration of ~5 µg/mL and 1% (v/v), respectively, and the aliquots were rotated end over end (~10 rpm, 30 min, 4°C) to ensure lysis was complete. The insoluble cellular material was then pelleted by centrifugation (10,000×g, 30 min, 4°C). Meanwhile, glutathione particles were aliquoted into a separate tube for each mutant, so that there was 37.5 µL of settled beads

for each aliquot of lysed cells. The particles were washed with PBS (4 × 800 µL) as outlined in Table 20. After centrifugation of the lysed cells, the soluble cellular protein was retrieved and added to the pre-equilibrated glutathione particles. The GST-fusion was allowed to bind to the particles for 2 h while the samples were rotated end over end (~10 rpm, 4°C) to improve mixing. The unbound soluble cellular protein was then retrieved and discarded. The particles bound to each GST-fusion were then washed six times to remove non-specifically bound protein. The first wash was 750 µL PBS with 1% (v/v) Triton X-100, and it was mixed by end over end rotation (~10 rpm, 5 min, 4°C). The other washes were with 750 µL PBS and were mixed as usual (Table 20). Finally, the particles were washed once with and then stored in PBS with 50% glycerol as a 12.5 % slurry (v/v) at –20°C.

Table 20. The procedure for washing magnetic glutathione particles with the aid of a magnetic stand. The stand used, a MagnaRack, is for up to 24 × 1.5 mL tubes with a removable component containing 12 magnets which attract the magnetic particles to the side-wall of the tubes.

Step	Process
1.	~ 800 µL PBS was added to each tube containing glutathione particles and all lids were closed.
2.	The tubes were each flicked and inverted to mix
3.	The tubes were placed in contact with the magnets on the MagnaRack, this attracted most of the magnetic particles to the side of the tube.
4.	The tubes were then carefully inverted while still in the rack to resuspend any remaining particles on each tube's wall and lid, allowing them to be attracted to the magnet.
5.	The PBS was carefully removed with a micropipette so as not to disturb the particles. The position of the particles against the side-wall of the tube allowed the pipette tip to reach to the bottom of the tube and remove virtually all of the PBS.

To allow semi-quantitative analysis of the amount of Fn bound by each GST-construct, the amount of GST-fusion present in each pull-down had to be approximately constant. Therefore, the GST-fusion bound to glutathione particles was normalized as follows. A sample of each GST-fusion bound to glutathione particles and also a serial dilution of purified GST-SfbI-5 at known concentration (as determined by A_{280}, Section 2.3.6) were separated by SDS PAGE and stained with Flamingo fluorescent stain (Bio-Rad Laboratories). The amount of GST-fusion present in each sample was then determined by analysis of the fluorescent intensity of the bands on the gel (Section 2.9.1.3). Based

on this analysis, the amount of GST-fusion bound to the particles was normalized to 2 μg GST-fusion per 1 μL settled particles by diluting the bound particles with 'empty' glutathione particles that had been washed with PBS and pre-equilibrated in PBS with 50% glycerol. These normalized samples were aliquoted as 10 μL settled beads per tube and then stored as a 12.5% (v/v) slurry at –20°C.

2.9.1.3 SDS PAGE analysis with fluorescent stain and quantification of band intensity

SDS PAGE was performed as outlined in Section 2.3.5 with the following exceptions. Protein from pull-down experiments was eluted from 10 μL of glutathione particles by addition of 100 μL of 1× sample buffer (Table 10) and heating at 70°C for 10 minutes. According to both the Flamingo stain's manufacturer and our own results, there is a linear relationship between the fluorescence of stained bands and amount of protein present in the bands when there is between 1 ng and 1,000 ng of protein per band, with the ideal amount being approximately 20–100 ng. To achieve this range the protein eluted from the particles was typically diluted with 1× sample buffer 1000-fold relative to the volume of particles. To obtain maximum separation of the protein of interest all gels were run in MOPS SDS running buffer (Table 10; 200 V, ~70 min). After running the gel, the gels were then stained with Flamingo stain following the manufacturer's instructions for optimal sensitivity, which included staining for at least 12 h. Because of the sensitivity of this fluorescent technique, all containers for staining and washing gels were cleaned with 5% Decon 90 before use and then rinsed thoroughly with HPW.

Images were acquired of fluorescently stained gels using a Molecular Imager FX (Bio-Rad laboratories) and analysed with Quantity One software (Bio-Rad Laboratories). Fluorescence was excited at 532 nm using a laser that emitted at both 532 and 1024 nm and a filter that blocked the 1024 nm emission. Emission from the gel was filtered with a 605 nm band-pass filter. The digital exposure time was optimized to get the highest signal possible from the imager without saturating any pixels in the image. All images were scanned at a resolution of 100 μm except for the image of the final pull-down experiment (Section 2.9.1.5), which was scanned at the higher resolution of 50 μm. The effect of small black dots in the image caused by fluorescence of particulates on the gel was reduced using either a 3×3 or 5×5 pixel median filter for some gel images. A 5×5 median filter was used for the image of the final pull-down experiment. Lanes in the gel image were selected and the background was subtracted using the program's

'rolling disc' method; a disc size of 16 was used for the image of the final pull-down experiment. The bands on the gel were then selected based on contours of intensity, which allowed bands that were not rectangular to be analysed most accurately. The fluorescent intensity of the bands was calculated as a 'contour quantity', that is, the volume of the band measured in counts $\times$ mm^2 where counts are a measure of the fluorescent signal. For standardization of the GST-construct, the quantity of GST-fusion present in each band on the gel was determined from the contour quantities of the bands of GST-SfbI-5 standards. This was achieved using linear regression analysis of the standards and unknowns in Excel (Microsoft).

2.9.1.4 Preliminary pull-down experiment with GST-SfbI-5 and the T558A mutant

Before performing a series of pull-downs with all GST-constructs, an experiment was performed to determine the ratio of GST-fusion to human plasma required for optimal sensitivity. GST-SfbI-5 and the T558A mutant were used for this experiment. The T558A mutant was chosen because it had the weakest K_d for binding to pNTD (Table 37, p 188). Additional glutathione particles with either GST-SfbI-5 or the GST-SfbI-5 T558A mutant were prepared for this preliminary experiment with 2 µg construct per 1 µL particles (Section 2.9.1.2). For this experiment, the ratio between the plasma and GST-fusion was varied by changing the volume of plasma and keeping the amount of GST-fusion constant. A series of calculations based on some assumptions were used to inform the range of plasma volumes tested in this experiment. These assumptions and calculations are explained in Section 2.9.1.4.1 and the procedure followed is explained in Section 2.9.1.4.2.

2.9.1.4.1 Outline of assumptions and calculations

First, it was assumed that the law of mass action (Equation 14) applied even though GST-SfbI-5 was bound to the glutathione particles and not free in solution. Second, given that the concentration of Fn dimer in the blood is in the range 0.65−0.87 µM (Mosher, 2006) the minimum and maximum effective Fn concentrations were rationalized as follows. The minimum effective Fn concentration used, 0.65 µM, was chosen by assuming that for steric reasons GST-SfbI-5 might only be able to bind one molecule in a Fn dimer. The maximum effective Fn concentration used, 1.7 µM, was chosen by assuming that not only could GST-SfbI-5 bind to both molecules in an Fn dimer, but that also the molecules in an Fn dimer behaved independently of each other.

The latter was considered a reasonable assumption because SfbI-5 binds to the N-terminus of Fn, and an Fn dimer is connected by its C-terminus (Figure 2). The result of the latter assumption is that only the effective concentration of Fn was affected, whether GST-SfbI-5 bound one or both molecules of Fn; the stoichiometry was considered 1:1.

$$K_d = \frac{[A]^m [B]^n}{[A_m B_n]}$$

Equation 14. The law of mass action applied to an equilibrium dissociation reaction. In a reaction at reversible equilibrium, where m molecules of A and n molecules of B interact to form the complex $A_m B_n$, such as $A_m B_n \leftrightarrow m A + n B$, then the K_d is defined according to the equilibrium concentrations of A, B and $A_m B_n$ as shown, where square brackets denotes the concentration.

Using these assumptions and the law of mass action, a series of calculations were performed in Excel with Equation 15, using a K_d range of 0.3–200 nM, a Fn concentration range of 0.65–1.7 µM, and a range of plasma volumes to give an Fn:GST-SfbI-5 molar ratio of between 1:0.01 and 1:2. These calculations contributed to experimental design in two key ways. First, they predicted that a Fn:GST-SfbI-5 molar ratio of 1:1 will result in the greatest 'resolution', that is, the largest change in the amount of Fn:GST-SfbI-5 complex formed over a K_d range of 0.3–200 nM. Second, given that varying the plasma volume varies the GST-SfbI-5 concentration, the plasma volume range required to test Fn:GST-SfbI-5 molar ratios near 1:1 was 370–960 µL. Therefore, based on all of the above considerations, the following five plasma volumes were chosen: 270, 360, 480, 640 and 900 µL.

$$[AB] = \frac{\left([A]_t + [B]_t + K_d\right) - \sqrt{\left([A]_t + [B]_t + K_d\right)^2 - 4[A]_t [B]_t}}{2}$$

Equation 15. Concentration of a bimolecular complex AB at equilibrium given the initial concentration of A and B. This equation is derived by combining the law of mass action and the formula for solving a quadratic equation. Annotation: $[A]_t$ and $[B]_t$, total concentration of A and B, respectively, that is either free or bound in the AB complex; $[AB]$, concentration of the AB complex at equilibrium.

2.9.1.4.2 Procedure

The preliminary pull-down experiment was performed as follows. Aliquots (10 μL) of glutathione particles bound to GST-SfbI-5 (6 aliquots) and GST-SfbI-5 T558A (6 aliquots), GST (1 aliquot) and also 'blank' glutathione particles with no GST bound (1 aliquot) were each pre-equilibrated with PBS (3×800 μL; Table 20). Each of these aliquots was a pull-down. Human plasma was thawed in a 37°C water bath, the protease inhibitor AEBSF was added to a final concentration of 1 mM and the plasma was centrifuged to pellet any particulates ($1,500 \times g$, 30 min, 4°C). The cleared human plasma was then added to the glutathione particles according to Table 21 and allowed to bind by rotating the container end over end (~10 rpm, ~3 h, 4°C). There were four control pull-downs: GST, 'blank' glutathione particles and a GST-SfbI-5 and T558A mutant pull-down to which 900 μL of PBS was added instead of human plasma (Table 21). After the binding steps the beads were each washed with PBS (5×800 μL; Table 20) and any bound protein was eluted with 100 μL of $1\times$ sample buffer and heating at 70°C for 10 min. The samples were then analysed by fluorescent SDS PAGE (Section 2.9.1.3).

Table 21. The pull-downs performed in a preliminary experiment to determine the optimal volume of plasma.

Pull-down	GST construct	Volume plasma (μL)	Pull-down	GST construct	Volume plasma (μL)
1	SfbI-5	270	8	T558A	640
2	T558A	270	9	SfbI-5	900
3	SfbI-5	360	10	T558A	900
4	T558A	360	11	SfbI-5	–
5	SfbI-5	480	12	T558A	–
6	T558A	480	13	GST	640
7	SfbI-5	600	14	–	640

2.9.1.5 Final pull-down experiment with all GST-fusions

This experiment was performed in triplicate. The procedure was the same as for the preliminary experiment except 270 μL of plasma was added to each of the ten pull-downs. These were: 'blank' glutathione particles, glutathione particles bound to GST-SfbI-5, GST-TickFnBR, GST, and GST-fusions of seven SfbI-5 mutants: E553A,

F554A, K556A, T558A, G564A and E573A. After the binding and wash steps the samples were eluted and analysed by fluorescent SDS PAGE (Section 2.9.1.3).

2.9.2 GST pull-downs with $^{1-5}$F1

These experiments were performed with Glutathione Sepharose 4 FastFlow agarose beads (GE Lifesciences) and Handee Mini-Spin Columns (Pierce).

Two pull-down experiments were performed, the first investigating binding between GST-TickFnBR and $^{1-5}$F1 and the second investigating binding of either GST-TickFnBR or GST-SfbI-5 to $^{1-5}$F1 in the presence or absence of SfbI-5 and TickFnBR. These two experiments had very similar protocols. Therefore, the following describes the protocol for both experiments, with parameters differing between the first and second experiments separated by an '*or*'. The glutathione beads were aliquoted to give a final settled volume of 25 *or* 20 µL. The beads were then washed with PBS (4 × 400 µL; Table 22). Solutions of GST-fusions or just PBS were added to the beads as summarized in Table 23 *or* Table 24. The GST construct was allowed to bind with gentle rocking on a gel rocker (~3 h, 4°C). The flow-through was removed by centrifugation (~1,200×*g*, 30 s, room temperature) and the beads were washed with PBS (4 × 400 *or* 600 µL; Table 22). $^{1-5}$F1 was added to the pull-downs (Table 23) *or* $^{1-5}$F1 with or without SfbI-5 or TickFnBR were added to the pull-downs (Table 24). The samples were allowed to bind with gentle rocking on a gel rocker (~75 min, 4°C). The pull-downs were washed with PBS (4 × 400 µL; Table 22) and any bound protein was eluted with 1× SDS PAGE sample buffer (50 *or* 40 µL) and heating at 70°C for 10 minutes. The samples were then analysed by SDS PAGE (Section 2.3.5).

Table 22. The procedure for washing glutathione beads glutathione particles with the aid of a mini-spin column.

Step	Process
1.	Small plastic frits were added to the bottom of each tube to stop gravity flow from the mini-spin columns.
2.	400–600 µL PBS was added to each tube containing glutathione particles and all lids were closed.
3.	The columns were each flicked and inverted to mix.
4.	The frit was removed and the columns were placed in a 1.5 mL tube and then the lids were loosened.
5.	The tubes were centrifuged (~1,200×g, 30 s, room temperature).
6.	The flow-through was emptied from the 1.5 mL tube.

Table 23. Solutions added in the binding steps in the first pull-down experiment with $^{1-5}$F1. In binding step 1, 800 µL of one of 9 µM GST-TickFnBR in PBS, 3 µM GST in PBS or just PBS (–) was added to each pull-down. In binding step 2, 800 µL of either 3.8 µM $^{1-5}$F1 in PBS or PBS (–) was added to each pull-down. The concentration of GST-fusions and $^{1-5}$F1 was determined by A_{280} (Section 2.3.6).

Pull-down	Binding step 1	Binding step 2
P	GST-TickFnBR	$^{1-5}$F1
C1	–	$^{1-5}$F1
C2	GST	$^{1-5}$F1
C3	GST–TickFnBR	–

Table 24. Solutions added in the binding steps in the second pull-down experiment with $^{1-5}$F1. In binding step 1, 600 µL of either a 9.1 µM solution of the indicated GST construct in PBS or just PBS (–) was added to each pull-down. In binding step 2, 600 µL of 6 µM $^{1-5}$F1 either with or without 12.1 µM of the indicated FnBR in PBS or just 600 µL PBS (–) was added to each pull-down. The concentration of GST-fusions and $^{1-5}$F1 was determined by A_{280} (Section 2.3.6). The concentration of TickFnBR and SfbI-5 peptides was determined from the weight of lyophilized peptides (Section 2.3.9).

Pull-down	Binding step 1	Binding step 2
P1	GST-SfbI-5	$^{1-5}$F1
P2	GST-TickFnBR	$^{1-5}$F1
Q1	GST-SfbI-5	$^{1-5}$F1, TickFnBR
Q2	GST-TickFnBR	$^{1-5}$F1, SfbI-5
C1	GST	$^{1-5}$F1, TickFnBR
C2	–	$^{1-5}$F1, TickFnBR
C3	GST-SfbI-5	–
C4	GST-TickFnBR	–
C5	–	$^{1-5}$F1, SfbI-5
C6	–	$^{1-5}$F1
C7	GST	$^{1-5}$F1, SfbI-5
C8	–	TickFnBR
C9	GST	TickFnBR
C10	GST-SfbI-5	TickFnBR
C11	–	SfbI-5
C12	GST	SfbI-5
C13	GST-TickFnBR	SfbI-5
C14	GST	$^{1-5}$F1

3   Characterization of the disordered state of SfbI

SfbI-5 is predicted to be disordered, to undergo a disorder-to-order transition, and bind NTD *via* a tandem β-zipper (Section 1.4; Schwarz-Linek *et al.*, 2003). The major goal of this thesis is to characterize this interaction (Section 1.9). NMR spectroscopy is an ideal technique with which to confirm and characterize the intrinsically disordered state of SfbI-5 and gain insight into the disorder-to-order transition that SfbI-5 undergoes on complex formation (Sections 1.5, 1.6). Therefore, as described below, SfbI-5 uniformly labelled with ^{15}N and ^{13}C has been produced and purified (Section 3.2); the amide proton (^{1}H^{N}), amide nitrogen (^{15}N^{H}), alpha, beta and carbonyl carbons (^{13}C^{α}, ^{13}C^{β} and ^{13}C′, respectively) resonances have been assigned (Section 3.3.2); and progress has been made towards characterization of the disordered state of SfbI-5 by NMR (Section 3.3.4). Furthermore, before embarking on *in vitro* experiments the disordered prediction of SfbI-5 has been verified by *in silico* sequence analysis of SfbI, as described in Section 3.1.

3.1   *In silico* sequence analysis predicts SfbI is an IDP

In silico sequence analysis of SfbI revealed that all the FnBR repeats in SfbI, including SfbI-5, are predicted to be disordered. The full-length sequence of both SfbI and SfbI-5 were analysed with three commonly-used disorder predictors: IUPred, VSL2 and VLXT (Table 25; reviewed in Dosztanyi & Tompa, 2008). Each of the three predictors predicts disorder using different techniques (Table 25), therefore regions where the predictors agree are more likely to be correctly predicted (Dosztanyi & Tompa, 2008). All methods output the probability of each amino acid being in a disordered region, with probability 0 for fully ordered, and probability 1 for fully disordered. Any region with a probability over 0.5 is called disordered. Based on the three predictions, it appears the N-terminal third of SfbI is folded and the remaining two-thirds is disordered except for the very C-terminus (Figure 15A). This can be rationalized based on the domain organization of SfbI (Figure 15A). Whether the N-terminal third of the protein contains a folded domain has not been experimentally determined, but a construct corresponding to the most N-terminal 279 residues of SfbI has been implicated in binding to fibrinogen (Katerov *et al.*, 1998), indicating this N-terminal region can function as a discrete unit and that it is feasible that this region could be folded. Although the exact N-terminal boundary of the disordered region is unclear from these predictions (Figure 15A), it is

clear that the disordered region encompasses both the proline rich repeat region and all five FnBRs. Furthermore, the C-terminus is probably predicted as ordered because it contains the Gram positive anchor, which includes a hydrophobic membrane-spanning region (Figure 15A). Also, when the SfbI-5 sequence is considered in isolation, it is still predicted to be mostly disordered (Figure 15B). Interestingly, both IUPred and VLXT predict two regions of relative order, one within the UR and first FnBR, and one within SfbI-5 (Figure 15). These are possibly molecular recognition features (MoRFs; Vacic *et al.*, 2007) and will be discussed further in Section 3.5.2. Therefore, *in silico* sequence analysis predicts all FnBRs in SfbI are disordered, including SfbI-5.

Table 25. Predictors used to predict the disorder of both full-length SfbI and SfbI-5.

Predictor	Reference(s)	Comments
IUPred	Dosztanyi *et al.*, 2005a; 2005b	Predicts disordered regions on the basis of amino acid composition. It was built from a database of folded proteins, from which the number of favourable interactions each residue type makes with each other residue type was calculated. This is used to estimate a pairwise interaction energy for each residue by considering its potential to interact with residues 2–100 positions in either direction.
VSL2	Obradovic *et al.*, 2005; Peng *et al.*, 2006	Predicts both short (4–30 residues) and long (>30 residues) disordered regions by combining a short disorder predictor with a long disorder predictor into a 'meta' predictor using a linear support vector machine. VSL2 combines information on a wide range of sequence features, including amino acid composition, position-specific evolutionary information from a PSI-BLAST search and also secondary structure predictions.
VLXT	Romero *et al.*, 2001	Combines three predictors: one built for detecting long-disordered regions, and one each for the N- and C-termini. It looks at a range of sequence attributes, which were optimized for each of the three predictors, including hydropathy, net charge, and amino acid composition. It has been found that VLXT is particularly good at detecting regions in IDPs that become ordered on complex formation (Mohan *et al.*, 2006).
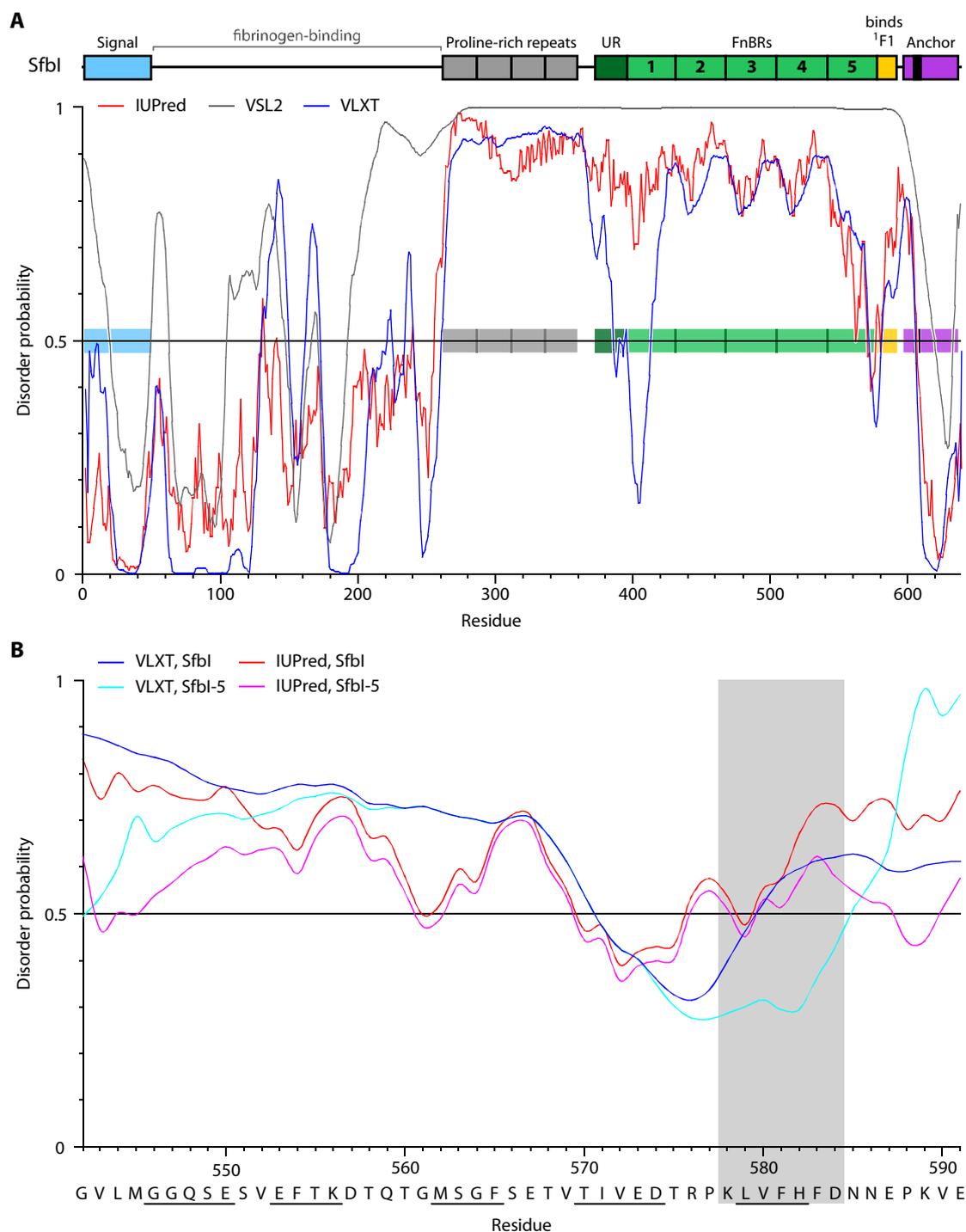
Figure 15. Disorder prediction of SfbI and SfbI-5. A. The IUPred, VSL2 and VLXT predictions of the disorder of SfbI are shown in grey, blue and red, respectively. Regions predicted to be disordered have a disorder probability of greater than 0.5. The domains of SfbI reported by Schwarz-Linek *et al.* (2006) are shown in the context of the predictions and also above the disorder plot with domains labelled (see Figure 3 for domain definitions). **B.** Disorder prediction of SfbI-5 by IUPred and VLXT made either in the context of full-length SfbI sequence (red and blue, respectively) or for the SfbI-5 sequence in isolation (magenta and cyan, respectively). Residues predicted to form β-strands on complex formation with [1-5]F1 based on comparison with known Fn-module:FnBP structures are underlined. The grey boxed region is predicted to have significant secondary structure propensity (Section 3.3.4.2; Figures 23, 24).

3.2   Production of ^{15}N, ^{13}C-labelled SfbI-5

^{15}N, ^{13}C-labelled SfbI-5 was expressed as a GST-fusion in *E. coli* and purified by glutathione affinity chromatography (Figure 16; Sections 2.3.1, 2.3.4.1). SDS-PAGE analysis of *E. coli* samples, taken before and after induction of GST-SfbI-5 overexpression, revealed that a protein that migrated with a MW corresponding to GST-SfbI-5 had been overexpressed, as had a contaminant that migrated with a MW between that of GST-SfbI-5 and GST (lanes −I, + I, Figure 16). Samples from lysis of the *E. coli* cells to separate the soluble cellular protein from the insoluble material showed that greater than 95% of GST-SfbI-5 remained in the soluble fraction (lanes T, S and I, Figure 16). Next, GST-SfbI-5 was purified from the soluble cellular protein and nearly all of the contaminant by glutathione affinity chromatography (*e.g.*, lanes E3.1, E3.2 and E3.3 in Figure 16). This indicates the overexpressed contaminant does not contain functional GST, either because it is misfolded or because it is a C-terminal truncation of GST-SfbI-5. Also, SDS PAGE analysis revealed that a significant amount of GST-SfbI-5 did not bind to the column, and remained in the flow-through from the first affinity chromatography step (lane F1, Figure 16). Therefore, the affinity chromatography was repeated a further two times, each time using the flow-through from the previous chromatography step, until the amount of GST-SfbI-5 left in the flow-through was similar to the amount of other cellular proteins (lane F3, Figure 16). These additional chromatography steps were worthwhile because they more than doubled the yield of GST-SfbI-5.

After dialysing the pooled GST-SfbI-5 fractions from affinity chromatography, the GST was removed from SfbI-5 by cleavage with Factor Xa. SDS-PAGE analysis of samples from a Factor Xa cleavage trial of ^{15}N, ^{13}C-labelled SfbI-5 revealed that cleavage was 95% complete after 16 h at a ratio of 10 μg Factor Xa per 5 mg GST-SfbI-5 (Figure 17; Section 2.3.7). Therefore, these conditions were used for all cleavage reactions of ^{15}N, ^{13}C-labelled GST-SfbI-5.

To separate SfbI-5 from GST and Factor Xa the cleavage reaction was then purified by rpHPLC in acidic conditions (Figure 18; Section 2.3.8). The pooled SfbI-5 fractions (Figure 18) were lyophilized and the yield determined by weight (Section 2.3.9). The yield was 5.3 mg, assuming 20% of the weight of the sample was water. ESMS of the pooled sample of ^{15}N, ^{13}C-labelled SfbI-5 revealed that it was uniformly labelled, with greater than 99.5% incorporation. (Experimental and theoretical MWs are 6098 Da and

6102.2 Da, respectively.) Furthermore, the sample was pure, as determined by SDS PAGE analysis (lane 'Final', Figure 17B). This sample was used to prepare an NMR sample (1 mM, pH 5.5, 10% D_2O, 0.02% NaN_3) which was used for three-dimensional NMR experiments to assign $^1H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C'$ resonances of SfbI-5 (Section 3.3).
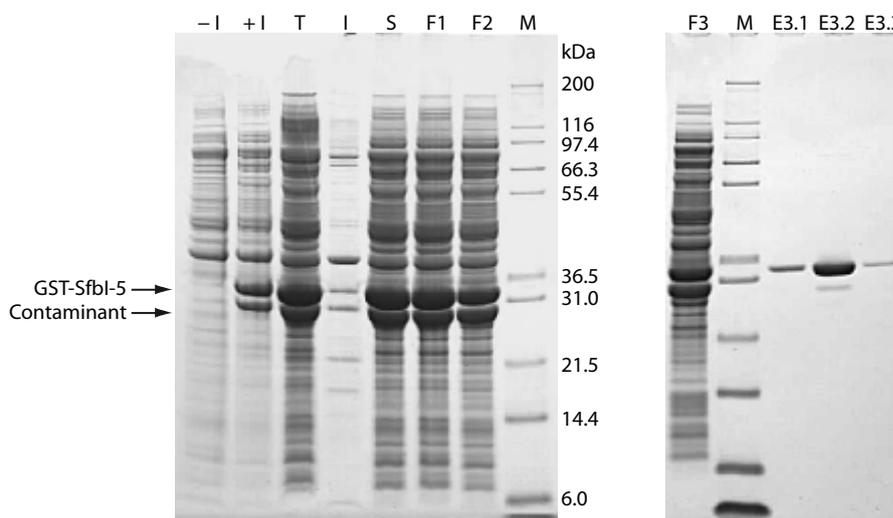
Figure 16. Analysis of samples from expression and glutathione affinity purification of ^{15}N, ^{13}C-labelled GST-SfbI-5. SDS-PAGE lane contents are indicated where abbreviations are: −I and +I, *E. coli* cells before and after induction of GST-SfbI-5 expression, respectively, normalised to the same cell density; T, total cellular protein; I, insoluble cellular protein; S, soluble cellular protein; F1–F3, flow-through after glutathione affinity purification of S, F1 and F2, respectively; E3.1–E3.3, three sequential eluate fractions from the glutathione affinity purification of F2, which was the third affinity step; M, Mark 12 protein MW marker. See Table 9 for full description of the samples. The positions of GST-SfbI-5 and an overexpressed contaminant are indicated on the gel on the left, as are the MWs of the marker. These bands appear in similar positions on the gel on the right.
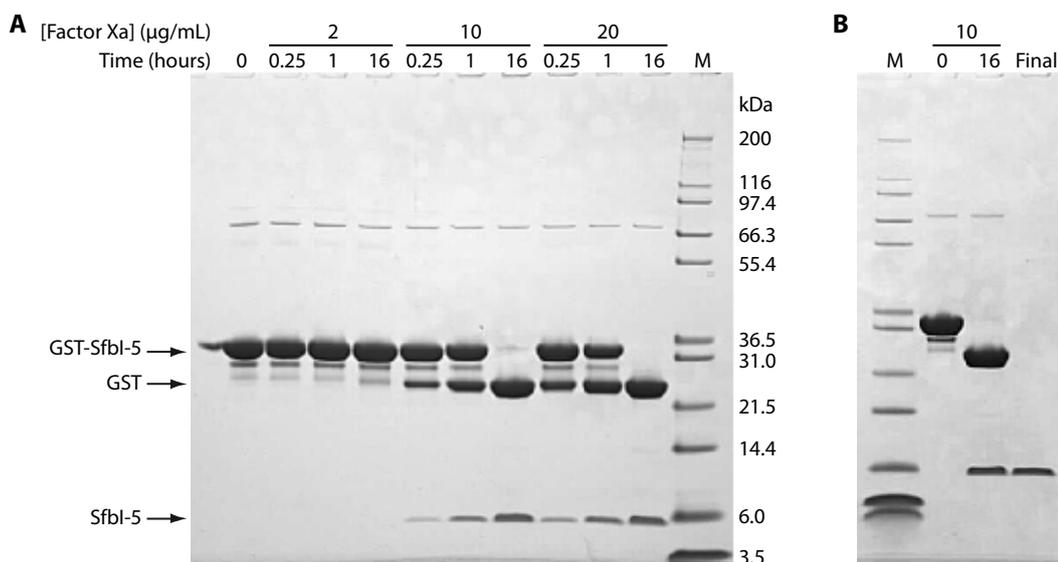
Figure 17. Analysis of ^{15}N, ^{13}C-labelled SfbI-5 samples from a cleavage trial and final cleavage with Factor Xa, and the sample used for the NMR spectroscopy. A. Factor Xa cleavage trial of ^{15}N, ^{13}C-labelled GST-SfbI-5. For each lane, the concentration of the Factor Xa in the sample and the cleavage time are shown. The concentration of GST-SfbI-5 in all samples was approximately 5 mg/mL as determined by the A$_{280}$. The MWs of the Mark 12 protein MW marker (M) are indicated, as are the positions of GST-SfbI-5, GST and SfbI-5. **B.** Samples from the Factor Xa cleavage of all of ^{15}N, ^{13}C-labelled GST-SfbI-5 at the concentration and after the time indicated. The 'Final' lane is a sample of the ^{15}N, ^{13}C-labelled SfbI-5 NMR sample. Markers, GST-SfbI-5, GST and SfbI-5 appear in similar positions as shown in the gel in A.
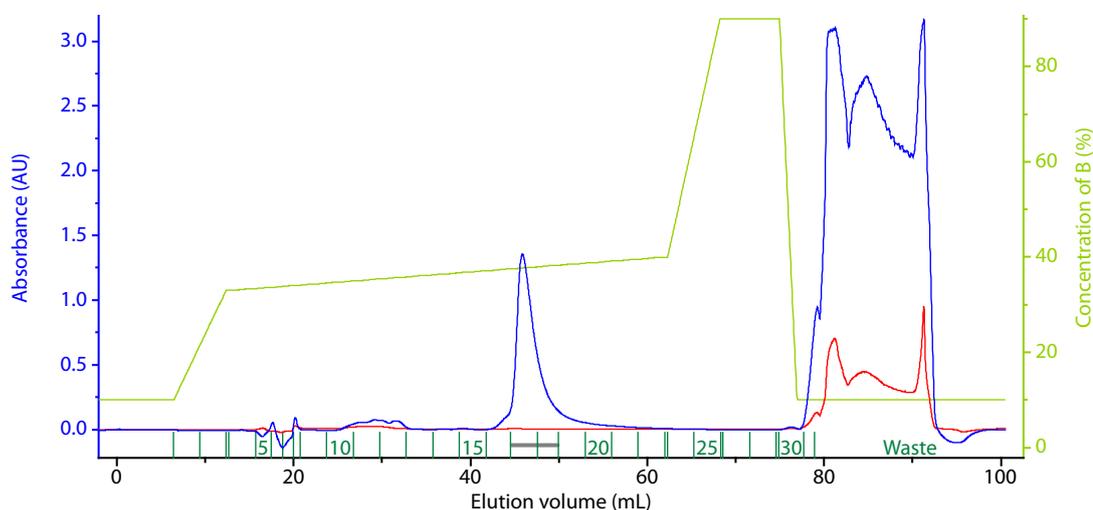
Figure 18. Acidic rpHPLC purification of cleaved ^{15}N, ^{13}C-labelled GST-SfbI-5. Shown is a representative chromatogram from rpHPLC of the Factor Xa cleavage reaction mixture of ^{15}N, ^{13}C-labelled GST-SfbI-5. The 800 µL sample was purified under acidic conditions on a C4 column with a gradient between 0.15% TFA in water and 0.1% TFA in acetonitrile (solvents A and B, respectively). Trace colours are: light green, concentration of solvent B; blue and red, absorbance at 220 and 280 nm, respectively; dark green, fractions collected during the run. The grey line indicates two fractions of SfbI-5 collected for lyophilizing; equivalent fractions were collected from other rpHPLC runs.

3.3   Characterization of ^{15}N, ^{13}C-labelled SfbI-5 by NMR spectroscopy

All NMR experiments with ^{15}N, ^{13}C-labelled SfbI-5 were acquired on a Bruker 700 MHz spectrometer at 298 K and processed using NMRPipe (Delaglio *et al.*, 1995). During processing the signal to noise ratio and resolution of the spectra were enhanced with functions summarized in Section 2.4.3. All spectra were referenced directly to DSS in the ^{1}H dimension and indirectly in the ^{15}N and ^{13}C dimensions (Section 2.4.3) and then analysed with CCPN analysis software (Vranken *et al.*, 2005).

3.3.1   NMR spectroscopy confirms SfbI-5 is an intrinsically disordered protein (IDP)

A heteronuclear single quantum correlation (HSQC) spectrum of ^{15}N, ^{13}C-labelled SfbI-5 confirmed that SfbI-5 is an IDP (Figure 19). This experiment shows correlations between an amide's nitrogen and hydrogen(s) chemical shifts. Thus, for every residue in a protein, except for prolines, there is usually a single peak in the HSQC spectrum[12]. The peaks in the SfbI-5 HSQC spectrum show limited chemical shift dispersion, particularly in the proton dimension (Figure 19), which is indicative of a lack of stable secondary and tertiary structure elements (Section 1.6.1.2; Dyson & Wright, 2001; Penkett *et al.*, 1998; Rehm *et al.*, 2002). Furthermore, at ~12 Hz the line-widths of the peaks at half their height in the ^{1}H dimension are relatively narrow compared to those of a folded protein. For example, ^{2}F1^{3}F1 has line-widths in the ^{1}H dimension of ~15 Hz. The line-widths are inversely proportional to the transverse relaxation time, T_2, which is generally longer for a disordered protein than an ordered protein of a similar MW (Dyson & Wright, 2001). Therefore, SfbI-5 appears to be intrinsically disordered, as indicated by both low dispersity and narrow line-widths of peaks in its HSQC spectrum.

---

[12] There are also peaks from sidechain amides, such as glutamine.
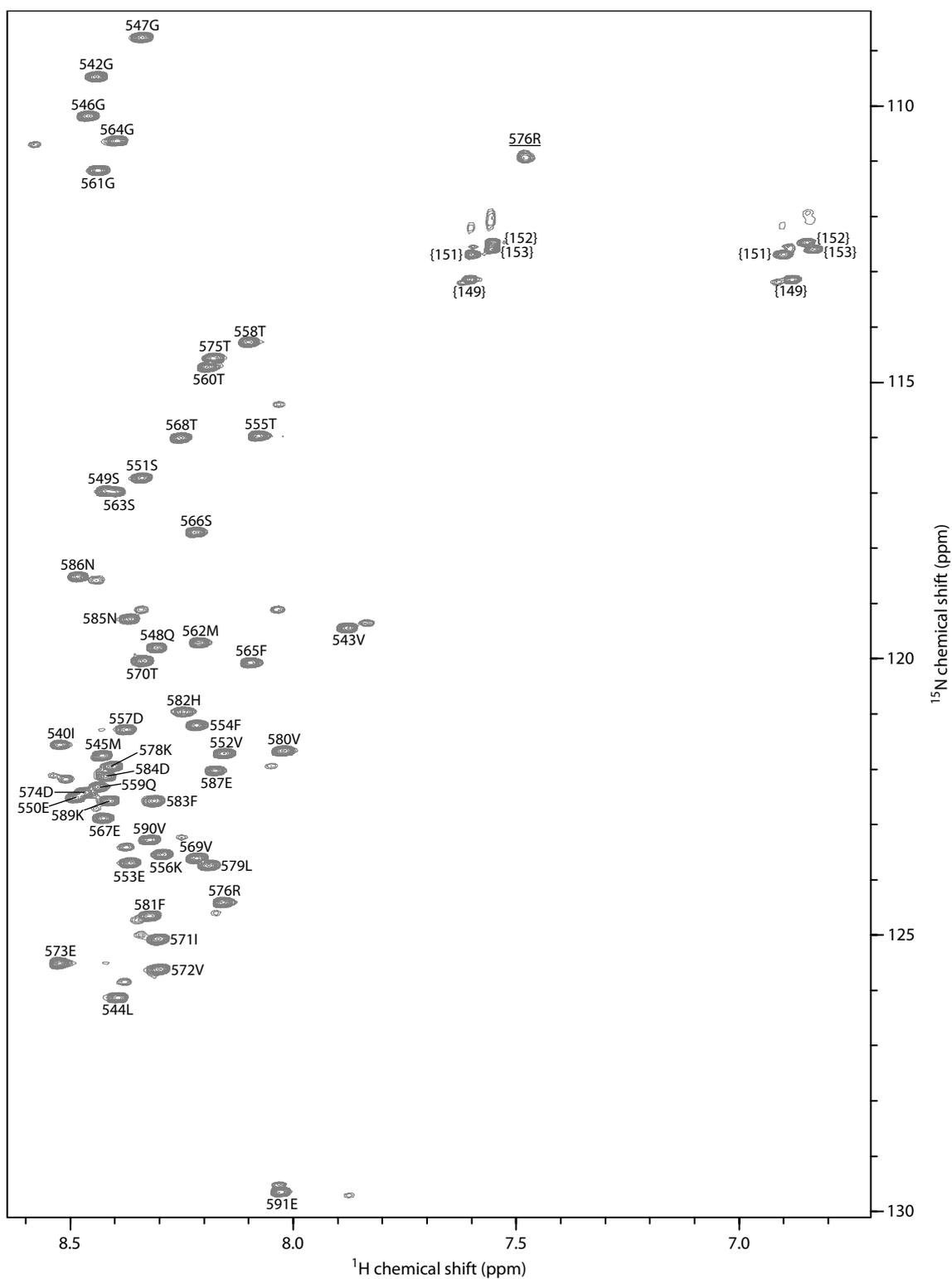
Figure 19. The assigned HSQC spectrum of ^{15}N, ^{13}C-labelled SfbI-5. The ^{1}H, ^{15}N HSQC spectrum of ^{15}N, ^{13}C-labelled SfbI-5 sample (1 mM, pH 5.5, 10% D$_2$O, 0.02% NaN$_3$) was acquired on a Bruker 700 MHz spectrometer at 298 K. ^{13}C decoupling was achieved with a 180° pulse. The peaks in the spectra were assigned to the ^{1}H^N and ^{15}N^H resonances of SfbI-5 residues. The assignment procedure is described in Section 3.3.2.2. Each peak is labelled according to its position in the full-length SfbI-5 sequence. The peak annotated I(540) is not native to full-length SfbI, but is directly N-terminal to P541 from native SfbI, and the peak annotated R576 is from the sidechain of R576. Peaks annotated with a number in curly brackets arise from the sidechains of glutamine and asparagine residues and were not unambiguously

assigned. The other smaller, unassigned peaks appear to be from residues on either side of a proline, and probably result from the proline being in a *cis*-isomer as opposed to the more abundant *trans*-isomer.

3.3.2 Assignment of ^{1}H^N, ^{15}N^H, ^{13}C$^\alpha$, ^{13}C$^\beta$ and ^{13}C′ resonances of ^{15}N, ^{13}C-labelled SfbI-5

3.3.2.1 HNCO/HN(CA)CO and CBCA(CO)NH/CBCANH spectra of ^{15}N, ^{13}C-labelled SfbI-5 were acquired

To determine the secondary structural propensity of SfbI-5 by chemical shift analysis it was first necessary to assign SfbI-5 resonances with respect to its primary sequence using three-dimensional (3D) triple-resonance NMR experiments. When choosing these experiments, the typical chemical shift dispersion of the resonances involved was considered. The chemical shifts of ^{13}C′ and ^{15}N^H resonances are dependent on the neighbouring primary sequence as well as on nearby secondary and tertiary structure, so are dispersed in both folded and disordered proteins. This contrasts with ^{13}C$^\alpha$ and ^{13}C$^\beta$ resonances, which are mostly dependent on structure, and hence, while important for prediction of secondary structure propensities, are less dispersed in disordered proteins (Schwarzinger *et al.*, 2001; Yao *et al.*, 1997). Therefore, two pairs of three-dimensional triple-resonance NMR spectra with ^{15}N, ^{13}C-labelled SfbI-5 were acquired: HNCO/HN(CA)CO and CBCA(CO)NH/CBCANH.

The HNCO experiment correlated the ^{1}H^N and ^{15}N^H resonances of residue i with the ^{13}C′ resonance of the preceding residue $i-1$ (Figure 20). The HN(CA)CO experiment provides complementary connectivity information: it correlates the ^{1}H^N and ^{15}N^H resonances of residue i with the C′ resonances of both residues i and $i-1$ (Figure 20). Similarly, the CBCA(CO)NH and CBCANH experiments correlated the ^{1}H^N and ^{15}N^H resonances of residue i with the ^{13}C$^\alpha$ and ^{13}C$^\beta$ resonances of either residue $i-1$ or both residues i and $i-1$, respectively (Figure 20). The CBCANH experiment had the added property that the phasing of the ^{13}C$^\alpha$ and ^{13}C$^\beta$ peaks are negative and positive, respectively, with the exception of glycine ^{13}C$^\alpha$ peaks, which are positive. Therefore, the CBCA(CO)NH and CBCANH experiments gave information on residue connectivity and allowed assignment of the ^{13}C$^\alpha$ and ^{13}C$^\beta$ resonances, which are more useful in determining secondary structural propensities. Furthermore, the phase properties of the CBCANH experiment helped to reduce ambiguity in the spectra. However, when spectral overlap led to difficulties in residue assignment, the greater

chemical shift dispersion of $^{13}C'$ compared to $^{13}C^\alpha$ and $^{13}C^\beta$ in IDPs was exploited by analysis of the HNCO/HN(CA)CO spectra. It was also found that when one pair of spectra were ambiguous the other pair were not[13]. Therefore, these experiments allowed complete sequence-specific assignment of the $^{1}H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C'$ resonances of SfbI-5. The assignment process is described in Section 3.3.2.2 and the assigned 3D spectra and chemical shifts are reported in Section 3.3.2.3.
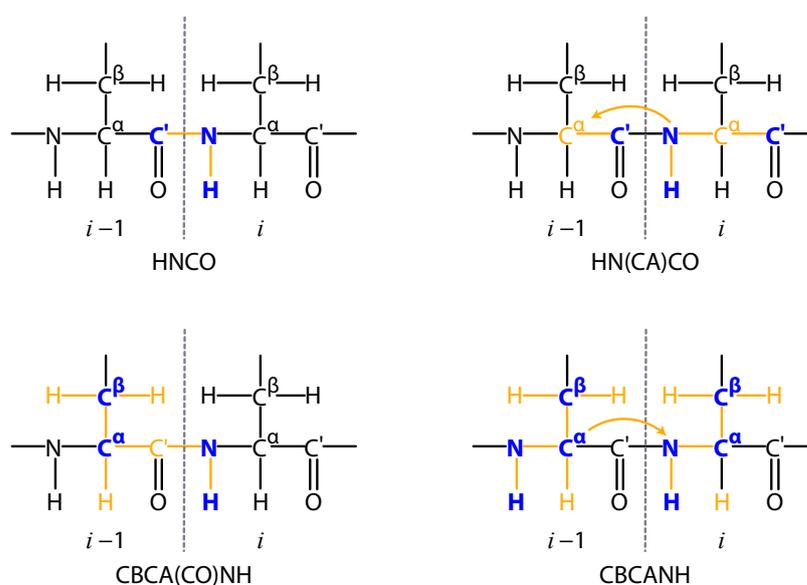
Figure 20. Schematics of the series of three-dimensional NMR experiments acquired on ^{15}N, ^{13}C-labelled SfbI-5. The name of the experiment is shown underneath its schematic. For each experiment, magnetization is transferred between nuclei, the chemical shift of some of these nuclei are detected in the recorded spectra. Detected nuclei are shown in blue, nuclei used in the transfer but not detected are yellow and the scalar couplings used in magnetization transfer are shown by yellow highlighted bonds or arrows (adapted from Table 7.3 from Cavanagh *et al.*, 1995).

3.3.2.2   Procedure for SfbI-5 resonance assignment

The assignments of the $^{1}H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C'$ resonances of SfbI-5 were achieved as follows. First, an HSQC peak was used to navigate to a corresponding strip of peaks in the three-dimensional spectra at the same $^{1}H^N$ and $^{15}N^H$ chemical shifts. When all the three-dimensional spectra were overlaid, peaks on each $^{1}H^N$, $^{15}N^H$ strip correspond to the inter-residue correlation between the $^{1}H^N$ and $^{15}N^H$ resonances of a residue and the $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C'$ resonances of the preceding residue; and also the

---

[13] For example, the peaks from V552 and E553 are ambiguous in the HNCO/HN(CA)CO experiments but not in the CBCA(CO)NH/CBCANH experiments (Figure 21, p 114).

intra-residue correlations between the $^1H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C'$ resonances. Inter-residue peaks were distinguished from intra-residue peaks by the presence of two overlaid peaks at the same $^1H^N$, $^{15}N^H$ and ^{13}C chemical shifts; one from each of the spectrum in a pair (Figure 21). Furthermore, the $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts were distinguished from each other by their difference in phase, as previously mentioned. Therefore, peaks in a strip were assigned as coming from the same or an adjacent 'spin system' (*i.e.*, residue) and also by resonance type. For example, a single peak at 8.21, 119.7 and 176.5 ppm was assigned as the intra-residue correlation between $^1H^N$, $^{15}N^H$ and $^{13}C'$ respectively from a single spin system (Figure 21). Second, the preceding spin system was identified by searching for a new $^1H^N$, $^{15}N^H$ strip of peaks that had inter-residue $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C'$ peaks at the same ^{13}C chemical shifts as corresponding intra-residue peaks in the current spin system (Figure 21). This process was continued until a series of connected spin systems was established, corresponding to sequential residues. A series stopped at either the N- or C-terminus, or at a proline; because a proline does not contain $^1H^N$, it does not have a $^1H^N$, $^{15}N^H$ strip. However, the residue type and number for each of the spin systems were unknown at this stage.

The residue types for some spin systems could be assigned based on characteristic $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts, and then these residues acted as 'anchors' from which the rest of the sequence could be assigned. Most spin systems have $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts in the range of approximately 50–65 ppm and 30–45 ppm, respectively. However, glycines, in addition to having a $^{15}N^H$ chemical shift of approximately 110 ppm (Figure 19), only have one intra-residue peak in the CBCANH spectra, a negative peak from the $^{13}C^\alpha$ resonance at a ^{13}C chemical shift of about 45 ppm (Figure 21). Also, threonines and serines have unusual $^{13}C^\beta$ chemical shifts of approximately 63 and 69 ppm, respectively (Figure 21). This allowed some spin systems to be assigned to residue types. Based on the distribution of these known, 'anchor' spin systems it was then possible to make sequence-specific assignments of the other spin systems based on the connectivities already established. For example, two sequential spin systems were identified as corresponding to the sequence Thr-Gly based on chemical shifts. This sequence occurs only once in SfbI-5, and immediately allowed the sequence to be assigned as T560 and G561 (Figure 21). Furthermore, the adjacent spin system C-terminal to G561 has to be M562 despite this spin system not having unusual $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts, because it was connected to G561 (Figure 21). Subsequently, this assignment was confirmed because, if correct, the next spin system would have resonances characteristic

of a serie, S563, which it did. In this way, sequence-specific assignments were made for the $^1H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C'$ resonances of SfbI-5.
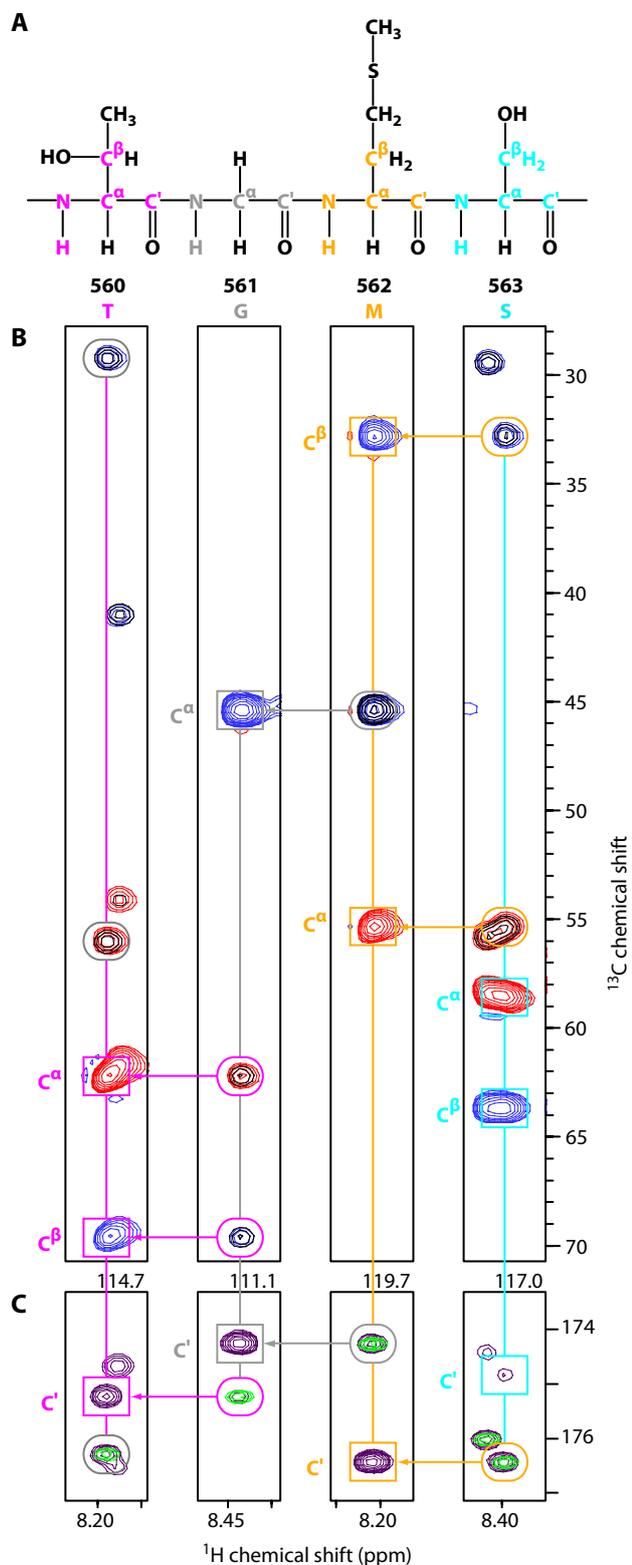
Figure 21. An example of sequence-specific, sequential assignment of SfbI-5 residues with the HNCO/HN(CA)CO, and CBCA(CO)NH/CBCANH spectra. Legend on p 128.

Figure 21. An example of sequence-specific, sequential assignment of SfbI-5 residues with the HNCO/HN(CA)CO, and CBCA(CO)NH/CBCANH spectra. The 3D heteronuclear experiments of ^{15}N, ^{13}C-labelled SfbI-5 (1 mM, pH 5.5, 10% D$_2$O, 0.02% NaN$_3$) were performed on a Bruker 700 MHz spectrometer at 298 K. **A.** A schematic of an SfbI-5 peptide fragment with sequence TGMS is shown, with the nuclei that are detected in the 3D experiments highlighted in magenta, grey, dark yellow and cyan, for residues T, G, M and S, respectively. As indicated, this sequence corresponds to residues 560–563 in SfbI-5. **B.** Four 'strips' of the CBCA(CO)NH and CBCANH spectra. Each strip is centred on the ^{1}H^N and ^{15}N^H chemical shifts of the residue shown directly above in A. The CBCA(CO)NH peaks are black, the CBCANH positive and negative peaks are blue and red, respectively. **C.** Four strips of the HNCO and HN(CA)CO spectra at the same ^{1}H^N and ^{15}N^H chemical shifts as in B. That is, these strips are also centred on the ^{1}H^N and ^{15}N^H chemical shifts of the residue shown directly above in A. The HNCO peaks are light green and the HN(CA)CO peaks are dark purple. The ^{15}N^H chemical shifts on which each strip is centred in both B and C is shown above the strips in C.

For both B and C, the ^{1}H^N, ^{15}N^H chemical shift for each residue is shown as a line, with the same colouring as in A. On each strip, intra-residue peaks resulting from the correlations between that residue's ^{1}H^N and ^{15}N^H resonances and its ^{13}C$^\alpha$, ^{13}C$^\beta$ and ^{13}C′ resonances are indicated by rectangles of the same colour, and are also labelled by resonance type to the left of each strip. The inter-residue peaks resulting from the correlation between that residue's ^{1}H^N and ^{15}N^H resonances and the preceding residues' ^{13}C$^\alpha$, ^{13}C$^\beta$ and ^{13}C′ resonances are indicated by ellipses, and coloured with the preceding residues' colour. The arrows link peaks arising from correlations of the same ^{13}C$^\alpha$, ^{13}C$^\beta$ and C′ resonances, thus showing how the inter-residue connectivity was established with these spectra. It is this connectivity that allows the sequential assignments to be made.

3.3.2.3 ^{1}H^N, ^{15}N^H, ^{13}C$^\alpha$, ^{13}C$^\beta$, and ^{13}C′ resonance assignment of SfbI-5

All SfbI-5 residues except for the three prolines and N-terminal glycine have been assigned chemical shifts for their ^{1}H^N, ^{15}N^H, ^{13}C$^\alpha$, ^{13}C$^\beta$ and ^{13}C′ resonances. For each proline, the ^{13}C$^\alpha$, ^{13}C$^\beta$ and ^{13}C′ resonances were assigned but not the amide nitrogen, because proline does not contain an amide proton, the amide nitrogen was not detected in this set of experiments. Additionally, only the ^{13}C$^\alpha$ and ^{13}C′ resonance of the N-terminal glycine were assigned and not the ^{15}N^H or ^{1}H^N resonances. This is probably because the amide protons of the glycine are labile and in exchange with the solvent. The assigned HSQC and 3D heteronuclear experiments are shown in Figure 19 and Figure 22, respectively. The chemical shifts are reported in Table 26, and will be analysed further in Section 3.3.4.
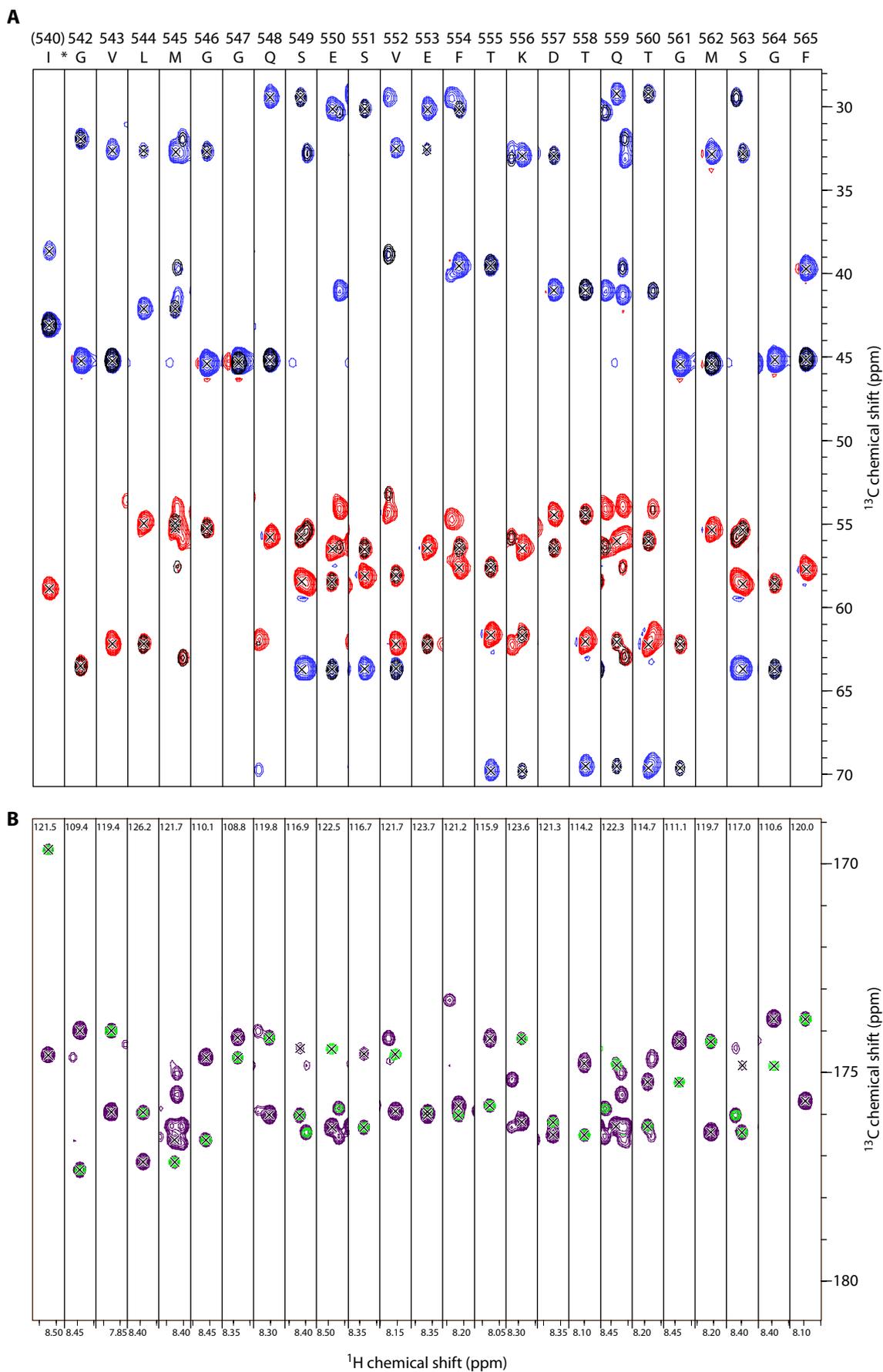
Figure 22. Series of strips from the CBCA(CO)NH/CBCANH and HNCO/HN(CA)CO spectra of
[15]N, ^{13}C-labelled SfbI-5 (Part 1 of 2). Legend on p 131.
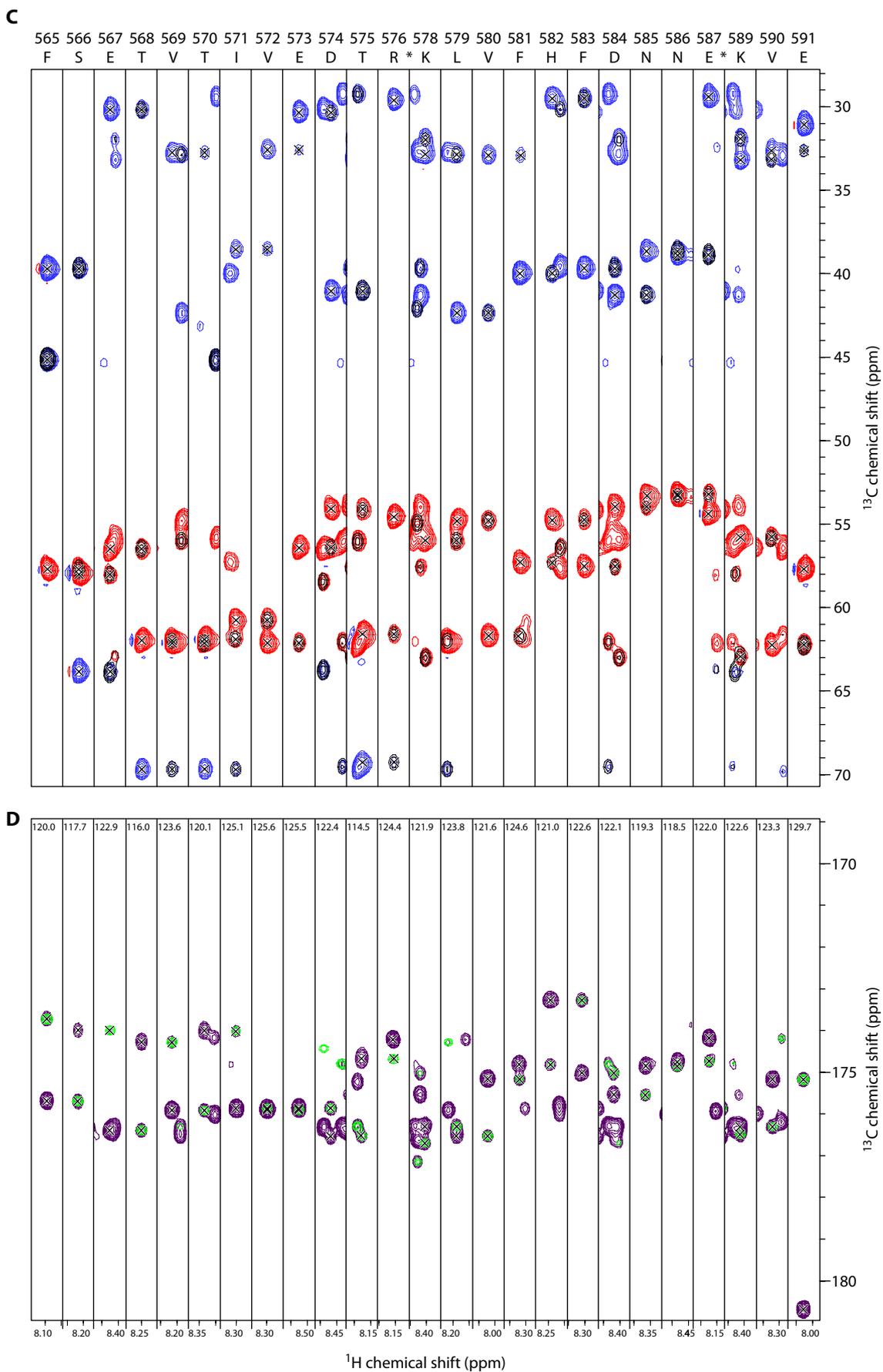
Figure 22. (Part 2 of 2). Legend on p 131.

Figure 22. Series of strips from the CBCA(CO)NH/CBCANH and HNCO/HN(CA)CO spectra of
^{15}N, ^{13}C-labelled SfbI-5. Shown are strips of the 3D spectra centred on the $^1H^N$ and $^{15}N^H$ chemical shifts for each residue in SfbI-5. The residue type and number each strip represents is shown above each strip, where I(540) is not native to full-length SfbI, but is directly N-terminal to P541 from native SfbI. SfbI-5 residues 541, 577 and 588 are prolines, and hence did not have a $^1H^N$ resonance. Each of their positions in the SfbI-5 sequence is indicated by an asterisk (*). **A.** CBCA(CO)NH and CBCANH strips for SfbI-5 residues 540–565. CBCA(CO)NH peaks are black, CBCANH positive and negative peaks are blue and red, respectively. **B.** HNCO and HN(CA)CO strips for SfbI-5 residues 540–565. HNCO peaks are green and HN(CA)CO peaks are dark purple. **C.** and **D.** As for A and B, respectively, except for SfbI-5 residues 565–591. In each strip, black crosses mark the centres of peaks assigned to the $^1H^N$ and $^{15}N^H$ resonances of the indicated SfbI-5 residue and to either intra or inter-residue $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C'$ resonances, depending on the spectrum. The $^{15}N^H$ chemical shift of each residue's strip is shown at the top of the strip in B and D.

Table 26. Chemical shifts of the $^1H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$ and C′ resonances of SfbI-5 at pH 5.5, and 298 K. Shifts are shown in ppm relative to DSS (Section 2.4.3). Residue numbering is shown relative to full-length SfbI. G(539) and I(540) are not native to SfbI, but are directly N-terminal to the native residue P541.

Number	Residue	$^1H^N$	$^{15}N^H$	$^{13}C^\alpha$	$^{13}C^\beta$	$^{13}C'$
(539)	G	–	–	43.17	–	169.75
(540)	I	8.523	121.56	58.97	38.75	174.68
541	P	–	–	63.62	32.03	177.43
542	G	8.442	109.47	45.31	–	174.09
543	V	7.877	119.44	62.27	32.72	176.05
544	L	8.392	126.13	55.03	42.20	177.24
545	M	8.428	121.75	55.36	32.78	176.71
546	G	8.460	110.18	45.47	–	174.74
547	G	8.339	108.76	45.29	–	174.26
548	Q	8.305	119.80	55.88	29.54	176.11
549	S	8.417	116.97	58.55	63.83	174.52
550	E	8.488	122.51	56.58	30.23	176.41
551	S	8.339	116.73	58.18	63.78	174.65
552	V	8.155	121.72	62.29	32.64	176.03
553	E	8.365	123.69	56.51	30.25	176.12
554	F	8.217	121.21	57.70	39.60	175.89
555	T	8.078	115.97	61.75	69.89	174.28
556	K	8.293	123.54	56.54	33.03	176.28
557	D	8.374	121.28	54.54	41.08	176.59

Number	Residue	$^1\text{H}^\text{N}$	$^{15}\text{N}^\text{H}$	$^{13}\text{C}^\alpha$	$^{13}\text{C}^\beta$	$^{13}\text{C}'$
558	T	8.100	114.27	62.16	69.60	174.89
559	Q	8.435	122.32	56.12	29.33	176.38
560	T	8.191	114.73	62.31	69.72	175.32
561	G	8.437	111.17	45.49	–	174.35
562	M	8.210	119.71	55.44	32.90	176.53
563	S	8.399	116.99	58.68	63.78	174.93
564	G	8.395	110.63	45.24	–	173.81
565	F	8.095	120.07	57.79	39.79	175.78
566	S	8.218	117.71	58.10	63.93	174.09
567	E	8.425	122.89	56.56	30.27	176.49
568	T	8.252	116.00	62.07	69.78	174.37
569	V	8.217	123.61	62.22	32.87	176.01
570	T	8.337	120.04	61.99	69.78	174.11
571	I	8.302	125.07	60.85	38.64	175.96
572	V	8.301	125.62	62.21	32.68	176.00
573	E	8.524	125.51	56.49	30.44	175.95
574	D	8.469	122.42	54.20	41.13	176.62
575	T	8.179	114.57	61.67	69.36	174.77
576	R	8.157	124.41	54.66	29.70	174.31
577	P	–	–	63.11	32.06	176.79
578	K	8.406	121.95	56.06	32.94	176.39
579	L	8.190	123.74	54.89	42.44	176.62
580	V	8.022	121.67	61.79	33.01	175.26
581	F	8.322	124.65	57.40	40.07	174.91
582	H	8.246	120.96	54.84	29.60	173.37
583	F	8.312	122.57	57.63	39.76	175.11
584	D	8.419	122.13	54.06	41.36	175.63
585	N	8.366	119.28	53.39	38.74	174.94
586	N	8.484	118.52	53.29	38.97	174.83
587	E	8.175	122.03	54.47	29.49	174.27
588	P	–	–	63.03	32.01	176.58
589	K	8.413	122.57	55.89	33.25	176.40
590	V	8.322	123.28	62.33	32.72	175.26
591	E	8.029	129.65	57.79	31.16	180.77

3.3.3   The secondary chemical shifts ($\Delta\delta$) of SfbI-5

The secondary chemical shifts ($\Delta\delta$) of SfbI-5 were calculated by subtracting sequence-corrected random coil chemical shifts from the corresponding experimentally observed chemical shifts (Table 26) as explained in Section 2.4.5.2. Two sets of $\Delta\delta$ values using two datasets of random coil shifts and sequence correction factors were calculated. The first dataset, WS, was experimentally determined by Wishart $et\ al.$ (1995) from Acetyl-Gly-Gly-X-Ala-Gly-Gly-NH$_2$ and Acetyl-Gly-Gly-X-Pro-Gly-Gly-NH$_2$ peptides, where X is each of the 20 naturally-occurring amino acids. The WS dataset was recorded in 1 M urea at pH 5 and at 298 K and referenced to DSS. To our knowledge there is only one other experimentally determined random coil shift dataset that also has corresponding sequence corrections for (at minimum) the ^{1}H^N, ^{15}N^H, ^{13}C$^\alpha$ and ^{13}C$'$ residues: that of Schwarzinger $et\ al.$ (2000; 2001) measured in 8 M urea at pH 2.3 and 293 K with Acetyl-Gly-Gly-X-Gly-Gly-NH$_2$ peptides. The WS dataset was chosen in preference to this dataset because it was much closer to the experimental conditions of water pH 5.5 at 298 K.

The other random coil dataset used for calculations of $\Delta\delta$ values, WJ2, was statistically-derived by Wang and Jardetzky (2002b) from a chemical shift database of 415 assigned proteins, where all residues were first categorized as being either random coil, α-helix or β-strand based either on their structures (326 proteins) or a probability-based statistical analysis of their chemical shifts (89 proteins) using the PSSI program developed from analysis of an earlier chemical shift database (Wang & Jardetzky, 2002a). Hence, the WJ2 dataset also contained accompanying chemical shifts for all residues in β-strand and α-helix. The $\Delta\delta$ values calculated from these experimental and statistical datasets are shown in Figure 23, and will be referred to as WS $\Delta\delta$ and WJ2 $\Delta\delta$ values, respectively. These $\Delta\delta$ values are useful for analysing the secondary structure propensity of SfbI-5, as discussed in Section 3.3.4.
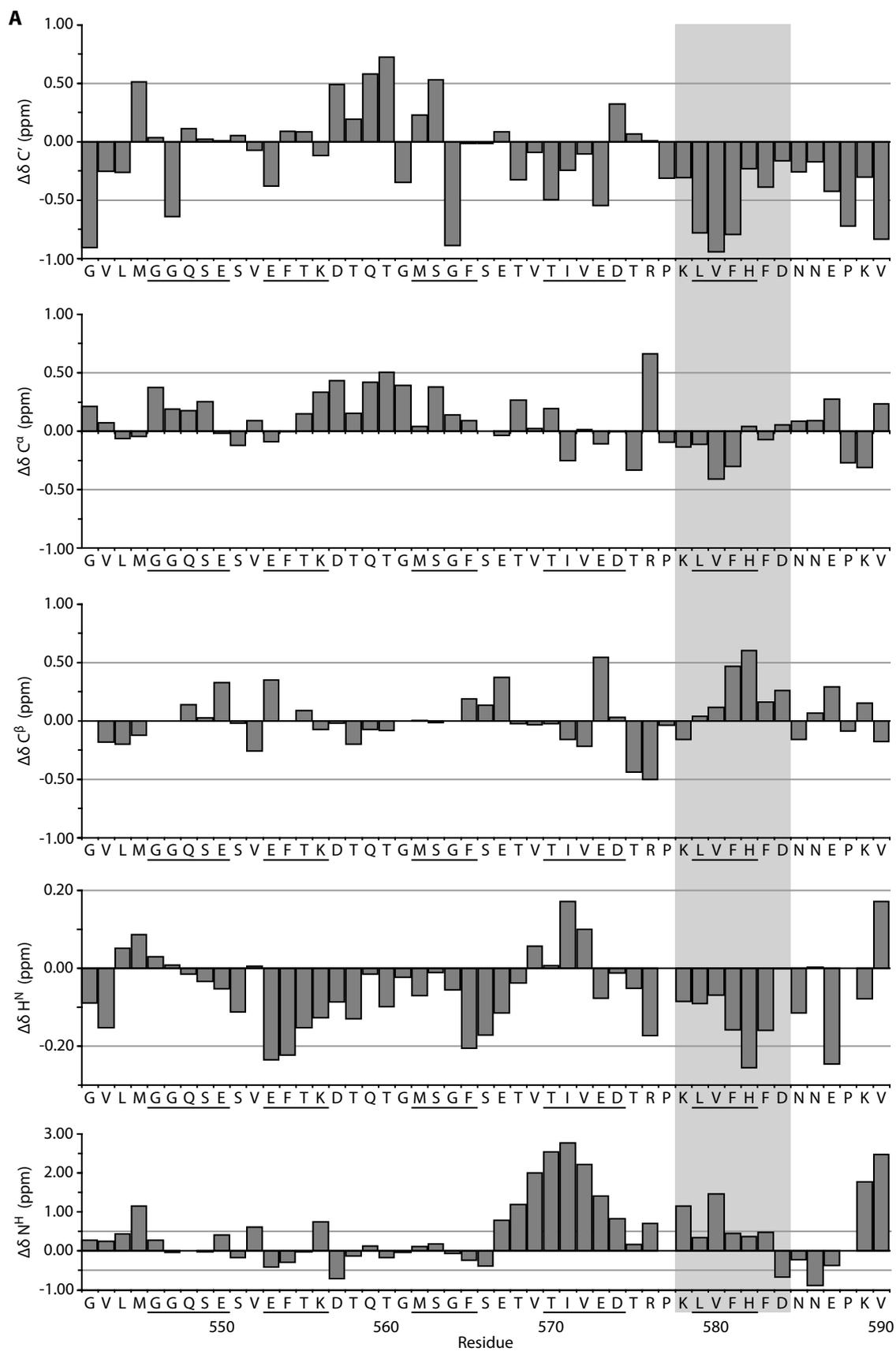
Figure 23. Secondary chemical shifts (Δδ) for the $^1H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C'$ nuclei of SfbI-5 native residues. (Part 1 of 2). Legend on p 136.

134
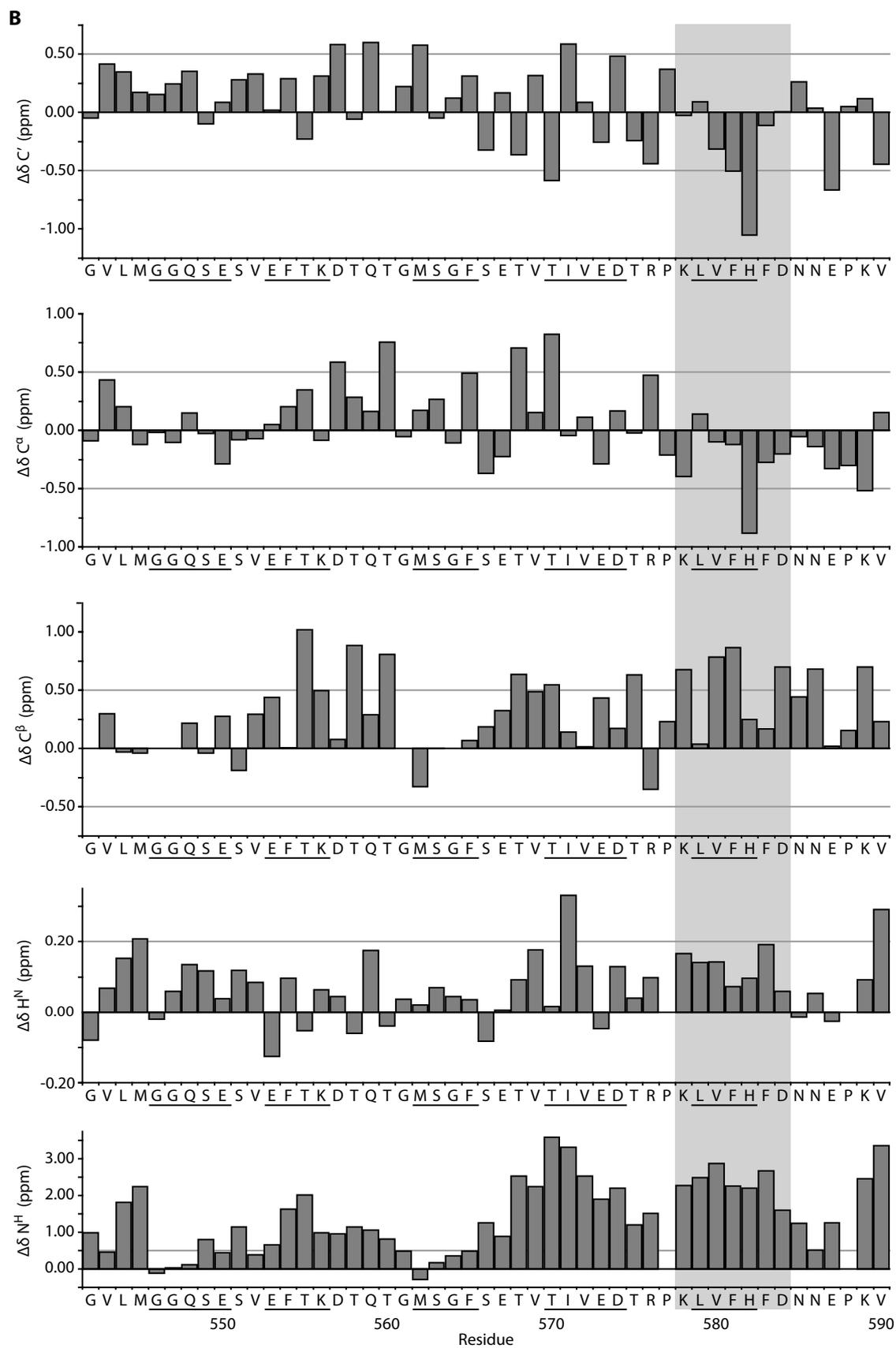
Figure 23. (Part 2 of 2). Legend on p 136.

Figure 23. Secondary chemical shifts (Δδ) for the $^1H^N$, $^{15}N^H$, $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C'$ nuclei of SfbI-5 native residues. Δδ values were calculated by subtracting sequence-corrected random coil chemical shifts from the corresponding experimentally observed chemical shifts of SfbI-5 (Table 26). Two sets of Δδ values are shown; those calculated using the random coil shifts and sequence correction factors of either **A.** Wishart *et al.* (1995; WS) or **B.** Wang and Jardetzky (2002b; WJ2). For both A and B, the chemical shift index (CSI) threshold values implemented in NMRView for the WS dataset are indicated by grey horizontal lines (Section 3.3.4; Johnson & Blevins, 1994; Schwarzinger *et al.*, 2000; Schwarzinger *et al.*, 2001). To our knowledge, no recommended CSI threshold values have been published for the Wang and Jardetzky dataset. Residues predicted to form β-strands on complex formation with $^{1-5}F1$ based on comparison with known Fn-module:FnBP structures are underlined. The grey boxed region appears to have some secondary structure propensity (SSP) as discussed in Section 3.3.4.2.

3.3.4 Secondary structure propensity of SfbI-5 calculated by chemical shift analysis

3.3.4.1 SfbI-5 behaves like a random coil as determined by the chemical shift index

To predict any regions of stable secondary structure in SfbI-5, the CSI for SfbI-5 was calculated using the WS Δδ values (Section 2.4.5.2). CSI threshold values for the WJ2 dataset have not been published, so a CSI calculation was not performed for WJ2 Δδ values. The CSI for each assigned SfbI-5 nuclei resulted in only the $^{15}N^H$ nuclei of residues 568–575 and the $^{13}C'$ nuclei of residues 580–582 being predicted to form β-strands (Figure 23A). Interestingly, both these regions are predicted to contain residues that are either known or predicted to form a β-strand on binding to $^{1-5}F1$, based on comparison with known F1-module:FnBR peptide structures. However, all other nuclei are predicted to be random coil. Therefore, the consensus CSI for all SfbI-5 residues is random coil; further confirming SfbI-5 is an IDP because it lacks stable secondary structure.

3.3.4.2 SfbI-5 appears to have some secondary structure propensity for β-strand conformations

It is possible that some SfbI-5 residues have secondary structure propensity (SSP) for either β-strand or α-helix conformations. That is, it is possible some SfbI-5 residues spend more time occupying φ and ψ angles typical of a β-strand or an α-helix than would be expected for a protein behaving like a random coil. In this case, it would be expected that the Δδ values would not be over the CSI threshold values because they are

not forming stable secondary structure. Instead, some SSP for β-strand conformations, for example, will be qualitatively indicated by a local density of Δδ values that are positive for $^1H^N$, $^{15}N^H$ and $^{13}C^\beta$ nuclei and negative for $^{13}C^\alpha$ and $^{13}C'$ nuclei. Analysis of SfbI-5 Δδ values reveals that only residues 578–584 appear to fulfil these criteria for the Δδ values calculated with both random coil datasets (grey boxed region in Figure 23). An exception to this trend is the $^1H^N$ nuclei from the WS Δδ values, which the $^1H^N$ WJ2 Δδ values directly contradict (compare Figure 23A and B). Therefore, the consensus of the Δδ values indicates residues 578–584 may have some SSP for β-strand conformations.

The SSP was quantified by calculating the SSP score for each SfbI-5 residue. The SSP score was developed by Marsh *et al.* (2006) to help analyse any SSP in IDPs (Section 1.6.2.2). The SSP score for SfbI-5 is shown in Figure 24, where it can be seen that the N-terminal two-thirds of SfbI-5 is essentially random coil, with almost all SSP scores between −0.1 and +0.1. Although most residues in this region have a slight propensity for β-strand, this is probably within experimental error. In contrast, residues from the C-terminal third of SfbI-5 exhibit significant propensity for β-strand; nearly all residues have SSP scores less than −0.1. In particular, residues 580–584 have an SSP score less than −0.3; indicating more than a third of the SfbI-5 molecules exhibit φ and ψ angles typical of β-strand for this region at any given time. Interestingly, residues 580–583 are predicted to form β-strand on binding to $^{1-5}F1$ and have also been qualitatively indicated from analysis of the WS Δδ values. Therefore, the majority of SfbI-5 has a small (possibly insignificant) propensity for adopting β-strand conformations, while the C-terminal third has a significant propensity for adopting β-strand conformations.

G V L M G G Q S E S V E F T K D T Q T G M S G F S E T V T I V E D T R P K L V F H F D N N E P K V E

Residue
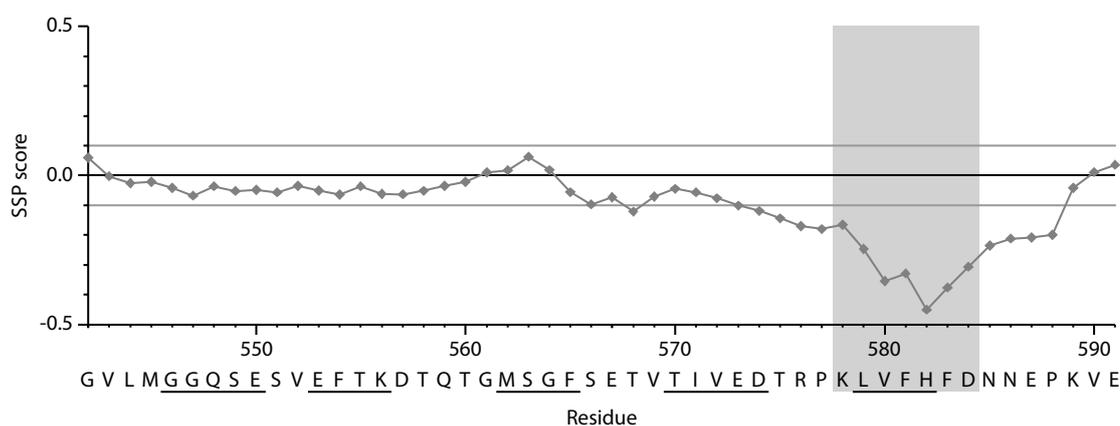
Figure 24. SSP of SfbI-5. The sequence-corrected SSP score was calculated for SfbI-5 using all available chemical shift information. It is worth noting that the Perl script kindly made freely available for download by Marsh *et al.* (2006) does not calculate sequence corrections for the random coil values. Therefore, the appropriate sequence corrections from the WJ2 dataset were added to the appropriate observed SfbI-5 chemical shifts prior to SSP score calculation (Section 2.4.5.3). The grey boxed region shows significant SSP when both the $\Delta\delta$ values (Figure 23) and SSP score are considered.

3.4 Preliminary characterization of ^{15}N-labelled SfbI-5 in complex with pNTD

3.4.1 SfbI-5, pNTD and the pNTD:SfbI-5 complex do not behave as globular proteins by size exclusion chromatography

The apparent MWs of pNTD, SfbI-5 and the pNTD:SfbI-5 complex were determined by SEC. These experiments were performed for two reasons. First, to further characterize SfbI-5 as a putative IDP, because a high apparent MW by SEC is an indication that a protein is disordered (Receveur-Brechot *et al.*, 2006). Second, the apparent MW of the pNTD:SfbI-5 complex can be used to predict how the complex might behave in NMR experiments, because a complex that behaves with high apparent MW might have broader linewidths because of fast transverse relaxation.

The apparent MWs were determined by SEC as follows. First, globular protein standards of known MWs were used to calibrate a Superdex 200 column, and create a calibration curve (Figure 25). The standards show a clear logarithmic relationship between MW and retention time ($R^2 = 0.96$; Figure 25B). Then, SEC was used to determine the retention volumes of pNTD, SfbI-5 and pNTD:SfbI-5, and from these their apparent MWs were calculated with respect to the globular protein standards (Figure 26; Table 27). The apparent MWs of SfbI-5, pNTD and pNTD:SfbI-5 are, respectively, about four, two and three times their theoretical MWs (Table 27). Thus, all

three species do not behave like globular proteins because their apparent and actual MWs are not approximately equal.

The Stokes radii of the protein species can be estimated by using the relationship between MW and the Stokes radius, which was determined by Uversky (1993) for both folded and denatured globular proteins using data from the literature (Equations 9, 10). From these relationships it was possible to estimate the experimental Stokes radii of the protein species from their apparent MWs, and also the theoretical Stokes radii of a fully denatured and a folded globular protein of the same MW as each of the protein species (Table 27). From this analysis SfbI-5 is predicted to have a Stokes radius of 22 Å, which is approximately equal to the theoretical Stokes radius 21 Å of SfbI-5 if it was fully denatured. This indicates SfbI-5 behaves like a random coil, which agrees with the chemical shift analysis (Section 3.3.4). The apparent MW of pNTD indicates this protein behaves in solution like a globular protein twice its size. This might be either because it dimerizes or because pNTD behaves more like a rod than a sphere. Interestingly, the pNTD:SfbI-5 complex behaves even less like a globular protein than pNTD, with an apparent MW approximately three times its actual MW (Table 27). This might be as a result of a higher order complex forming (*e.g.,* a molar ratio of 2:2) or because SfbI-5 stabilizes the flexible linkers between F1-modules, making the pNTD:SfbI-5 complex more rigid and rod-like relative to pNTD. A SEC MALLS experiment indicated that the majority of the complex was a 1:1 heterodimer (Section 3.4.2). Therefore, pNTD:SfbI-5 appears to have a high apparent MW because it is rod-like. Furthermore, these experiments indicated the characterization of SfbI-5 in complex with pNTD using NMR spectroscopy might prove difficult because of the high apparent MW of the complex.
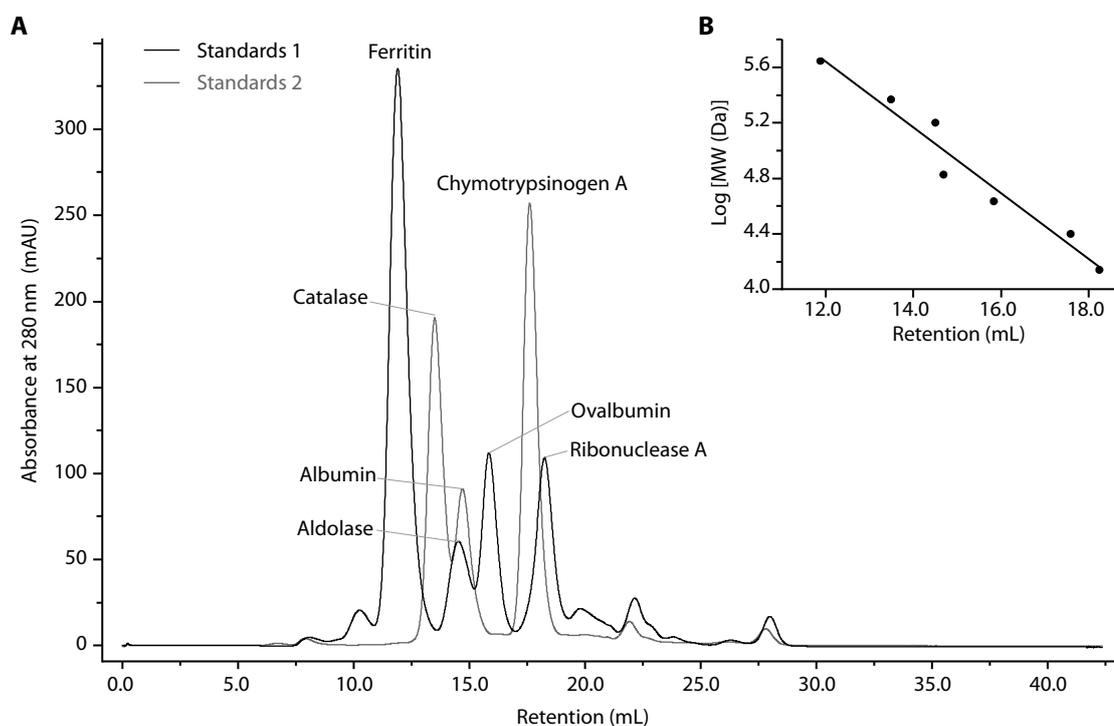
Figure 25. SEC calibration with globular protein standards. A. Overlaid chromatograms of the A_{280} readings from two globular protein standard runs on a Superdex 200 column in PBS at room temperature (Section 2.5.1). Peaks are labelled by protein species and corresponding MW. **B.** MW calibration curve showing the logarithmic relationship between retention volume and the known MWs of the globular protein standards. The line of best fit as calculated with linear regression is shown, and is described by Equation 8 (Section 2.5.2).

Table 27. Apparent and theoretical MW of pNTD, SfbI-5 and pNTD:SfbI-5. The apparent MWs were calculated from SEC with respect to globular protein standards (Section 2.5.2). The theoretical MWs were calculated from the primary amino acid sequences of the species. ESMS and N- and C-terminal sequencing of pNTD was used to confirm the appropriate primary sequence was used. pNTD:SfbI-5 complex was assumed to be 1:1, an assumption that was later confirmed (Section 3.4.2). The three Stokes radii values were calculated for the protein species based on their theoretical and apparent MWs (Section 2.5.2). See text for definitions of the three types of Stokes radii.

Protein species	MW (kDa)		Stokes radius (Å)		
	Theoretical	**Apparent**	**Theoretical, denatured**	**Theoretical, folded**	**Experimental**
SfbI-5	5.79	20.7	21	14	22
pNTD	28.9	55.7	50	25	31
pNTD:SfbI-5	34.7	95.4	55	26	38
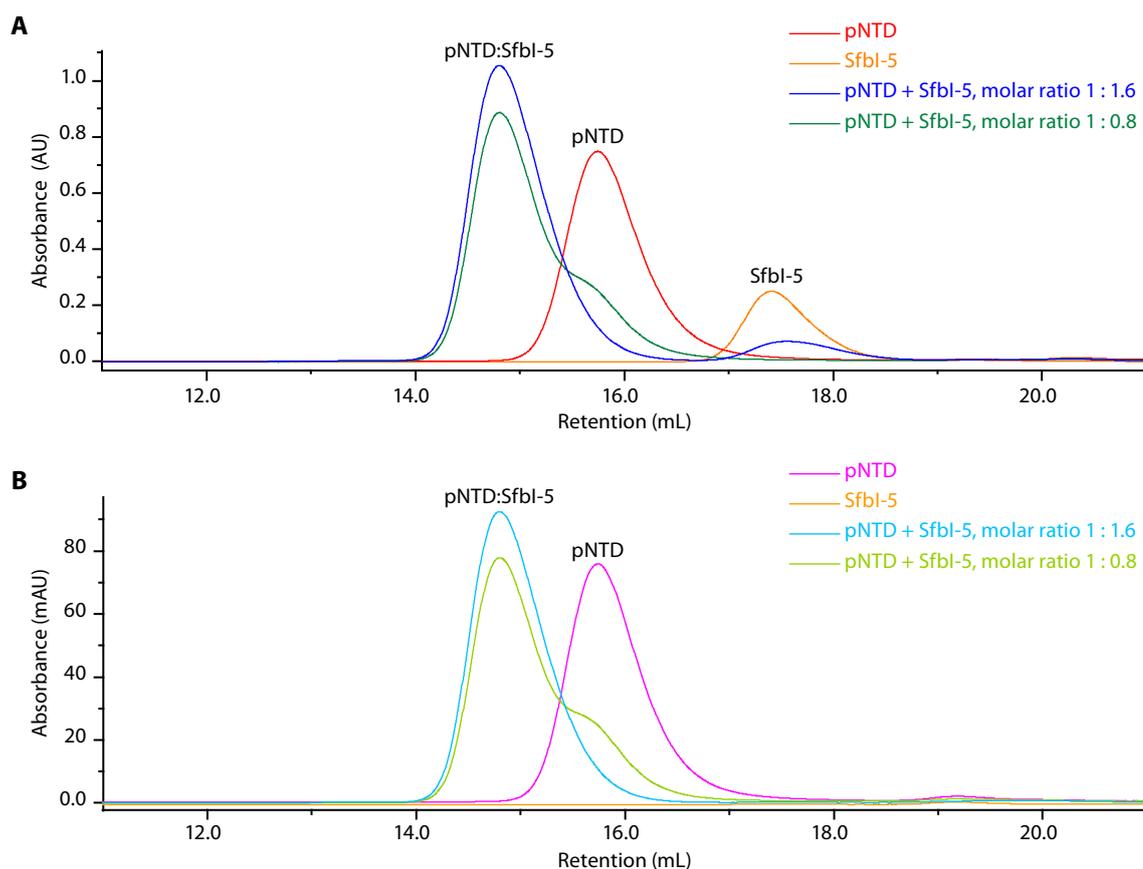
Figure 26. Overlaid SEC traces of pNTD, SfbI-5 and pNTD:SfbI-5. Absorbance at **A.** 215 nm and **B.** 280 nm for the four samples: pNTD, SfbI-5, pNTD and SfbI-5 at a molar ratio of 1:1.6, pNTD and SfbI-5 at a molar ratio of 1:0.8. All samples were analysed by SEC on a Superdex 200 column in PBS at room temperature (Section 2.5.1). The colours corresponding to the absorbance of each sample are indicated, and the peaks are labelled with protein species. Note that SfbI-5 does not absorb at 280 nm so is not seen in B.

3.4.2 SfbI-5 forms a 1:1 complex with pNTD by SEC MALLS

A SEC MALLS experiment (Section 2.5.3) was performed with an aliquot of the pNTD and SfbI-5 sample used in the preliminary NMR experiments reported in Section 3.4.3, which had been diluted to give a final concentration of ~30 µM for both ^{15}N-labelled SfbI-5 and pNTD in PBS. The aims of this experiment were to determine the MW of the pNTD:SfbI-5 complex, and hence, the molar ratio of the pNTD:SfbI-5 complex and to confirm that the majority of SfbI-5 was bound to pNTD.

There were three peaks detected in this experiment (Peaks I – III, Figure 27). The experimental MW of Peak I corresponds to the theoretical MW of either a 2:1 or a 2:2

complex between pNTD and SfbI-5 (Table 28). Peak II is the major species in the sample and corresponds to the MW of either a 1:1 or a 1:2 complex between pNTD and SfbI-5 (Table 28), and Peak III did not scatter enough light for the MW to be accurately determined. However, based on where this species elutes, it is probably unbound ^{15}N-labelled SfbI-5.

The experimental MW of both Peaks I and II differs from possible theoretical MWs by ~3 kDa (Table 28). This difference is a measure of the experimental error. Compared to the possible theoretical MWs of the pNTD:SfbI-5 complex, this experimental error is relatively small (less than 9%). However, because SfbI-5 has a low MW (5.9 kDa), the error is approximately 50% of the MW of SfbI-5. This leads to ambiguity in the molar ratio of SfbI-5 within the pNTD:SfbI-5 complex for both Peaks I and II. Ideally, the correct molar ratio would be found by replicating the SEC MALLS experiment to measure the mean experimental MW. The standard deviation from the mean could then be used as a measure of the experimental error. Unfortunately, this SEC MALLS experiment was not repeated, so the most probable molar ratio of pNTD:SfbI-5 for Peak II was determined using other information.
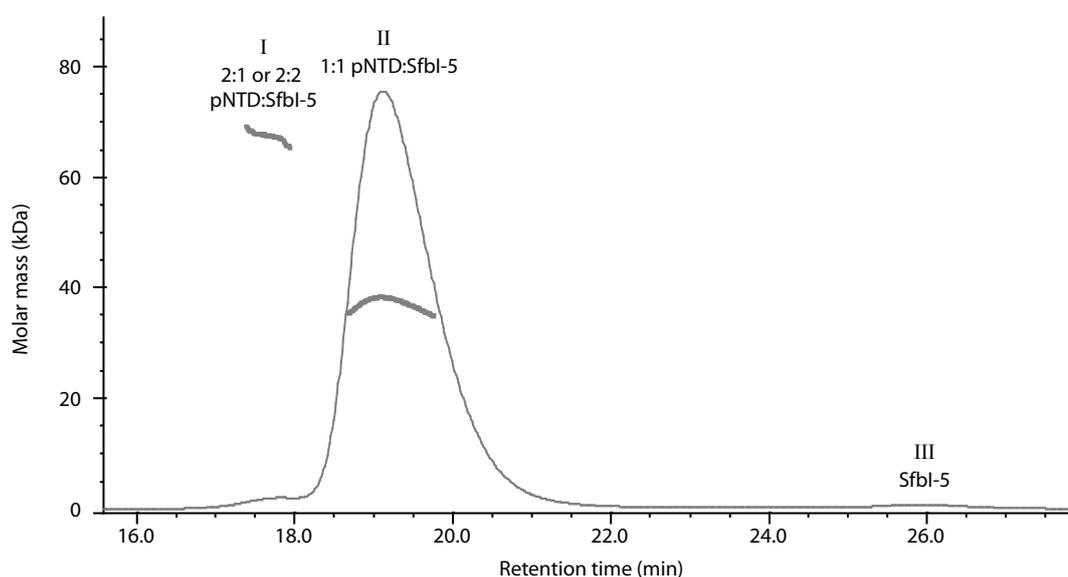
Figure 27. SEC MALLS of ^{15}N-labelled SfbI-5 in complex with pNTD. SEC was performed on a Superdex 75 column in PBS at room temperature and MALLS data acquired and analysed as described in Section 2.5.3. The refractive index of the sample is shown as a thin grey line. The molar mass calculated from analysis of the MALLS data is shown as thick grey lines. Indicated above each of the peaks (labelled I – III) is the protein species responsible for that peak, which was determined from comparison between theoretical MWs and those determined experimentally from the MALLS data (Table 28) along with additional considerations discussed in the text.

Table 28. Experimental MWs from SEC MALLS for pNTD in complex with ¹⁵N-labelled SfbI-5 compared to possible theoretical MWs for different ratios of pNTD to SfbI-5. Experimental MWs were determined by MALLS analysis of Peaks I and II from SEC of pNTD with SfbI-5 (Figure 27, Section 2.5.3). For each experimental MW, the two molar ratios of pNTD to SfbI-5 that have a theoretical MW closest to the experimental MW are reported (A and B), as is the corresponding MW. For each theoretical MW, the error is then reported as: |Experimental MW − Theoretical MW|. The theoretical MWs of pNTD and SfbI-5 were calculated from their primary amino acid sequences. SDS PAGE analysis of the SEC MALLS sample showed that pNTD and SfbI-5 were not degraded, and that each species migrated with a MW that was in agreement with its theoretical MW (data not shown).

Peak	MW (kDa)			Molar ratio of pNTD to ¹⁵N-labelled SfbI-5		Error (kDa)	
	Experimental	Theoretical					
		A	B	A	B	A	B
I	37.2	34.8	40.6	1:1	1:2	2.4	3.4
II	67.2	63.6	69.5	2:1	2:2	2.3	3.6

It is highly likely that Peak II (Figure 27) is a pNTD:SfbI-5 complex at a molar ratio of 1:1, because the fractional molar ratio of binding (n) of pNTD/SfbI-5 was shown to be 1 in ITC binding experiments (Schwarz-Linek *et al.*, 2004b; Table 18). Furthermore, the SEC MALLS sample was prepared at a pNTD:SfbI-5 molar ratio of approximately 1:1. Therefore, if Peak II was a 1:2 complex of pNTD:SfbI-5, not only should *n* be 0.5, but pNTD would also be in excess in the SEC MALLS experiment, and observed as a separate peak[14], except no such peak was observed (Figure 27). Hence, it is most probable that Peak II, the major species in the SEC MALLS experiment, was the 1:1 complex between pNTD and SfbI-5.

Three key observations can be made from the SEC MALLS experiment. First, given that the major species in the SEC MALLS experiment was a 1:1 complex, then the major species in the NMR experiments was probably also the 1:1 complex. Second, that there was only a slight excess of SfbI-5 (Peak III, Figure 27), confirming that the majority of SfbI-5 was bound to pNTD. Third, while there was only a small amount of either a 2:1 or 2:2 complex of pNTD:SfbI-5 in Peak I relative to the 1:1 complex in

[14] It can be assumed that the Superdex 75 column used in the SEC MALLS experiment would resolve pNTD and pNTD:SfbI-5 both because the Superdex 200 column has been shown to resolve pNTD and pNTD:SfbI-5 (Figure 25), and because the Superdex 75 column has a higher resolution than the Superdex 200 column at the MWs of pNTD and pNTD:SfbI-5.

Peak II, this ratio may be different at the conditions in the NMR experiment, where the protein concentration was almost four-times higher and the sample was in water as opposed to PBS. An equilibrium between a 1:1 complex and a higher order complex of pNTD:SfbI-5 during the NMR experiments might lead to peak broadening caused by intermediate exchange on the NMR timescale.

3.4.3 The pNTD:SfbI-5 complex was detectable in a preliminary NMR experiment

A longer term goal is to characterize the SfbI-5 residues involved in binding $^{1-5}$F1. This might be achieved by assigning the spectrum of ^{15}N, ^{13}C-labelled SfbI-5 bound to pNTD. Given that the high apparent MW of SfbI-5:pNTD could lead to broad linewidths in NMR spectra, the feasibility of assigning the spectrum of pNTD-bound SfbI-5 was tested by acquiring an HSQC experiment of ^{15}N-labelled SfbI-5 in complex with pNTD at a molar ratio of 1:1. This spectrum was kindly acquired by Christina Redfield (Biochemistry Department, University of Oxford) at 310 K on a 500 MHz Bruker spectrometer equipped with a cryoprobe. The cryoprobe improves the signal to noise ratio in the spectrum by cooling the electrical coils that generate the radio frequency pulses during the experiment. Furthermore, acquiring at 310 K will increase the tumbling rate of pNTD:SfbI-5, which should make the peaks narrower by reducing the transverse relaxation rate.

A comparison of the HSQC spectra of SfbI-5 free and in complex with pNTD revealed that the complex could be detected by NMR, with most peaks in the HSQC spectrum of pNTD:SfbI-5 appearing to have moved relative to corresponding peaks in the SfbI-5 spectrum (Figure 28). This indicates the chemical environment of most residues has changed. Furthermore, there are two features of the pNTD:SfbI-5 spectrum that are particularly noteworthy.

First, seven peaks in the pNTD:SfbI-5 spectrum do not have obvious counterparts in the SfbI-5 spectrum (Figure 28). Two of these, [1] and [2], probably arise from the resonance of sidechain nuclei, and the remaining five probably arise from backbone resonances. Interestingly, peaks [6] and [7] increase the dispersion of that region of the spectrum (Figure 28), indicating these peaks are not in a random coil, but in a unique chemical environment that may be the result of the of SfbI-5 forming secondary and tertiary structures when in complex with pNTD.
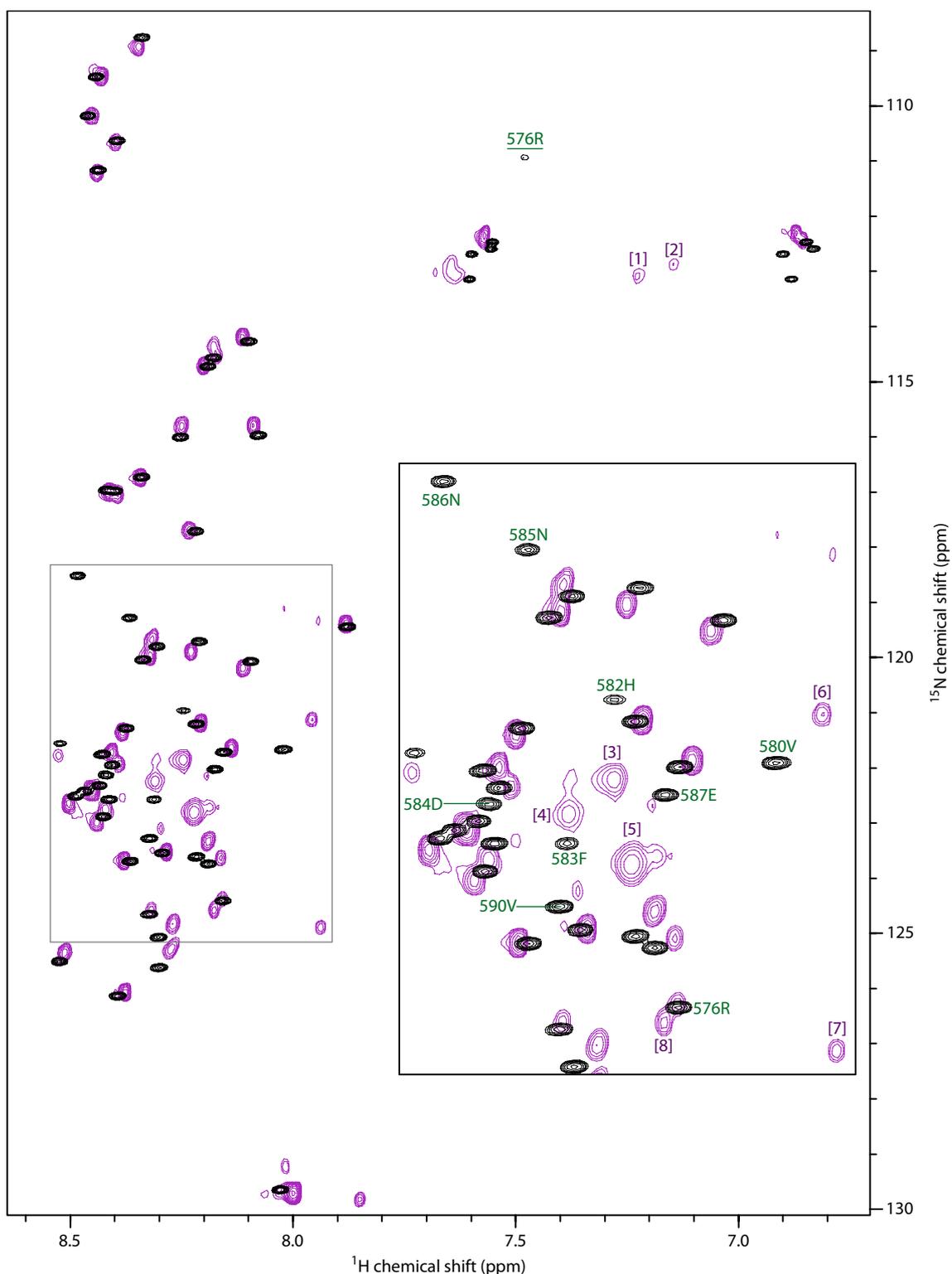
Figure 28. The HSQC spectra of ^{15}N-labelled SfbI-5 in the presence and absence of a 1:1 molar ratio of pNTD. The spectrum of ^{15}N, ^{13}C-labelled SfbI-5 (black) acquired at 298 K is described in Figure 19. The HSQC spectrum of ^{15}N-labelled SfbI-5 in complex with pNTD (~110 μM ^{15}N-labelled SfbI-5, ~110 μM pNTD, 10% D_2O, 0.02% NaN_3, pH 5.5) was acquired at 310 K on a 500 MHz spectrometer equipped with a cryoprobe and is shown in purple. The HSQC spectrum of bound SfbI-5 was overlaid on the HSQC spectrum of free SfbI-5 to maximize overlap between the two spectra. This was necessary because the spectra were acquired at different temperatures. Peaks in either HSQC spectrum that do not appear to have a corresponding peak in the other spectrum are labelled by either residue type and number

in the assigned HSQC spectrum of SfbI-5 or by a number in square brackets for peaks in the spectrum of the SfbI-5:pNTD complex. Inset is an enlargement of the grey boxed region.

Second, peaks corresponding to the ^{1}H^N and ^{15}N^H frequencies of residues V580, H582, F583, D584, N585, N586, E587 and V590 in the SfbI-5 spectrum do not have obvious counterparts in the pNTD:SfbI-5 spectrum (Figure 28). This may be because these resonances are in intermediate exchange on the NMR timescale and are, therefore, not detectable (Section 1.6.3) or because their chemical shifts have changed significantly and they correspond to one of the five peaks with no obvious counterpart already mentioned and labelled in Figure 28. Interestingly, all the resonances with no obvious counterparts in the pNTD:SfbI-5 spectrum are from the ^{1}F1-binding region of SfbI-5, which was shown to have secondary structural propensity in Section 3.3.4.2 (Figure 25). In summary, despite the high apparent MW and potential for dimerization of pNTD:SfbI-5, this preliminary experiment is promising and warrants further investigation.

3.5 Discussion

3.5.1 SfbI is probably an IDP

There are multiple lines of evidence in support of the FnBRs in SfbI being intrinsically disordered, and hence, SfbI being an IDP. Two lines of evidence existed before the work reported here. First, a region of SfbI corresponding to the C-terminal half of SfbI-1, all of SfbI-2–SfbI-4 and the N-terminal half of SfbI-5 was shown to contain no significant secondary structure by far-ultraviolet circular dichroism (House-Pompeo *et al.*, 1996). Second, Penkett *et al.* (1998) performed a thorough NMR analysis on peptide "D1–D4" consisting of the C-terminal half of the eighth FnBR and all of FnBRs 9–11 from *S. aureus* protein FnBPA (Figure 5A; Meenan *et al.*, 2007; Schwarz-Linek *et al.*, 2003). This analysis showed that the FnBRs were intrinsically disordered (Penkett *et al.*, 1998). Given the functional and sequence similarities between FnBRs in FnBPA and SfbI it was reasonable to assume SfbI FnBRs are also disordered.

The results presented here support the hypothesis that the FnBRs in SfbI are disordered. It has been shown by sequence analysis of SfbI with three different disorder predictors that the FnBRs are all predicted to be disordered (Figure 15; Section 3.1). On the basis of this evidence, ^{15}N, ^{13}C-labelled SfbI-5 (Section 3.2) was produced. An HSQC

experiment of this construct confirmed the disorder prediction, since the peaks in the spectra showed the poor peak dispersion and narrow linewidths typical of intrinsic disorder (Figure 19; Section 3.3.1). Furthermore, chemical shift analysis revealed that SfbI-5 had no region of stable secondary structure, with all residues having a consensus CSI value of random coil (Section 3.3.4.1). Finally, it has been shown that in PBS SfbI-5 is predicted to have the same Stokes radius as a protein of its size in denaturing conditions (Section 3.4.1), indicating SfbI-5 is behaving as if it is denatured in PBS. Therefore, by sequence analysis, NMR spectroscopy and SEC, SfbI-5 appears to be disordered. Given the sequence similarity between SfbI-5 and the other FnBRs in SfbI (Figure 29), it is highly likely that the other FnBRs are also disordered. Therefore, SfbI is probably an IDP.

```
               550        560        570        580        590
        ...|....|....|....|....|....|....|....|....|....|....|.
SfbI-1  E-VYGNQQNPVDIDKKL-PNETGFSGNMVETEDTKEP-------------
SfbI-2  GVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEP-------------
SfbI-4  GVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEP-------------
SfbI-3  GVLMGGQSESVEFTKDTQTGMSGQTASQVETEDTKEP-------------
SfbI-5  GVLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVE
B3T     -----------------------GSTTEVEDSKPKLSIHFDNEWPKED
Clustal    : *.*.:.*:: *.  .. :* :   . .***:
```

Figure 29. Sequence alignment between the five SfbI FnBRs. Residues are coloured based on chemical composition. Conserved residues are indicated below the alignment, where an asterisk (*) colon (:) and full stop (.) indicate identical, conserved and semi-conserved residues, respectively. Residue numbering for SfbI-5 with respect to full-length SfbI is shown above the alignment. Residues predicted to form β-strand on binding to [1–5]F1 as predicted from F1-module:FnBR complex structures are underlined (Bingham *et al.*, 2008). Residues predicted to form β-strand by either the ALB or PROF programs are highlighted or boxed, respectively (Ptitsyn & Finkelstein, 1983; Rost & Sander, 1993). B3T is also shown, it is the sequence of the *S. dysgalactiae* FnBR peptide whose structure has been solved in complex with [1]F1[2]F1 (Schwarz-Linek *et al.*, 2003).

3.5.2   The [1]F1-binding region of SfbI-5 is a preformed contact site

SSP analysis by calculation of an SSP score revealed that while the N-terminal two-thirds of SfbI-5 had slight, possibly insignificant, β-strand propensity, the remaining C-terminal third had significant β-strand propensity, particularly residues 580–584 (Section 3.3.4.2; Figure 24). This is interesting because this region was predicted to be more ordered than the rest of SfbI-5 by the VLXT predictor (Figure 15B), and its SSP score indicates this prediction was correct. Also, based on alignment between SfbI-5 and the *S. dysgalactiae* FnBR peptide B3T, whose structure has been solved in complex

with ^{1}F1^{2}F1 (Schwarz-Linek *et al.*, 2003), SfbI-5 residues 579–582 are predicted to form β-strand on binding to the ^{1}F1-module in $^{1-5}$F1 (Figure 29). All four of these residues have high β-strand propensity relative to the rest of SfbI-5 (Figure 24). This indicates the ^{1}F1-binding region of SfbI-5 is a primary contact site (PCS), as discussed below.

Short regions of IDPs that undergo a disorder-to-order transition on binding to their partner(s) are called, variously, molecular recognition features (MoRFs), primary contact sites (PCSs) or preformed structural elements (PSEs; Csizmok *et al.*, 2005; Fuxreiter *et al.*, 2004; Mohan *et al.*, 2006). These 'short recognition motifs' are all associated with IDPs and have related, but not identical, definitions as summarized in Table 29. Based on these definitions, each F1-binding region in SfbI-5 is predicted to be a β-MoRF. Further, the ^{1}F1-binding region of SfbI-5, which is predicted to be the only ^{1}F1-binding site in SfbI, appears to be a β-MoRF, a PSE and a PCS.

Table 29. Definitions of the short recognition motifs important in molecular recognition in IDPs

Name	Reference	Definition and comments
Molecular recognition feature (MoRF).	Mohan *et al.*, 2006	A short region (~20 residues) in an IDP that undergoes a disorder-to-order transition on binding its partner. They are classed according to the secondary structure they adopt on complex formation. For example, a β-MoRF forms β-strand on complex formation.
		Possible MoRFs can be identified as a region of relative order flanked by disorder, as judged by disorder predictor VLXT. However, not all MoRFs fulfil this criterion.
Preformed structural element (PSE)	Fuxreiter *et al.*, 2004	A MoRF with inherent secondary structure propensity for the secondary structure the MoRF obtains on binding to its partner.
		PSEs can be identified by secondary structure predictions on the sequence of interest. The accuracy of secondary structure prediction by the ALB (Ptitsyn & Finkelstein, 1983) and PROF (Rost & Sander, 1993) programs is as high for PSEs as it is for globular proteins.
Primary contact site (PCS)	Csizmok *et al.*, 2005	A MoRF and a PSE which, as a result of some SSP, is exposed and is therefore likely to make the first contact with its partner, leading to a native-like encounter complex. PCSs can be identified by experimental evidence of SSP, and/or by identification of regions preferentially cleaved by very low concentrations of proteases, indicating the region is exposed.

The theory behind PCSs and PSEs is kinetic, rather than thermodynamic, and is a combination of the two models that represent the two possible extremes in IDP protein recognition (Csizmok et al., 2005; Fuxreiter et al., 2004; Mohan et al., 2006; Wright & Dyson, 2009). The first extreme is called conformational selection (or inherent-structure selection) and occurs when a preformed structural element is selected from the rapidly inter-converting structural ensemble of an IDP by its recognition partner because of its close resemblance to the bound form. That is, folding then binding of the IUP occurs. The second extreme is called induced folding (or the induced-structure mechanism), whereby the IDP is disordered until it binds its partner, and then association induces folding. That is, binding then folding of the IDP occurs. The theory behind PCSs and PSEs combines these two extremes, by suggesting that the first contact site is conformationally selected and that this then induces folding (Fuxreiter et al., 2004). This model has recently been called the synergistic model (Espinoza-Fonseca, 2009).

The synergistic model gives a good explanation of the current information available on SfbI-5, because the chemical shift analysis revealed that only the [1]F1-binding motif in SfbI-5 of the five predicted β-MoRFs appears to have SSP. This led to the prediction that the [1]F1-binding region is a PCS, and so might increase the binding 'on' rate by acting as the first contact site for binding to [1−5]F1 because it already has some propensity for its bound structure. Once bound, the PCS might then help induce folding by bringing the rest of the FnBR β-MoRFs closer to their binding sites on [1−5]F1, thus reducing the 'dimensionality' of the search for more binding sites (Fuxreiter et al., 2004). In this sense, SfbI-5 might actually bind to the F1-modules like a zipper, one F1-module after another, starting from the [1]F1-binding region.

Interestingly, the N-terminal half of SfbI-1 and the UR is also predicted to be relatively ordered compared to the surrounding regions (Figure 15A). As can be seen from Figure 29, the amino acid composition of SfbI-1 differs significantly from that of the other four SfbI FnBRs in this region. It is tempting to speculate that this is also a PCS, and allows both SfbI-1 and SfbI-5 to 'zip up' different molecules of [1−5]F1 from either end. Perhaps the binding of [1−5]F1 to either SfbI-5 and/or SfbI-1 alters the conformation of the neighbouring FnBRs, helping [1−5]F1 bind to these FnBRs too. However, it is worth noting that the UR and SfbI-1, along with SfbI-2–SfbI-4, are not predicted to form secondary structure by either ALB or PROF (Figure 29, data for the UR not shown; Ptitsyn & Finkelstein, 1983; Rost & Sander, 1993). In contrast, SfbI-5 residues in the

[1]F1-binding region are predicted to form β-strand by ALB and residues in the [2]F1-binding regions are predicted to form β-strand by both ALB and PROF (Figure 29). This model for the interaction of SfbI with [1−5]F1 molecules would require extensive experimental work to confirm. Perhaps the easiest experiment is limited proteolysis of SfbI-5 to see whether the [1]F1-binding motif is exposed, as would be predicted for a PCS (Csizmok *et al.*, 2005).

3.5.3 Preliminary characterization of the SfbI-5:pNTD complex by NMR

In addition to characterizing the disordered state of SfbI-5, a longer term goal of our laboratory is to characterize the disorder-to-order transition of SfbI-5, and the interaction between SfbI-5 and [1−5]F1. Initial characterization of the SfbI-5:pNTD complex with SEC looked unpromising for NMR spectroscopy because the pNTD:SfbI-5 complex had a high apparent MW even though the complex was at a ratio of 1:1 (Table 27, Section 3.4.2). However, despite this, the preliminary NMR experiments appeared successful, with nearly all SfbI-5 resonances changing in the spectrum of pNTD:SfbI-5 compared to the spectrum of SfbI-5 alone (Figure 28). Of particular note was the 'disappearance' of nearly all peaks in the [1]F1-binding region (Section 3.4.3). This indicates that these resonances are either in intermediate exchange leading to peak broadening and/or their resonances have changed significantly on complex formation. It is interesting that this is the same region predicted to form a PCS, so this region is likely to have an important role in the mechanism of binding to pNTD that warrants further investigation.

The most obvious experiment to perform next is a NMR chemical shift perturbation assay of ^{15}N-labelled SfbI-5 titrated with [1−5]F1 or pNTD. This will determine whether there is slow, fast or intermediate exchange on the NMR timescale between nuclei in free and bound SfbI-5 (Section 1.6.3). If there is fast exchange, the assigned SfbI-5 peaks can be tracked as increasing amounts of pNTD is added, facilitating the assignment of bound SfbI-5 resonances. Alternatively, if there is slow or intermediate exchange, the resonances of bound SfbI-5 will need to be assigned directly using techniques similar to those reported here for assignment of free SfbI-5. The K_d for the [1−5]F1:SfbI-5 complex is 2 nM (Schwarz-Linek *et al.*, 2003). Hence, it is likely that the complex will have a slow dissociation rate, and so, it is most probable that that there will be either slow or intermediate exchange on the NMR timescale. Directly assigning bound SfbI-5 resonances will require a sample of ^{15}N, ^{13}C-labelled SfbI-5 with pNTD,

in which both components are ~0.5 mM and ideally 1 mM. The solubility of the pNTD:SfbI-5 complex may make preparing this sample difficult: a light precipitate was observed in the ~110 μM sample of the pNTD:SfbI-5 complex used in the preliminary HSQC experiment reported here. If assignment of bound SfbI-5 is successful, this will allow secondary chemical shift analysis to be performed, which will help to elucidate the residues involved in the interaction with pNTD. In conjunction with the experiments reported in this Chapter the interaction between SfbI-5 and $^{1-5}$F1 was characterized by mutation and binding studies and by crystallography, as discussed in the next Chapter.

4 The role of conserved FnBR residues in SfbI-5

Which FnBR residues are important for binding to $^{1-5}$F1 and what is their function? When this work started, the only known structure of an Fn:FnBP complex was of the complex between *S. dysgalactiae* FnBP peptide B3T and ^{1}F1^{2}F1 (Figure 4; Schwarz-Linek *et al.*, 2003). While this structure went some way to answering this question by identifying FnBR residues important for binding to ^{1}F1 and ^{2}F1 (Schwarz-Linek *et al.*, 2003), residue-specific information for the interactions of FnBRs with $^{3-5}$F1 had yet to be obtained. Recently, an analysis of 39 heterodimer complex structures, in which one of the proteins in each complex was an experimentally verified IDP, found that conserved residues in IDPs are more likely to be involved in the intermolecular interface than non-conserved residues (Mészáros *et al.*, 2007). SfbI-5 residues that are important for binding to $^{1-5}$F1 are, therefore, likely to be conserved. Thus, in the work presented here, conserved FnBR residues have been identified (Section 4.1) and the effect of individually mutating conserved FnBR residues in SfbI on binding to both pNTD and Fn has been measured (Section 4.3). Additionally, a high-resolution crystal structure of an SfbI-5 peptide in complex with ^{2}F1^{3}F1 has been determined (Section 4.2). Taken together, these results give insights into the role of conserved FnBR residues in SfbI-5 and which SfbI-5 residues are important for binding to $^{1-5}$F1 (Section 0).

4.1 Identification of FnBRs in the UniProt database

To rigorously identify conserved FnBR residues it is necessary to have a comprehensive MSA representing, as far as possible, all known FnBRs. Such an MSA did not exist, so to make a comprehensive MSA, the UniProt protein database was searched for all FnBRs.

4.1.1 Initial attempts to find FnBRs were unsuccessful

Initial attempts to find FnBRs within the UniProt database using either the position-specific iterated basic local alignment search tool (PSI-BLAST; Altschul *et al.*, 1997) or the multiple excitation maximization for motif elicitation server (MEME; Bailey & Elkan, 1994) were unsuccessful. The results obtained from a search with SfbI-5 using the PSI-BLAST algorithm on the BLAST server (blast.ncbi.nlm.nih.gov/Blast.cgi) did not identify staphylococcal FnBRs, possibly because the search identified a large

number of streptococcal FnBR sequences, which could have biased the search model (data not shown). The input for the MEME server (meme.nbcr.net/meme4_1_1/intro.html) was a group of full length FnBP sequences from which the MEME program identified possible domains. However, the MEME program did not identify the correct, functional, repeat boundaries for the initial search model (data not shown), probably because FnBRs are short (~35 residues), often tandemly arrayed, and can also contain a UR or a [1]F1-binding region (Figure 3). However, the correct repeat boundaries are functionally important, as is clearly demonstrated by the significantly different abilities of FnBRs with and without the correct repeat boundaries to bind to pNTD (Section 1.4.3; Schwarz-Linek *et al.*, 2004b). Also, although an FnBR family called 'Fn_bind' exists in the Pfam database (Finn *et al.*, 2008), the domain boundaries in the model are not the functional repeat boundaries (Figure 30; Schwarz-Linek *et al.*, 2004b), which may mean isolated, functional FnBRs are not found. For example, the Pfam database does not identify the single FnBR in BBK32, for which there is strong biophysical evidence (Section 1.4.3). Therefore, an alternative strategy was adopted using profile HMMs (Sections 1.7, 2.6), which can be created from a user chosen MSA from currently known FnBRs with functional repeat boundaries.
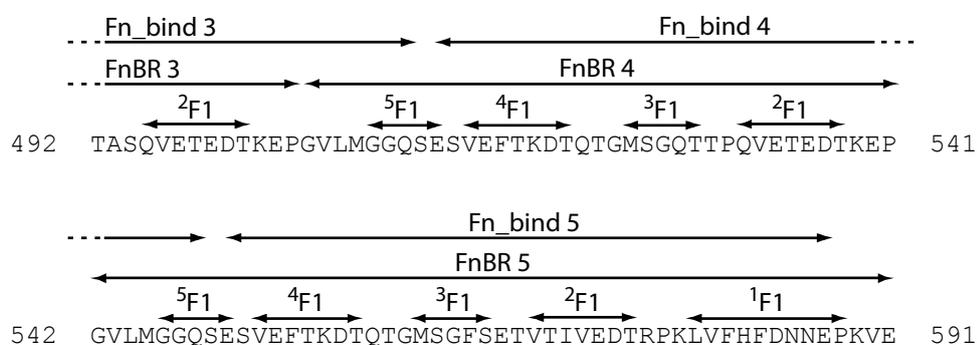
Figure 30. Comparison between functional FnBR boundaries and those defined by Pfam family Fn_bind. The sequence of part of SfbI is shown along with residue numbers corresponding to full length SfbI. Directly above the sequence, the regions predicted to bind the indicated F1-modules are shown. Above this, the functional FnBR repeats (FnBR 3, FnBR 4, *etc.*) and the Fn-binding domain as defined by Pfam family Fn_bind are shown (Fn_bind 3, Fn_bind 4, *etc.*). The Fn_bind domains align with the [2–4]F1-binding region of one FnBR and then also either a [5]F1-binding region of the C-terminally adjacent FnBR or a [1]F1-binding region, if either of these is present.

4.1.2 HMM searching and building

Two separate HMM building and searching 'rounds' were performed, one before any Fn:FnBP crystal structures were solved, and one after six crystal structures had been solved (Section 2.6.1). These two rounds will now be discussed in turn.

4.1.2.1 First round

The FnBRs chosen for the starting MSA in the first round were all from known staphylococcal and streptococcal FnBPs, collated from the work of Schwarz-Linek *et al.* (2003; 2004a), with the addition of the BBK32 FnBR (Raibaud *et al.*, 2005; Section 2.6.1). From this starting MSA of FnBRs, two new MSAs were generated by removing sequences that had either 70% or 90% pairwise identity to any other sequence in the MSA. These MSAs were called nr70 and nr90 (Table 14), respectively, and they were used for two separate cycles of HMM building and searching of the UniRef90 database as described in Section 2.6.2 and Table 15.

Thus, the protocol for identifying FnBPs in the first round involved two steps to remove bias from the identified domains. The first step was to search UniRef90, a subset of the UniProt database consisting of sequence clusters each with a representative sequence. A cluster consists of all sequences in the combined UniProt and UniParc databases with at least 90% sequence identity to the longest sequence in the cluster. The sequence that represents the cluster is chosen based on the entry quality, name, organism and length of sequence, where entries that are from UniProt, without names containing descriptives such as 'hypothetical' or 'putative', from model organisms, and the longest sequences are preferred. Effectively, this means no sequence in UniRef90 should have more than 90% identity to any other sequence in UniRef90, thereby minimizing sample bias when searching with the HMM. The second step was to remove similar FnBRs at either 70% or 90% pairwise identity as described in Table 15. This means sample bias should be minimized within the HMM itself. These steps were deemed necessary for two reasons. First, some closely-related bacterial strains had been characterized more thoroughly than others, resulting in large groups of related sequences that are almost identical (*e.g.,* Towers *et al.*, 2003). Second, most FnBPs contain multiple, highly homologous FnBRs (discussed further in Section 4.4.2).

The same 77 FnBPs were identified in both the nr70 and nr90 HMM searches, although the number of FnBRs identified in each FnBP varied slightly. This indicates that in this case, the identified FnBPs are relatively robust to changes in the HMM used to identify them. These 77 sequences were then searched with two new HMMs, one to identify upstream Fn-binding regions (UR) and one to identify [1]F1-binding regions. The latter was necessary because not enough FnBRs contain [1]F1-binding regions for this region to be included in the nr70 and nr90 HMMs. Finally, to show the predicted domain architecture of the 77 FnBPs, the sequences were annotated with FnBPs as found in the final nr70 search, and also UR, [1]F1-binding regions, predicted signal sequences and Pfam domains (not including Fn_bind). From this analysis, it was revealed that based on domain organization there are 11 sub-families of FnBPs (Figure 31), a result which is similar to that reported by Schwarz-Linek *et al.* (2006) from analysis of FnBPs reported in the literature. The predicted functions of the Pfam domains are summarized in Table 30.

All but three of the identified FnBPs have been annotated in the UniProt database as being a known or predicted FnBP, or a 'serum opacity factor' (SOF). SOFs are proteins which cause serum to become opaque by binding to high-density lipoprotein. Streptococcal SOFs are also known to bind Fn (Katerov *et al.*, 2000). The one new sub-family (consisting of three sequences), with a fourth related sequence identified from a BLAST search, comes from the saliva of the deer tick *Ixodes scapularis* (Ribeiro *et al.*, 2006). The putative FnBR from one of these tick proteins was cloned, expressed and shown to be an FnBR in Chapter 5. These results lend merit to this technique for identifying FnBRs.
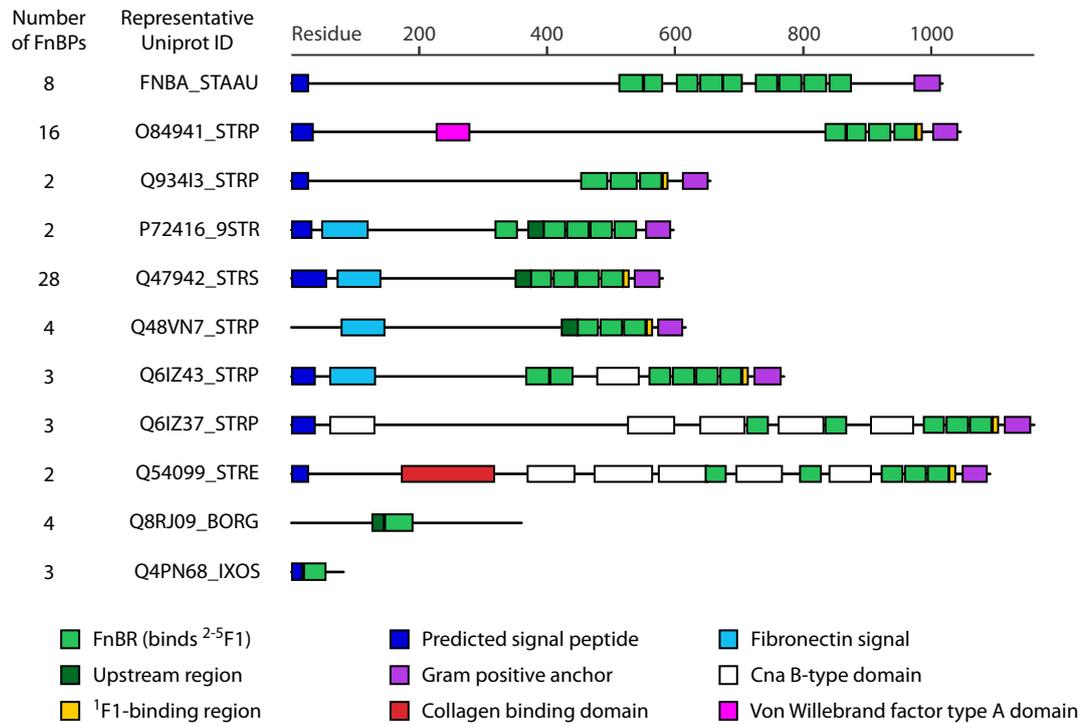
Figure 31. Representative FnBP from each sub-family, as classified by domain organization. FnBPs were identified in the UniRef90 database from iterative searches with HMMs derived from the nr70 MSA (Sections 2.6.1; 2.6.2). Identified FnBPs were each annotated with FnBRs, and, where applicable, the [1]F1-binding region, the UR and Pfam domains, as described in the text. The number of FnBPs belonging to each sub-family and the UniProt ID of the representative FnBP used in the schematic are indicated. Two FnBPs were not classified because they are truncated at their N- and C-termini, thus making it impossible allocate them to a sub-family. Some sequences had the same domain organization apart from the presence or absence of a predicted Gram positive anchor. This difference was not used to distinguish between sub-families, because the difference could exist either because the protein was truncated or because Pfam did not find a Gram positive anchor even when one is known to be there (*e.g.,* for SfbI). See Table 30 for predicted functions of Pfam domains.

Table 30. The predicted function of domains from Pfam found in FnBPs. This information was found at the Pfam website: www.pfam.sanger.ac.uk. It is worth noting that a domain in this context is a conserved region of a protein, not necessarily an autonomously folding region.

Domain	Function
Gram positive anchor	This domain consists of the cell-wall anchor LPXTG and a hydrophobic membrane spanning region, followed by some basic residues. The Pfam HMM does not identify all of these anchors, even when they are known to be present (*e.g.,* SfbI, Figure 3, p 31). However, not all gram-positive FnBPs contain this anchor, but are secreted rather than cell-wall attached (*e.g.,* FNE from *Streptococcus equi* subspecies *equi*, Lindmark *et al.*, 2001).
Collagen binding domain	This domain binds collagen; it is a globular protein with a jelly-roll fold (Symersky *et al.*, 1997).
Fn binding protein signal sequence	This Pfam domain is putatively a streptococcal signal sequence. However, the reference cited for this domain by the Pfam database (Ramachandran *et al.*, 2004) only mentions a DNA primer designed to target the bacterial secretion signal sequence in streptococcal FnBPs (as identified by the signal prediction server, SignalP; Bendtsen *et al.*, 2004). Corresponding to a translated sequence of 'IGLSGVS' the primer does target residues 27–33 of the secretion signal sequence (residues 1–37) from *S. pyogenes* FnBP Protein F2. However, the Pfam HMM 'Fn signal' corresponds to residues 61–131 of Protein F2. Therefore, it appears that the reference is incorrectly cited, leaving the function of this conserved streptococcal domain unknown.
Cna B-Type domain	This domain is predicted to function as a 'stalk'; that is, it is predicted to project ligand-binding domains away from the bacterial cell surface (Deivanayagam *et al.*, 2000). It is a globular domain, with a β-sandwich fold (Deivanayagam *et al.*, 2000).
Von Willebrand factor type A domain	This domain is common in extracellular or transmembrane proteins including collagens and integrins (Colombatti *et al.*, 1993). It appears to be involved in macromolecular assemblies. Proteins with this domain are involved in many functions, including cell adhesion (Colombatti *et al.*, 1993). Therefore, it is possible that the FnBPs contain this domain to either mimic the host and/or adhere to cells.

4.1.2.2   Second round

The second round was deemed necessary because not only would knowledge from the Fn:FnBP crystal structures change the starting MSA, and hence, possibly change which FnBRs were identified, but as two years had passed since the first search, it was also possible new FnBPs had been discovered. The first round's starting MSA was used in

the second round, except that the FnBR from BBK32 was omitted. This was for two reasons. First, there is a ^{2}F1^{3}F1:BBK32 peptide structure (G. Harris, unpublished), in which a BBK32 fragment designed to bind to just ^{2}F1 does so though a β-zipper interaction, but with a different β-strand register than would be predicted both from alignment of the sequence to other FnBRs and when compared to the other ^{2}F1^{3}F1:FnBR structures. It is unclear whether this β-strand register will differ when a longer sequence is crystallized with ^{2}F1^{3}F1. Second, it was hoped that subsequent HMM searches would identify BBK32 FnBP family members without having them in the starting MSA to begin with, as a measure of the effectiveness of the HMM search method. This technique of omitting sequences to see if the HMM can still find them worked for the more closely related streptococcal and staphylococcal sequences. That is, an HMM built with only the known streptococcal FnBRs identified known staphylococcal FnBRs. Therefore, the three remaining FnBRs for which there was structural information were used to inform the starting MSA 's100' (Table 14), as described in Section 2.6.1.

An FnBR in BBK32 (UniProt ID Q9L890) was identified by this round of HMM searching, but only by looking at hits with worse (larger) E-values than the default cut-off value of ten (see Section 1.7 for an explanation of E-values). By the second round's third iterative search, no new FnBRs were identified with an E-value greater than ten. However, the FnBR from BBK32 Q9L890 was identified with an E-value of 420. Since this is a known FnBR, it was decided that it was worth considering whether any sequences with better E-values than Q9L890 could also be FnBRs. This led to the identification of two more putative FnBRs amongst 23 sequences that are predicted to be false positives on the basis of their originating organism, sequence annotations, and/or lack of a predicted signal sequence (Section 2.6.2). Additionally, a putative new FnBP, identified in the first cycle with a relatively good E-value (0.011), had been deposited in UniProt database between the first and second rounds. The merit of these three putative FnBPs will now be discussed.

The first sequence identified (UniProt ID A8SJY5) was from *Parvimonas micra* (previously called *Peptostreptococcus micros,* Tindall & Euzeby, 2006), an oral Gram-positive cocci species found in the human mouth that is associated with various oral diseases (reviewed in Murdoch, 1998). This sequence appears highly likely to be a functional FnBP. The second sequence (UniProt ID Q4PN66) is related to the three

previously identified sequences from the deer tick *I. scapularis* (Section 4.1.2.1). This new sequence had been identified during the first round from a BLAST search with the *I. scapularis* sequences, but not by an FnBR HMM search. Given that one of these sequences is shown to contain a functional FnBR in Chapter 5, it is likely that all four of the *I. scapularis* sequences are real FnBPs. The third sequence (UniProt ID Q95P73) was from a protein called 'thrombostasin', which is an anti-clotting protein found in the saliva of the blood-feeding horn fly *Haematobia irritans* (Figure 32; Zhang *et al.*, 2002). Thrombostasin is the C-terminal half of pro-thrombostasin; it is released when pro-thrombostasin is cleaved three residues past the putative FnBR through an unknown mechanism by the horn fly proprotein convertase (Figure 32). Thrombostasin is then capable of binding thrombin, inhibiting the blood clotting cascade (Zhang *et al.*, 2004). It is unclear what the N-terminal half of the protein does, and it is possible that it could bind Fn to inhibit wound healing. That the predicted *I. scapularis* FnBR has been shown to be functional, lends merit to the N-terminal half of thrombostasin also being a functional FnBR.

```
            Predicted signal peptide
   1   NINIMKHFVVIGILALSAVCQAQNVLSGRRQHGAQGLSGYSGDNDWGYYGEAGAPGSDYS  60

                      Putative FnBR
  61   GSSGQWAPLDFDYNSLPGLSGYNHEQQDYEEDSYRHVR SAGPITLQLDDDDDDDSGIPIF  120
                                             HFPC cleavage

 121   EMDDEDEDSNDNQKFPLSFERFPENEKNQVGLRARFNKFMAKFTSLFGRRRGVNVPNAA  179
```

Figure 32. Pro-thrombostasin protein sequence from the horn fly *Haematobia irritans.* Shown are the predicted signal peptide identified by SignalP (Bendtsen *et al.*, 2004), the putative FnBR identified by an FnBR HMM search and the cleavage site for horn fly proprotein convertase (HFPC) identified by Zhang *et al.* (2004). Cleavage by HFPC releases thrombostasin (italicized), which has anti-thrombin activity.

4.1.2.3   Comparison between FnBRs identified in the first and second rounds

Comparison between the first and second rounds revealed that some FnBPs identified in the first round from the UniRef90 database downloaded on 27/02/2006 were no longer identified in the second round of searching the UniRef90 database downloaded on 10/03/2008. However, further investigation revealed that this was because these sequences were not in the 2008 UniRef90 database, but they were present in the UniRef90 database downloaded on 22/01/2009. The reason for the removal and return

of the sequences is unclear, but the result is that neither the first or second round represents all FnBRs in the UniRef90 database. Therefore, a final search was conducted of the UniRef90 database from 22/01/2009 with the final nr70 and s100 HMMs from the first and second rounds (Section 2.6.3).

FnBP sequences found by each search of the 2009 UniRef90 database were classified as true or false positives (Figure 33; Section 2.6.3). False positives were identified by considering if a putative FnBP made biological sense. That is, the putative FnBP had to be from an organism that could feasibly use an FnBP, and the putative FnBP had to have access to Fn by being extracellular, as indicated by sequence annotation or by the presence of a predicted signal sequence or membrane-spanning region. Information for the 111 predicted true positive FnBPs are listed in Table 31. Nearly all FnBP sequences were identified by both searches with either the final nr70 or s100 HMMs (Figure 33). Interestingly, only one sequence found by both searches was predicted to be a false positive, the other false positives were found by only one of the two HMMs. This suggests that searching a database with both HMMs could help identify 'false positives', that is, those sequences that have been identified by an FnBR HMM, but are unlikely to be a functional FnBP.
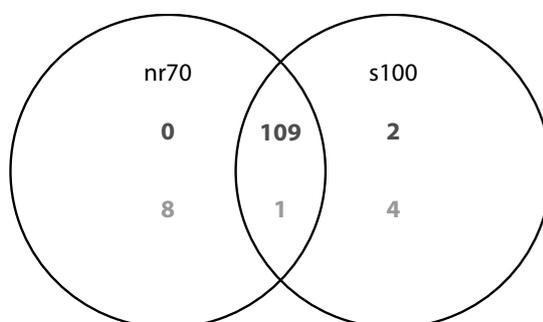
Figure 33. A Venn diagram of the FnBP sequences identified by either the final nr70 HMM or the final s100 HMM in a search of the UniRef90 database. The search was conducted on the UniRef90 database downloaded on 22/01/2009. Sequences predicted to be true and false positives are shown in dark and light grey, respectively. Putative FnBRs are predicted to be false positives based on selection criteria described in Sections 2.6.2 and 2.6.3.

Table 31. All FnBPs identified in the UniRef90 database These sequences were the combined result from two FnBR HMM searches of the UniRef90 database downloaded on 22/01/2009. For each sequence, the UniProt ID, description, gene name and species is reported as present in the UniProt database. When two sequences had the same description, gene name and species, their UniProt IDs were grouped together. Details of sequences with UniParc IDs (starting UPI) were found in the NCBI protein database. Sequences which have not previously been identified as FnBPs before this work are marked with an asterisk (*). Sequences were considered as previously identified if their sequence description included the words 'Fn-binding' or they were named after a protein previously identified as an FnBP; that is BBK32 (also known as Immonogenic Protein P35), Protein F (another name for SfbI); Protein F2 (Jaffe *et al.*, 1996), serum opacity factor (Katerov *et al.*, 2000).

Description	Gene name	Species	UniProt/UniParc ID(s)
BBK32	bbk32	*Borrelia afzelii.*	Q25BS9, Q8RJE4
Fn binding protein BBK32	bbk32	*Borrelia bissettii.*	Q9L890
Immunogenic protein P35		*Borrelia burgdorferi* 80a	UPI00017F33F3
BBK32 immunogenic protein P35		*Borrelia duttonii* (strain Ly)	B5RPD7, B5RN90, B5RN52
BBK32	bbk32	*Borrelia garinii.*	Q8RJ09
Putative uncharacterized protein*		*B. garinii.*	Q5XZ32
BBK32-like protein		*Borrelia hermsii* (strain DAH)	Q1CNX8
Thrombostasin*		*Haematobia irritans* (Horn fly)	Q95P73
Putative secreted collagen-like peptide*		*Ixodes scapularis* (Black-legged tick) (Deer tick)	Q4PMW3, Q4PN68
Putative secreted protein*		*I. scapularis*	Q4PN03 Q4PN66
Putative uncharacterized protein*		*Parvimonas micra* ATCC 33270	A8SJY5
Fn-binding protein A	FnbA	*Staphylococcus aureus* (strain bovine RF122 / ET3-1)	Q2YW62
Fn-binding protein A	FnbA	*S. aureus* (strain MRSA252)	Q6GDU5

Description	Gene name	Species	UniProt/UniParc ID(s)
Fn-binding protein A	FnbA	*S. aureus* (strain MSSA476)	Q6G6H3
Fn-binding protein	FnbB	*S. aureus* (strain MSSA476)	Q6G6H4
Fn-binding protein A	FnbA	*S. aureus* (strain Mu3 / ATCC 700698)	A7X6I5
Fn-binding protein precursor		*S. aureus* (strain USA300 / TCH1516)	UPI00016995F0
Fn-binding protein B		*S. aureus* (strain USA300 / TCH1516*)*	UPI00016995FF, UPI0001699608
Fn-binding protein A	FnbA	*S. aureus* (strain USA300)	Q2FE03
Fn binding protein A	FnbA	*S. aureus* subsp. *aureus*	A8W7G3
Fn binding protein A	FnbA	*S. aureus*	A1KDI9, A1KDJ1, A1KDS5, A7DY70, A7DY71, A7DY73, A7DY75, A7DY78, A7DY80, Q2UWU9, Q2UWV0 Q2UWW9, Q2UWX5, Q6EVM1
Fn binding protein B	FnbB	*S. aureus*	Q53682, Q6EVM0
Fn-binding protein		*S. aureus*	Q1PDE6
Fn-binding protein	Fnb	*S. aureus*	Q9AEP9
Fibronectin binding protein	FnBA	*Streptococcus dysgalactiae*	Q06556
Fn binding protein	FnBB	*S. dysgalactiae*	Q53971
Fn-binding protein	FneB	*Streptococcus equi* subsp. *equi*	Q5EF36
Fn-binding protein	Fne	*S. equi* subsp. *equi*	Q93ED6
Fn-binding protein Fnz	Fnz	*S. equi* subsp. *zooepidemicus* (strain MGCS10565)	B4U102
Fn-binding protein	Fnz	*S. equi subsp. zooepidemicus*	Q2HYU5

Description	Gene name	Species	UniProt/UniParc ID(s)
Fn-binding protein	Fnz2	*S. equi subsp. zooepidemicus*	Q76K22
Fn-binding protein	Fnz	*S. equi*	P72416
Fn-binding protein	Fnb	*Streptococcus equisimilis*	Q54099
Fn-binding protein	gbfA	*S. equisimilis*	Q6EWI8
Fn-binding protein		*S. pyogenes* serotype M12 (strain MGAS2096)	Q1J9J8
Fn binding protein 1	sfb1	*S. pyogenes* serotype M28	Q48VN7
Fn-binding protein		*S. pyogenes* serotype M4 (strain MGAS10750)	Q1J4G3, Q1J8U7
Methyl-accepting chemotaxis protein*		*S. pyogenes* serotype M49 591	UPI00004C2603
Fn-binding protein		*S. pyogenes* serotype M6	Q5XE71
Fn binding protein SfbX	sfbX	*S. pyogenes*	Q934I3
Fn/fibrinogen binding protein F	prtF15	*S. pyogenes*	O33709
Fn-binding protein I	sfbI	*S. pyogenes*	Q01924, Q711A7, Q711A8, Q711B0 Q711B3, Q711B6, Q711B7, Q711B8 Q711B9, Q711C0, Q711C1, Q711C3 Q711C6, Q711C9, Q711D0, Q711D2, Q711D3, Q711D5, Q711E2, Q711E4 Q711E8, Q711F0, Q711F2, Q711F3, Q711F6
Fn-binding protein II	sfbII	*S. pyogenes*	P72532
Protein F	prtF	*S. pyogenes*	Q54953
Protein F	prtF1.12	*S. pyogenes*	Q8RP54
Protein F2	prtF2	*S. pyogenes*	P72534, A2T8Z5, Q6IZ30, Q6IZ34, Q6IZ43

Description	Gene name	Species	UniProt/UniParc ID(s)
Serum opacity factor VT2.2		*S. pyogenes*	Q93T52
Serum opacity factor VT21		*S. pyogenes*	Q93T51
Serum opacity factor VT3.1		*S. pyogenes*	Q93T53
Serum opacity factor VT3.2		*S. pyogenes*	Q93T54
Serum opacity factor VT37.1		*S. pyogenes*	Q93T50
Serum opacity factor	sof2	*S. pyogenes*	O84941
Serum opacity factor	sof63	*S. pyogenes*	Q9L953
Serum opacity factor	sof1207	*S. pyogenes*	Q9RFJ4
Serum opacity factor	sof2967	*S. pyogenes*	Q9RPZ2
Serum opacity factor	sof9	*S. pyogenes*	Q9S3T2
Serum opacity factor	sof3875	*S. pyogenes*	Q9S4J3
Serum opacity factor	sof44	*S. pyogenes*	Q9S4K0
Serum opacity factor	sof49	*S. pyogenes*	Q9X2V2
Serum opacity factor	sof	*S. pyogenes*	Q9X3R6
Group G streptococcal Fn binding protein	gfbA	*Streptococcus sp. group G*	Q47942

4.1.3 Conserved residues identified with the HMM search

The conserved residues in FnBRs were identified as follows. The final s100 HMM was used to search for FnBRs in the FnBP sequences in Table 31, and it identified 417 FnBRs. Of these, 173 were unique, and these unique FnBRs were used to build a new HMM called 'FnBR-final', and this was then represented as a HMM 'logo' to visualize the conserved residues in FnBRs (Figure 34; Schuster-Bockler *et al.*, 2004). It can be seen from this 'logo' that there is a highly conserved glycine in the ^{3}F1-binding region (HMM position 21) and a highly conserved 'E-(D/E)-(T/S)' motif (HMM positions 32–34) in the ^{2}F1-binding region. The other two motifs which are moderately conserved are a hydrophobic-(E/D)-hydrophobic motif and another E-(D/E)-(T/S) motif, both in the ^{4}F1-binding region (HMM positions 9–11 and 13–15, respectively). All these regions make interactions with F1-modules in the Fn:FnBR crystal structures (Bingham *et al.*, 2008) and will be discussed in Section 4.4.4.6.

4.2 Crystallization of SfbI-5 peptides with F1-module pairs

Attempts were made to crystallize both ^{2}F1^{3}F1 and ^{4}F1^{5}F1 with appropriate SfbI-5 peptides; that is, PyTT5 (SfbI residues 560–577) and PyFF5 (SfbI residues 542–560), respectively. If successful, these crystal structures would complement the mutation data, reported in Section 4.3, in helping to identify functionally important residues within SfbI-5. Attempts with ^{4}F1^{5}F1 and PyFF5 were unsuccessful, and will not be discussed further. The successful crystallization of ^{2}F1^{3}F1 with PyTT5 is reported below.

4.2.1 Crystallization of ^{2}F1^{3}F1:PyTT5 at pH 5.6

Initial attempts to crystallize ^{2}F1^{3}F1:PyTT5 resulted in a 'hit' from the Index screen (Section 2.7.2). Within four days of setting up the crystal tray, small split crystals growing amongst micro crystals and crystalline precipitate had appeared in Index screen condition 93 (50 mM zinc acetate, 20% w/v polyethylene glycol (PEG) 3350; Figure 35A). These crystals were optimized by screening a range of PEG concentrations and pH. Since there was no buffer in the original condition, pH was screened by the addition of 100 mM citric acid bis-tris propane buffer to the condition. A data set was kindly collected by Dr Johan Turkenburg and Mr Sam Hart (YSBL, Chemistry Department, University of York) at the ESRF from crystal with the highest quality diffraction of those tested (Section 2.7.4). This crystal was grown in 100 mM citric acid bis-tris

propane pH 5.6, 50 mM zinc acetate, 14% w/v PEG 3350 (Figure 35B). The resolution of this data set was 2.6 Å. It was processed and integrated in MOSFLM (Leslie, 2006), scaled with SCALA (Evans, 2006) and then solved by molecular replacement with Phaser (McCoy *et al.*, 2007; Sections 2.7.5, 2.7.6).
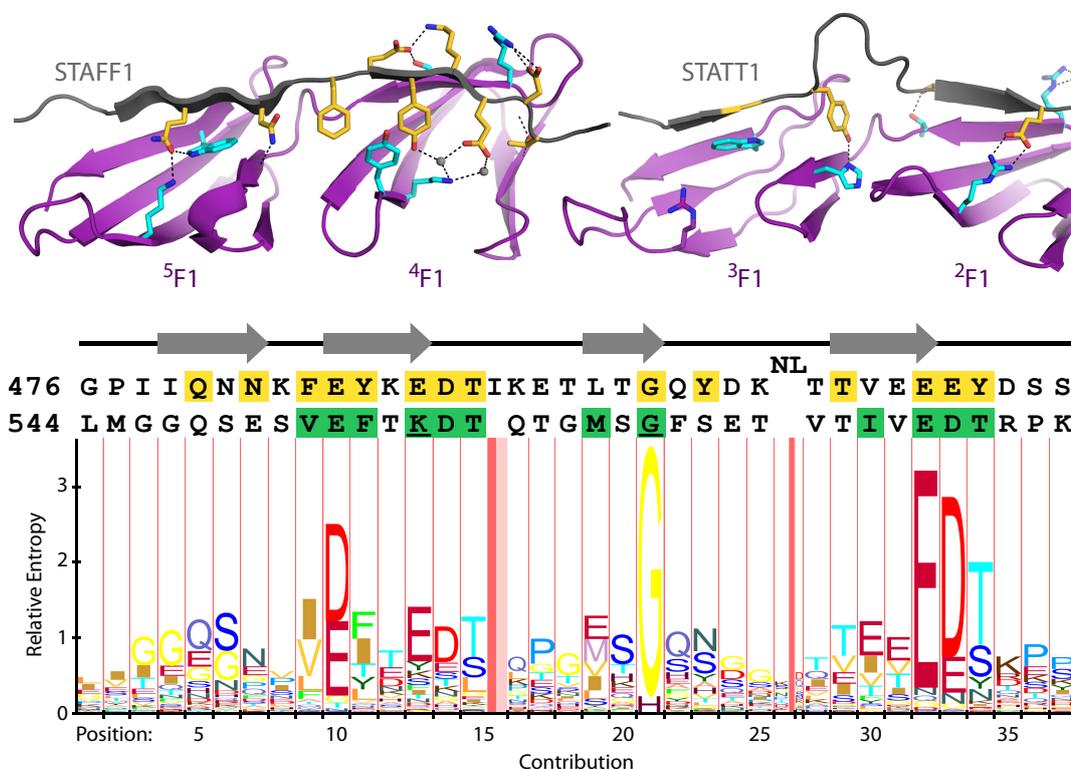
Figure 34. HMM 'logo' of FnBRs identified with an iterative HMM search of the UniRef90 database. Each logo position represents a Match state in the HMM of an MSA (Figure 8 on p 51). Column height is proportional to the information content of that position, relative to the assumed background. Residue height is proportional to the emission probability of that residue at that position (Figure 8B), ordered from most to least probable. The probability of a deletion is represented by the width of the stack: a narrower column corresponds to a higher probability of a deletion at that position, *i.e.,* a higher $M_n \rightarrow D_{n+1}$ transition probability (Figure 8B). Insertions are represented by light and dark pink areas. The width of the dark pink area represents the probability of at least one residue being inserted after the adjacent match state, *i.e.,* the $M_n \rightarrow I_n$ transition probability (Figure 8B). The width of a pink area (light and dark) represents how many residues are likely to be inserted, *i.e.* the $I_n \rightarrow I_n$ transition probability (Figure 8B).

Shown above the logo are the sequences of FnBPA-1 and SfbI-5 aligned to the columns in the logo. The crystal structures of ²F1³F1 and ⁴F1⁵F1 (purple) bound to FnBPA-1 peptides (grey) are shown in ribbon representation above the sequences (Bingham *et al.*, 2008). Key FnBPA-1 and F1-module sidechains important in the interaction are shown in stick representation and are coloured yellow or cyan, respectively. Hydrogen bonds are shown as black dashed lines and two bridging water molecules as grey spheres. The sidechain of R125 in ³F1, which is referred to in Section 4.3.1, is shown coloured purple. The important FnBPA-1 sidechains are also highlighted yellow in the FnBPA-1 sequence, and SfbI-5

residues chosen for mutation to alanine are highlighted green. Two SfbI-5 residues, K556 and G564, are underlined: these were also mutated separately to glutamate and glutamine, respectively. Regions of FnBPA-1 that form β-strand when bound to F1 modules are indicated above the sequences by grey arrows. The logo was drawn with the LogoMat-M server, www.sanger.ac.uk/Software/analysis/logomat-m/ (Schuster-Bockler *et al.*, 2004).
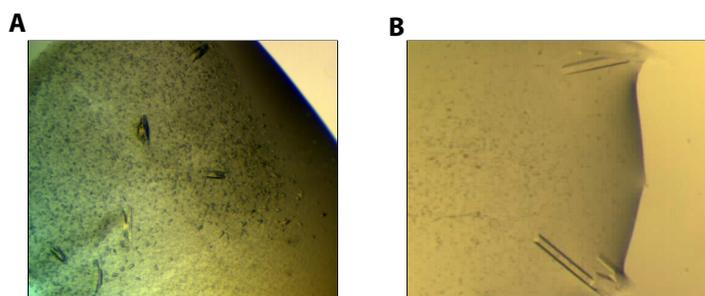
Figure 35. Photos of initial and optimised conditions with crystals of ^{2}F1^{3}F1:PyTT5. A. Initial 'hit' with ^{2}F1^{3}F1:PyTT5. The starting drop condition was a 1:1 mixture of well solution (50 mM zinc acetate, 20% w/v PEG 3350) and ^{2}F1^{3}F1:PyTT5 solution (300 μM ^{2}F1^{3}F1, ~3.5 mM PyTT5, pH 7.7). **B.** The optimized starting drop condition was a 1:1 mixture of well solution (0.1 M citric acid bis-tris propane pH 5.6, 14% w/v PEG 3350, 50 mM zinc acetate) and ^{2}F1^{3}F1:PyTT5 solution (as in A).

Initial attempts at molecular replacement in Phaser were unsuccessful with a model based on the ^{2}F1^{3}F1:STATT1 structure (pdb code 2rkz) with the peptide and some of the larger loops in ^{2}F1^{3}F1 deleted or a ^{2}F1 model. Molecular replacement was successful using Phaser with a ^{2}F1^{3}F1 model based on a model from a BALBES run (Long *et al.*, 2008; Section 2.7.6). BALBES is a molecular replacement pipeline that searches for structures in the Protein Databank with similar sequences to the sequence of the input data. The BALBES result used separate search models for ^{2}F1 and ^{3}F1; both models were based on the ^{2}F1^{3}F1 structure with pdb code 2cg6 (Rudiño-Piñera *et al.*, 2007). This ^{2}F1^{3}F1 model has a significantly different domain orientation than ^{2}F1^{3}F1 in the ^{2}F1^{3}F1:STATT1 structure (Figure 36E), which is probably the reason why some of the initial molecular replacement attempts were unsuccessful; a molecular replacement solution was also found in Phaser using separate ^{2}F1 and ^{3}F1 search models.
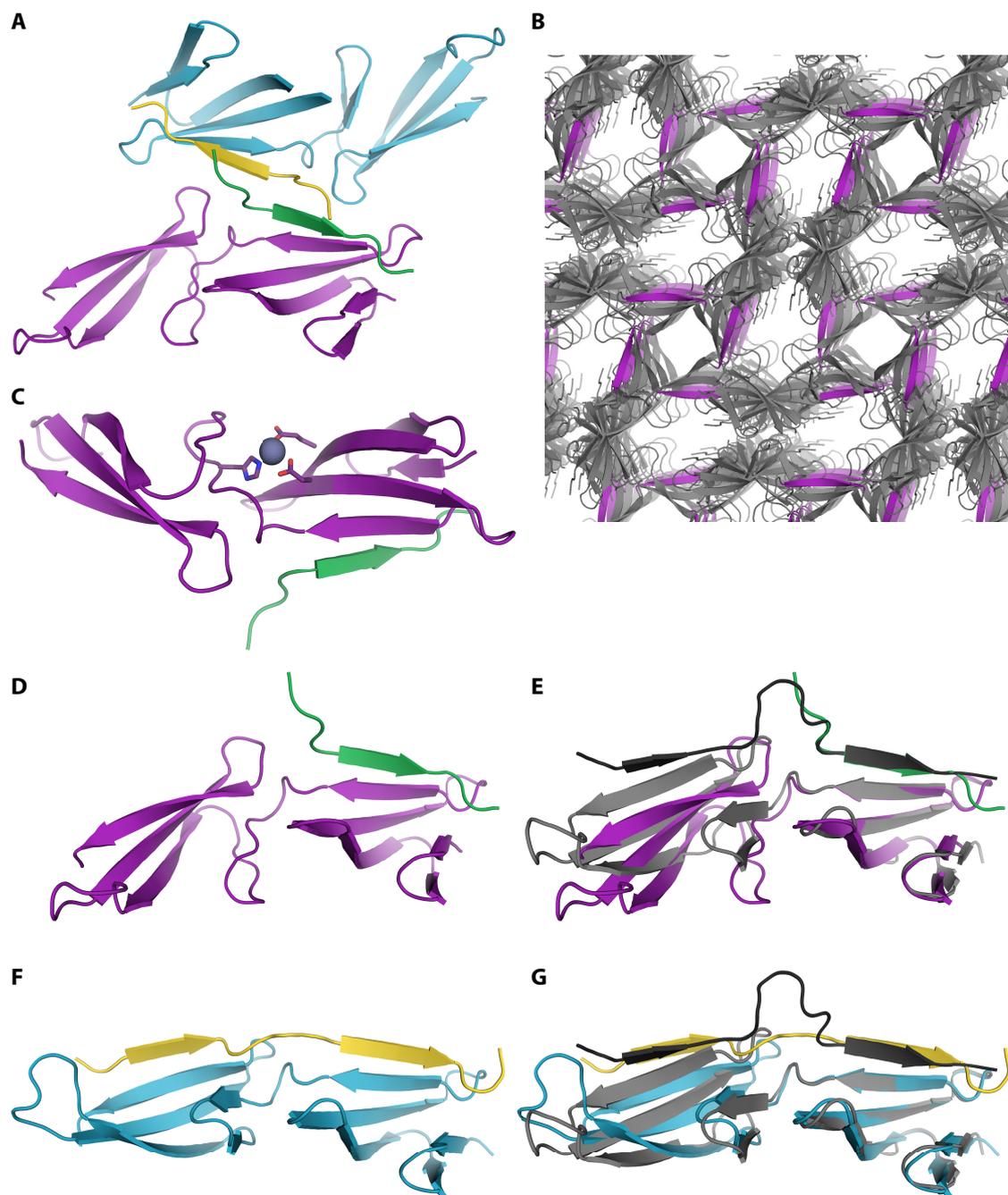
Figure 36. Representations of ²F1³F1:PyTT5 structures I and II. A. The asymmetric unit of Structure I. **B.** Crystal packing in Structure I. The E-strand of ³F1, highlighted in purple, is next to a solvent channel in all ²F1³F1 molecules and is not involved in any crystal contacts. **C.** Chelation of zinc by E82, D91, on ²F1 and H111 on ³F1 in Structure I. Zinc chelating sidechains are shown in stick representation and the zinc atom as a grey sphere. A fourth site is occupied by water (not shown). **D.** One of the two ²F1³F1:PyTT5 complexes in the Structure I asymmetric unit. PyTT5 (green) is binding along the E-strand of ²F1 in ²F1³F1 (purple). **E.** Comparison between Structure I (green and purple) and ²F1³F1:STATT1 (dark and light grey) aligned by residues in ²F1. The relative orientation of ³F1 in the two structures differs. **F.** One of the two ²F1³F1:PyTT5 complexes in the asymmetric unit in Structure II (Section 4.2.2). PyTT5 (yellow) binds along the E-strands of both modules in ²F1³F1 (blue). **G.** Comparison between ²F1³F1:PyTT5 Structure II (blue and yellow) and ²F1³F1:STATT1 (dark and light grey) aligned to each other by residues ²F1. The relative orientations of ³F1 with respect to ²F1 in the two structures are similar.

The initial model revealed that there are two ^{2}F1^{3}F1 molecules per asymmetric unit (Figure 36A). Importantly, unmodelled electron density for PyTT5 was clearly visible along ^{2}F1. This model was improved by iterative model-building and refinement in Coot (Emsley & Cowtan, 2004) and Refmac 5 (Murshudov *et al.*, 1997), respectively (Section 2.7.7). Based on the position of unmodelled sidechain density, the orientation of PyTT5 was identified as antiparallel. To avoid phase bias, this peptide density was originally modelled as alanines, and the sidechains were then built in. The register of the peptide was determined from the bulkier side chains. Also, key interacting PyTT5 sidechains were deleted and remodelled to avoid phase bias. The structure has been refined to give an R_{work} of 22.2 % and an R_{free} of 27.5% (Structure I, Table 32; Figure 36).

The most striking feature of the crystal structure is that PyTT5 is only bound to ^{2}F1 (Figure 36D). The reason for this does not appear to be crystal packing, since there are no crystal contacts blocking the E-strand of ^{3}F1 (Figure 36B). The reason could be a change in module orientation of ^{2}F1 with respect to ^{3}F1 when compared to the ^{2}F1^{3}F1:STATT1 structure (Figure 36E), which could be caused by the chelation of a zinc atom at the F1-module pair interface by E82 and D91 in ^{2}F1 and H111 in ^{3}F1 (Figure 36C). That the ^{2}F1-binding region rather than the ^{3}F1-binding region of PyTT5 binds ^{2}F1^{3}F1 in this structure would be predicted based on the conserved residues in the two F1-binding regions, and also from the ITC experiments of Schwarz-Linek *et al.* (2004b). In their work, they showed that the SfbI-5 ^{2}F1-binding region bound with micromolar affinity, but binding of the SfbI-5 ^{3}F1-binding region to ^{2}F1^{3}F1 was too weak for data fitting. Given that H111 is at the domain interface, it is possible pH also contributes to F1-module orientation. The other two ^{2}F1^{3}F1:FnBR structures are at close to neutral pH. Since PyTT5 does not bind ^{3}F1 in this structure, further crystallography was attempted with PyTT5, focussing on conditions without divalent cations and at neutral pH.

Table 32. Statistics for three crystal structures of ^{2}F1^{3}F1 bound to PyTT5. Highest resolution shell data is shown in brackets where applicable. Crystallization was achieved using a sitting drop in which the condition shown was mixed with a solution of 0.3 mM ^{2}F1^{3}F1, ~3.5 mM PyTT5, pH 7.7 at a ratio of 1:1. Abbreviation: HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

Parameter	Structures		
	I	II	III
Crystallization			
Condition	0.1 M citric acid bis-tris propane pH 5.6, 14% (w/v) PEG 3350, 50 mM zinc acetate	20 mM MgCl$_2$, 0.1 M HEPES pH 7.5, 22% (w/v) polyacrylic acid	0.1 M Tris pH 7.0, 1.5 M (NH$_4$)$_2$SO$_4$
Cryoprotectant	20% glycerol	20% glycerol	30% glycerol
Data collection			
Wavelength (Å)	0.9765	1.542	0.9804
Crystal to detector distance (mm)	360.5	250.0	233.8
Exposure time (s)	0.5	240	0.5
Oscillation angle (°)	0.5	0.5	0.5
Number of images	360	360	300
Space group	P2$_1$2$_1$2$_1$	P4$_1$	P4$_1$
Cell dimensions	a = 64.38 Å, b = 69.54 Å, c = 75.73 Å, α = β = γ = 90°	a = b = 50.04 Å, c = 92.78 Å, α = β = γ = 90°	a = b = 50.06 Å, c = 93.50 Å, α = β = γ = 90°
Resolution (Å)	51.23–2.60 (2.74–2.60)	50.06–2.90 (3.06–2.90)	28.22–1.70 (1.79–1.70)
Number of unique reflections	10946 (1556)	4995 (714)	25333 (3669)
Mean I/σ(I)	16.6 (3.9)	20.3 (5.8)	22.4 (3.6)
R$_{sym}$	0.102 (0.425)	0.109 (0.359)	0.054 (0.406)
Completeness (%)	99.9 (99.9)	97.7 (96.7)	100.0 (100.0)
Redundancy	6.8 (7.1)	7.6 (7.7)	6.1 (6.1)
Wilson B-factor (Å^2)	52.4	51.3	22.6
Refinement			
Resolution (Å)	51–2.6 (2.677–2.600)	93–2.9 (2.976–2.900)	26–1.7 (1.744–1.700)
R$_{work}$	0.222 (0.281)	0.185 (0.245)	0.177 (0.268)
R$_{free}$	0.275 (0.403)	0.297 (0.434)	0.225 (0.308)

Parameter	Structures		
	I	II	III
Number of atoms			
^{2}F1^{3}F1 (chains A, B)	1303	1332	1360
PyTT5 (chains C, D)	80	258	280
Water	23	1	236
B factors			
^{2}F1^{3}F1 (Å^2)	33.02	24.14	22.02
PyTT5 (Å^2)	32.43	27.02	21.48
Water (Å^2)	25.92	9.62	34.93
R.M.S. deviations			
Bond lengths (Å)	0.016	0.015	0.012
Bond angles (°)	1.898	1.638	1.268

4.2.2 Crystallization of ^{2}F1^{3}F1:PyTT5 at pH 7.0

Crystals were obtained in three different conditions at approximately neutral pH (Figure 37), and crystals from all these conditions were tested in-house to determine the quality of their diffraction pattern; that is, the resolution and mosaicity of the crystal and whether two diffraction images 90° apart could be indexed. All crystals were indexed, and they all had space group P41 and the same cell dimensions (a = b = 50 Å, c = 93 Å, α = β = γ = 90°, Table 33), indicating that they probably would have similar packing, and therefore, similar structures. Interestingly, the only similarity between all conditions was neutral pH (Table 33); indicating pH has a significant effect on crystallization of this complex.

Figure 37. Photos of crystals of ^{2}F1^{3}F1:PyTT5 at approximately neutral pH. All drops were a 1:1 mixture of well solution with ^{2}F1^{3}F1:PyTT5 solution (0.3 mM ^{2}F1^{3}F1 ~3.5 mM PyTT5, pH 7.7). **A. – C.** corresponds to crystals A – C (Table 33), respectively.

Table 33. A comparison of ^{2}F1^{3}F1:PyTT5 crystals showed that multiple conditions gave highly similar cell dimension and the same space group. Also shown are the crystallization time and diffraction of the crystals when shot at the YSBL and, where applicable, the ESRF. Photos of crystals are in Figure 37.

Crystal	Conditions	Crystallization time	Diffraction (Å)		Cell dimensions	Space group
			YSBL	ESRF		
A	20 mM MgCl$_2$, 0.1 M HEPES pH 7.5, 22% w/v polyacrylic acid	1 month	2.9	–	a = b = 50.04 Å, c = 92.78 Å, α = β = γ = 90°	P41
B	0.1 M Tris pH 7.0, 1.5 M (NH$_4$)$_2$SO$_4$	9 days	2.3	1.70	a = b = 50.06 Å, c = 93.50 Å, α = β = γ = 90°	P41
C	0.1 M Tris pH 7.0, 1.8 M Li$_2$SO$_4$	9 days	2.3	–	a = b = 50.09 Å, c = 93.10 Å, α = β = γ = 90°	P41

Given that PyTT5 did not bind ^{3}F1 in ^{2}F1^{3}F1:PyTT5 Structure I, a dataset was collected from crystal A in-house to confirm that PyTT5 was binding both F1-modules of ^{2}F1^{3}F1. Although the crystallization conditions for crystal A contained divalent cations, the fact that the cell dimensions and space group were the same as crystals B and C, which were not crystallized in conditions with divalent cations, meant that it was unlikely that the magnesium would be affecting F1 module orientation (Table 33).

This low resolution data set was solved to a resolution of 2.9 Å (Structure II, Table 32) by searching separately with models of ^{2}F1 and ^{3}F1 from ^{2}F1^{3}F1:STATT1 after removal of the peptide, major loops and the residues linking the two modules (Section 2.7.6). The model from molecular replacement showed that the module orientation had changed, and was now more similar to the other ^{2}F1^{3}F1 crystal structures (Figure 36G). Also, there was clear density corresponding to PyTT5 bound across both F1 modules. To avoid phase bias, PyTT5 was modelled in as alanines, and then the sidechains were built. The orientation and register of PyTT5 were determined by fitting the bulky sidechains: it was anti-parallel, with the same register as Structure I in the ^{2}F1-binding region. Therefore, Structure II shows PyTT5 bound along the triple-stranded β-sheets of

both F1 modules as anti-parallel β-strands — a tandem β-zipper interaction (Figure 36F; Section 1.4.1).

However, the resolution of Structure II was not sufficient to reveal the role of the residue sidechains. Therefore, crystallization conditions which gave crystals A, B and C were further optimized by varying the precipitant and pH ranges of each condition. The diffraction patterns of the resultant crystals were acquired in-house, and data was collected at the ESRF by Dr Johan Turkenburg and Mr Sam Hart for the four best crystals, including crystal B. It was crystal B that gave the best dataset, and it was subsequently solved to a resolution of 1.70 Å using phases from Structure II without peptide as a starting point for refinement. Again, there was clear unmodelled density that corresponded to PyTT5 (Figure 38C). For this structure, the strand orientation of PyTT5 could be determined as anti-parallel from the backbone carbonyl and amide groups and then confirmed using the sidechains. The final structure was refined to give an R-factor of 17.7% and an R_{free} of 22.5% (Structure III, Table 32; Figure 38).

Validation of Structure III was performed with the MolProbity server (Davis *et al.*, 2007). This server evaluates protein geometry on a number of criteria: the distribution of residues in the favourable, allowable and unfavourable regions of the Ramachandran plot; C^{β} deviations; number of unfavourable rotamers, angles and bonds. It also evaluates the contacts between all atoms, including hydrogen atoms that it models in, and reports a 'clashscore' based on the number of sterically unfavourable contacts per 1000 atoms. It also reports a 'MolProbity' score which combines the clashscore with a score for the protein geometry, and for each score ranks the input structure compared to other structures of similar resolution. Structure III had a clashscore of 2.87 and a MolProbity score of 1.29, which placed it in the 99th and 97th percentile, respectively, for structures in the resolution range 1.45–1.95 Å (the 100th percentile indicates structures with the best scores). Structure III is analysed in Section 4.4.4
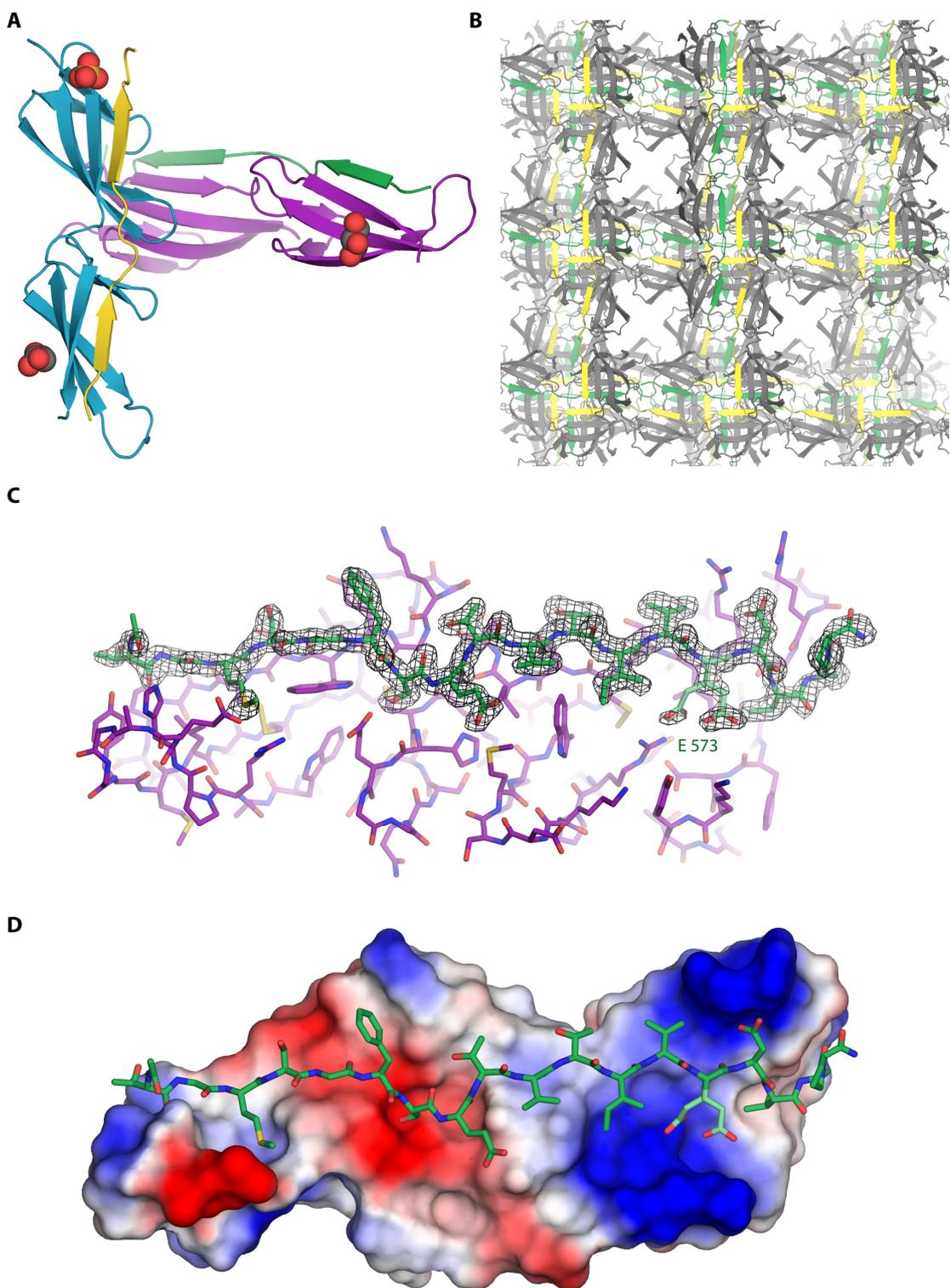
Figure 38. Representations of ^{2}F1^{3}F1:PyTT5 Structure III. A. Asymmetric unit of Structure III. Three additional ligands are shown as spheres: two glycerol molecules (grey) and sulfate group (sulfur in yellow). **B.** Crystal packing of Structure III, with ^{2}F1^{3}F1 shown in grey and the two PyTT5 molecules in each asymmetric unit in green and yellow. **C.** PyTT5 (green) binding to ^{2}F1^{3}F1 (purple) showing the difference density at a level of 1 sigma (grey) when the PyTT5 molecule is deleted. **D.** Electrostatic surface representation of ^{2}F1^{3}F1 shaded from negative (red) through neutral (white) to positive (blue). PyTT5 is shown in stick representation (green). E573A exists in two conformations, which are shown as pale green in both D and E.

4.3 Effect of mutating conserved FnBR residues on SfbI-5 function

4.3.1 Residues chosen for mutation

Having identified conserved FnBR residues from the HMM analysis described in Section 4.1, it was interesting to investigate the effect of mutating conserved residues on SfbI-5 binding to pNTD and Fn. This work aimed to elucidate which conserved residues were most important for binding, and whether residue importance could be predicted from the existing crystal structures (Bingham *et al.*, 2008).

The SfbI-5 residues chosen for mutation were those in the conserved motifs identified in Section 4.1.3, and are shown in comparison to the HMM logo in Figure 34. In addition to the most conserved residues, two less conserved residues, which were predicted to interact with Fn *via* backbone-backbone hydrogen bonding, were also mutated. This was for comparison with the more conserved motifs.

The chosen residues were individually mutated to alanine, to investigate how removal of the sidechain affected binding. The conserved glycine G564 was mutated to alanine to determine the effect of a methyl group on this residue. G564 appears to be conserved primarily to facilitate a β-strand interaction with tryptophan W146 in the E-strand of ^{3}F1 by minimizing steric hindrance with the bulky sidechain of W146 (Figure 39A; Bingham *et al.*, 2008).

G564 was also mutated to glutamine; the rationalization for this mutation is as follows. ^{3}F1 is homologous to ^{5}F1, and ^{5}F1 also has a tryptophan, W237, at a position homologous to W146 in ^{3}F1 (Figure 39). However, the ^{5}F1-binding region of SfbI-5 and of most other FnBPs do not have a glycine at the position corresponding to G564 in the ^{3}F1-binding region, but have a glutamine instead (HMM position 5, Figure 34). In the ^{4}F1^{5}F1:STAFF1 complex, this glutamine (Q516) reaches across W237 in ^{5}F1 and forms hydrogen bonds with both K217 and W237 in ^{5}F1 (Figure 39). The sidechain χ1 angle of W237 is different when compared to W146, which facilitates this interaction (Figure 39). It was possible, therefore, that the χ1 angle of W146 might change to accommodate a similar interaction between W146 and R125 in ^{3}F1 and a glutamine at position 546 in the ^{3}F1-binding region of SfbI-5. If the χ1 angle can change, it is predicted that binding to pNTD would be tighter in the G564Q mutant than in the G564A mutant; if the χ1 angle does not change, binding should be weaker.

175

Also, K556 was mutated to glutamate, because although this position is a lysine in some streptococcal FnBRs (including all those in SfbI), it is a glutamate in most FnBRs (HMM position 13, Figure 34). Thus, a total of 14 mutants were created to analyse the effect of residue conservation on binding to pNTD and Fn: 12 mutations to alanine, one to glutamate and one to glutamine. These mutants were named, for example, K556A, to indicate residue K556 in SfbI was mutated to alanine in SfbI-5.
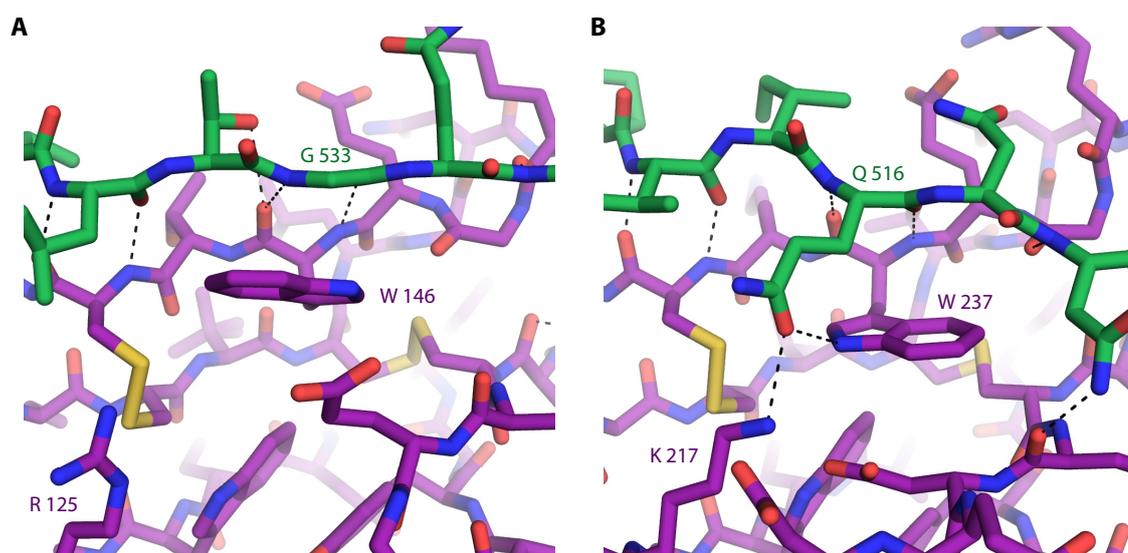
Figure 39. The χ1 angle of a conserved tryptophan differs between ^{3}F1 and ^{5}F1 in the complexes between FnBPA-1 peptides and either ^{2}F1^{3}F1 or ^{4}F1^{5}F1. A. and **B.** Either ^{3}F1 or ^{5}F1 (purple), respectively, bound to FnBPA-1 peptides (green). The conserved tryptophan residues in ^{3}F1 and ^{5}F1 are indicated (W146 and W237, respectively), as are other residues discussed in the text.

4.3.2 Creating, expressing and purifying mutants

All mutants were created from site-directed mutagenesis of pGEX-SfbI-5 as confirmed by DNA sequencing in both the forward and reverse directions (data not shown; Section 2.2.3).

The mutants and wild-type SfbI-5 were expressed and purified in three groups (Table 34). Group 1 *E. coli* cultures were induced to overexpress protein while growing at 37°C. The cultures grew to an OD_{600} of ~1.5 within one hour of induction. Cultures were grown for a further three hours; however the OD_{600} of the cultures remained approximately the same or even decreased slightly (Table 35; Figure 40). Given the comparative success and good yields of the cultures overexpressing ^{15}N, ^{13}C-labelled GST-SfbI-5 at 30°C (Section 3.2), Group 2 cultures were induced to overexpress

protein while growing at 30°C and were then grown for a further 17 hours to a final OD_{600} of ~4 (Table 35; Figure 40). Importantly, the amount of soluble GST-fusion expressed per cell was approximately constant regardless of expression conditions (compare 'S' lanes in Figure 41). Therefore, the higher cell density in Group 2 cultures corresponded directly to a higher yield of purified GST-SfbI-5 construct (Table 35). However, the purity of the GST-SfbI-5 construct in Group 2 was lower than in Group 1, with a significant contaminant migrating just below the GST-fusion (Figure 41B, C, D). This contaminant appears to be mainly insoluble (Figure 41B), but some remains soluble and is present in the GST-fusion eluates (E1, E2 lanes, Figure 41C, D). The contaminant is either an N- or C-terminal truncation of the GST-fusion that retains an intact Factor Xa cleavage site, since it is still cleaved by Factor Xa (Figure 42, p 181). Its presence may lead to an overestimate of the GST-fusion yield by A_{280}. However, the relatively low level of this impurity in the GST-fusion eluates compared to the overall increase of yield in Group 2 means that the final yield of pure SfbI-5 construct after removal of GST-fusion was still higher in Group 2. Therefore, inducing overexpression of GST-SfbI-5 constructs in *E. coli* grown at 30°C rather than at 37°C significantly improved the yield of the purified construct.

Given the improved yield of GST-SfbI-5 constructs from Group 2 with respect to Group 1, Group 3 constructs were overexpressed under similar conditions to Group 2, except that cultures were induced at a higher OD_{600} and were grown for 16 hours after induction. Cultures in Group 3 grew to a significantly higher final OD_{600} than Group 2 (Table 35). However, this did not lead to higher yields (Table 35), indicating that the average soluble GST-SfbI-5 construct overexpressed per cell had decreased. Therefore, a higher final cell density during overexpression of a GST-SfbI-5 construct does not always result in a higher yield.

Table 34. SfbI-5 and mutant constructs were expressed and purified in the three groups shown. The two mutants with an asterisk (*) were expressed and purified in Group 3 but their binding to pNTD was not determined because of limited availability of pNTD.

Group	Constructs
1	SfbI-5, K556A, K556E, E573A, D574A and T575A
2	V552A, E553A, F554A, D557A and T558A
3	Q548A*, M562A, G564A, G564Q, I571A and double mutant (K556E, T558A)*

Table 35. Comparison of yield and various growth conditions during the overexpression of SfbI-5 mutant Groups 1, 2 and 3. The minimum and maximum values within a group are reported for each growth condition for a 1L culture of each mutant. The yield was calculated for all mutants using the extinction coefficient and MW of GST-SfbI-5 (Section 2.3.6). Only one of the two cultures of mutant F554A expressed GST-fusion, so its values were not included when determining the values for Group 2.

Property	Group 1	Group 2	Group 3
Overexpression temperature (°C)	37	30	30
OD_{600} on transfer to 30°C	N/A	0.38–0.52	0.55–0.73
OD_{600} on induction with IPTG	1.0–1.1	0.8–1.1	1.0–1.4
Final OD_{600}	1.4–1.6	3.4–4.2	4.8–5.1
Yield from 1st affinity step (mg)	11–14	27–34	27–39
Yield from 2nd affinity step (mg)	6–8	13–23	16–19
Total yield of GST-SfbI-5 construct (mg)	17–21	43–57	43–57
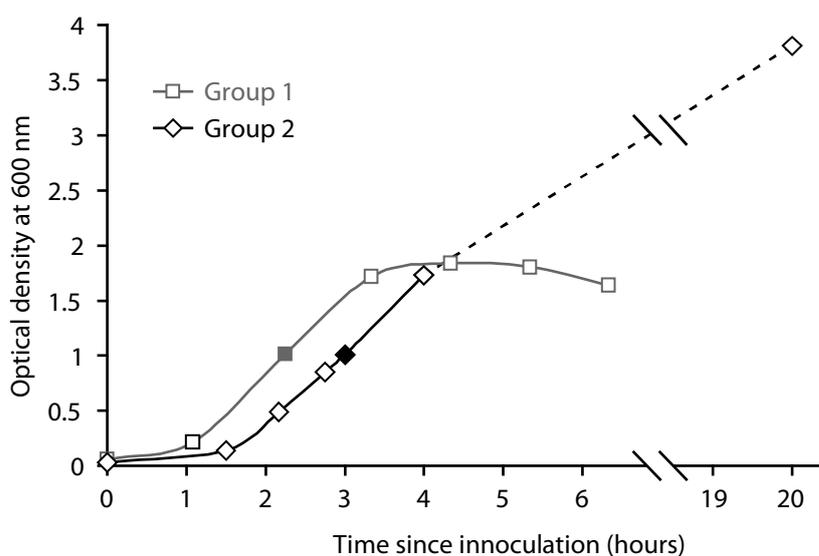
Figure 40. Growth curves of *E. coli* overexpressing a GST-SfbI-5 mutant either at 37°C from Group 1 or at 30°C from Group 2. Culture growth was monitored by OD_{600}. The growth curves for SfbI-5 mutants K556E from Group 1 and V552A from Group 2 are shown. These two curves are both representative of the growth curves measured for all cultures in their respective groups. Group 2 cultures were transferred from 37°C to 30°C just after taking the second time point. The filled symbol indicates when the culture was induced. A dashed line is shown between the second-to-last and last time points for the Group 2 curve since time points were not taken regularly enough to predict the growth in between these times.
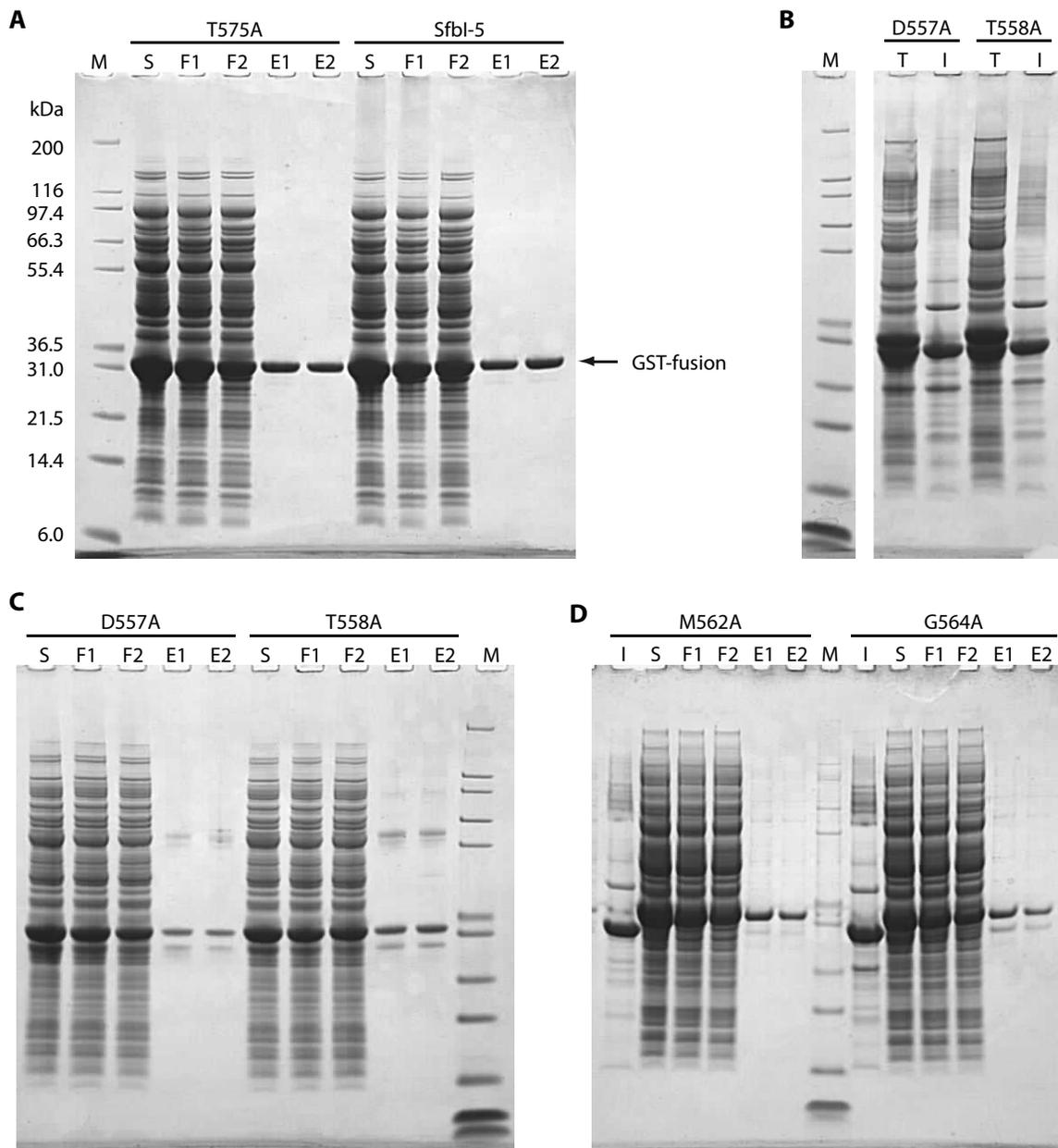
Figure 41. Analysis of the purification of GST-SfbI-5 and mutants from Groups 1, 2 and 3. Representative gels from each group are shown. The wild-type SfbI-5 or SfbI-5 mutant purified as a GST-fusion in these samples is indicated over the appropriate lanes. Lane contents are indicated by the following abbreviations: M, Mark 12 protein MW marker; S, soluble cellular protein; F1, flow-through from first purification of S; F2, flow-through from re-purification of F1; E1, pooled eluate fractions from first purification of S, E2, pooled eluate fractions from re-purification of F1;T, total cellular protein; I, insoluble cellular protein (Table 9). **A.** Purification of two of the GST-fusions from Group 1, the MWs of the marker are shown, and so are the bands corresponding to the GST-fusion. These bands appear in similar positions in B, C and D. **B.** Total and insoluble cellular protein from two of the GST-fusions purified in Group 2. **C.** and **D.** Purification of the GST-fusions from Group 2 or Group 3, respectively.

GST-SfbI-5 constructs were purified from soluble *E. coli* protein with two glutathione affinity chromatography steps; the second step was worthwhile because it gave a

significant increase in yield. The constructs were purified by a batch method (Section 2.3.4.2) rather than by fast performance liquid chromatography because this is more efficient when purifying multiple constructs. SDS PAGE analysis showed that for all purifications there was a significant amount of overexpressed GST-fusion that did not bind to the glutathione agarose and remained in the flow-through after the first affinity chromatography step ('F1' lanes, Figure 41). Therefore, for all purifications the flow-through from the first affinity chromatography step was re-purified with a second affinity chromatography step (Figure 41; Section 2.3.4.3). This second step was worthwhile because the yield was ~60% the yield of the first step, resulting in a significant increase on overall yield (Table 35).

After purification of the GST-SfbI-5 constructs, Factor Xa protease was used to cleave GST from the GST-SfbI-5 construct using the reaction time and protease concentration determined for ^{13}C, ^{15}N-labelled GST-SfbI-5 (Section 3.2). All cleavage reactions were close to 100% complete as shown by SDS PAGE (Figure 42). The SDS PAGE analysis also showed a clear difference in migration for the various SfbI-5 mutants when compared to each other and wild-type SfbI-5 (Figure 42). Each SfbI-5 construct was purified from the cleavage reaction by rpHPLC with acidic conditions (Section 2.3.8). The peak shape for all mutants was similar to that of SfbI-5 (Figure 18), except for the K556A and K556E mutants, whose peaks had an unusually large tail (*e.g.,* Figure 43). For both the K556A and K556E mutants, the tails were repurified and they had the same migration as the mutant in the original rpHPLC purification (*e.g.,* Figure 43), indicating the majority of the tail was the main species, and it was therefore pooled with the main peak from the original rpHPLC purifications. After rpHPLC, appropriate fractions were pooled, lyophilised and yield was determined by weight by assuming that the freeze dried material consisted of 20% water by weight (Section 2.3.9). This correction was made based on experience in our laboratory with other FnBR peptides.

Figure 42. Analysis of the Factor Xa cleavage of GST-SfbI-5 and mutants. The SfbI-5 or mutant in these samples is indicated over the appropriate lanes. Lane contents are indicated by the following abbreviations: M, Mark 12 protein MW marker; − and + , samples before and after cleavage, respectively, of GST from the SfbI-5 mutant by Factor Xa. **A.** Factor Xa cleavage of two mutants from Group 1. MWs of the marker are shown, and so are the bands corresponding to GST-fusion, GST, and the SfbI-5 mutants. These bands appear in similar positions in B, C and D. **B.**, **C.** and **D.** Factor Xa cleavage of GST-SfbI-5 mutants from the remainder of Group 1, all of Group 2 and all of Group 3, respectively.

Figure 43. Acidic rpHPLC purification and repurification of the SfbI-5 K556E mutant.
A. Chromatogram from a purification of cleaved GST-SfbI-5 K556E mutant sample by acidic rpHPLC
with a gradient between water and acetonitrile. Trace colours are: light green, concentration of solvent B
(acetonitrile with 0.1% TFA); blue, absorbance at 220 nm; dark green, fractions collected during run.
Fractions of the K556E mutant collected for lyophilisation with similar fractions from other rpHPLC runs
are indicated with a light grey line. Fractions from the tail of the peak collected for repurification are
indicated with a dark grey line. These fractions were pooled with similar fractions from another run and
re-purified. **B.** Chromatogram from the re-purification of K556E tail fractions. The trace colours are the
same as in A.

The purity of the rpHPLC purified SfbI-5 constructs was analysed by SDS PAGE
(Figure 44). It can be seen that nearly all mutants have a contaminant migrating just
above and/or just below the purified mutant. These impurities are also visible in the
samples prior to rpHPLC (Figure 42). This means the rpHPLC method used was unable
to separate out these contaminants. As previously mentioned, the SfbI-5 constructs
migrate significantly differently in SDS PAGE despite differing by at most two amino
acids. This indicates that small changes in amino acid composition can have a relatively
large effect on the migration pattern peptides of this MW in SDS PAGE. It is difficult to

tell, therefore, whether these contaminating species have a smaller or larger MW with respect to the corresponding SfbI-5 construct. However, based on band intensity, the contaminants are less than 10% of the total sample. Given the likely losses in yield that would occur in attempting to remove these impurities, it was decided that this is an impurity level that is acceptable for subsequent biophysical characterization of SfbI-5 constructs.
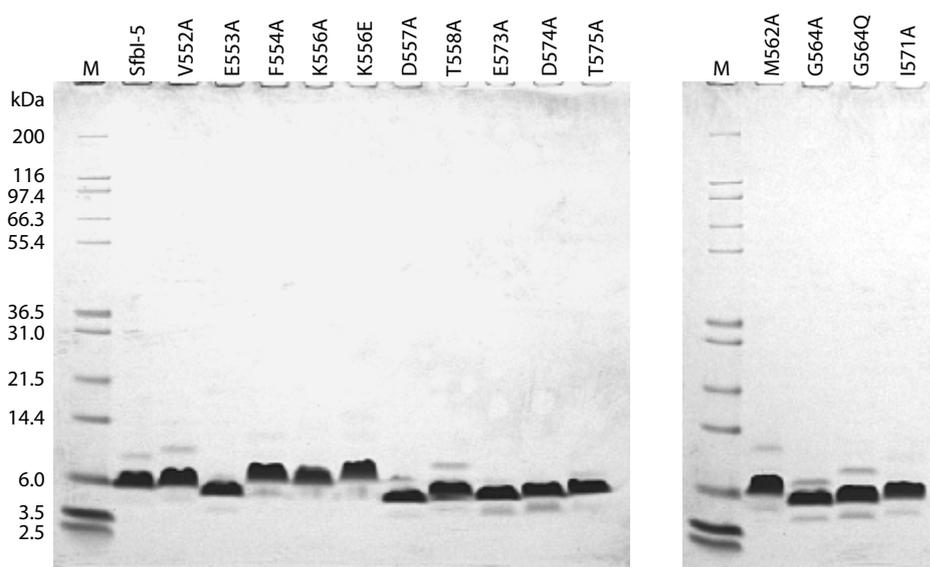
Figure 44. Analysis of purified SfbI-5 constructs. The SfbI-5 or mutant in these samples is indicated over the appropriate lanes. The MWs of the Mark 12 protein MW marker (M) are shown. This SDS PAGE analysis was kindly performed by Ms Dominika Gruszka (Potts group, Department of Biology, University of York).

The SfbI-5 mutations were confirmed by ESMS, all experimental MWs are within 0.8 Da of their respective theoretical MW, which is within experimental error (data not shown). Generally, there were no other significant species identified by ESMS other than potassium adducts (+38 Da). The exception to this was for the K556A and K556E mutants, which besides potassium adducts also contained two additional adducts, an additional 15 or 16 Da with and without potassium. This probably corresponded to oxidation of methionine, and is possibly a result of the difficulty in purifying these peptides by rpHPLC, leading to a longer time elapsing before these peptides were lyophilized. The other significant contaminant identified by ESMS was in the D557A sample. The contaminant's ESMS signal intensity was about 10% that of full length D557A and it had an experimental MW of 4287.3 Da. This is possibly a C-terminal truncation of D557A; the theoretical MW of D557A ending at L579 is 4286.8 Da. The

contaminants visible in SDS PAGE were not obvious in the ESMS results, indicating that these contaminants did not ionize similarly to the SfbI-5 constructs.

4.3.3 ITC of SfbI-5 mutants with pNTD

4.3.3.1 Rationale for ITC experimental design and data analysis

To investigate the effect of mutating conserved SfbI-5 residues on the binding of SfbI-5 to pNTD, ITC experiments were performed with SfbI-5 and with all SfbI-5 mutants at 37°C in PBS. These conditions are biologically relevant, and are also the conditions used previously by Schwarz-Linek *et al.* (2004b) for SfbI-5 binding to pNTD. For each SfbI-5 construct, an ITC experiment consisted of a binding titration and a control titration; the SfbI-5 construct titrated into PBS with or without pNTD, respectively. Furthermore, for the T575A mutant experiment an additional two buffer controls were performed; PBS into pNTD and PBS into PBS (Figure 45). In this experiment, the integrated heats throughout the three control titrations and at the end of the T575A binding titration are approximately equal to each other (Figure 45). This indicates that once the pNTD binding sites have been titrated with T575A, the majority of the measured heat from each injection arises from mechanical phenomena (Velazquez-Campoy *et al.*, 2004), such as the force of injecting the syringe solution into the cell, with no significant heat in these titrations arising from the dilution of SfbI-5 T575A or pNTD with buffer or from mixing SfbI-5 T575A with pNTD. Therefore, it appears either the heats at the end of the binding titration or the heats from the control titration are a suitable control for heat effects other than the binding of SfbI-5 constructs to pNTD.

Ideally, the same control for heat effects other than binding would be used for all ITC experiments. However, there were two sources of error that meant this was not possible. First, unlike the experiment shown in Figure 45, in many experiments the integrated heats in the control titration (SfbI-5 construct into PBS) were not similar in size to the integrated heats at the end of the binding titration (SfbI-5 construct into pNTD). This indicates the presence of experimental error that was not controlled for between the binding and control titrations. For these experiments, the end of the binding titration was the best control. Second, for some ITC experiments, the heats at the end of the binding titration continued to decrease at a rate which was not consistent with the rest of the data. This 'feature' of a binding titration was repeated within the corresponding

control titration; for these experiments the heats from the control titration were the best control. Furthermore, the repeat experiments for some SfbI-5 constructs showed that the mismatch between the control and the end of a titration, or titrations with 'features' were not specific to particular constructs, with one or the other occurring in only one of the two experiments for a construct (*e.g.,* Figure 46, experiments with SfbI-5, F554A and T558A). In fact, titrations with similar 'features' were performed in a similar time frame, indicating the problem was possibly mechanical in some way. Therefore, the control used for each ITC experiment was the one that gave the best fit to the experimental data, as judged by the χ^2 value obtained in non-linear least squares regression analysis to fit the single binding site model to the data (Table 18; Section 2.8.3).
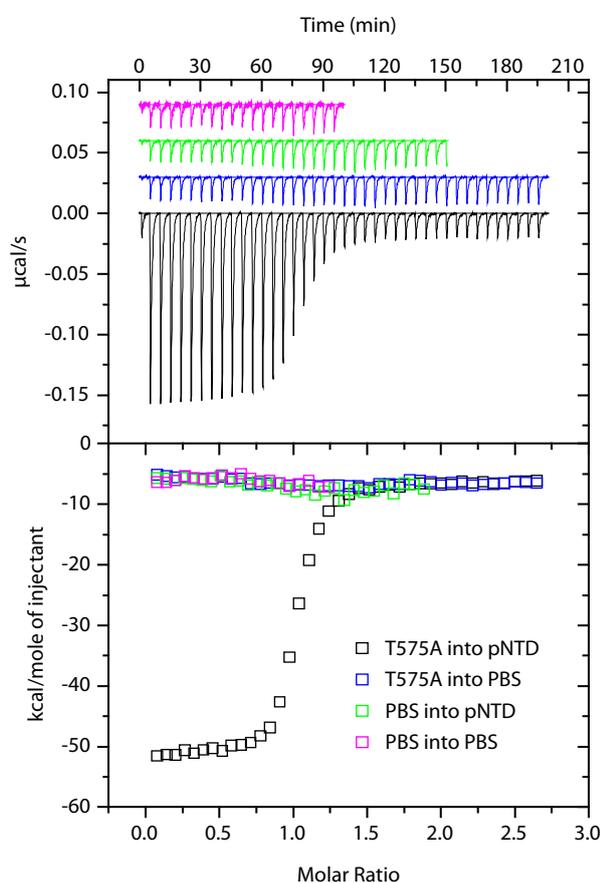
Figure 45. ITC experiments with SfbI-5 T575A and pNTD and corresponding controls. Top panel shows experimental traces from four titrations performed in PBS at 37°C: black, binding titration of SfbI-5 T575A solution titrated into pNTD solution; blue, control titration of SfbI-5 T575A solution titrated into PBS; green, control titration of PBS titrated into pNTD solution; and magenta, control titration of PBS titrated into PBS. The three controls are displaced relative to the binding titration by increments of −0.03 µcal/s. The bottom panel shows the processed molar heats of binding normalised to the ratio between SfbI-5 construct to pNTD. The same colouring scheme is used as in the top panel.

While fitting the ITC data to the single-site binding model, the SfbI-5 construct concentration was corrected so that the molar ratio of binding (n) was one, unless amino acid analysis was carried out on that sample. This correction was reasonable because when amino acid analysis was carried out on the sample n was always within 12% of 1.0 (Table 18). For some SfbI-5 constructs a large correction to the concentration was needed. However, Table 36 shows clearly that compared to amino acid analysis, setting n to one resulted in a more accurate estimate of the concentration than using the weight of the lyophilised peptides. A possible reason for this is that each construct may not have been completely soluble at the concentration of its stock solution, and thus, the assumed concentration of the stock solution was inaccurate, which led to an inaccurate concentration of the construct in the syringe. For example, a fine precipitate was observed in the stock solution of the K556A construct before dilution. It might also arise because of the impurities seen in the samples by SDS PAGE analysis (Figure 44).

Table 36. Comparison between concentration values of some SfbI-5 construct samples derived by different methods. Shown are the concentrations calculated for all ITC experiments involving an SfbI-5 construct where the concentration of the syringe sample in one of those ITC experiments was determined by amino acid analysis.

Construct	Titration	Concentration (µM) using different methods:		
		weight	setting n to 1	amino acid analysis
SfbI-5	1	30	25	–
	2	30	25	24
F554A	1	30	32	28
	2	30	30	–
K556A	1	30	22	21
	2	30	24	–
K556E	1	30	26	–
	2	30	27	24
T558A	1	30	24	–
	2	30	24	–
	3	60	41	42
D574A	1	30	25	–
	2	60	54	50

4.3.3.2 Is the effect of mutating conserved residues in SfbI-5 biologically relevant?

All the ITC experiments with pNTD and SfbI-5 constructs are shown in Figure 46, the thermodynamic parameters and their associated errors are reported in Table 37, and compared in Figure 47. Wild-type SfbI-5 bound to pNTD with a K_d of 3.5 nM, which is in good agreement with the 2 nM K_d measured previously for this interaction under the same conditions (Schwarz-Linek *et al.*, 2004b). The results from ITC analysis of SfbI-5 mutants binding to pNTD showed that all mutants had K_ds less than 55 nM (Table 37). Given the concentration of Fn dimers in the plasma is 0.65–0.87 µM (Mosher, 2006) and that the concentration of SfbI in the plasma is likely to be at least an order of magnitude less than this, then the effect of the mutation on the amount of SfbI-5 in complex with Fn can be estimated from Equation 15 (p 111) by assuming both that the law of mass action applies (Equation 14) and that the K_d for SfbI-5 binding to Fn in plasma is the same as the K_d for SfbI-5 binding to pNTD in PBS. The lower limit for the effective concentration of Fn is 0.65 µM, which assumes SfbI-5 forms a 1:1 complex with the Fn dimer for steric reasons, rather than a 2:1 complex. In this scenario, approximately 99.4% of the total amount of wild-type SfbI-5 would be in complex with Fn, whereas 91.4% of the weakest SfbI-5 mutant, T558A, would be in complex with Fn. This means that the mutation with the largest effect on SfbI-5 binding to pNTD is predicted to change the amount of SfbI-5 bound to Fn by only 8%, which is unlikely to confer a significant evolutionary advantage to *S. pyogenes*. If the effective concentration of Fn is actually 1.7 µM because SfbI-5 can bind to both molecules of Fn in a dimer, then the effect of the mutation is even smaller, at 3%.

The hypothesis behind these ITC experiments is that the function of SfbI-5 is to bind to $^{1-5}$F1 in Fn and, thus, the most likely reason conserved residues would be under selective pressure is to allow SfbI-5 to bind to Fn. That is, mutations in SfbI-5 that lead to a significant loss of binding to Fn would be selected against. However, the above calculations imply that mutation of the conserved FnBR residues in SfbI-5 would not significantly affect binding of SfbI-5 to Fn in the biologically relevant context of plasma. Therefore, either the assumptions in the calculations of the effect of K_d on complex formation are incorrect; or SfbI-5 residues are conserved for reasons other than binding to Fn; or the ITC experiments did not reflect the true effect of the mutations in a biological context. Hence, the effects of the mutations in the more biological context of SfbI-5 binding to Fn in plasma were measured (Section 4.3.4).

Table 37. Thermodynamic parameters measured with ITC for the interaction between pNTD and SfbI-5 mutants in PBS at 37°C. Abbreviations: K_d, dissociation constant; $\Delta H°$, change in enthalpy on binding; $\Delta S°$, change in entropy on binding; $\Delta G°$, Gibbs energy on binding; $\Delta\Delta G°$, change in Gibbs energy on binding of SfbI-5 mutants to pNTD with respect to SfbI-5 binding to pNTD. The associated errors for these parameters are also shown. However, because $\Delta S°$ is calculated from two dependent variables ($\Delta G°$ and $\Delta H°$), it is not possible to use a simple equation to calculate the associated error, so this error is not reported. The association constant and $\Delta H°$ for each ITC experiment performed with each construct are shown in Table 18, p 103.

Construct	K_d (nM)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal mol^{-1} K^{-1})	$\Delta G°$ (kcal/mol)	$\Delta\Delta G°$ (kcal/mol)
SfbI-5	3.5 ± 0.2	-45.4 ± 0.7	-108	-11.99 ± 0.04	0.00 ± 0.06
V552A	10.6 ± 0.8	-44.3 ± 0.2	-106	-11.30 ± 0.05	0.69 ± 0.06
E553A	8.4 ± 0.8	-45.2 ± 0.3	-109	-11.45 ± 0.06	0.54 ± 0.08
F554A	31.5 ± 4.9	-42.9 ± 0.7	-104	-10.63 ± 0.10	1.36 ± 0.10
K556A	1.6 ± 0.2	-45.7 ± 1.4	-107	-12.46 ± 0.08	-0.47 ± 0.09
K556E	0.8 ± 0.2	-49.0 ± 2.4	-116	-12.90 ± 0.18	-0.90 ± 0.18
D557A	13.8 ± 1.0	-43.2 ± 0.2	-103	-11.14 ± 0.05	0.85 ± 0.06
T558A	55.4 ± 2.7	-37.9 ± 1.6	-89.0	-10.29 ± 0.03	1.70 ± 0.05
M562A	8.2 ± 0.4	-40.0 ± 0.1	-92.2	-11.46 ± 0.03	0.53 ± 0.05
G564A	28.7 ± 2.1	-38.5 ± 0.2	-89.6	-10.69 ± 0.04	1.30 ± 0.06
G564Q	49.6 ± 2.5	-34.7 ± 0.3	-78.5	-10.35 ± 0.03	1.64 ± 0.05
I571A	7.7 ± 0.3	-44.5 ± 0.1	-106	-11.50 ± 0.03	0.49 ± 0.05
E573A	13.9 ± 0.8	-43.6 ± 0.2	-105	-11.14 ± 0.03	0.85 ± 0.06
D574A	41.0 ± 5.9	-46.3 ± 1.2	-116	-10.47 ± 0.09	1.52 ± 0.10
T575A	11.1 ± 0.5	-44.9 ± 0.4	-108	-11.27 ± 0.03	0.72 ± 0.05
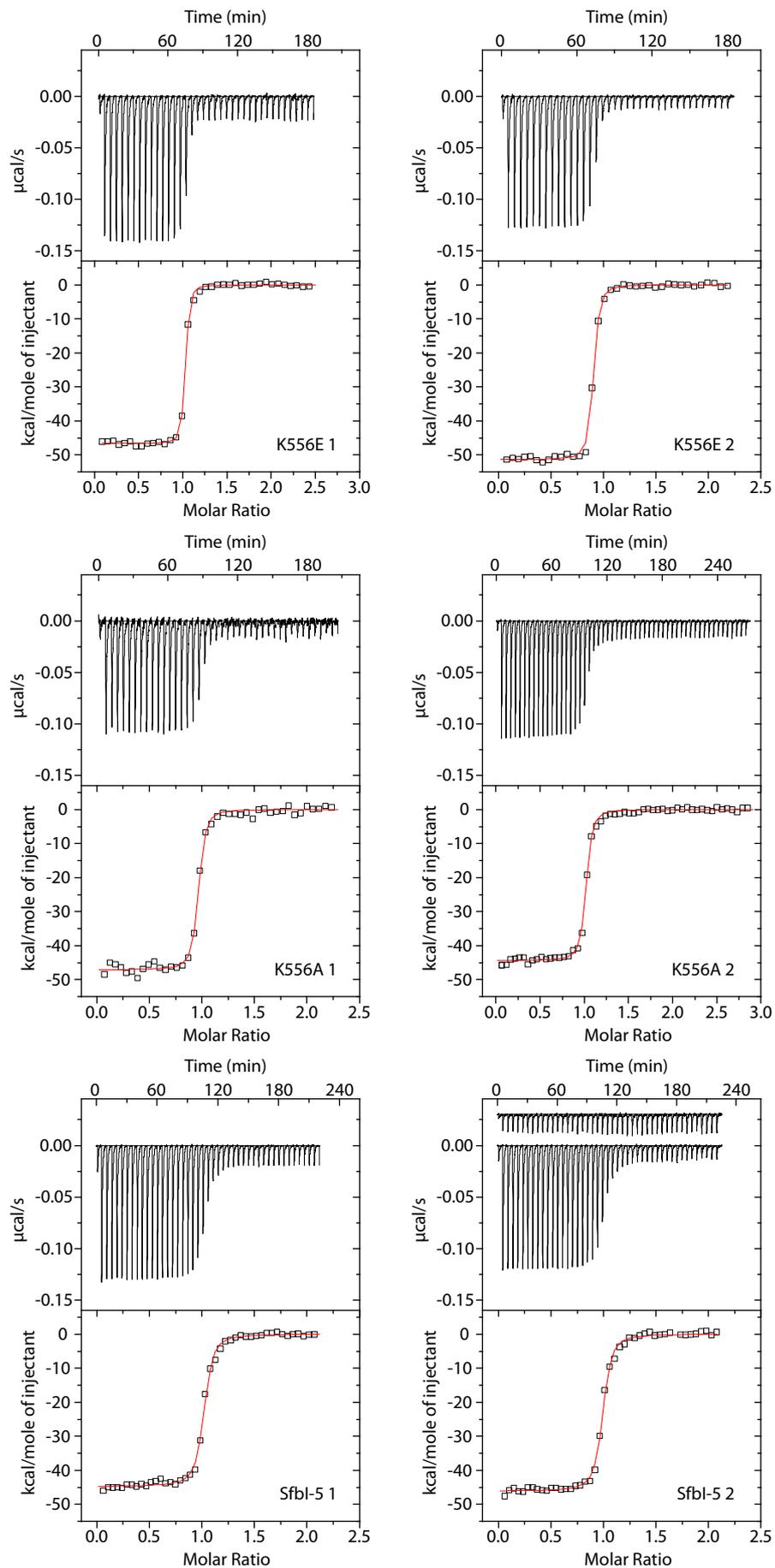
Figure 46. ITC experiments of pNTD with several SfbI-5 mutants in PBS at 37°C. (Part 1 of 4).

Legend on p193.
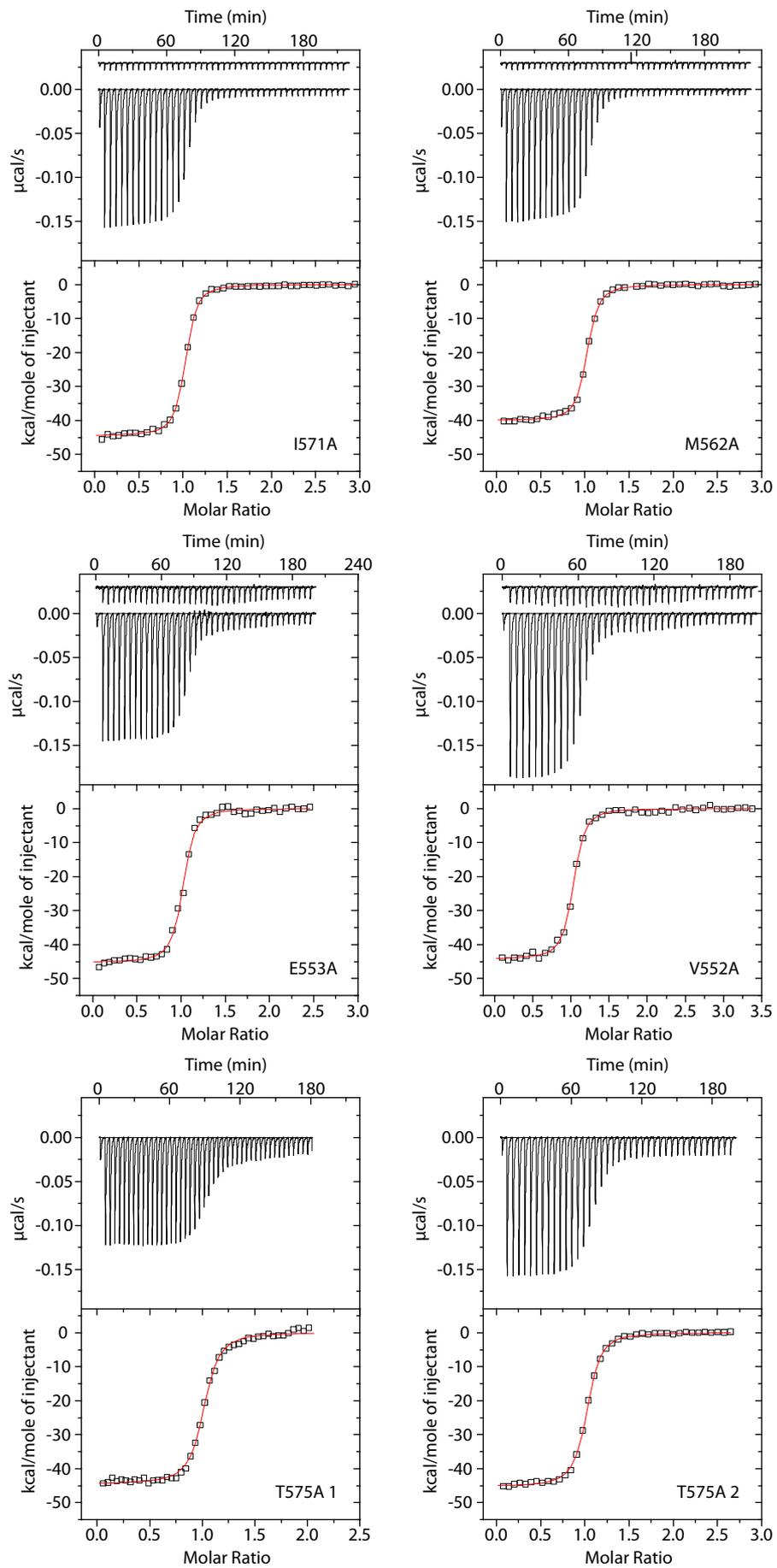
Figure 46. (Part 2 of 4). Legend on p193.
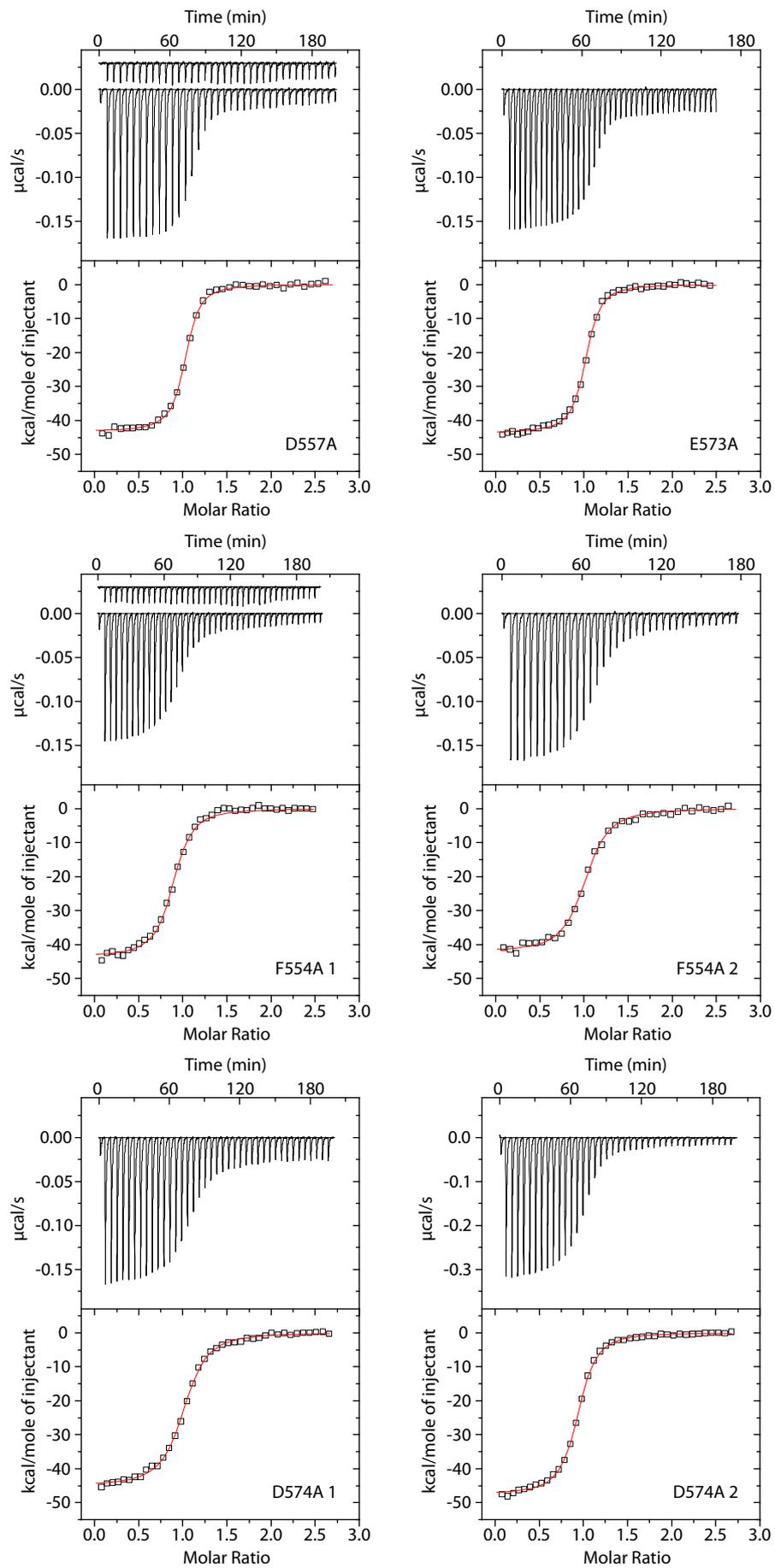
Figure 46. (Part 3 of 4). Legend on p193.
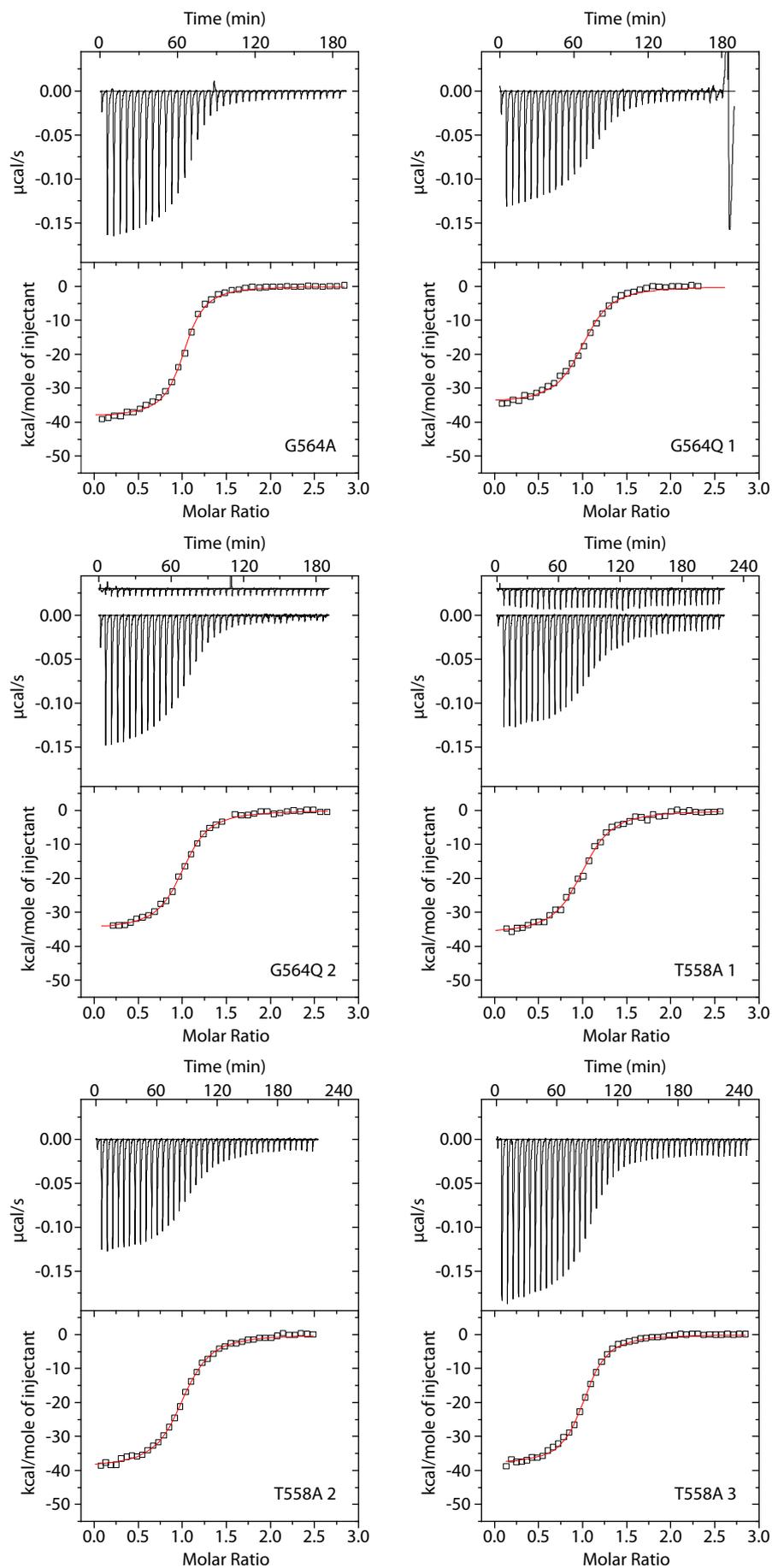
Figure 46. (Part 4 of 4). Legend on p193.

Figure 46. ITC experiments of pNTD with several SfbI-5 mutants in PBS at 37°C. The name of the SfbI-5 construct is indicated (with sequence numbering for full length SfbI). Top panel shows experimental trace as a solution of SfbI-5 construct was titrated into a pNTD solution. When a control titration of the same SfbI-5 construct into dialysis PBS was used as the control for that ITC experiment, it is also shown, but displaced by −0.03 μcal/s. When a control titration is not shown, the last five or ten data points were used as a control, as explained in Section 2.8.3. The bottom panel shows the processed molar heats of binding normalised to the ratio between SfbI-5 construct to pNTD. The curve from non-linear least squares regression analysis, using a single-site binding model (Section 2.8.3), is shown in red for each titration. Thermodynamic parameters derived from these experiments are in Table 37.
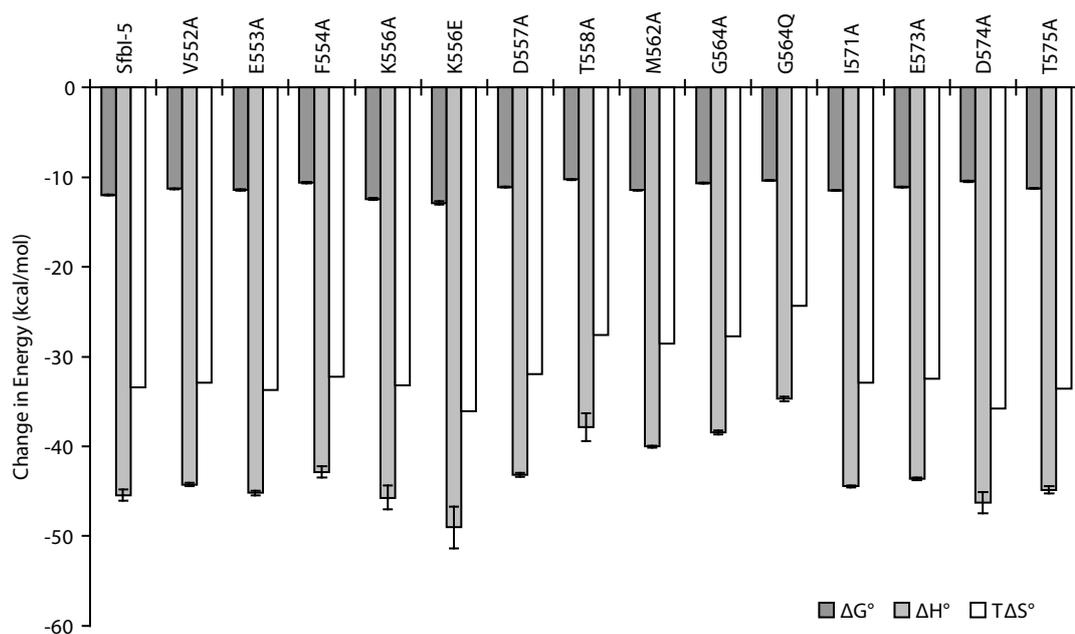
Figure 47. Change in Gibbs energy, enthalpy and entropy on binding of SfbI-5 constructs to pNTD. Abbreviations are $\Delta G°$, $\Delta H°$ and $\Delta S°$ for Gibbs energy, enthalpy and entropy, respectively. Note that the change in entropy has been multiplied by the temperature (T) for ease of comparison.

4.3.4 Effect of mutating conserved SfbI-5 residues in on binding to Fn in human plasma

4.3.4.1 Rationale

It is possible that the effect of mutating conserved FnBR residues in SfbI-5 on SfbI-5 binding to pNTD could differ significantly from their effect on SfbI-5 binding to Fn in the more biologically relevant context of human plasma. For example, full-length Fn or other molecules in plasma could be competing for the same [1−5]F1 binding site as SfbI-5. Therefore, to test whether these mutations would have a measurable effect on binding of SfbI-5 to Fn in plasma, a series of GST pull-down experiments were performed between GST-SfbI-5 constructs and human plasma (Figure 14; Section 2.9.1).

In this experiment, the overall aim was to measure semi-quantitatively the amount of Fn some of the GST-SfbI-5 constructs bound in human plasma by analysing the relative band intensity of Fn in pull-down samples on fluorescently stained SDS-PAGE gels. Fluorescent stain was used since it is sensitive and shows a linear correlation between staining and amount of protein over a large range (0.5–1000 ng; Berkelman, 2006). It was predicted that the experimental error with this method would be relatively high, but that if the binding of the mutants was significantly weaker to Fn in the context of human plasma compared to pNTD in PBS then the difference might still be detectable.

4.3.4.2 Preliminary GST pull-down experiments

Two preliminary experiments were performed to optimize various experimental parameters before doing the final pull-down experiment. First, GST-SfbI-5, seven GST-SfbI-5 mutants and GST were overexpressed in *E. coli* at 30°C for 18 hours (Section 2.9.1.1). Then, the GST constructs were purified with magnetic glutathione-linked particles. The amount of each GST construct bound to the particles was quantified by fluorescent SDS PAGE analysis (Figure 48) and then normalized to be 2 µg construct/µL particles (Section 2.9.1.2). This analysis showed that the GST-SfbI-5 constructs were mostly pure, except for a smaller contaminant migrating between the MW of GST-fusion and GST (Figure 48A). Furthermore, some GST-fusions, particularly D574A had a contaminant directly beneath the band corresponding to full-length construct (Figure 48A). When normalizing the amount of D574A based on band intensity this contaminant was not included.

Second, the ratio of GST-SfbI-5 construct to human plasma was optimized for maximum 'sensitivity,' that is, the maximum difference between amount of Fn bound by SfbI-5 and an SfbI-5 mutant. Sensitivity was optimized using two GST-SfbI-5 constructs: wild-type SfbI-5 and T558A, which is the mutant that had the highest K_d for binding to pNTD by ITC (Table 37). The ratio of GST-SfbI-5 construct to plasma Fn was changed by keeping the amount of GST construct bound to glutathione particles constant in all the pull-downs and varying the volume of human plasma (Section 2.9.1.4). This meant that only the concentration of SfbI-5 construct varied, and the concentration of components in human plasma remained constant.
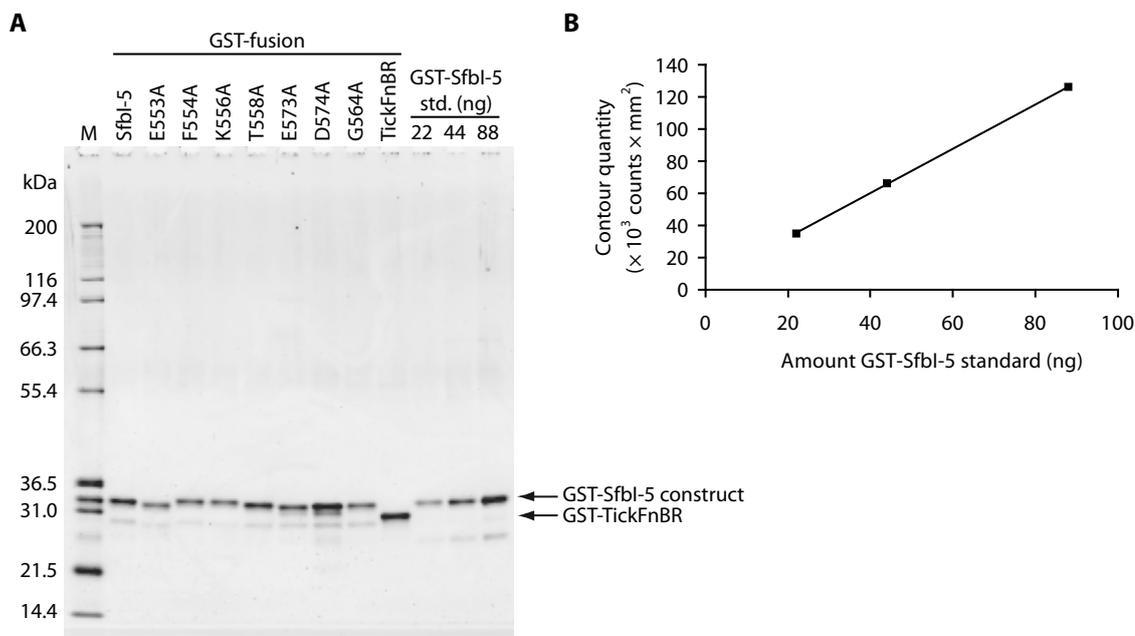
Figure 48. Analysis of purified GST-FnBR constructs. A. SDS PAGE analysis of GST-FnBR constructs purified with glutathione magnetic particles compared to a serial dilution of GST-SfbI-5 standard (std.) of known concentration based on $A_{280-320}$. The construct in each lane is indicated, labelled SfbI-5 for wild-type SfbI-5, or by the mutation with numbering relative to full-length SfbI-5. A GST-fusion of TickFnBR, a putative FnBR from *I. scapularis* investigated further in Chapter 5, was also purified and used in a pull-down. The MWs of the Mark 12 protein MW marker (M) are indicated, as is the position of bands corresponding to GST-SfbI-5 constructs and GST-TickFnBR. The marker was 'spiked' with GST-SfbI-5, which migrates between 31.0 and 36.5 kDa. The gel was stained with fluorescent stain and the band intensities were analysed (Section 2.9.1.3). **B.** The contour quantities of the bands in the serial dilution of GST-SfbI-5 were correlated with the known amounts of GST-SfbI-5 loaded on the gel using Quantity One software (Section 2.9.1.3). Based on band intensity, approximately 12% of the GST-SfbI-5 standard had partly degraded to a band migrating as the same MW as GST (*e.g.,* lane '88' in A). Therefore, a correction was made to the amount of GST-SfbI-5 in each standard calculated from $A_{280-320}$. The correlation coefficient for this analysis was 0.9999. This standard curve was used to determine the concentration of each construct bound to glutathione particles and the construct-bound particles were then standardized to 2 μg construct/μL particles by the addition of glutathione particles with no protein bound. SDS-PAGE analysis and standardization of GST was performed similarly but on a different gel (data not shown).

The minimum and maximum volume of plasma used in this experiment was calculated assuming the law of mass action (Equation 14) applied even though SfbI-5 would be bound to glutathione particles *via* GST and not free in solution. Initial calculations revealed that maximum sensitivity would theoretically occur when the molar ratio between the SfbI-5 construct and Fn was one to one (Section 2.9.1.4.1). Given this and also that the Fn dimer concentration range in plasma is 0.65–0.87 μM (Mosher, 2006),

the minimum volume of plasma was calculated assuming SfbI-5 could bind bound both molecules in a Fn dimer such that the upper concentration limit of Fn monomer in plasma was 1.7 µM. It was also considered possible that for steric reasons SfbI-5 might only bind one molecule of Fn in the blood, so the lower concentration limit used in calculations was 0.65 µM of Fn. These calculations resulted in a range of 360–900 µL plasma to be covered in this experiment. Additionally, 270 µL plasma was also tested in this experiment. This is because it was likely that the calculated amount of GST-SfbI-5 present on the beads was an upper limit, since some bound GST-SfbI-5 might be inactive or inaccessible to the Fn for steric reasons.

The SDS-PAGE analysis from this experiment indicated that both SfbI-5 constructs specifically bound Fn, that the background was low, and that neither the GST-bound particles nor particles by themselves bound Fn (Figure 49A). Furthermore, analysis of the band intensity indicated that the smallest volume of human plasma (270 µL) gave maximum sensitivity (Figure 49B). Based on the theoretical calculations discussed above, this result indicates that GST-SfbI-5 was probably able to bind both molecules in a Fn dimer, and that the concentration of GST-SfbI-5 able to bind Fn was less than 2 µg/µL particles. With this volume of plasma, T558A bound approximately 21% less Fn than SfbI-5. Thus, this preliminary experiment confirmed that SfbI-5 and a SfbI-5 mutant bound significantly different amounts of Fn, and that this technique was sensitive enough to detect the difference. In addition, it showed SfbI-5 and SfbI-5 T558A specifically bound Fn, because neither construct bound significant amounts of other plasma proteins (Figure 49A).

4.3.4.3 Analysis of the ability of GST-SfbI-5 constructs to bind to Fn in plasma

The final pull-down experiment was performed in triplicate (experiments 1–3), analysed by SDS PAGE and the amount of Fn and GST-SfbI-5 construct present in each pull-down determined by analysing fluorescent band intensity (Figures 50, 51). All GST-SfbI-5 constructs specifically bound similar amounts of Fn, and neither GST-bound particles or free particles interacted with Fn (Figure 50). Thus, no SfbI-5 mutation stopped SfbI-5 binding to Fn being detected. This is in agreement with the ITC results.
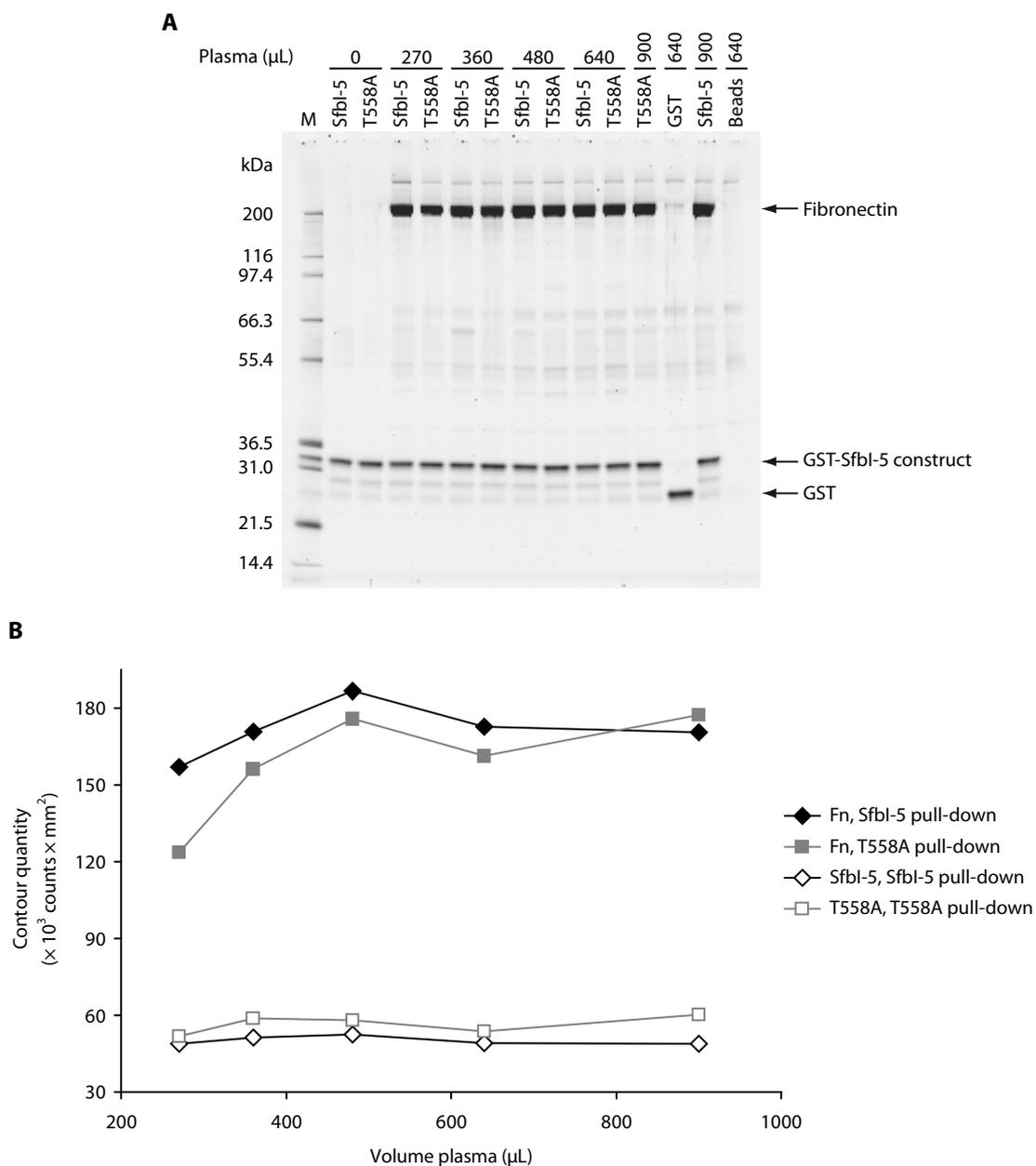
Figure 49. Analysis of GST-SfbI-5 and GST-SfbI-5 T558A pull-downs with varying volumes of human plasma. A. SDS PAGE analysis of GST-SfbI-5 and GST-SfbI-5 T558A pull-downs each performed with 20 µg GST-fusion bound to 10 µL glutathione magnetic particles and varying volumes of plasma. The construct in each pull-down and the volume of plasma used are indicated above the corresponding lanes in the gel. The two controls were GST with no FnBR fusion (GST) or glutathione magnetic particles with no protein bound (Beads). The MWs of the Mark 12 protein MW marker (M) are indicated, as is the position of bands corresponding to Fn, GST-SfbI-5 constructs and GST. The marker was 'spiked' with GST-SfbI-5, which migrates between 31.0 and 36.5 kDa. The gel was stained with fluorescent stain and the band intensities were analysed as described in Section 2.9.1.3. **B.** The contour quantity of the bands on the gel corresponding to Fn (closed symbols) or GST-SfbI-5 construct (open symbols) for all SfbI-5 and T558A pull-downs (black or grey, respectively) were determined using Quantity One software (Section 2.9.1.3).
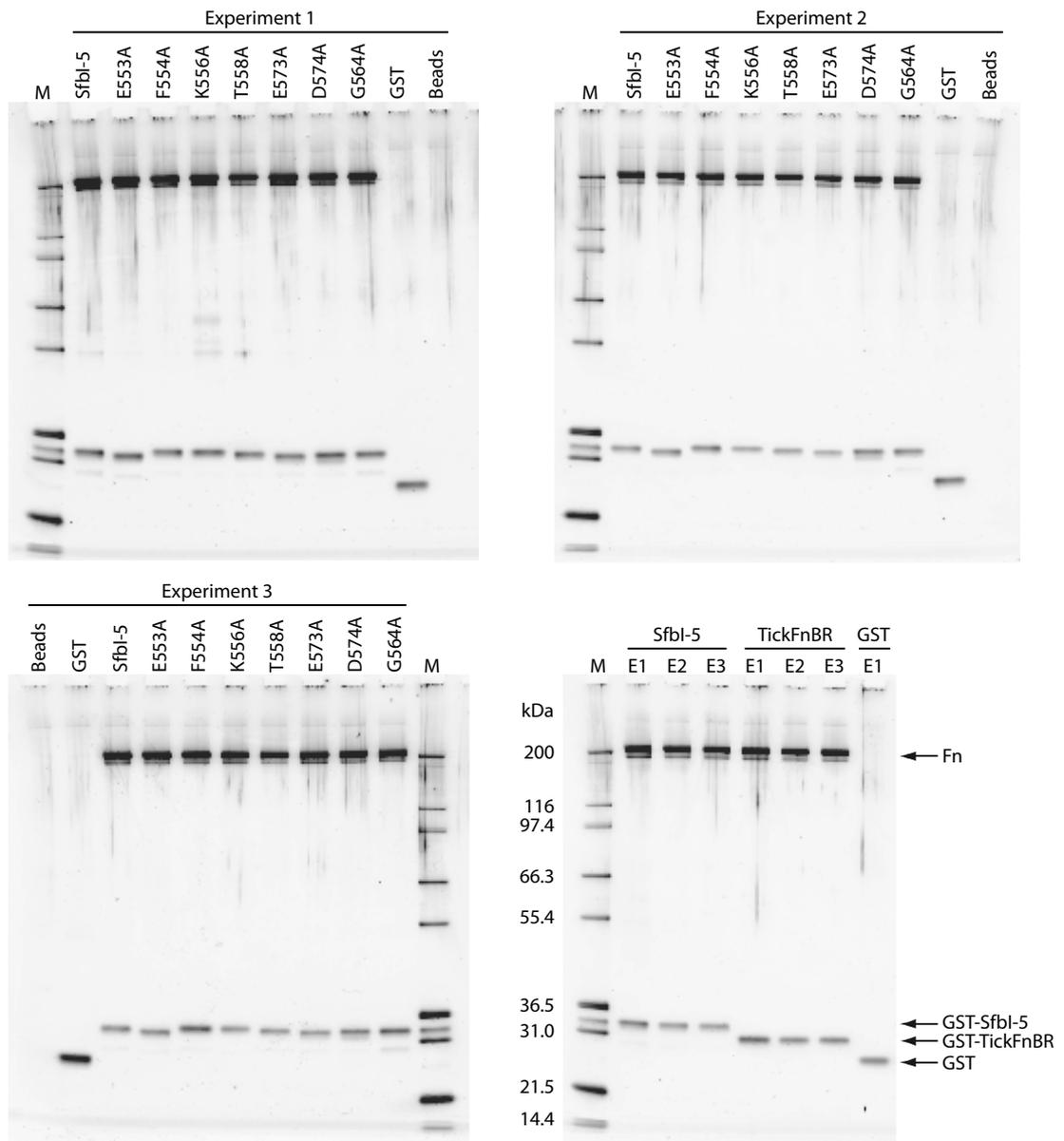
Figure 50. SDS-PAGE analysis of GST pull-down of Fn from human plasma by GST-FnBR constructs. Each experiment consisted of a group of 11 pull-downs, including controls, and was performed in triplicate (Experiments 1–3, or E1–E3). The 'prey' in each pull-down was Fn in human plasma; the 'bait' in each pull-down varied, and is indicated above each lane (Figure 14). In all cases the SfbI-5 construct was a GST-fusion, and is labelled as either SfbI-5 for wild-type SfbI-5, or by the mutation with numbering relative to full-length SfbI-5. The two controls were GST with no FnBR fusion (GST) or glutathione magnetic particles with no bait attached (Beads). A GST-fusion of TickFnBR, a putative FnBR from *I. scapularis* investigated further in Chapter 5, was also bait in a pull-down. It will be discussed in Chapter 5 but is shown here for comparison with the other pull-downs. Each gel also has Mark 12 protein MW marker that have been 'spiked' with GST-SfbI-5 (M). To help evaluate error in the analysis of band intensity, the SfbI-5 samples from each experiment were loaded a second time on the bottom right gel (E1–E3). Also indicated on this gel are the MWs of the marker and bands corresponding to Fn, GST-SfbI-5, GST-TickFnBR and GST. These bands occur in similar positions in the other gels. All gels were run at the same time, stained with fluorescent stain, and photographed in a single image (Section 2.9.1.3). Then, the Fn and GST-FnBR construct band intensities were analysed (Figure 51).
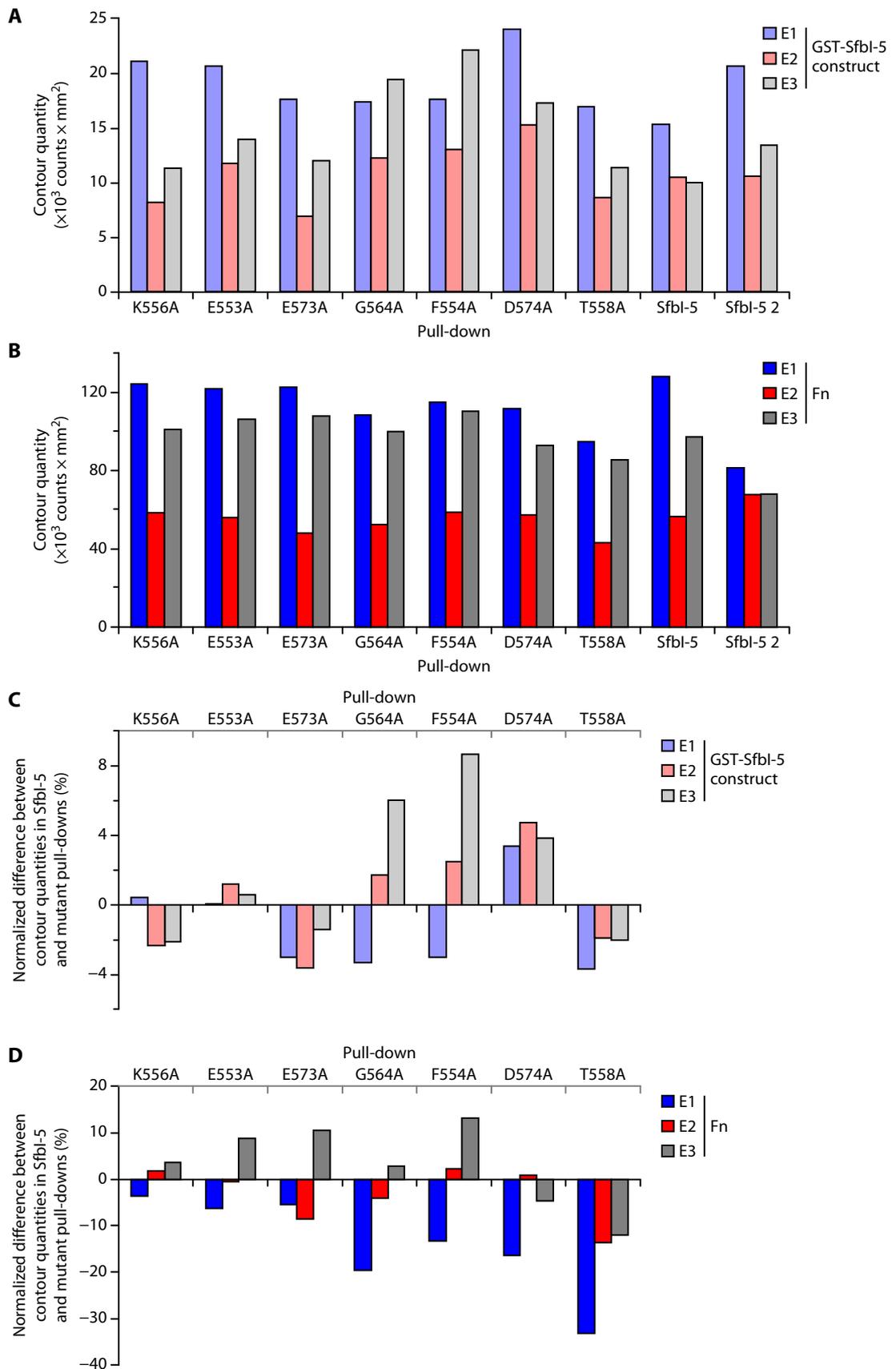
Figure 51. Band intensity analysis demonstrates GST-SfbI-5 mutants do not pull-down a significantly different amount of Fn from human plasma than GST-SfbI-5. Legend on p 200.

Figure 51. Band intensity analysis demonstrates GST-SfbI-5 mutants do not pull-down a significantly different amount of Fn from human plasma than GST-SfbI-5. A. Gels of each of the three pull-down experiments (E1–E3; Figure 50) were analysed with Quantity One software, and the contour quantity for the band corresponding to the GST-SfbI-5 construct was determined (Section 2.9.1.3). The results from experiments 1–3 are shown in blue, red and grey, respectively. The samples corresponding to each pull-down are indicated by the GST-SfbI-5 construct used in the pull-down. Wild-type SfbI-5 samples from each experiment were run a second time on another gel, these are shown separately as SfbI-5 2. Mutant constructs are ordered, right to left, from lowest to highest dissociation constant for binding to pNTD as determined by ITC (Table 37). **B.** As for A, except the contour quantities are for the bands corresponding to Fn. **C.** The normalized difference between the contour quantities of each GST-SfbI-5 mutant and GST-SfbI-5 from the same experiment. Normalized values for each mutant pull-down were calculated with the equation: $(C_{mutant} - C_{SfbI-5})/C_{SfbI-5} \times 100$ where C_{SfbI-5} and C_{mutant} are the contour quantities of the GST-SfbI-5 construct for the wild-type and mutant pull-downs, respectively. A negative value indicates less of the GST-SfbI-5 mutant was present compared to wild-type in the same experiment. **D.** As for C, except the normalized difference between the contour quantities of Fn in each mutant pull-down and the wild-type pull-down from the same experiment is shown.

The analysis of band intensity indicated that, unfortunately, experimental error was greater than the differences these experiments were attempting to identify (Figure 51). At least some of this error was from the SDS-PAGE analysis itself rather than differences between experiments, because the same SfbI-5 samples loaded on different gels had different band intensities (compare SfbI-5 and SfbI-5 2 in Figure 51A, B). Therefore, for each experiment, the differences in band intensities between mutant and wild-type pull-downs were normalized with respect to the corresponding band (either Fn or GST-SfbI-5) in that experiment's wild-type pull-down (Figure 51C, D). For a comparison between the amount of Fn bound by a mutant and wild-type to be meaningful, the amount of mutant and wild-type GST-SfbI-5 present in each pull-down needs to be approximately constant. Unfortunately, the normalized difference between band intensities in mutant and wild-type pull-downs shows that this is not the case, with the GST-SfbI-5 mutant band intensities varying, on average, by 20% from the corresponding GST-SfbI-5 band (Figure 51C). Therefore, because experimental error is high, it is unclear whether the mutations have a significantly different effect on SfbI-5 binding to Fn in plasma compared to SfbI-5 binding to pNTD in PBS.

4.4 Discussion

4.4.1 Domain organization of FnBPs

The domain organizations of FnBPs identified in the first round of searches with the nr70 HMM were analysed (Figure 31). This analysis confirms that most bacterial FnBPs contain the common features already mentioned with respect to SfbI (Figure 3; Section 1.3.3). Furthermore, it appears many FnBPs contain at least one other domain that is predicted to bind extracellular macromolecules (*e.g.,* collagen binding domain, von Willebrand factor type A domain, Table 30). This observation agrees with an analysis of domain organization of literature-derived FnBP sequences using the SMART protein domain family database (Schwarz-Linek *et al.*, 2006).

While some domain organizations are more prevalent than others (Figure 31), it is difficult to draw definite conclusions about the significance of this finding. This is because similar domain organizations are clustered by species, and it is possible that some domain organizations are more prevalent just because more closely related strains have been sequenced for one species rather than another. For example, 22 of the 28 sequences that have the same domain organization as UniProt ID Q47942 were sequenced in a study of evolution of SfbI in *S. pyogenes* strains (Figure 31; Towers *et al.*, 2003). If a similar analysis were to be conducted for other species, their prevalence would also increase. For example, an evolutionary analysis of FnBPA conducted after the first round of HMM building and searching (Loughman *et al.*, 2008) contributed five additional FnBPA sequences to the final FnBR data set. Thus, although FnBPs have domains with adhesive properties, whether some domain organizations are more common because they confer an evolutionary advantage is unclear.

4.4.2 FnBRs have probably evolved through homologous recombination

That many FnBPs contain multiple, highly homologous FnBRs (Figure 31) is clear. For example, approximately 60% of the FnBRs identified in the final set of FnBP sequences (Table 31) were identical to at least one other FnBR. FnBPs with tandemly repeated FnBRs have probably evolved by repeat expansion, which Tompa (2003) proposes to be a common mechanism for IDP evolution. Significantly more IDPs have been found to have repeat regions compared to globular proteins (Tompa, 2003): 39% of protein

sequences from a dataset of known IDPs contained repeat regions, compared to 14% of sequences in the Swiss-Prot protein database (part of UniProt).

That repeat expansion by recombination can occur in FnBPs has been shown by a study on the evolution of SfbI by Towers *et al.* (2003), who analysed the DNA sequences of SfbI from 34 *S. pyogenes* strains. They found that SfbI has evolved by frequent intragenic homologous recombination events. This led to a variable number of repeats, but they found the repeats themselves were highly conserved, with >91% identity in pairwise DNA alignments of the FnBR repeat regions. It is likely that the other FnBPs with tandem FnBRs have evolved similarly through homologous recombination.

4.4.3 The putative *Ixodes scapularis* FnBPs are eukaryotic and not prokaryotic

The HMM searches identified a new, putative FnBP family from *I. scapularis* (Table 31). The DNA sequences for these putative FnBPs were discovered in an analysis of Expressed Sequence Tags (ESTs) from the salivary glands of adult female ticks (Ribeiro *et al.*, 2006). ESTs are DNA sequences that arise from sequencing the 5′ and/or 3′ end of messenger RNA (mRNA) by first transcribing a complementary DNA (cDNA) strand of the RNA.

The Lyme disease-causing spirochete, *B. burgdorferi*, lives in the salivary glands of *I. scapularis*, and is transmitted to the ticks' host during feeding (Sonenshine, 1993). *B. burgdorferi* is the only bacterial species apart from staphylococci and streptococci known to have FnBPs (Section 1.4.3). Therefore, it is possible that the putative *I. scapularis* FnBPs are not from *I. scapularis* but from *B. burgdorferi*. However, there are five reasons why it is likely these sequences are eukaryotic and not prokaryotic.

First, Ribeiro *et al.* (2006) state that tick salivary glands used in their EST analysis were isolated from 'pathogen-free' adult female ticks, and therefore should not be infected with *B. burgdorferi*, or any other bacterium. In fact, one of the libraries in their study was designed specifically to assess the effect of infecting young ticks (nymphs) with *B. burgdorferi*.

Second, the ESTs were isolated based on an mRNA processing step specific to eukaryotes: polyadenylation of the 3′ end of the mRNA (Ribeiro *et al.*, 2006). All but

one of the ESTs from the initial study that identified the four putative FnBPs contained this tail and the associated signal sequence, which is necessary for polyadenylation (Table 38). The EST that did not contain the polyadenylation signal and tail is truncated at its 3′ end with respect to the 18 other sequences in its cluster, indicating this is probably a sequencing error rather than a lack of mRNA processing.

Third, the ESTs do not contain a Shine-Dalgarno sequence upstream of their start codons, which is the prokaryotic ribosomal binding site (reviewed in Kozak, 1999). However, all ESTs contain the same sequence directly upstream of their start codon: AAGATG. This is consistent with the most important requirement of the Kozak sequence, which is to have a purine, preferably A, 3 bp upstream of the start codon; that is, (A/G)NNATG. The Kozak sequence is the eukaryotic signal for starting translation of an mRNA (reviewed in Kozak, 1999).

Fourth, the *B. burgdorferi* genome has been sequenced, yet no *B. burgdorferi* protein or DNA sequences are pulled up by a BLAST search (Altschul *et al.*, 1990) with the protein or DNA sequences of any of the putative *I. scapularis* FnBPs.

Finally, despite both BBK32 and the four putative FnBPs containing putative FnBRs identified by the HMM search, these FnBR protein sequences are dissimilar, as are the full-length protein sequences (Figure 52). Therefore, it appears highly likely that the *I. scapularis* proteins are eukaryotic, not prokaryotic. Thus, the presence of putative FnBRs in both *I. scapularis* and *B. burgdorferi* species appears to be an example of convergent evolution, and an indication of the usefulness of FnBPs at the site of tick feeding.

A

```
BBK32    EEPIESNEIDLTIDSDLRPKSSLQGIAGSNSISYTDEIEEEDYDQY
Q4PN68   --EVVGGNVEFDIHSTGG------GISGHNQGP---QTEVEDTGES
Q4PMW3   --EVVGGHVDYDIHSGGG------GISGQNQGP---QTVEEDTGES
Q4PN03   GAGGGGGVVEYEEHTGG-------GSSGHHHGP---LTEYEDSQPG
Q4PN66   GSLSGSGAVEYETDTGPG-----TGMFGGHGGG---PVEGEDSRLP
            ..  ::    .:           *  * :         **
```

B

```
BBK32    MKKVKSKYLALGLLFGFISCDLFIRYEMKEESPGLFDKGNSILETSEESIKKPMNKKGKG
Q4PN68   ---MKSTLVAICFLLAVVHT--MV------------------------------------
            :**. :*: :*:..:     ::
```

```
BBK32    KIARKKGKSKVSRKEPYIHSLKRDSANKSNFLQKNVILEEESLKTELLKEQSETRKEKIQ
Q4PN68   ------------------------------------------------------------
```

```
BBK32    KQQDEYKGMTQGSLNSLSGESGELEEPIESNEIDLTIDSDLRPKSSLQGIAGSNSISYTD
Q4PN68   -------------------------EVVGGNVEFDIHSTGG------GISGHNQGP---
                                     : ..::::: *.*        **:* *. .
```

```
BBK32    EIEEEDYDQYYLDEYDEEDEEEIRLSNRYQSYLEGVKYNVDSAIQTITKIYNTYTLFSTK
Q4PN68   QTEVEDTGES--GQAGPPAPPGGRLPG-----LSG-------------------------
          : * ** .:    .: .        **..       *.*
```

```
BBK32    LTQMYSTRLDNFAKAKAKEEAAKFTKEDLEKNFKTLLNYIQVSVKTAANFVYINDTHAKR
Q4PN68   ------------------------------------------------------------
```

```
BBK32    KLENIEAEIKTLIAKIKEQSNLYEAYKAIVTSILLMRDSLKEVQGIIDKNGVWY
Q4PN68   ------------------------------------------------------
```

Figure 52. Alignment between the *I. scapularis* FnBPs and *B. burgdorferi* BBK32. A. Alignment between the FnBRs in BBK32 and the four putative *I. scapularis* FnBPs. **B.** Alignment between full-length BBK32 and the full-length FnBP from which TickFnBR is derived. For both A and B, residues are coloured based on chemical composition. Conserved residues are indicated below the alignment, where an asterisk (*) colon (:) and full stop (.) indicate identical, conserved and semi-conserved residues, respectively.

Table 38. Putative FnBPs from *I. scapularis* identified from an HMM search of the UniRef90 database. Shown for each of the four putative FnBPs identified from the UniProt database is the corresponding UniGene cluster. Each cluster consists of 'mRNA' and EST sequences. The ESTs have been experimentally identified from one of four *I. scapularis* cDNA libraries, and are automatically clustered on the basis of sequence similarity. Each cluster contains at least one mRNA that is derived from an alignment of the ESTs of that cluster. This mRNA sequence consists only of an open reading frame (ORF), it does not contain the 5′ and 3′ untranslated regions. Translation of the mRNA gives the associated UniProt protein sequence. Also shown are: the length of each mRNA; whether an EST has any insertions and/or deletions relative to the mRNA in its cluster; and whether an EST has a polyadenylation signal and/or tail in its 3′ untranslated region, where a signal is the sequence AATAAA and a tail is two or more adenines at the 3′ end.

Identification number			mRNA or EST	*Ixodes* cDNA library[15]	ORF length (bp)[16]	Polyadenylation	
UniProt	UniGene	GenBank				Signal	Tail
Q4PN68	Isc.2034	DQ065905	mRNA		243		
		DN972841	EST	1	+2	Yes	Yes
Q4PMW3	Isc.2034	DQ066010	mRNA		243		
		DN972608	EST	1	0	Yes	Yes
		DN971345	EST	2	+1	Yes	Yes
Q4PN66	Isc.2843	DQ065907	mRNA		210		
		DN972910	EST	1	+1	Yes	Yes
		EL516390	EST	3	0	Yes	Yes[17]

[15] Four different *I. scapularis* cDNA libraries were used to generate the EST sequences in the three clusters. These are: (1) A salivary gland cDNA library taken from female adults 6–12 hours after attachment to a rabbit (Ribeiro *et al.*, 2006); (2) A salivary gland cDNA library taken from unfed female adults (Ribeiro *et al.*, 2006); (3) A SMART cDNA library from the synganglia of feeding adult females (Huang *et al.*, 2007); (4) A pooled cDNA library composed of non blood fed and various blood fed and replete whole *I. scapularis* and *I. scapularis* life-cycle stages (Hill *et al.*, 2007).

[16] The length of the mRNA sequence is reported. For each EST sequence, insertions or deletions in the ORF relative to its corresponding mRNA sequence is reported. Insertions are positive, deletions are negative. The large deletions in ESTs DN972887, EW864813, EW864814, EW920079 and EW920080 are truncations from the 5′ end relative to the mRNA DQ065970.

[17] There is no polyadenylation tail present in the submitted EST sequence. However, the sequence is annotated as having a tail. Since all four sequences from cDNA library 3 are like this, it is likely that the researchers have trimmed away the polyadenylation tail from the ESTs before submission.

Identification number			mRNA or EST	*Ixodes* cDNA library	ORF length (bp)	Polyadenylation	
UniProt	UniGene	Genbank				Signal	Tail
Q4PN03	Isc.2489	DQ065970	mRNA		252		
		DN972169	EST	1	0	Yes	Yes
		DN972197	EST	1	0	Yes	Yes
		DN972280	EST	1	0	Yes	Yes
		DN972344	EST	1	0	Yes	Yes
		DN972429	EST	1	0	No	No[18]
		DN972460	EST	1	0	Yes	Yes
		DN972494	EST	1	0	Yes	Yes
		DN972499	EST	1	$+3, -1$	Yes	Yes
		DN972544	EST	1	0	Yes	Yes
		DN972567	EST	1	0	Yes	Yes
		DN972598	EST	1	0	Yes	Yes
		DN972887	EST	1	-16	Yes	Yes
		EL516037	EST	3	0	Yes	Yes[17]
		EL516566	EST	3	0	Yes	Yes[17]
		EL516981	EST	3	0	Yes	Yes[17]
		EW864813	EST	4	-213	Yes	Yes
		EW864814	EST	4	-213	Yes	Yes
		EW920079	EST	4	-213	Yes	Yes
		EW920080	EST	4	-213	Yes	Yes

4.4.4   Analysis of ^{2}F1^{3}F1:PyTT5 Structure III

4.4.4.1   Asymmetric unit

There are two ^{2}F1^{3}F1:PyTT5 complexes in the asymmetric unit (AU; Figure 38A). Additionally, there are two glycerol molecules from the cryoprotectant that bind to equivalent points on ^{3}F1 in both complexes of the AU. The glycerol molecules bind on the opposite side of ^{3}F1 compared to PyTT5 (Figure 38A), so do not appear to impact on peptide binding. There is also a sulfate molecule from the crystallization conditions

---

[18] The 3′ untranslated region of this sequence aligns with the other ESTs, but is truncated at the 3′ end relative to them, such that the polyadenylation tail and signal are not present.

binding to R99 and R101 of ^{2}F1 in one of the two complexes (Figure 53). This results in a change in the orientation of these arginines, and in turn, D574 in PyTT5 forms hydrogen bonds with these residues differently in the two complexes in the AU. Therefore, the structural analysis reported below has been performed on the other complex in the AU.
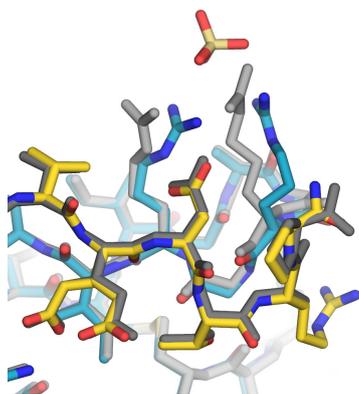
Figure 53. The binding of a sulfate group appears to change the orientation of nearby sidechains in one of two ^{2}F1^{3}F1:PyTT5 complexes in the AU. R99 and R101 of ^{2}F1 in one complex (coloured) chelates a sulfate group from the crystallography conditions. The other complex (grey) does not have a sulfate group at this site.

4.4.4.2 PyTT5 forms a large, extended interface with ^{2}F1^{3}F1

The extent of the interface between PyTT5 and ^{2}F1^{3}F1 was analysed using ProFace (Saha *et al.*, 2006), which calculated the amount of buried surface area (BSA) at the interface using Equation 16. The total amount of BSA, 1950 Å^2, classifies this interface as 'large' as defined by Janin and colleagues (Lo Conte *et al.*, 1999; Chakrabarti & Janin, 2002; reviewed in Janin *et al.*, 2008). The interface is not contributed to evenly by both ^{2}F1^{3}F1 and PyTT5, with PyTT5 contributing approximately 13% more BSA to the interface. The interface can be further divided into two categories at both the atom and residue level (Lo Conte *et al.*, 1999; Chakrabarti & Janin, 2002).

At the atom level, the interface is divided into buried and accessible interface atoms, where buried atoms lose all accessible surface area (ASA) and accessible atoms lose some ASA. The ProFace analysis showed that every residue of PyTT5 is involved in the interface with ^{2}F1^{3}F1, with at least one atom of each residue in PyTT5 having some BSA at the interface (Figure 54A; Table 39).

At the residue level, the interface is divided into core and rim residues, with core residues containing at least one buried interface atom. Rim residues contain at least one accessible interface atom and no buried interface atoms. By this definition, 14 of the 18 residues in PyTT5 were core residues and the remaining residues were rim residues

(Figure 54B; Table 39). In comparison, ^{2}F1^{3}F1 had 20 core residues and 18 rim residues out of a total of 89 residues (Figure 54B; Table 39). This indicates that each residue of PyTT5 contacts, on average, more residues of ^{2}F1^{3}F1 than *vice versa*. It also indicates that a much greater percentage of PyTT5 is involved in the interface compared to ^{2}F1^{3}F1. This is further shown by the fact that 40% of the ASA of PyTT5 is buried at the interface, compared to 17% of the ASA of ^{2}F1^{3}F1. Thus, this analysis reveals that PyTT5 is a very efficient ligand, contributing more contacts and a greater percentage of the interface than ^{2}F1^{3}F1 with fewer atoms and residues than ^{2}F1^{3}F1. These properties of PyTT5 have been highlighted as common properties of disordered proteins when they form complexes (Deane *et al.*, 2004; Gunasekaran *et al.*, 2003; Mészáros *et al.*, 2007).

$$BSA = ASA_A + ASA_B - ASA_{A:B}$$

Equation 16. Buried surface area (BSA) for a complex A:B is calculated from the accessible surface area (ASA) of A, B and A:B (Janin *et al.*, 2008).
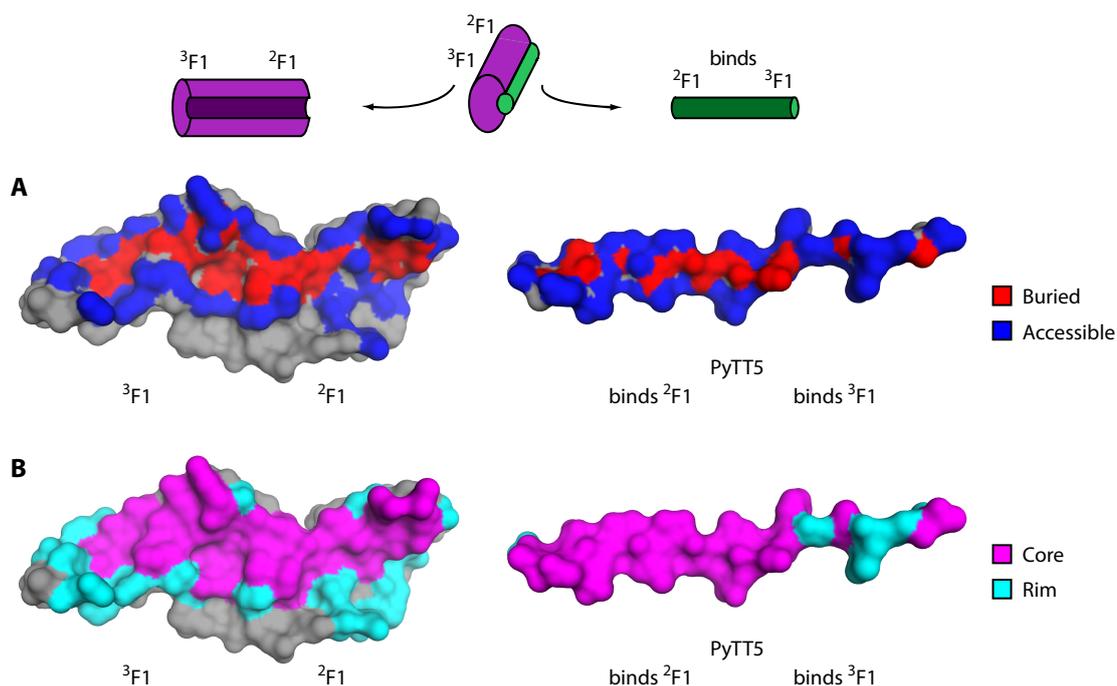
Figure 54. The interface between ^{2}F1^{3}F1 and PyTT5. Surface representation of ^{2}F1^{3}F1 and PyTT5 (left and right respectively) that shows the interface between ^{2}F1^{3}F1 and PyTT5. Indicated are the ^{2}F1 and ^{3}F1 modules, as are the regions of PyTT5 that bind ^{2}F1 and ^{3}F1. The schematic above the interfaces shows how the complex was opened like a book to reveal the interface. **A.** The interface is shown at the atom level and is represented as buried and accessible atoms, which are coloured red and blue, respectively. Buried atoms lose all ASA in the interface between PyTT5 and ^{2}F1^{3}F1, accessible atoms lose some, but not all, ASA in the interface. **B.** The interface is shown at the residue level. The interface is represented as

core and rim residues, which are coloured magenta and cyan, respectively. Core residues contain at least one buried interface atom. Rim residues contain at least one accessible interface atom and no buried interface atoms. Residue and atom accessibility was calculated from the output of ProFace (Saha *et al.*, 2006) using chains B and D from Structure III.

Table 39. Analysis of the interfaces in ^{2}F1^{3}F1:PyTT5 Structure III by change in ASA on complex formation. Statistics for the interface between ^{2}F1^{3}F1 chain B and PyTT5 chain D and the interface between PyTT5 chains C from one AU and D from a neighbouring AU (Section 4.4.4.4) were calculated with the ProFace server (Saha *et al.*, 2006). The terms core, rim, buried and accessible are defined in the text. Local density is a measure of the atomic packing at the interface (Bahadur *et al.*, 2004).

Complex	^{2}F1^{3}F1:PyTT5			PyTT5:PyTT5		
Molecule, chain	^{2}F1^{3}F1, B	PyTT5, D	Complex	PyTT5, C	PyTT5, D	Complex
Interface area ($\mathring{A}^2$)						
Total	851.36	1095.61	1946.97	466.77	459.76	926.53
Core	653.14	852.99	1506.13	181.25	246.16	427.41
Rim	241.49	242.62	484.11	285.52	213.6	499.12
Interface area / Surface area	0.17	0.41	0.24	0.18	0.19	0.19
Number of interface atoms						
Total	122	89	211	49	49	98
Buried	44	22	66	5	6	11
Accessible	78	67	145	44	43	87
Number of interface residues						
Total	38	18	56	11	11	22
Core	20	14	34	4	5	9
Rim	18	4	22	7	6	13
Fraction of non-polar atoms	0.57	0.60	0.58	0.67	0.65	0.66
Non-polar interface area ($\mathring{A}^2$)	443.49	592.27	1035.76	301.07	290.54	591.61
Fraction of fully buried atoms	0.36	0.25	0.31	0.12	0.12	0.12
Local Density	36.56	32.70		22.57	22.45	

4.4.4.3 SfbI-5 is predicted to form a very large, extended interface with $^{1-5}$F1

It is intriguing to consider the extent of the interface between SfbI-5 and $^{1-5}$F1. ProFace analysis of a model of $^{1-5}$F1 bound to SfbI-5 (Figure 55) predicts that the total BSA of the interface would be ~5150 $\mathring{A}^2$. This is a very large interface for a heterocomplex. In fact, comparison between the predicted interface area of $^{1-5}$F1:SfbI-5 and the observed

interface areas from a dataset of heterodimer complex structures (Mészáros *et al.*, 2007) reveals that the predicted interface is larger than any of the interfaces in the dataset, and more than 1500 $Å^2$ larger than the largest complex involving an IDP (Figure 56). Furthermore, SfbI-5 contributes 9% more area to the interface than $^{1-5}$F1 and the interface is predicted to involve 45 of 50 residues in SfbI-5 and 91 of 225 residues in $^{1-5}$F1, with 40% of the ASA of SfbI-5 buried, compared to 18% of the ASA of $^{1-5}$F1. Thus, the reasons PyTT5 is an efficient ligand for binding to ^{2}F1^{3}F1 are also predicted to make SfbI-5 an extremely efficient ligand when in complex with $^{1-5}$F1. For a globular protein to bury a similar amount of surface area it would have to be approximately two to three times bigger (Gunasekaran *et al.*, 2003). Therefore, a clear biological advantage of the intrinsic disorder of SfbI-5 is tight binding for fewer resources committed on the part of *S. pyogenes*.
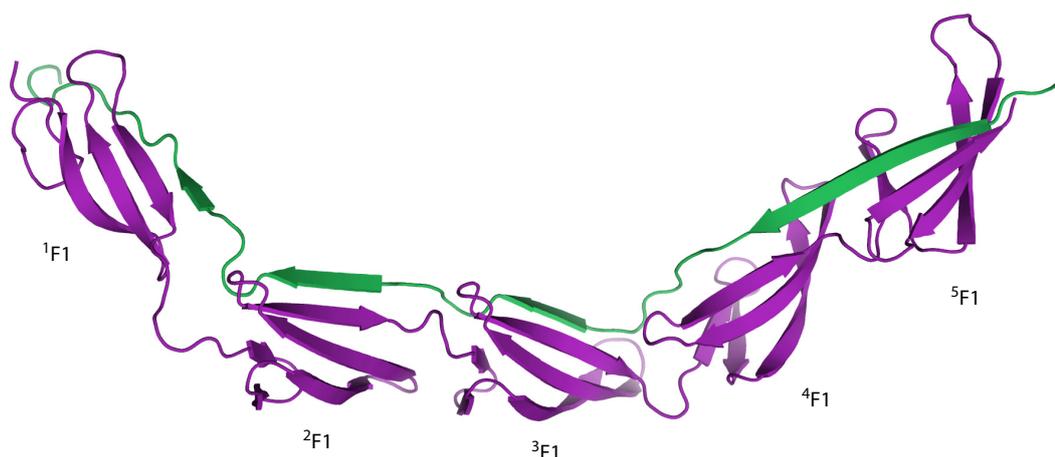
Figure 55. Model of the $^{1-5}$F1:SfbI-5 complex. Ribbon representation of model of SfbI-5 (green) in complex with $^{1-5}$F1. The model was built from ^{2}F1^{3}F1:PyTT5, an alignment between the two ^{4}F1^{5}F1 and *S. aureus* FnBPA peptide structures (pdb codes 2rky and 2rl0) and ^{1}F1 and the ^{1}F1-binding region from the structure of ^{1}F1^{2}F1 in complex with B3T (pdb code 1o9a) using the program Modeller (Sali & Blundell, 1993).
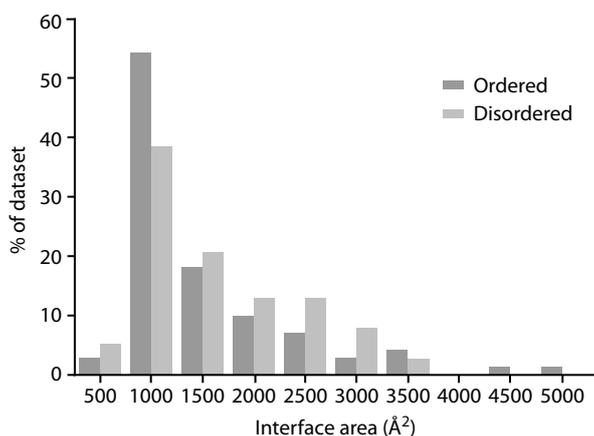
Figure 56. Comparison of the interface areas of heterodimeric complexes in which either both components are ordered or one component is an IDP. The dataset contained 39 'disordered complexes', which each had one component that had been experimentally shown to be disordered before complex formation. The dataset also contained 72 ordered complexes, the components of which had individually well-defined structures. Taken from Figure 2(a) in Mészáros *et al.*, 2007.

4.4.4.4 Interaction between two PyTT5 molecules in ^{2}F1^{3}F1:PyTT5 Structure III are probably crystal contacts

The crystal packing in ^{2}F1^{3}F1:PyTT5 Structure III is such that a PyTT5 molecule from one AU interacts with another PyTT5 molecule from a neighbouring AU (Figure 38B, Figure 57A). This interface was analysed with ProFace (Table 40) to determine whether it is likely to be a specific or non-specific protein-protein interaction. That is, whether the PyTT5 dimer is independent of or dependent on the surrounding crystal. At 927 Å^2, the PyTT5:PyTT5 interface area is classified as small for a specific protein-protein interaction, but still within the range observed (Figure 56; Lo Conte *et al.*, 1999). Non-specific dimer interfaces larger than 800 Å^2 formed by crystal contacts have been compared to specific homo- and hetero-dimers (Bahadur *et al.*, 2004). It was found that two parameters were useful in distinguishing non-specific dimers from specific dimers. First, the number of fully buried interface atoms expressed as a fraction of the total number of interface atoms (buried and accessible) was, on average, 0.22 ± 0.09 for non-specific dimers and 0.34 ± 0.09 for specific dimers (Bahadur *et al.*, 2004); for PyTT5:PyTT5 this parameter is 0.12 (Table 40), indicating PyTT5:PyTT5 is probably a non-specific dimer. Second, the ligand density, which is a measure of the interface's atomic packing, was on average 32 ± 6 and 42 ± 6 for non-specific and specific dimers, respectively (Bahadur *et al.*, 2004). The ligand density for both PyTT5 molecules in the PyTT5 dimer was about 22.5 (Table 39), also indicating PyTT5:PyTT5 is a non-specific dimer. Therefore, it appears highly likely that the PyTT5:PyTT5 interaction is non-specific and results from crystal contacts.
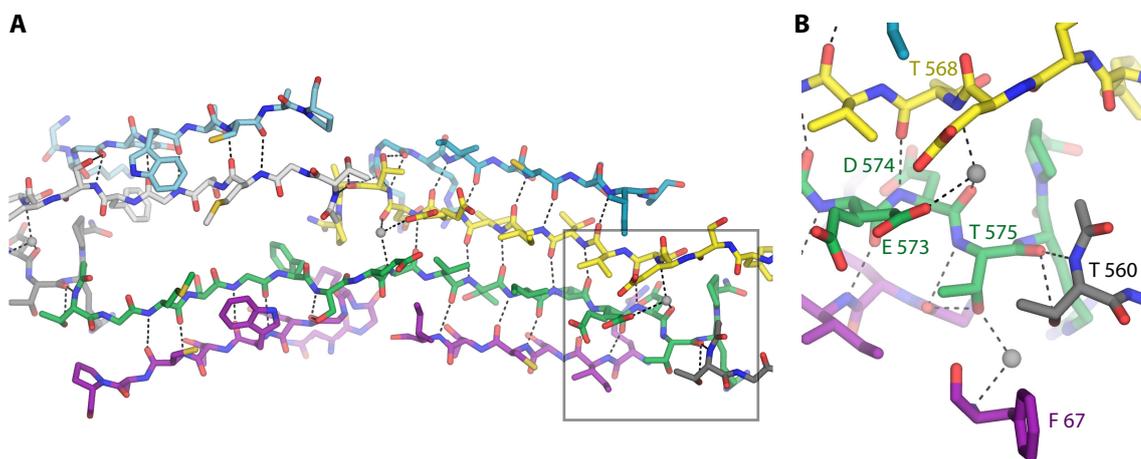
Figure 57. Key crystal contacts that may affect the interaction between ^{2}F1^{3}F1 and PyTT5 in Structure III. A. PyTT5 chains C and D and the E-strands of ^{2}F1^{3}F1 from chains A and B (blue and purple, respectively), with hydrogen bonds shown as black dashes. The ^{2}F1-binding region of a PyTT5 molecule (green) interacts with the ^{2}F1-binding region of an adjacent PyTT5 molecule from a symmetry-related complex (yellow). Furthermore, the N-terminal residues of one PyTT5 molecule (green or yellow) interact with the C-terminal residues of an adjacent PyTT5 molecule (dark or light grey). Key water molecules at the interface between symmetry related protein molecules are shown as grey spheres. **B.** Enlargement of the boxed region in A, with residues involved in crystal contacts either directly or *via* an invariant bridging water molecule labelled. Residues are coloured as in A. The yellow residues are part of the linker region of a PyTT5 molecule.

4.4.4.5   Important mainchain-mainchain interactions between ^{2}F1^{3}F1 and PyTT5

Both the mainchain and sidechains of PyTT5 interact with both F1 modules through extensive polar and van der Waals contacts (Figures 38D, 58). PyTT5 binds across the E-strands of both F1-modules, extending each β-sheet by an additional anti-parallel β-strand (Figure 38A). PyTT5 interacts *via* six and five mainchain-mainchain hydrogen bonds with ^{2}F1 and ^{3}F1, respectively; with residues 562–565 and 570–574 adopting φ and ψ angles typical of β-strand (Figure 58). Therefore, PyTT5 interacts with ^{2}F1^{3}F1 *via* a tandem β-zipper.

4.4.4.6   Interactions with ^{2}F1^{3}F1 involving PyTT5 sidechains

For this analysis, PyTT5 residues 560–565 are the ^{3}F1-binding region, residues 566–569 are the linker region and residues 570–577 are the ^{2}F1-binding region. The interactions involving the sidechains of PyTT5 in each of these regions will now be discussed in turn.

The sidechains of the PyTT5 ^{3}F1-binding region interact with ^{3}F1 primarily *via* van der Waals contacts, with M562 contacting L134 and R125 in ^{3}F1, and F565 stacking against the aliphatic sidechain carbons of K143 and E145 (Figures 58, 59A). Furthermore, a glycine at position 564 allows PyTT5 to adopt a β-strand formation while avoiding a steric clash with W146 in ^{3}F1, which would be thermodynamically unfavourable (Figures 58, 59A). Therefore, the conserved glycine identified in Section 4.1.3 performs the same function in this structure as in the two structures of ^{2}F1^{3}F1 with *S. aureus* peptides (^{2}F1^{3}F1:STATT1 and ^{2}F1^{3}F1:STATT5; Bingham *et al.*, 2008).

The PyTT5 linker region does not loop away from ^{2}F1^{3}F1 as in the ^{2}F1^{3}F1:STATT1 and ^{2}F1^{3}F1:STATT5 structures (Figures 5B, 36G), although it is the same length as the linker region in STATT5. Instead, the PyTT5 linker region binds to the interface between ^{2}F1 and ^{3}F1 primarily *via* hydrogen bonds (Figures 58, 59B). S566 forms hydrogen bonds with the sidechain of E112 and backbone of H111 in ^{3}F1 *via* its hydroxyl group, with this group becoming completely buried on complex formation (Figure 59B). E567 forms hydrogen bonds with three bridging water molecules; with one water molecule *via* its sidechain and with two water molecules *via* its mainchain. These water molecules are 'invariant', that is, they are in the same positions in both complexes in the AU. As a result, E567 inserts into two hydrogen bond networks, and is thus connected to I106 and R109 in ^{2}F1 and H111 in ^{3}F1 (Figures 58, 59B).

The PyTT5 ^{2}F1-binding region binds to ^{2}F1 *via* both van der Waals and hydrogen bond interactions. The two key van der Waals contacts are the I571 and D574 sidechains interacting with the sidechains of W90 and R99, respectively. T570 forms a hydrogen bond with the sidechain of T105, forming a favourable staggered conformation, which has been shown to be a determinant of strand register (Hutchinson *et al.*, 1998). The sidechain of E573 exists in two conformations in both complexes in the AU with approximately 50% occupancy in each conformation. This result was confirmed by examining the difference density map for Structure III in the presence and absence of each of the two conformations (Figure 60). One of the two sidechain conformations forms a salt bridge to the sidechain of R83 in ^{2}F1 (Figures 58, 59C). This is the conformation of the corresponding sidechain in both complexes of *S. aureus* peptides with ^{2}F1^{3}F1 (Bingham *et al.*, 2008). The second E573 sidechain confirmation forms a hydrogen bond to the carbonyl group of D574 *via* a bridging water molecule (Figures 58, 59C). The sidechain of D574 forms a salt bridge with both R101 and R99, and T575

forms hydrogen bonds to the backbone atoms of G100 and F67. The latter hydrogen bond is *via* a bridging water molecule (Figures 58, 59C). Therefore, the PyTT5 residues corresponding to the conserved 'E-(D/E)-(T/S)' motif identified in Section 4.1.3 form key interactions in the structure.

The interaction of the 'E-(D/E)-(T/S)' motif in PyTT5 is similar to that of the corresponding residues in the ^{2}F1^{3}F1:STATT1 and ^{2}F1^{3}F1:STATT5 structures (Bingham *et al.*, 2008), except for two key differences. First, the residues corresponding to E573 in these structures do not have two conformations. Second, the STATT1 and STATT5 residues corresponding to T575 both form a hydrogen bond directly with F67 rather than *via* a bridging water molecule (Bingham *et al.*, 2008). Therefore, including the difference in the linker region previously mentioned, there are three key differences between the complexes of ^{2}F1^{3}F1 with *S. aureus* peptides and PyTT5.

4.4.4.7   Effect of crystal contacts on the interaction between ^{2}F1^{3}F1 and PyTT5

The three differences between ^{2}F1^{3}F1:PyTT5 Structure III and the two *S. aureus* structures could be caused by crystal contacts. The crystal packing in ^{2}F1^{3}F1:PyTT5 Structure III is such that the ^{2}F1-binding region in PyTT5 forms mainchain-mainchain hydrogen bonds with another PyTT5 ^{2}F1-binding region from a neighbouring AU (Figures 38B, 57A). Also, the N-terminal acetyl group and T560 of PyTT5 from one AU forms hydrogen bonds with and stacks against T575 at the C-terminal end of an adjacent AU (Figure 57B). These crystal contacts mean that there is no space for either the PyTT5 linker to loop out, or for T575 to interact directly with F67 in ^{2}F1 (Figure 57). The water molecule that forms hydrogen bonds with both D574 and an E573 conformation also forms a hydrogen bond with T568 from the linker region of a symmetry-related PyTT5 molecule. Thus, crystal contacts may affect one of the two E573 conformations, the linker region and T575. However, as previously mentioned, the linker region forms extensive hydrogen bonds with the ^{2}F1^{3}F1 module interface. Thus, it is not possible to tell whether the conformation of the linker region is caused by these favourable interactions, the crystal contacts, or a combination of the two. Interestingly, in Structure I the crystal packing is different: T575 forms a direct hydrogen bond with F67, and E573 exists in one conformation, forming a salt bridge with R83. This implies the crystal contacts probably affect the conformation of T575 and E573 and may affect the conformation of the linker region.
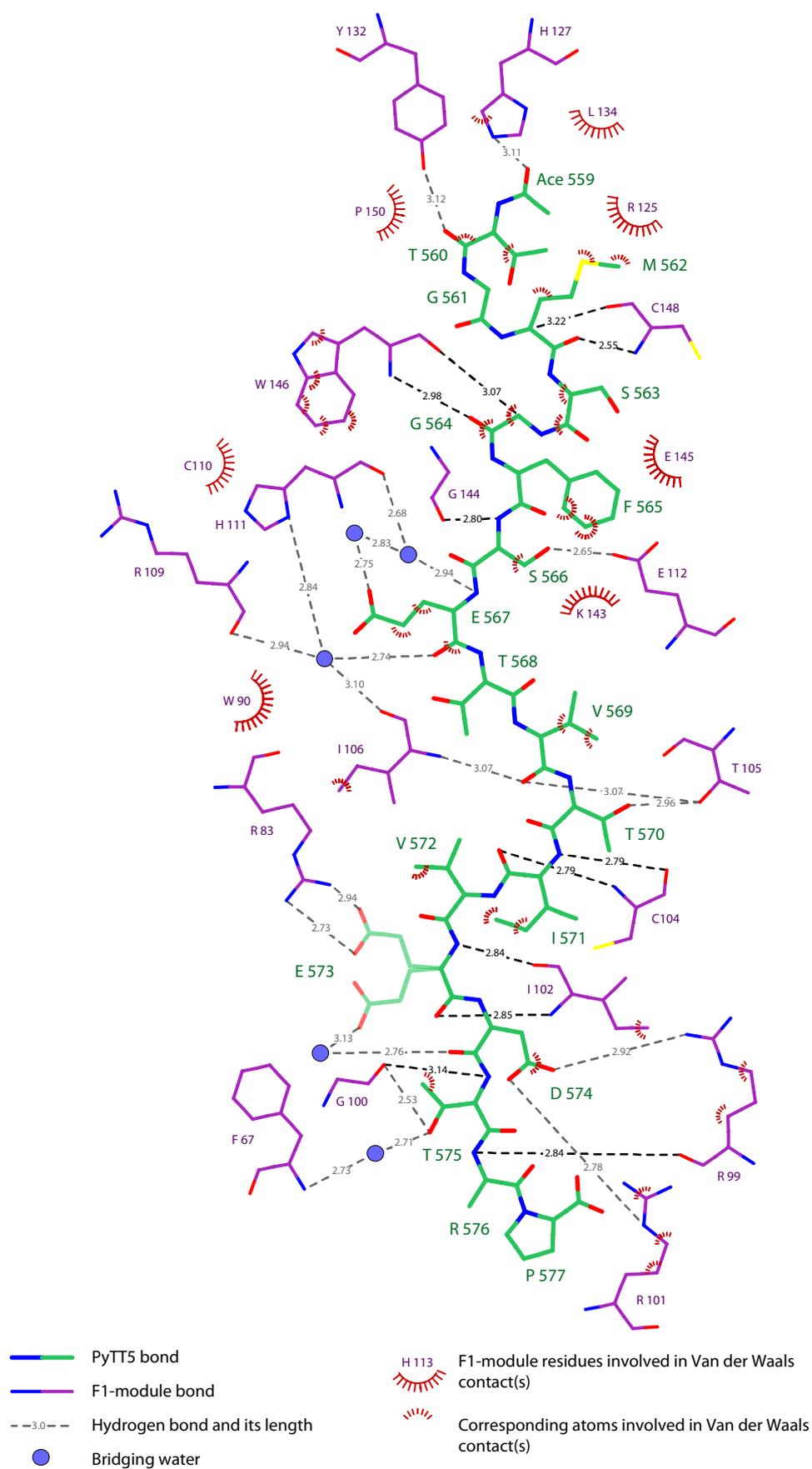
Figure 58. Schematic of interaction of PyTT5 with ²F1³F1. Residues in ^{2}F1^{3}F1 (purple) which form van der Waals contacts and/or hydrogen bonds with PyTT5 (green) are shown, as are bridging water molecules (blue circles). Hydrogen bonds between mainchain atoms on both PyTT5 and ^{2}F1^{3}F1 are black,

all others are grey. E573 exists in two conformations, both of which are shown (in lighter green). Van der Waals contacts and hydrogen bonds were calculated by the LIGPLOT program (Wallace *et al.*, 1995), which also created the diagram.
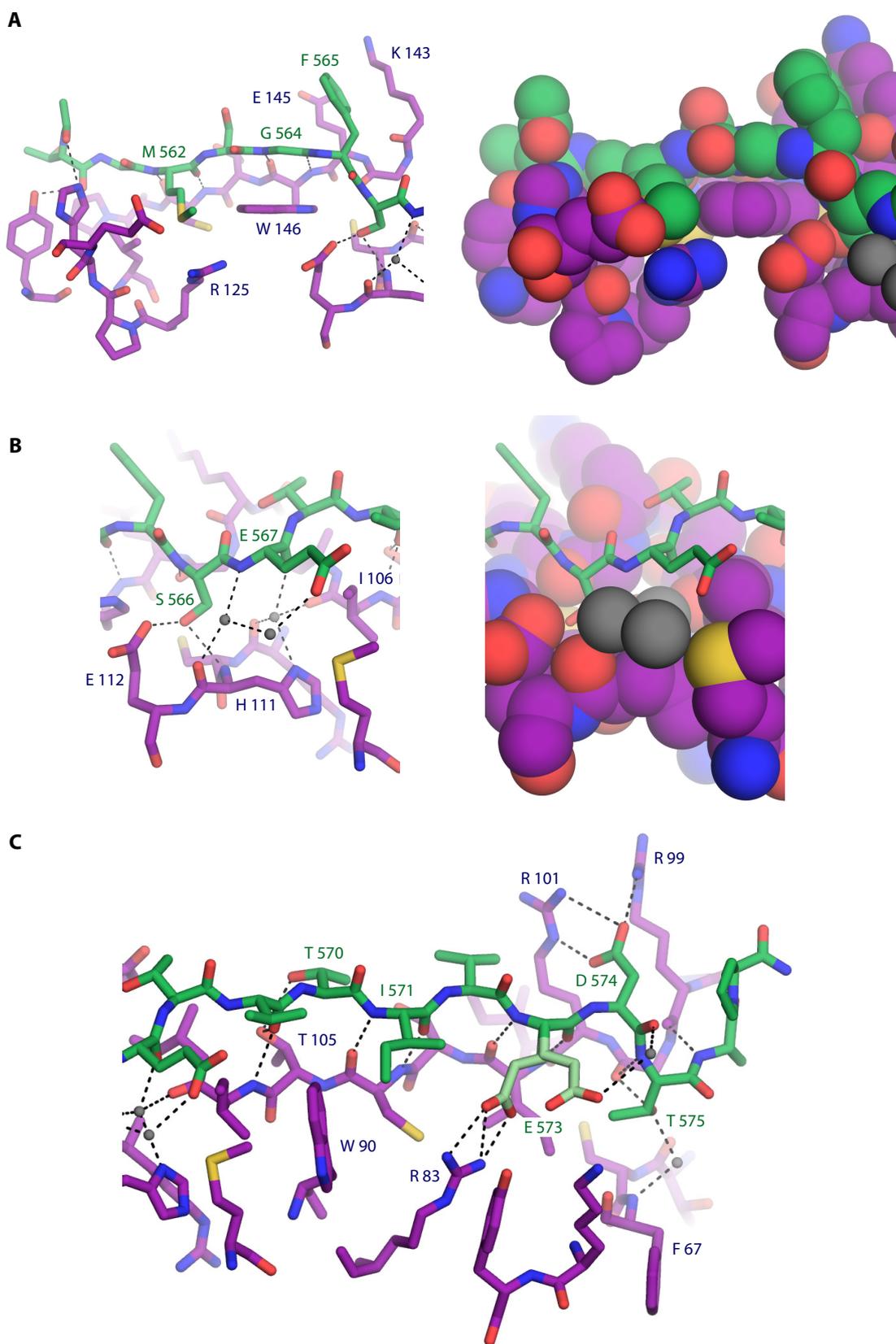
Figure 59. The interactions between $^2F1^3F1$ and PyTT5. Legend on p 217.

Figure 59. The interactions between ^{2}F1^{3}F1 and PyTT5. ^{2}F1^{3}F1 and PyTT5 interface residues are shown in purple and green, respectively. Interface residues have at least one atom with some surface area that contributes to the interface between ^{2}F1^{3}F1 and PyTT5. Hydrogen bonds are black dashes and invariant bridging water molecules are shown as grey spheres. Residues mentioned in the text are labelled. **A.** The interaction between ^{3}F1 and the ^{3}F1-binding region of PyTT5, shown as a stick representation and a ball representation on the left and right, respectively. **B.** The interaction between ^{2}F1^{3}F1 and the linker region of PyTT5, where PyTT5 is shown in stick representation and ^{2}F1^{3}F1 is shown in stick representation on the left and ball representation on the right. **C.** The interaction between ^{2}F1 and the ^{2}F1-binding region of PyTT5 shown as a stick representation.
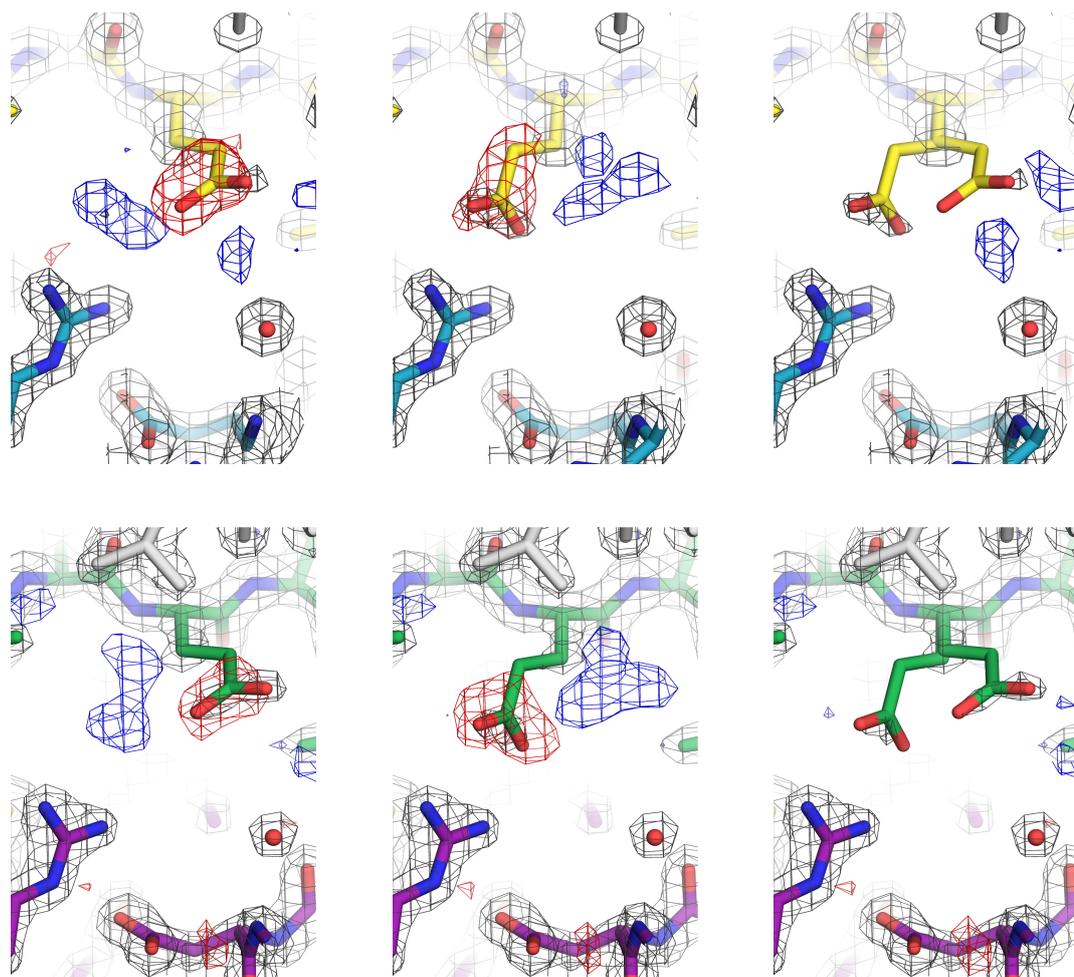
Figure 60. E573 exists in two conformers in both ^{2}F1^{3}F1:PyTT5 complexes in the AU. ^{2}F1^{3}F1 chains A and B are shown in light blue and purple, respectively. PyTT5 chains C and D are shown in yellow and green, respectively. Symmetry related molecules are shown in light grey. Electron density at 2 sigma is shown in grey, positive difference density at +3 sigma is shown in blue and negative difference density at −3 sigma is shown in red. Left and middle columns show E573 with 100% occupancy in conformation A or B, respectively. Right column shows E573 in conformations A and B, both with 50% occupancy.

217

4.4.5 Analysis of the thermodynamic parameters for SfbI-5 constructs binding to pNTD

4.4.5.1 Significant enthalpy entropy compensation does not appear to occur for SfbI-5 mutants binding to pNTD

There is a negative correlation between change in enthalpy and entropy on binding of SfbI-5 mutants to pNTD with respect to wild-type SfbI-5: where enthalpy has become more favourable, entropy has become more unfavourable, and *vice versa* (Figures 47, 61). This correlation, termed 'enthalpy-entropy compensation' (EEC; Sharp, 2001), might occur for one of two reasons. First, it can be the result of the linear relationship between Gibbs energy, enthalpy and entropy: $\Delta G° = \Delta H° - T\Delta S°$. When the range of Gibbs energy for each mutant is small, then this becomes $\Delta H° - T\Delta S° \approx$ constant, resulting in apparent EEC. Second, it could be an underlying thermodynamic property of the system. A test of whether EEC is occurring, proposed by Sharp (2001), is to compare the correlation between the experimental entropies and enthalpies to the correlation between randomly generated enthalpies and the entropies calculated from these enthalpies and the experimental Gibbs energy (Figure 61). If the correlation between enthalpies and entropies in both datasets is similar, then the observed EEC is not statistically significant. The correlation between the enthalpies and entropies as measured by the R^2 value from linear regression analysis was 0.97 for both datasets. This indicates that EEC is not statistically significant because the experimental error associated with the enthalpies is too large with respect to the size of Gibbs energy (Sharp, 2001).
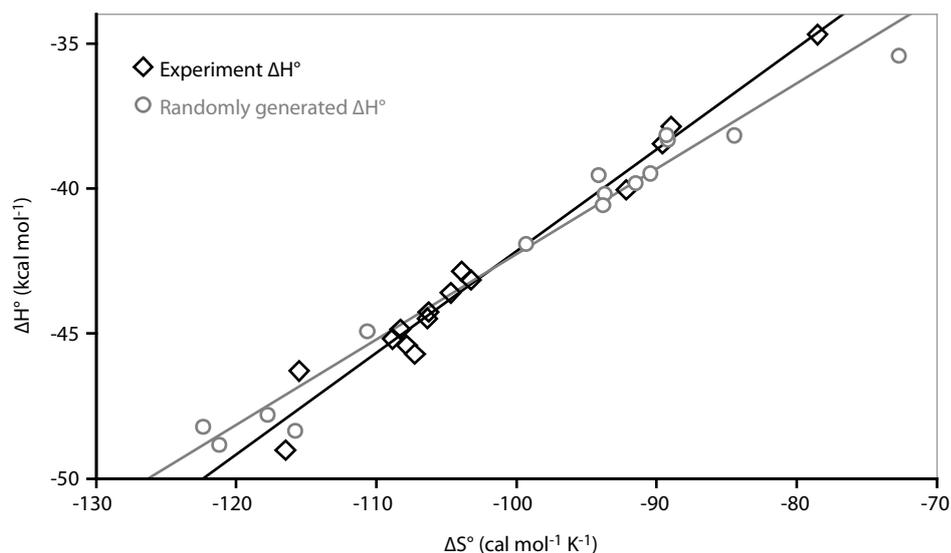
Figure 61. Relationship between change in entropy and enthalpy of binding of SfbI-5 constructs to pNTD for experimental and randomly generated enthalpies. Abbreviations are $\Delta H°$ and $\Delta S°$ for enthalpy and entropy, respectively. Experimental entropies and enthalpies are shown by black diamonds. For enthalpies and entropies shown by grey circles, each entropy was calculated from an experiment Gibbs energy and a randomly generated enthalpy. The range of randomly generated enthalpies fell within the same range as the experiment enthalpies. The lines-of-best fit are shown for the experiment and randomly generated enthalpies in black and grey, respectively. This analysis was proposed by Sharp (2001) as a test for EEC.

4.4.5.2 SfbI-5 does not appear to have binding 'hot-spots'

The effect of mutating conserved SfbI-5 residues on the binding of SfbI-5 to pNTD was determined by calculating the change in Gibbs energy of binding ($\Delta\Delta G°$) of the SfbI-5 mutants relative to wild-type SfbI-5 (Figure 62A). All $\Delta\Delta G°$ values were relatively small (less than 2.0 kcal/mol) compared to similar analyses on mutating interfaces in globular complexes. Thus, this interaction does not have a classical 'hot spot' of binding typical of globular protein complexes (Bogan & Thorn, 1998). To examine why this might occur, the mutated residues were considered in the context of the ^{2}F1^{3}F1:PyTT5 structure. It has been observed that hot spots tend to be core residues, and tend not to be rim residues (Janin *et al.*, 2008). It was hypothesized that this is because the rim residues act like an 'o-ring' and shield the core residues from water. However, the extended nature of PyTT5, when in complex with ^{2}F1^{3}F1, means that PyTT5 does not contain an 'o-ring' of rim residues (Figure 54B). It would be predicted that intact SfbI-5 when in complex with $^{1-5}$F1 would also lack o-ring residues, because the interactions of SfbI-5 with ^{1}F1, ^{4}F1 and ^{5}F1 are also likely to be extended. The reason this might lead

to the absence of hot spots can be better understood by considering the interface from the level of atoms.

At the atom level, Lo Conte *et al.* (1999) have made two key observations regarding protein-protein interactions. First, that the mainchain is not affected by site-directed mutagenesis. Second, that when accessible interface atoms are removed by mutation, they can be replaced by water molecules at a lower energetic cost than buried atoms. By this rationale, therefore, hot spot residues are likely to contain a buried interface atom that is deleted when the residue is mutated to alanine. That is, the buried atom cannot be a mainchain or a C^β atom. Only two of 22 (9.1%) buried interface atoms in PyTT5 fulfil these criteria. In comparison, 14 of 44 (32%) buried interface atoms in $^2F1^3F1$ fulfil these criteria. Therefore, it appears the extended nature of the interface between SfbI-5 and pNTD, which is facilitated by the intrinsic disorder of free SfbI-5, is likely to preclude hot spots in SfbI-5.

Absence of hot spots might be a general property of complexes containing an IDP, since most IDPs in complexes have extended interfaces that are unlikely to shield important interacting residues from water (Mészáros *et al.*, 2007). An extensive review of the literature did not reveal any other quantitative mutational analysis of a complex containing an IDP. This suggests that the work presented here is a novel analysis in the relatively new field of characterizing the interactions between IDPs and globular proteins.

Future work in this field is needed to establish if and how ideas used to understand globular protein interactions, such as hot spots or core and rim residues, apply to IDP-containing complexes.
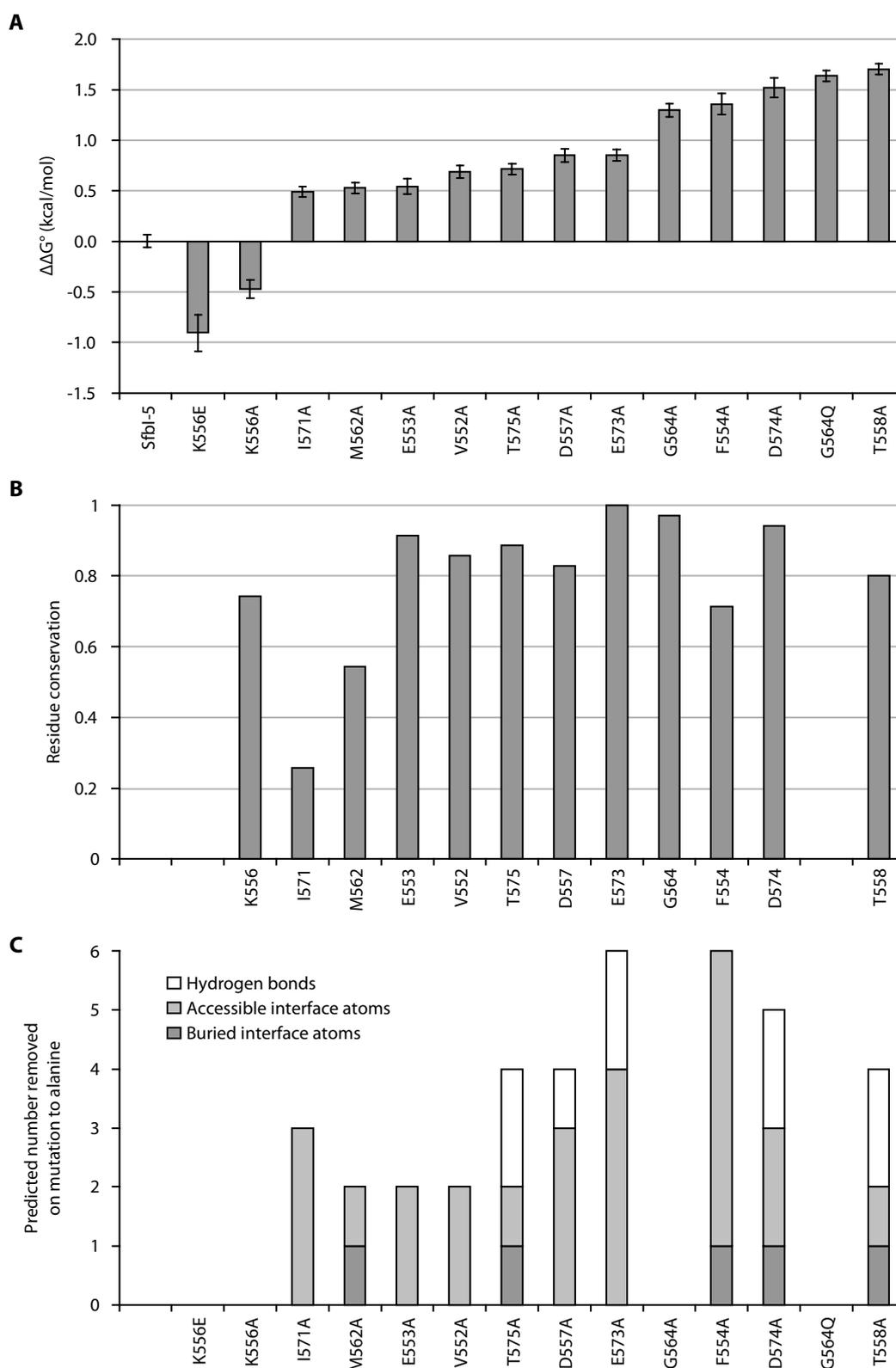
Figure 62. The effect of mutating SfbI-5 residues on binding to pNTD compared to conservation of the mutated residues and the predicted interface atom and hydrogen bond changes on mutation to alanine. A. ΔΔG° values for SfbI-5 mutants binding to pNTD compared to wild-type SfbI-5, from most to least favourable change in ΔΔG° (left to right). Errors shown are standard errors. **B.** Residue conservation for the residues that were mutated in the SfbI-5 constructs. These values were calculated on the Scorecons server (Valdar, 2002) using the final alignment of all non-identical FnBRs. Fractional

conservation levels are reported, such that the most conserved FnBR residue had a conservation score of one, and the least conserved residue had a conservation score of zero. **C.** Predicted number of accessible or buried interface atoms and hydrogen bonds removed on mutation to alanine shown in light grey, dark grey and white, respectively. These values were calculated from appropriate F1-module:FnBR peptide structures using the results from ProFace analyses of the structures (Saha *et al.*, 2006). Since no available structure has a lysine at a position equivalent to K556, values for this residue were not calculated. Graph columns in B and C are aligned with respect to the columns in A for easier comparison.

4.4.5.3   Analysis of ΔΔG° values

In an effort to rationalize the effect of mutations in SfbI-5 on the ΔΔG° for binding to pNTD, residue conservation and propensity for disorder were considered, as were the following in the context of F1-module:FnBR peptide structures: BSA, number of buried atoms and removal of hydrogen bonds as a result of mutations (Figure 62).

The six residues mutated in the ^{4}F1-binding region in SfbI-5 were VEF and KDT (Figure 34). At the corresponding positions in STAFF1 and STAFF5, STAFF5 contains a VEF motif and STAFF1 contains an EDT motif. Therefore, the ^{4}F1^{5}F1:STAFF5 and ^{4}F1^{5}F1:STAFF1 structures were used to help rationalize the effect of mutating VEF and DT, respectively. The seven residues mutated in the ^{2}F1^{3}F1-binding regions were analysed in the context of the ^{2}F1^{3}F1:PyTT5 structures I and III. The effect of each mutation is discussed below.

I571A, M562A and E553A mutants had the smallest unfavourable ΔΔG° values, all approximately 0.5 kcal/mol (Table 37; Figure 62A). I571 and M562 were mutated because they are less well conserved (HMM positions 19 and 30, Figure 34; Figure 62B) and were predicted to form β-strand interactions with an F1 module *via* backbone-backbone hydrogen bonds, a prediction that was confirmed by ^{2}F1^{3}F1:PyTT5 Structure III (Figure 59C). The small ΔΔG° values for these mutations indicate that the sidechains of these two residues are relatively unimportant in binding, as would be predicted from their lower conservation. Therefore, these mutations indicate the 'background' effect of mutating residues with low conservation, and suggest that changes above 1 kcal/mol are probably significant in the context of this system.

However, the similarly small effect of the E553A mutation is somewhat surprising given the high conservation of a negatively charged residue at this position relative to

other residues in FnBRs (HMM position 10, Figure 34; Figure 62A and B). The entropy and enthalpy of binding for this mutant are also similar to those of wild-type SfbI-5 (Table 37), implying the mutation caused no significant changes to the thermodynamics of binding. There is a glutamate at the position corresponding to E553 in both STAFF1 and STAFF5. Bingham *et al.* (2008) observed that this glutamate is near R197 in the linker between [4]F1 and [5]F1, implying an interaction with R197 may be the reason for a conserved negatively charged residue at this position. However, there is no hydrogen bond formed between the conserved glutamate and R197, despite no obvious crystal contacts stopping this interaction in. For example, chains K and L in the [4]F1[5]F1:STAFF5 structure or chains A and D in the [4]F1[5]F1:STAFF1 structure. Therefore, both the [4]F1[5]F1:FnBR peptide structures and the mutant binding data imply that E553, although conserved, is unlikely to form a key role in the interaction between SfbI-5 and pNTD.

The $\Delta\Delta G°$ values for the V552A and F554A mutations of 0.69 and 1.4 kcal/mol, respectively, can be explained by the loss of favourable van der Waals contacts (Figure 62A and C). The residues that correspond to V552 and F554 in the [4]F1[5]F1:FnBR peptide complexes interact with two hydrophobic pockets in [4]F1 (Bingham *et al.*, 2008). The mutation of these residues would remove these favourable van der Waals interactions. This is particularly the case for F554, which is predicted to lose six interface heavy atoms on mutation to alanine, including one heavy atom that is predicted to be completely buried (Figure 62C). Of all the SfbI-5 mutants, this is the most interface atoms predicted to be removed on mutation to alanine, which may be the reason this mutation has the fourth largest effect on binding.

K556A had a small but favourable $\Delta\Delta G°$ of −0.47 kcal/mol and K556E had a more favourable $\Delta\Delta G°$ of −0.90 kcal/mol. This result would be predicted from conservation: glutamate is the most conserved residue at this position; lysine is the second most conserved. It would also be predicted from the [4]F1[5]F1:STAFF1 structure, where a glutamate at the position equivalent to K556 forms hydrogen bonds indirectly with K172 in the C-strand of [4]F1 *via* two invariant bridging water molecules. Thus, a lysine would be predicted to have unfavourable electrostatic interactions at this position. Given that glutamate is predicted to be more energetically favourable at this position, both from the [4]F1[5]F1:STAFF1 structure and the $\Delta\Delta G°$ value for the K556E mutation, why is lysine the second most conserved residue type at this position, rather than, for

example, aspartate? There are two possible explanations. First, there might not be an unfavourable electrostatic interaction between K556 in SfbI-5 and K172 in ^{4}F1. In contrast to the ^{4}F1^{5}F1:STAFF1 structure, in the ^{4}F1^{5}F1:STAFF5 structure K172 does not interact with the conserved glutamate at the position corresponding to K556, but instead points away from the glutamate and interacts with D157 on the ^{4}F1 B-strand either directly *via* a salt-bridge or indirectly *via* a bridging water molecule. Therefore, it may be that the F1-module:FnBR interaction tolerates the juxtaposition of two lysines by K172 pointing away from K556 and interacting with D157 instead. Second, the lysine at this position may be conserved because, like glutamate and unlike aspartate, it is more common in disordered proteins than ordered proteins, indicating it is disorder promoting (Dunker *et al.*, 2008). Therefore, the selection pressure for having a disorder promoting residue at this position might outweigh the selection pressure for tighter binding to pNTD.

D557A had a relatively moderate unfavourable ΔΔG° of 0.85 kcal/mol and D574A had a relatively large unfavourable ΔΔG° of 1.5 kcal/mol. These two residues are both predicted to form hydrogen bonds with arginines in the E-strands of ^{4}F1 and ^{2}F1, respectively. The aspartate at an equivalent position to D557A in the ^{4}F1^{5}F1:STAFF5 structure forms a hydrogen bond with R191 in ^{4}F1. In ^{2}F1^{3}F1:PyTT5 Structures I and III, D574 either forms two hydrogen bonds with R101 in the E-strand of ^{2}F1 or hydrogen bonds with both R101 and R99, which is in the loop between the D- and E-strand (Figure 59C). Not only does the mutation of D557 to alanine have less of an effect on binding than D574, but D557 is also less conserved than D574 (compare HMM positions 14 and 33 in Figure 34). This is probably because the ^{2}F1-binding region that interacts with the conserved aspartate is more positively charged than the corresponding ^{4}F1-region: there is a serine in ^{4}F1 at the equivalent position to R99 in ^{2}F1. Thus, the results of the D557A and D574A mutation are consistent both with the conservation and the available structural information for these residues.

T558A had the largest unfavourable ΔΔG° out of the SfbI-5 mutants, 1.7 kcal/mol, and T575A had a relatively moderate ΔΔG° of 0.7 kcal/mol. This result is particularly surprising for two reasons. First, T558 and T575 appear to form equivalent interactions with ^{4}F1 and ^{2}F1, respectively, so would be predicted to have a similar ΔΔG° values on mutation to alanine. T558 is predicted to form hydrogen bonds with the amide group of F156 and the carbonyl group of G190 in ^{4}F1, and T575 forms equivalent hydrogen

bonds with F67 and G100 in ^{2}F1 (Figures 58, 59C). In doing so both residues have the same number of accessible and buried interface atoms and their sidechains contribute similar amounts of BSA to the interface. Second, based on residue conservation it would be predicted that T575 would have a larger $\Delta\Delta G°$ than T558; since T575 is more conserved than T558 (compare HMM positions 15 and 34, Figure 34). A possible explanation for this surprising result is that the T558A mutation in combination with the energetically unfavourable K556 leaves D557 to form the only favourable sidechain interactions in the KDT motif, perhaps this interaction is not enough to keep this motif bound.

G564A and G564Q both had relatively large effects on $\Delta\Delta G°$ of 1.3 and 1.6 kcal/mol, respectively. The effect of the G564A mutation was predicted both from the very high level of conservation of glycine at this position (HMM position 21, Figure 34) and from the ^{2}F1^{3}F1:FnBR *S. aureus* peptide structures (Section 4.3.1). The ^{2}F1^{3}F1:PyTT5 Structure III confirmed that this glycine is conserved to allow β-strand formation along the E-strand of ^{3}F1 while minimizing steric clashes with W146 in ^{3}F1. That this would otherwise be thermodynamically unfavourable is shown by the relatively large change in $\Delta\Delta G°$ on mutation to alanine. The G564Q mutant had a less favourable $\Delta\Delta G°$ value compared to the G564A mutant. Therefore, the hypothesis that the χ1 angle of W146 can vary to allow the glutamine to reach across and interact favourably with R125 in the C-strand of ^{2}F1 (Figure 39; Section 4.3.1) is probably incorrect. Thus, the G564A and G564Q mutant $\Delta\Delta G°$ values are consistent with residue conservation and further illustrate the role played by conserved residues in the interaction between ^{3}F1 and FnBR peptides.

The E573A mutation had a moderate $\Delta\Delta G°$ of 0.85 kcal/mol, which is surprising given both the high level of conservation of this residue (HMM position 32, Figure 34) and its role in interacting with ^{2}F1. In the ^{2}F1^{3}F1:FnBR *S. aureus* structures and the ^{2}F1^{3}F1:PyTT5 Structure I, this highly conserved glutamate forms a favourable end-on-end salt-bridge with R83 in the A-strand of ^{2}F1 (Bingham *et al.*, 2008). Its relatively modest effect on binding may be because the favourable gain in enthalpy on forming two salt bridges with R83 is offset by the loss of enthalpy caused by breaking hydrogen bonds with water molecules. Its high conservation may be because glutamate is enriched in disordered proteins compared with ordered proteins, and so probably promotes disorder (Dunker *et al.*, 2008).

In conclusion, the relative effects of most mutations are consistent with residue conservation and can be explained by analysis of the available F1-module:FnBR peptide structures. There appears to be a general trend that mutation of residues that are predicted to lose a buried interface atom and/or the ability to form hydrogen bonds as a result of mutation has a more unfavourable effect on binding to pNTD than mutation of residues that just lose accessible interface atoms (Figure 62).

Where the effect of the mutation of a residue is not consistent with its conservation, it might be playing a role in maintaining the disordered state of SfbI-5. Alternatively, these residues might have a role in hindering competing interactions in their biological context of full length SfbI binding to Fn.

4.4.6 Effect of SfbI-5 mutations in a more biologically relevant context

In addition to adhering to host tissues *via* Fn in the ECM, *S. pyogenes* may come into contact with Fn in plasma during the course of infection. For example, in superficial infections, *S. pyogenes* causes inflammation, which may result in vascular leakage and exposure of the bacteria to plasma. Furthermore, *S. pyogenes* will contact plasma in severe invasive infections as a result of tissue damage, which may also result in bacteraemia (reviewed in Tart *et al.*, 2007).

A series of GST pull-down experiments between GST-SfbI-5 constructs and plasma were used in an attempt to measure the effect of SfbI-5 mutations on the ability of SfbI-5 to bind to Fn (Section 4.3.4). These experiments were more biologically relevant than the ITC experiments for the following reasons. First, the experiments examined SfbI-5 binding full-length Fn in plasma rather than pNTD in buffer. It is possible that SfbI-5 residues are conserved because these residues help SfbI-5 bind $^{1-5}$F1 by disrupting interactions between Fn and either itself, other Fn molecules or other plasma macromolecules; this hypothesis was tested by measuring binding of SfbI-5 to Fn in plasma. Second, in the pull-down experiment GST-SfbI-5 is bound to a bead *via* glutathione, mimicking the biological situation where full-length SfbI is covalently tethered to the outside of a bacterial cell (Section 1.3.3). Therefore, these experiments were a simple model for *S. pyogenes* binding Fn *via* SfbI-5 in blood, such as might occur in bacteraemia.

The SfbI-5 construct pull-down experiments were in agreement with the ITC results, in that no mutation blocked SfbI-5 binding Fn. Assuming there are no significant inter- or intra-molecular interactions involving Fn in plasma, the maximum decrease between the amount of Fn bound by SfbI-5 and a mutant is predicted to be ~16%. This assumes that: the dissociation constants found by ITC for SfbI-5 and the T558A mutant binding to pNTD applied to SfbI-5 and the T558A mutant binding to Fn in plasma; two molecules of SfbI-5 bind one molecule of Fn; the concentration of Fn dimer in plasma is 0.9 μM; the concentration of SfbI-5 in the pull-down is 1.8 μM; and the wash steps following the pull-down had a negligible relative effect on the amount of Fn bound by wild-type and mutant SfbI-5. These approximate calculations confirm the initial assumption that only if the amount of Fn bound by mutants was significantly affected by a component of plasma would the difference between wild-type and mutant be detectable by the pull-down experiments. Unfortunately, however, the experimental error in the final pull-down experiments was too high to allow any conclusions to be drawn about the effect of other mutated SfbI-5 residues on binding to Fn, except that the pull-down results do not disagree with the ITC results.

4.4.7 Summary and alternate explanations for SfbI-5 residue conservation

Mutating SfbI-5 conserved residues to alanine had relatively little effect on binding to pNTD, and no hot spot residues were identified (Figure 62; Section 4.4.5.2). We predict that the absence of hot spot residues is a direct result of the extended interface between SfbI-5 and pNTD, which results in no sidechain residues becoming completely buried on interface formation. Therefore, the energetic cost of losing the interaction made by a sidechain on a residue's mutation to alanine is probably compensated for by sidechain atoms being replaced by water (Section 4.4.5.2).

The relative effect of most mutations has been rationalized based on residue conservation, and the role of corresponding residues in the available F1-module:FnBR structures (Section 4.4.5.3). However, although there is a plausible explanation for the absence of hot spots in SfbI-5, their absence raises an interesting question: what selective pressure is leading to SfbI residue conservation? This question is raised by two facts: the weakest mutant by ITC had a dissociation constant of 60 nM; and the concentration of Fn dimers in plasma is 0.6–0.9 μM. Therefore, even the weakest SfbI-5 mutant is likely to be almost completely bound by Fn at equilibrium (Section 4.3.3.2).

This conclusion assumes that the dissociation constant measured for SfbI-5 binding to pNTD in PBS is valid in the biological context of *S. pyogenes* in plasma binding to Fn *via* SfbI. If this is the case, it appears unlikely that there would be selective pressure on *S. pyogenes* to keep these residues conserved. If these residues are not conserved to maintain tight binding to Fn, why are they conserved? There are at least two plausible explanations.

First, it is possible that mutation of conserved residues may have a greater effect on SfbI-5 binding to full-length Fn in blood or the ECM in tissues than they do in binding pNTD in PBS. This hypothesis was examined in the pull-down experiments, which showed that within high experimental error SfbI-5 mutants did not bind significantly less Fn in plasma than would be predicted from the ITC results of SfbI-5 mutants binding pNTD (Sections 4.3.4.3, 4.4.6). Although by no means conclusive, these results suggest that mutation of conserved SfbI-5 residues does not drastically alter binding of SfbI-5 to plasma Fn.

Second, mutation of an SfbI-5 conserved residue may significantly change the association and dissociation rates, and it may be that one of the rates is under selective pressure rather than the equilibrium dissociation constant. A mutation could have a large effect on the association and dissociation rates without affecting the Gibbs energy of complex formation ($\Delta G°$) only if the ratio between the dissociation and association rates remains constant. This occurs when a mutation affects the energy of the transition state between free SfbI-5 and SfbI-5 bound to $^{1-5}$F1 without affecting the energy of either free or bound SfbI-5 (Figure 63; Chilkoti *et al.*, 1995). Given this, the possible biological reasons for either the association or dissociation rate to be under selective evolutionary pressure, rather than the equilibrium constant, were considered.

The work in this chapter suggests that the bacterial FnBPs might act as mechanical tethers between the bacteria and Fn, and therefore, the dissociation rate is under selective evolutionary pressure. *S. pyogenes* and *S. aureus* are exposed to forces while colonizing the host: for example, shear force, either from blood flow while the bacteria are in the blood circulatory system, or from saliva in a pharyngeal infection, when the host swallows. Therefore, it is likely that FnBPs and other bacterial adhesins need to resist force, to remain adhered to host tissue.

$[S_m:N]^\ddagger$ $\Delta\Delta G^\ddagger$ $[S_{wt}:N]^\ddagger$ $[S_M:N]^\ddagger$

$\Delta G_m^{\ddagger} \neq \Delta G_{wt}^{\ddagger}$ mutation m $\Delta G_{wt}^{\ddagger}$ mutation M $\Delta G_M^{\ddagger} = \Delta G_{wt}^{\ddagger}$

$[S_m]+[N]$ $[S_{wt}]+[N]$ $[S_M]+[N]$

$\Delta G_M \neq \Delta G_{wt}$

$[S_M:N]$

$\Delta\Delta G$

$\Delta G_m = \Delta G_{wt}$ ΔG_{wt}

$[S_m:N]$ $[S_{wt}:N]$
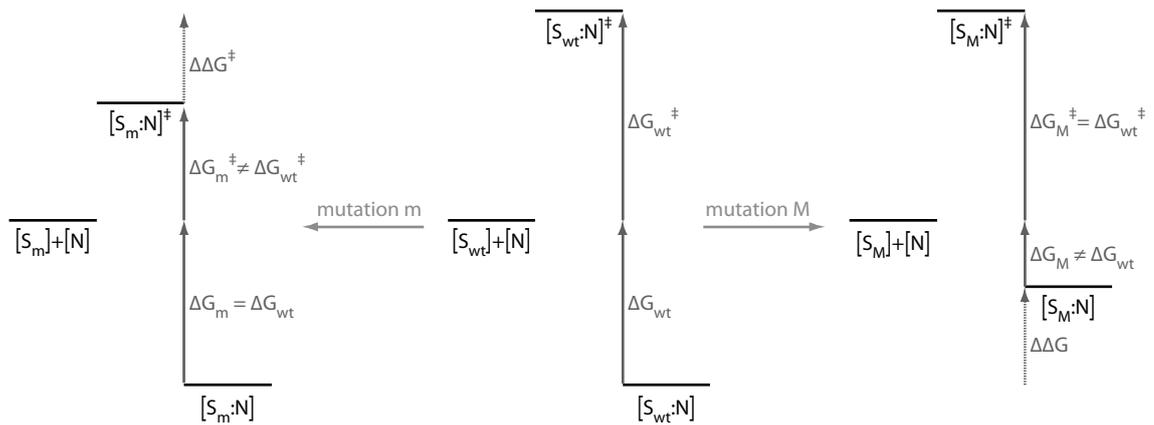
Figure 63. Schematic of the effects a mutation can have on the thermodynamics of protein complex formation. Two proteins, wild-type SfbI-5 (S_{wt}) and NTD (N), form a complex S_{wt}:N with a favourable Gibbs energy of ΔG°_{wt} *via* a transition state $[S_{wt}:N]^\ddagger$ with an activation Gibbs energy of $\Delta G^\circ_{wt}{}^\ddagger$ (middle panel). The thermodynamic effect of any SfbI-5 mutation can be thought of as consisting of two parts shown here as two limiting cases. In one limiting case, represented by SfbI-5 mutation M (S_M; right panel), the Gibbs energy of S_M and N relative to the complex S_M:N changes by $\Delta\Delta G^\circ$ (*i.e.*, $\Delta G^\circ_M \neq \Delta G^\circ_{wt}$), but the activation Gibbs energy stays the same (*i.e.*, $\Delta G^\circ_M{}^\ddagger = \Delta G^\circ_{wt}{}^\ddagger$). In the other limiting case represented by mutation m (S_m, left panel) the activation Gibbs energy changes by $\Delta\Delta G^{\circ\ddagger}$ (*i.e.*, $\Delta G^\circ_m{}^\ddagger \neq \Delta G^\circ_{wt}{}^\ddagger$), but the Gibbs energy of S_m and N relative to the complex S_m:N stays the same (*i.e.*, $\Delta G^\circ_m = \Delta G^\circ_{wt}$). In the latter case (mutation m), the association and dissociation rates either both increase or both decrease such that the ratio between them, the equilibrium dissociation constant, does not change (*i.e.*, $\Delta\Delta G^\circ = 0$). Similar schematics can be drawn for situations in which the affected thermodynamic parameter is not Gibbs energy, but instead enthalpy or entropy. This figure is adapted from Figure 3 in Chilkoti *et al.*, 1995.

There are studies that suggest cocci FnBPs are capable of acting as mechanical tethers, both *in vivo* and *in vitro*. Most studies involve *S. aureus*, but Lembke *et al.* (2006) investigated *S. pyogenes* clinical isolates and found that serotype M2 was able to bind to and form a biofilm on Fn-coated glass under flow conditions, although the shear rate experienced by the cells in these experiments is not specified. These researchers further found that the likely adhesins involved were probably MSCRAMMs rather than M protein. Although they did not investigate which MSCRAMMs are involved, at least two FnBPs have been identified in the HMM searches presented here with FnBRs from serotype M2 (Table 31). Furthermore, *in vivo* studies of a mouse model of mastitis (an infection of the mammary glands) with two isogenic *S. aureus* strains with and without FnBP expression have shown that the former was better at colonizing glands subjected to suckling by pups compared to non-suckled glands (Brouillette *et al.*, 2003). These results indicate FnBPs help *S. aureus* resist the shear force from milk flow. The ability

of Fn to bind to whole *S. aureus* cells expressing FnBPs has also been studied with atomic force microscopy (AFM; Mitchell *et al.*, 2008). A rupture bond strength (unbinding force) of ~1000±500 pN at a pulling rate of 2 μm/s was measured. This is an unusually large force, with an unusually large range, which the researchers propose is a result of the predicted extended, multivalent interaction formed between Fn and *S. aureus* FnBPs (Mitchell *et al.*, 2008). Therefore, there is much evidence to suggest that FnBPs can act as mechanical tethers between bacteria and Fn.

Assuming there is evolutionary pressure for FnBPs to stay bound to Fn during an applied force, such as blood or milk flow, then the dissociation rate rather than the equilibrium constant will be under selective pressure (reviewed in Robert *et al.*, 2007). This is because rupture bond strength is theoretically proportional to the logarithm of the dissociation rate, and not proportional to the equilibrium dissociation constant. Furthermore, the dissociation rate is not a constant, but depends on the applied force: a larger applied force will lead to faster dissociation. While dissociation of a mechanical tether is force-dependent, association is distance-dependent, and depends on molecule length and flexibility (Robert *et al.*, 2007). In this sense, the disordered state of SfbI will probably be an advantage, helping complex formation.

Thus, it is plausible that association and/or dissociation rates of SfbI-5 binding to Fn might be under selective pressure, rather than the equilibrium constant, leading to residue conservation that cannot be readily explained from the magnitude of $\Delta\Delta G°$ values measured by mutating conserved SfbI-5 residues by ITC. These theories could be tested by measuring the dissociation rates by either AFM or surface plasmon resonance.

5 Characterization of a putative eukaryotic FnBR

5.1 Introduction and justification

It is plausible that the four putative eukaryotic FnBPs from the deer tick *I. scapularis* identified in the HMM searches are functional FnBPs (Figure 64; Table 31; introduction to *I. scapularis* in Section 1.8). If these putative FnBPs are functional, then they are the first eukaryotic examples of FnBPs. Also, if they are shown to be functional it helps to validate the functional relevance of the other novel FnBRs identified in HMM searches.
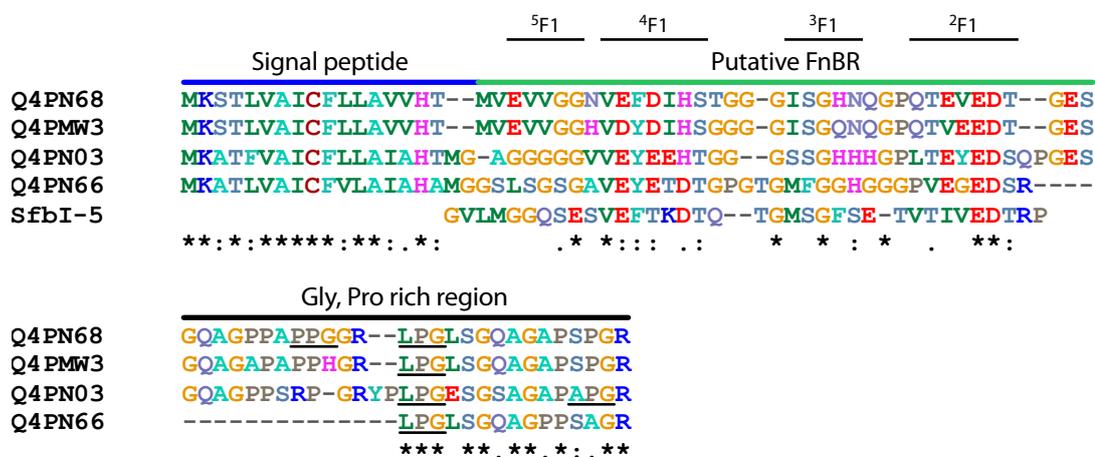
Figure 64. Alignment between the four putative *I. scapularis* FnBPs and the [2-5]F1-binding region of SfbI-5. Indicated above the sequences are: the signal peptides predicted by the signal P server (Bendtsen *et al.*, 2004); the putative FnBRs including the predicted F1-module binding regions based on previously determined F1-module:FnBR peptide structures (Bingham *et al.*, 2008); and a region rich in glycine and proline residues. Putative prolyl hydroxylase sites are underlined. TickFnBR (see text) is the putative FnBR from Q4PN68, but it also has an N-terminal Gly-Ile di-peptide from the expression plasmid (not shown). The [2-5]F1-binding region of SfbI-5 is shown for comparison, with residues aligned with respect to each other, using the structures of [2]F1[3]F1 in complex with an SfbI-5 peptide (Section 4.2.2) or a TickFnBR peptide (Section 5.5.3). Residues are coloured based on chemical composition. Conserved residues in the *I. scapularis* FnBRs are indicated below the alignment, where an asterisk (*) colon (:) and full stop (.) indicate identical, conserved and semi-conserved residues, respectively.

For a putative FnBP to be functional, three criteria need to be met: i) the putative FnBP can specifically bind Fn; ii) the putative FnBP has access to Fn, either the originating organism's or a host's; and iii) an FnBP's usefulness can be justified based on the biology of the originating organism. The latter two criteria are justified by experimental evidence that indicates these proteins are expressed in the saliva of a tick while it is

attached to a mammalian host, and feeding on the host's blood (Ribeiro *et al.*, 2006). Therefore, it is likely that the putative FnBPs are exposed to Fn in the ECM and blood of the host, where it is plausible they might aid tick attachment and/or interfere with wound healing. Support for these hypotheses will be discussed further in Section 5.5.1.

The first criterion, that the putative FnBPs can bind Fn, is predicted by their sequence similarity to known FnBRs (*e.g.,* Figure 64). In the work presented here, this prediction was tested experimentally for 'TickFnBR', a predicted FnBR from an *I. scapularis* FnBP (UniProt ID Q4PN68).

5.2   TickFnBR cloning, expression and purification

The cDNA for TickFnBR was created synthetically from PCR of two overlapping oligonucleotides (Section 2.2.4). This technique not only allowed the TickFnBR cDNA to be appended with the additional 5′ and 3′ sequences necessary for cloning into the pGEX vector, but also allowed the TickFnBR cDNA to be codon-optimized for synthesis in *E. coli*. Using this synthetic cDNA a vector, pGEX-TickFnBR, was created and verified by DNA sequencing (data not shown). This vector allows expression of a GST-fusion of TickFnBR, GST-TickFnBR, that contains a specific cleavage site between the GST and TickFnBR sequences for the protease Factor Xa, which facilitates removal of the GST after purification. pGEX-TickFnBR was used to express GST-TickFnBR in *E. coli* at 37°C for 4 hours. GST-TickFnBR was purified from other soluble *E. coli* proteins by glutathione affinity chromatography. Purity was greater than 95% as determined by SDS-PAGE (Figure 65). After purification, Factor Xa protease was used to specifically cleave GST from TickFnBR (*e.g.,* '−' and '+' lanes, Figure 66). Next, TickFnBR was purified from the TickFnBR, GST and Factor Xa mixture.

Further purification of TickFnBR by 'acidic' rpHPLC was unsuccessful. RpHPLC was performed on a C4 column with a gradient between water and acetonitrile that contained 0.15% and 0.1% TFA, respectively, *i.e.,* acidic rpHPLC. A chromatogram from acidic rpHPLC purification of a sample of the TickFnBR, GST and Factor Xa mixture revealed an unexpectedly small peak relative to the GST peak (Figure 67A). SDS PAGE confirmed this small peak was TickFnBR (data not shown). The small peak size was noticeable particularly by comparison with the SfbI-5 peak height in rpHPLC purification of a sample of cleaved GST-SfbI-5 (Figure 67B). SDS-PAGE analysis of steps in the TickFnBR purification indicated that ~85% of the TickFnBR from the

cleavage reaction was lost during rpHPLC (data not shown). This was unexpected given that both the MW and predicted isoelectric point (4.00) of TickFnBR are comparable to other FnBRs successfully purified in our laboratory by acidic rpHPLC.
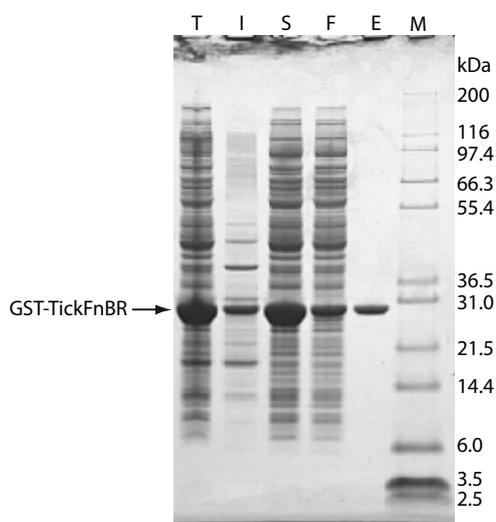
Figure 65. Analysis of GST-TickFnBR expression and purification. SDS-PAGE was performed on samples taken during the purification of GST-TickFnBR. Lane contents are indicated by the following abbreviations: T, total cellular protein; I, insoluble cellular protein; S, soluble cellular protein; F, flow-through from affinity purification of S; E, pooled eluate fractions; M, Mark 12 protein MW marker. See Table 9 for descriptions of these samples. The positions of the GST-TickFnBR band and MWs of the marker are indicated.
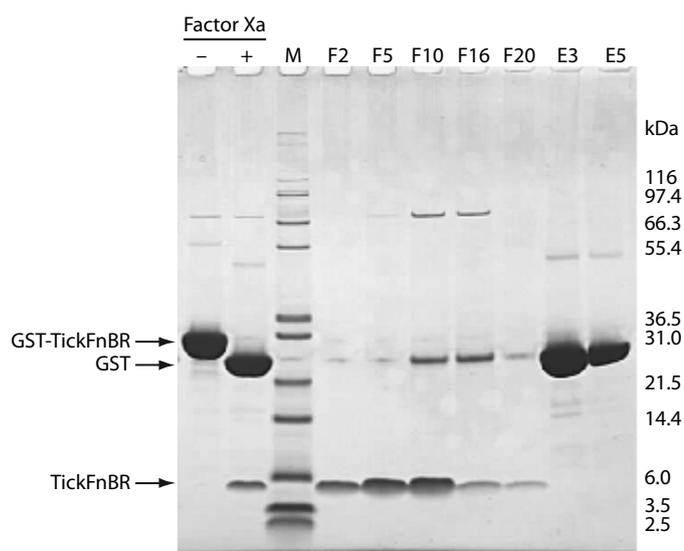
Figure 66. Cleavage of GST-TickFnBR with Factor Xa, followed by removal of most of the GST from TickFnBR with glutathione affinity chromatography. SDS PAGE analysis of samples of GST-TickFnBR before (−) and after (+) cleavage with Factor Xa and also samples of the fractions from glutathione affinity chromatography in which most of the cleaved GST was removed from TickFnBR. Lane contents are indicated, where abbreviations are: M, Mark 12 protein MW marker; F2, F5, F10, F16, F20; the second, fifth, tenth, 16th and 20th fraction, respectively, of the flow-through from the glutathione affinity chromatography, which contains material that did not bind to the glutathione agarose column. E3, E5, the third and fifth fraction, respectively, from the elution of GST from the glutathione agarose column with glutathione elution buffer.
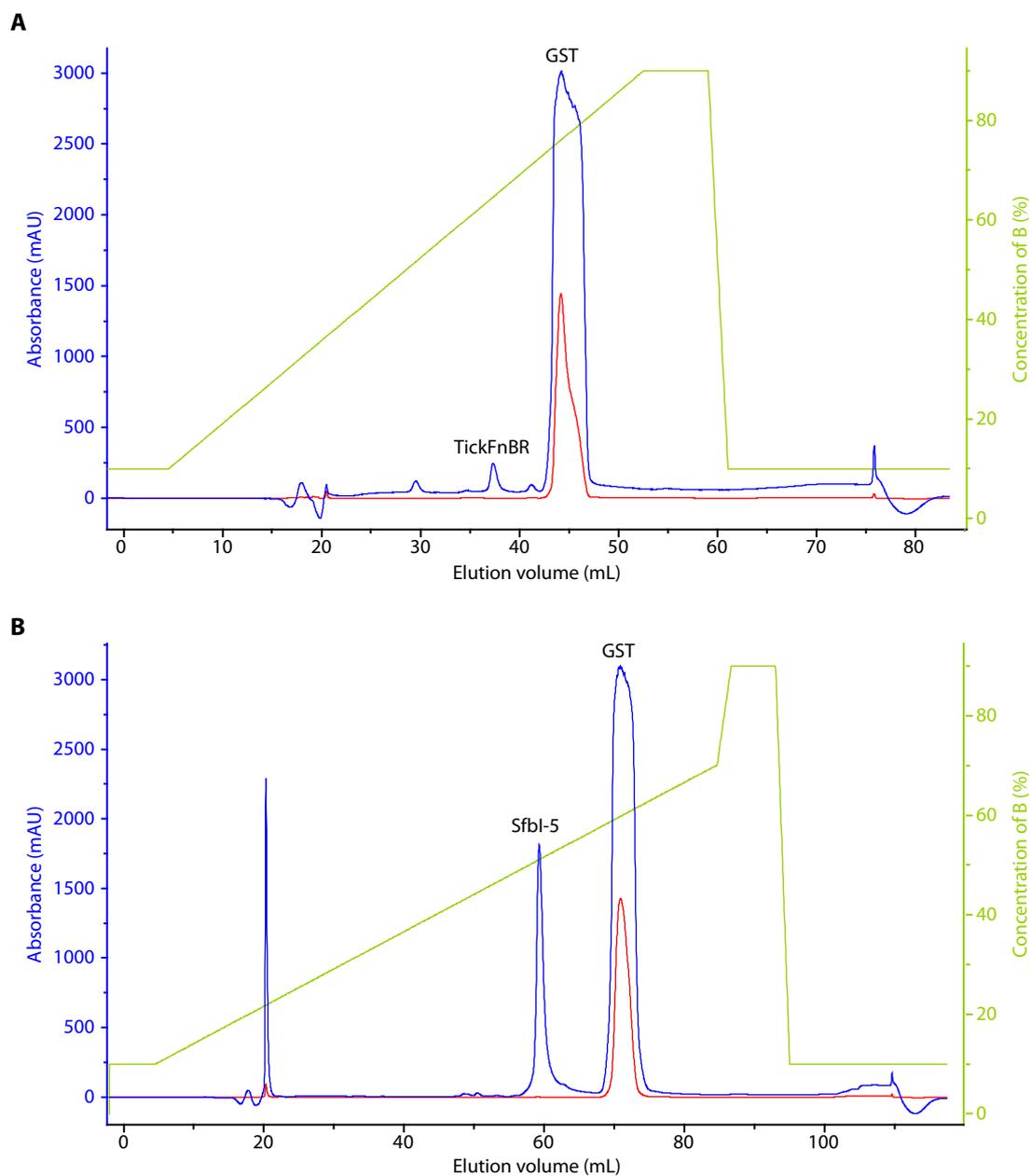
Figure 67. Comparison between acidic rpHPLC purification of GST-TickFnBR and GST-SfbI-5 samples after cleavage with Factor Xa protease to remove GST from the FnBR. Chromatograms shown are samples cleaved by Factor Xa of **A.** GST-TickFnBR (~2.5 mg fusion) and **B.** GST-SfbI-5 (~1 mg fusion). Both samples were purified under acidic conditions on a C4 column with a similar gradient between water and acetonitrile. Trace colours are: green, concentration of solvent B (acetonitrile with 0.1% TFA); blue and red, absorbance at 220 and 280 nm, respectively. Peaks are labelled with corresponding protein species.

Subsequently, purification of TickFnBR from GST and Factor Xa after cleavage was optimized to give a single product of the correct MW by ESMS. The key step is the final purification by rpHPLC on a C12 column with basic rather than acidic solvents: a gradient between 10 mM ammonium carbonate in water and acetonitrile was used

(Figure 68; Section 2.3.8). This 'basic' rpHPLC procedure is used in our laboratory to purify chemically-synthesized fragments of FnBRs. Not only does this step remove Factor Xa, but also, and more importantly, it allows the TickFnBR peptide to be lyophilized and weighed, which is useful in yield determination given that TickFnBR has a low extinction co-efficient at 280 nm. However, because we use a C12 column rather than a C4 column for basic rpHPLC, there was a risk that the GST would bind irreversibly to the column (Footnote 9, p 76). Therefore, most of the GST was removed from the cleaved GST-TickFnBR sample before basic rpHPLC with an additional glutathione affinity chromatography step (Figure 66; Section 2.3.4.4). It was found that the purity of TickFnBR was improved if both glutathione affinity purification steps were performed at 4°C rather than at room temperature (Figure 68). The final yield from the optimized protocol was 7.5 mg of TickFnBR from 1L of culture.

5.3 TickFnBR binds $^{1-5}$F1 and Fn

5.3.1 GST-TickFnBR binds to $^{1-5}$F1

FnBRs bind to $^{2-5}$F1 in Fn (Section 1.4.2). Therefore, whether GST-TickFnBR could bind to $^{1-5}$F1 was investigated using a GST pull-down (Figure 14; Section 2.9.2). In this experiment, $^{1-5}$F1 was mixed with GST-TickFnBR that was bound to glutathione agarose. If binding between TickFnBR and $^{1-5}$F1 occurred, a glutathione agarose:GST-TickFnBR:$^{1-5}$F1 complex would be formed (Figure 14). After washing away unbound $^{1-5}$F1, the GST-TickFnBR:$^{1-5}$F1 complex was eluted from the glutathione agarose with SDS PAGE sample buffer and analysed by SDS PAGE. This analysis revealed that $^{1-5}$F1 bound to GST-TickFnBR (P lane, Figure 69). The controls indicate there was some non-specific binding of $^{1-5}$F1 to the glutathione agarose and to GST (C1 and C2 lanes, Figure 69). However, the $^{1-5}$F1 band in lane P is too intense to be explained by non-specific binding. Therefore, this preliminary experiment indicates that the TickFnBR binds to $^{1-5}$F1, the same region of Fn to which streptococcal and staphylococcal FnBRs bind.
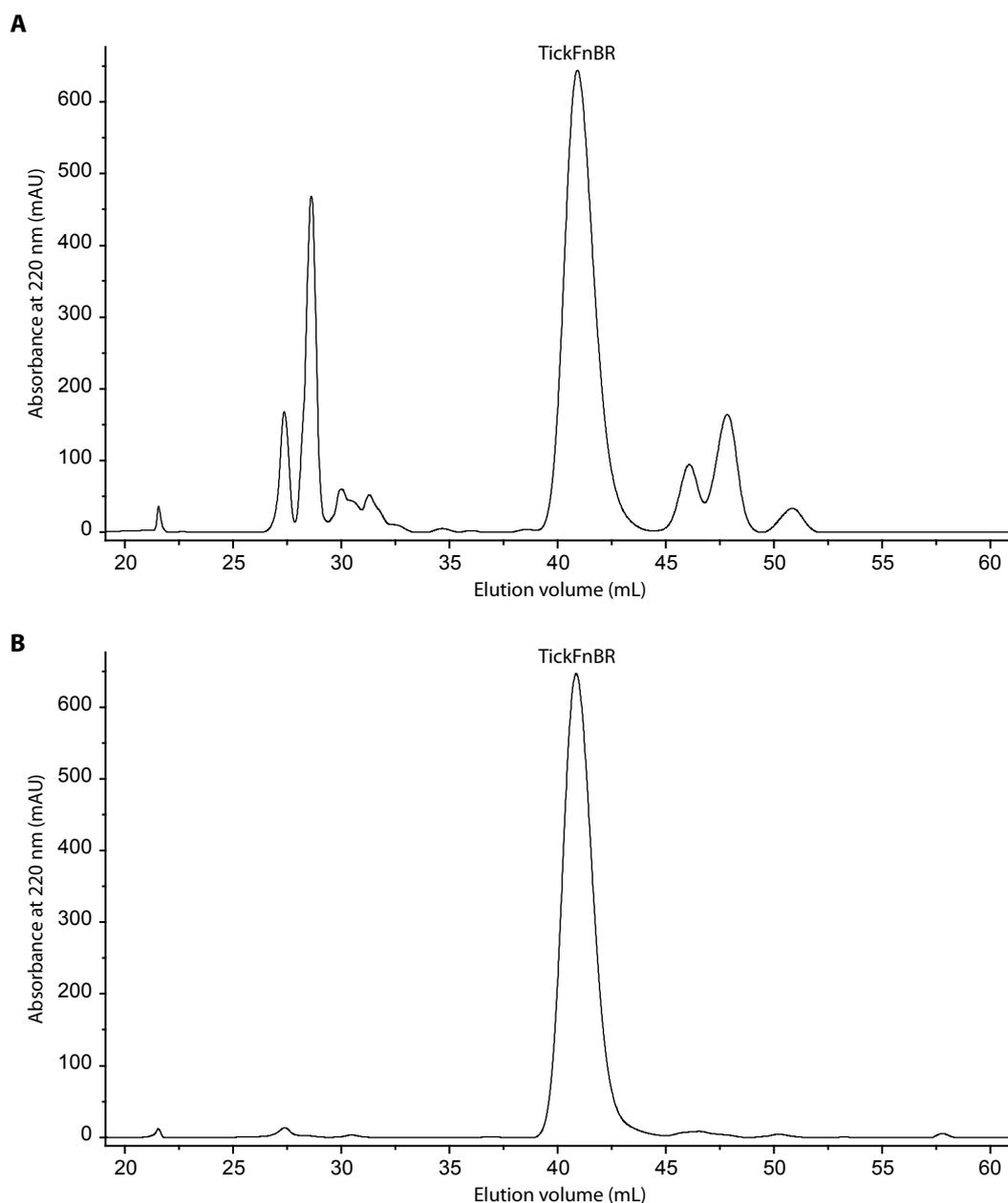
Figure 68. Basic rpHPLC purification of cleaved GST-TickFnBR samples indicated higher purity was achieved when affinity chromatography steps were performed at 4°C rather than at room temperature. Chromatograms showing the basic rpHPLC purifications of TickFnBR samples that had previously undergone two glutathione affinity chromatography steps at either **A.** room temperature or **B.** 4°C. The first glutathione affinity chromatography step was to remove GST-TickFnBR from the soluble cellular *E. coli* proteins (Figure 65) and the second was to remove GST from a GST-TickFnBR sample that had previously been cleaved with Factor Xa (Figure 66). Both samples were purified under basic conditions on a C12 column with the same method. The peak corresponding to TickFnBR is labelled.
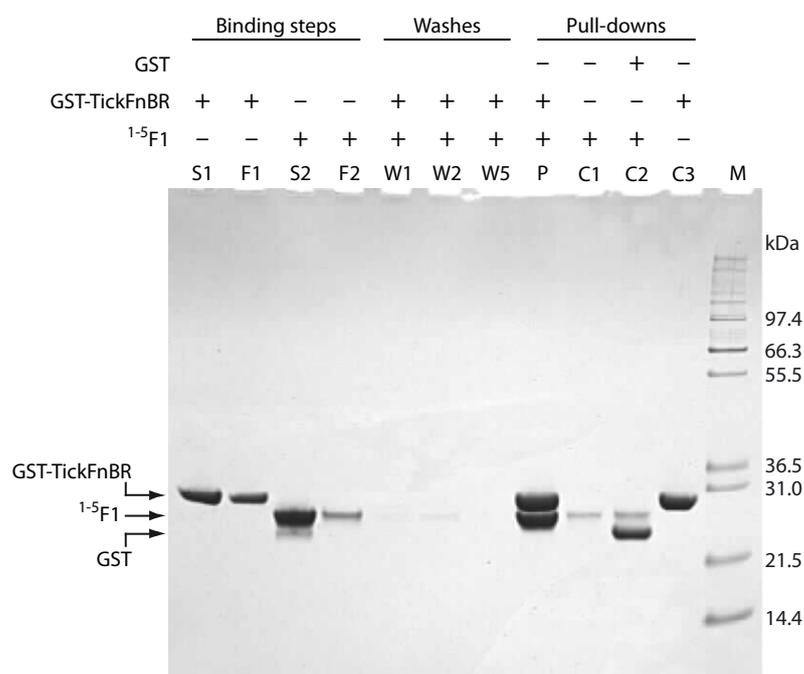
Figure 69. Analysis of a GST-TickFnBR pull-down of $^{1-5}$F1. SDS PAGE of samples from a GST-TickFnBR pull-down of $^{1-5}$F1 with three controls. The controls were $^{1-5}$F1 added to either glutathione-agarose or GST bound to glutathione-agarose to determine non-specific binding of $^{1-5}$F1; and also GST-TickFnBR that was not bound to $^{1-5}$F1, to ensure the sample did not degrade during the experiment. Lane contents are indicated, where '+' indicates the corresponding protein species had been added to the sample in that lane, and a '−' indicates the corresponding protein species had not been added to the sample in that lane. Abbreviations are: S, sample added to glutathione agarose for a binding step; F, flow-through collected after the binding step; W1, W2 and W5, PBS washes 1, 2 and 5, respectively, of pull-down 'P' after $^{1-5}$F1 binding step; P, pull-down of $^{1-5}$F1 by GST-TickFnBR; C, control pull-down experiments; M, Mark 12 protein MW marker. Bands corresponding to GST-TickFnBR, $^{1-5}$F1 and GST are indicated, as are the MWs of the marker.

5.3.2   SfbI-5 inhibits TickFnBR binding to $^{1-5}$F1

A GST pull-down experiment was performed in which both TickFnBR and SfbI-5 competed for binding to $^{1-5}$F1 (Figure 70; Section 2.9.2). This experiment also acted as a repeat for the initial pull-down experiment shown in Figure 69. For this experiment, the amount of $^{1-5}$F1 added to each pull-down was approximately equimolar to the estimated amount of GST-fusion bound to the glutathione agarose and the FnBRs were added so that they were at a molar ratio of 2:1 with $^{1-5}$F1 to ensure that all $^{1-5}$F1 was bound (Table 24). This experiment showed that, when added individually, both SfbI-5 and TickFnBR bound $^{1-5}$F1 (lanes P1, P2, S1–2, F1 and F2, Figure 70). However, when both SfbI-5 and TickFnBR are present, only SfbI-5 binding to $^{1-5}$F1 was detected (lanes Q1, Q2, S3, F3, S4 and F4, Figure 70). Thus, it is likely that SfbI-5 and TickFnBR have

overlapping binding sites on $^{1-5}$F1 and that SfbI-5 binds to $^{1-5}$F1 more tightly than TickFnBR.

The latter would be predicted, because SfbI-5 contains a ^{1}F1-binding region and TickFnBR does not. ITC experiments in PBS have shown that the K_d for SfbI-4 binding to pNTD is 62 nM at 25°C, whereas the K_d for SfbI-5 binding to pNTD at 25°C is too small to measure under the experimental conditions (Schwarz-Linek *et al.*, 2004b). SfbI-4 does not contain a ^{1}F1-binding region, but otherwise its sequence is 73% identical to the $^{2-5}$F1 binding region of SfbI-5 (Figure 29). Therefore, the difference in K_d between SfbI-4 and SfbI-5 for binding to $^{1-5}$F1 is likely to occur primarily as a result of the ^{1}F1-binding region in SfbI-5. This evidence supports the prediction that SfbI-5 will bind $^{1-5}$F1 more tightly than TickFnBR.

The conclusions made from these experiments assume that SfbI-5 and TickFnBR are at the same or very similar concentrations in their respective pulldowns Q1 and Q2 (Figure 70). While comparison between the band intensities of TickFnBR and SfbI-5 in lanes S3 and S4 in Figure 70 suggests that there was considerably more SfbI-5 present than TickFnBR during the binding steps, TickFnBR stains very poorly compared to SfbI-5 (Figure 71).

In summary, the results from this experiment support the hypothesis that TickFnBR is an FnBR that binds $^{1-5}$F1 at the same site and in a similar way as other FnBRs by binding to $^{2-5}$F1. These results also suggest that TickFnBR has a lower affinity than SfbI-5 for $^{1-5}$F1.
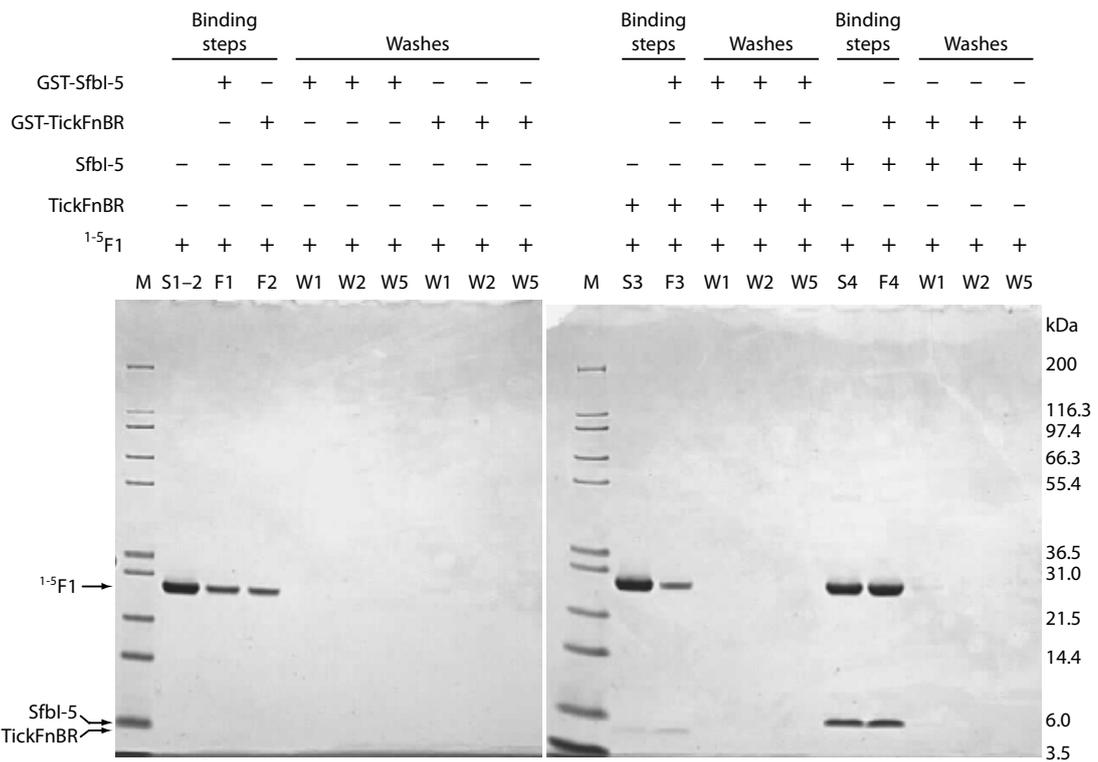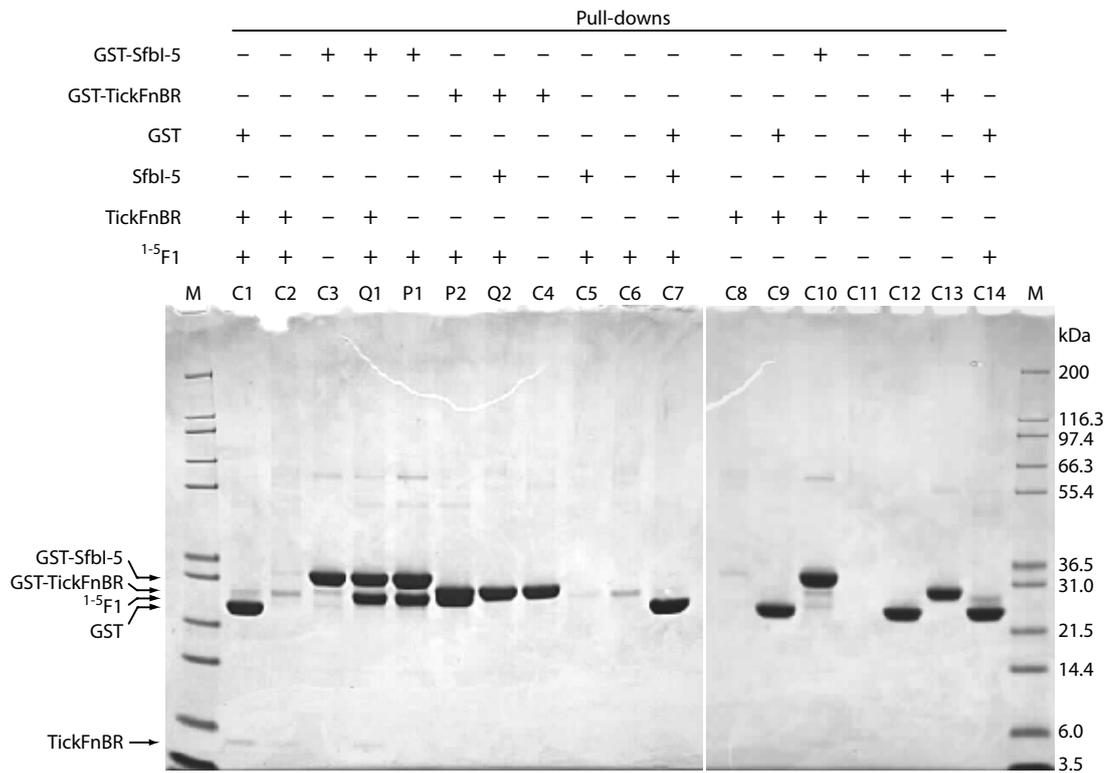
Figure 70. Analysis of a competition GST pull-down between SfbI-5, TickFnBR and $^{1-5}$F1. Lane contents are indicated, where '+' indicates the corresponding protein species had been added to the sample in that lane, and a '−' indicates the corresponding protein species had not been added to the sample in that lane. Abbreviations are: S, sample added to glutathione agarose for a binding step; F, flow-through collected after the binding step; S1–2, sample added to both pull-downs P1 and P2, the flow-through from which is shown in F1 and F2, respectively; W1, W2, W5, PBS washes 1, 2 and 5, respectively, of a pull-down after a $^{1-5}$F1 binding step; P1, P2, pull-down of $^{1-5}$F1 by either GST-SfbI-5 or GST-TickFnBR,

respectively; Q1, Q2; competition pull-down of $^{1-5}$F1 by either GST-SfbI-5 or GST-TickFnBR in the presence of either TickFnBR or SfbI-5, respectively. C1–C14, control pull-down experiments; M, Mark 12 protein MW marker. Bands corresponding to GST-TickFnBR, GST-SfbI-5, $^{1-5}$F1, GST, SfbI-5 and TickFnBR that appear on either gel are indicated on the leftmost gel and the MWs of the marker are indicated on the rightmost gel.
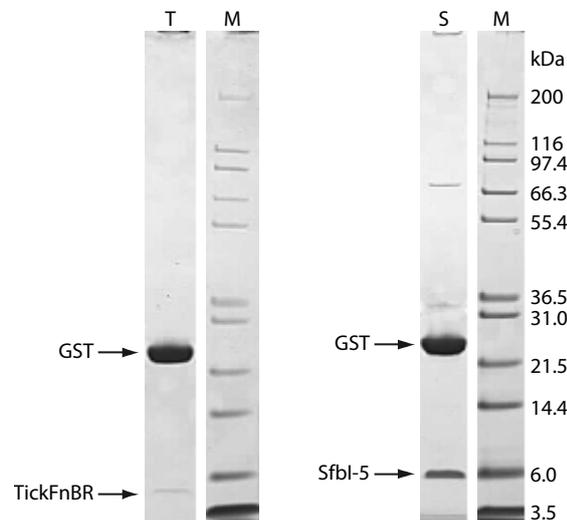
Figure 71. Similar molar amounts of SfbI-5 and TickFnBR have very different band intensities on SDS PAGE stained with Coomassie. Lane contents are indicated, where abbreviations are: T, cleaved GST-TickFnBR sample (0.27 nanomoles) and S, cleaved GST-SfbI-5 sample (0.24 nanomoles); M, Mark 12 protein MW marker. The concentration of both samples was determined by the A_{280} of the GST-FnBR sample before cleavage with Factor Xa. The intensity of the GST band also indicates that a similar amount of the cleaved GST-FnBR was loaded.

5.3.3 GST-TickFnBR specifically binds to Fn in plasma

A final series of pull-down experiments were performed to investigate whether GST-TickFnBR could bind to Fn in plasma. These experiments, performed in triplicate, showed that GST-TickFnBR specifically bound to Fn in plasma (Figure 72A). Importantly, this indicates that TickFnBR interacts with full-length Fn and not just $^{1-5}$F1 in isolation. In the same experiment GST-SfbI-5 also specifically bound to Fn in plasma (Figure 72A). Interestingly, band intensity analysis of the amount of Fn pulled-down by GST-SfbI-5 compared to GST-TickFnBR in each of the three experiments indicates that both FnBRs bound similar amounts of Fn (Figure 72B). This suggests that despite the absence of a ^{1}F1-binding region TickFnBR binds Fn tightly, with a binding affinity for Fn that is similar to that of SfbI-5. These experiments also show that TickFnBR can interact with Fn when many other macromolecules are present. That is, the Fn:TickFnBR interaction appears to be of both high affinity and specificity. In

conclusion, taken together the pull-down experiments strongly support the hypothesis that TickFnBR is an FnBR that tightly and specifically binds to the $^{1-5}$F1 region of Fn.
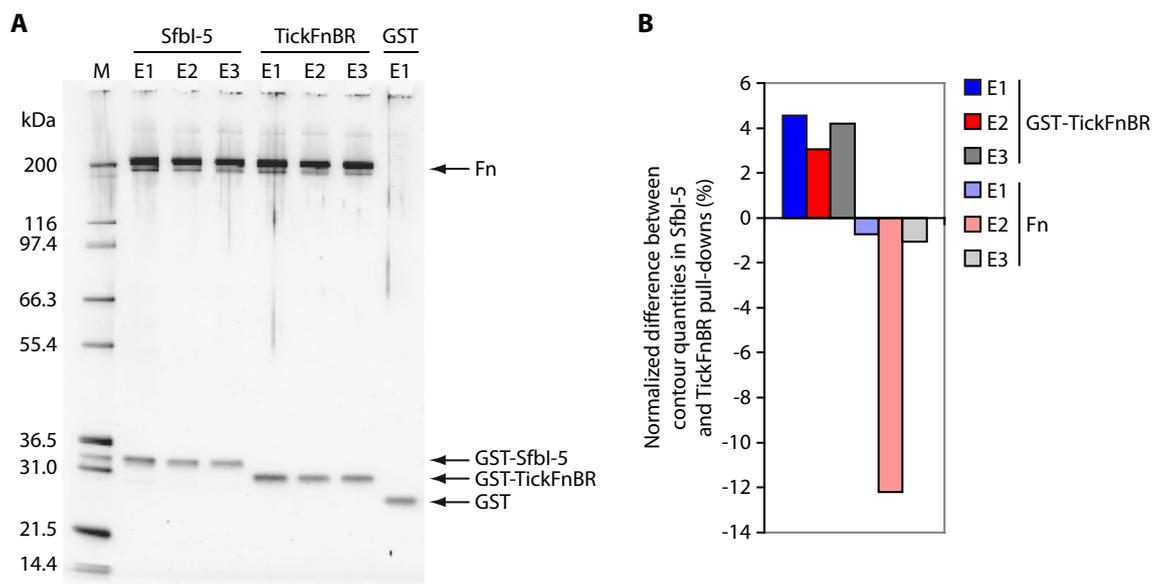
Figure 72. GST-TickFnBR bound a similar amount of Fn from plasma as GST-SfbI-5. A. Each experiment consisted of pull-downs with GST-TickFnBR, GST-SfbI-5 and controls, and was performed in triplicate (E1–E3). In each pull-down a GST-construct, indicated above each lane, was used to 'pull-down' Fn from human plasma. The two controls were GST with no FnBR fusion (GST) or glutathione magnetic particles with no bait attached. On this gel only one GST control is shown. This gel is also shown in Figure 50 (p 198), as are the other controls, which were run on different gels. The gel also has Mark 12 MW marker that has been 'spiked' with GST-SfbI-5 (M). Indicated are the MWs of the marker and bands corresponding to Fn, GST-SfbI-5, GST-TickFnBR and GST. The gel was stained with fluorescent stain to facilitate analysis of the band intensities (Section 2.9.1.3). **B.** For each experiment, the normalized difference between the intensities (*i.e.,* contour quantities) of either the GST-TickFnBR band or the Fn band relative to the corresponding band in the GST-SfbI-5 pull-down. The normalized values were calculated as described in the legend of Figure 51. A negative normalized value for either GST-TickFnBR or Fn means that less of that protein was present in the GST-TickFnBR pull-down than in the GST-SfbI-5 pulldown from the same experiment.

5.4   Structural characterization of the $^{1-5}$F1:TickFnBR interaction

The pull-down experiments (Section 5.3) support the hypothesis that TickFnBR interacts with the $^{1-5}$F1 region of Fn. Next, this interaction was characterized to determine whether the properties of the $^{2-5}$F1:TickFnBR interaction were similar to other $^{2-5}$F1:FnBR interactions. Specifically, the hypothesis that TickFnBR behaves like SfbI-5 was tested. That is, whether TickFnBR is disordered and undergoes a disorder-to-order transition to bind $^{2-5}$F1 *via* a tandem β-zipper.

5.4.1 TickFnBR appears to be disordered by NMR

To investigate whether TickFnBR was disordered or ordered in solution, a one-dimensional ^{1}H NMR spectrum was performed with TickFnBR (0.14 mM, pH 6.4) at 298 K. TickFnBR is either unfolded or disordered at pH 6.4 because the spectrum shows low chemical shift dispersion, with no peaks with a chemical shift greater than 8.6 ppm or less than 0.7 ppm (Figure 73). Peaks in these two regions are diagnostic of a folded protein, because they indicate the amide and aliphatic protons are in diverse chemical environments (Rehm *et al.*, 2002; Section 1.6.1.2). Given that FnBRs are intrinsically disordered (Chapter 3; House-Pompeo *et al.*, 1996; Penkett *et al.*, 1997), this result is in agreement with TickFnBR being an FnBR.
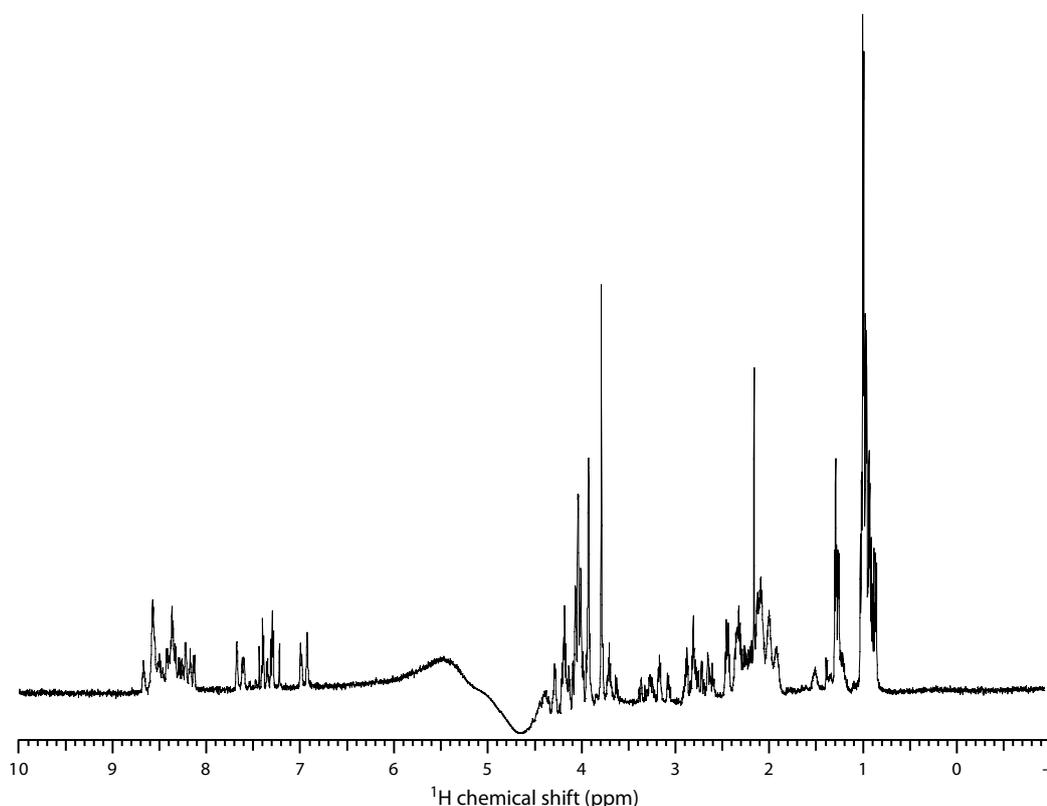
Figure 73. One-dimensional ^{1}H NMR spectrum of TickFnBR (0.14 mM, pH 6.4). Spectrum was recorded on a Bruker 700 MHz spectrometer at 298 K.

5.4.2 Attempts to characterize the interaction between TickFnBR and F1-modules or NTD were unsuccessful

An NMR chemical shift perturbation assay between TickFnBR and ^{4}F1^{5}F1 was attempted with the aim of measuring the binding between these two species. In this type

of assay, a series of HSQC spectra are acquired of a ^{15}N-labelled protein as increasing amounts of unlabelled ligand is titrated into the sample. ^{1}H^{N} and ^{15}N^{H} chemical shifts are highly dependent on the chemical environment of the nucleus, because electrons in nearby atoms shield the applied magnetic field to differing extents (Section 1.6.1.2). Therefore, because binding of a protein partner changes the chemical environment of the interacting residues, HSQC experiments can be used to detect protein-protein interactions. The perturbation of the chemical shifts of the ^{15}N-labelled protein during the titration, as increasing amounts of protein:ligand complex are formed, can be used both to determine the binding affinity between the protein and the ligand and also to determine the nuclei affected by ligand binding.

For this experiment, HSQC spectra were recorded of ^{15}N-labelled ^{4}F1^{5}F1 (0.15 mM, pH 6.3) before and after the addition of a pre-weighed aliquot of lyophilized TickFnBR (pH 6.5 before lyophilization) giving a final molar ratio of ^{4}F1^{5}F1:TickFnBR of 1:0.2 (Figure 74). However, on addition of TickFnBR a precipitate formed. Addition of another aliquot of TickFnBR led to a decrease in the intensity of all the signals in the ^{4}F1^{5}F1 HSQC spectrum (black spectrum, Figure 74), indicating that ^{4}F1^{5}F1 was precipitating. Given that both the TickFnBR sample before lyophilization and also the ^{4}F1^{5}F1 sample were soluble before they were mixed together, and that addition of TickFnBR to ^{4}F1^{5}F1 caused ^{4}F1^{5}F1 to precipitate in a concentration dependent manner, a binding event is taking place, although it may be non-specific. However, attempts to optimize solubility of the ^{4}F1^{5}F1:TickFnBR complex with a solubility screen over a wide pH and buffer range were not successful (data not shown; Jancarik *et al.*, 2004).

An attempt at measuring binding between pNTD and TickFnBR in PBS at 25°C by ITC was also unsuccessful. Only very small heats were produced and no significant binding event was observed between TickFnBR and pNTD during the titration, although TickFnBR was in excess by the end of the experiment, with a final molar ratio between pNTD and TickFnBR of 1:3 (Figure 75A). Furthermore, unlike the TickFnBR and ^{4}F1^{5}F1 HSQC experiments, TickFnBR and pNTD did not precipitate during the ITC experiment. This was ascertained both from visual inspection and SDS PAGE analysis of the samples (Figure 75B). The weakest dissociation constant that theoretically could be measured in this experiment was approximately 1 μM. Thus, given the apparent similarities of binding of SfbI-5 and TickFnBR to either pNTD or Fn in the pull-down experiments (Figures 70, 72), it was surprising no binding was observed.
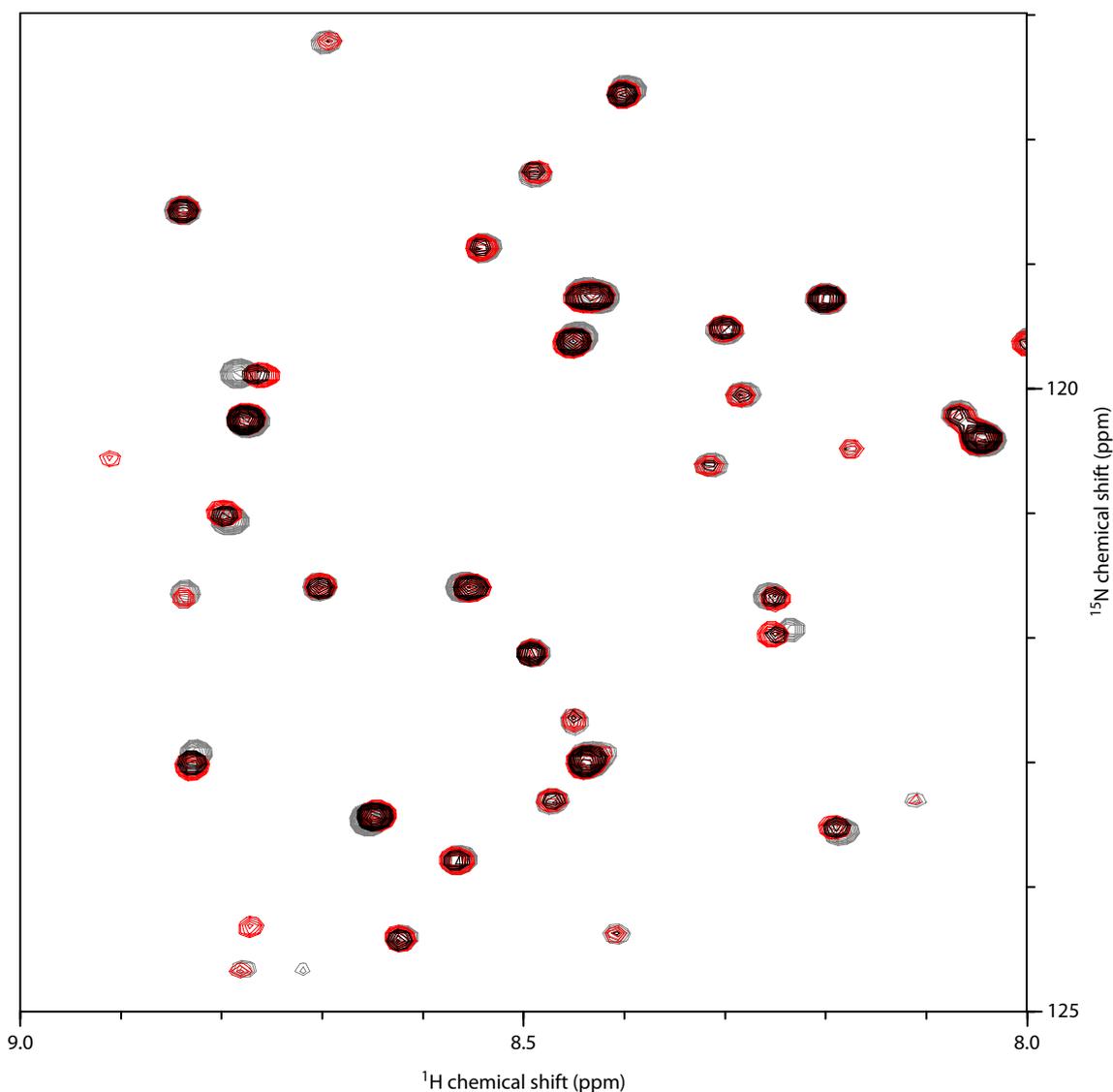
Figure 74. Overlay of a representative section of HSQC spectra of ^{15}N-labelled ^{4}F1^{5}F1 (0.15 mM, pH 6.3) with 0 (grey), 0.2 (red) and 0.4 (black) molar equivalent of TickFnBR. Spectra were recorded on a Bruker 700 MHz spectrometer at 298 K.

There are at least three possible explanations for the discrepancy between the ITC results and the GST pull-down experiments. First, that TickFnBR is not a true FnBR, but can only bind to $^{1-5}$F1 as a GST-fusion. This is unlikely given that neither $^{1-5}$F1 nor Fn interact significantly with GST (Figures 69, 70, 72). Second, TickFnBR is an FnBR that binds to $^{1-5}$F1 much more weakly than other FnBRs. This would be surprising given that both TickFnBR and SfbI-5 bound similar amounts of either $^{1-5}$F1 or Fn in the pull-down experiments (Figures 70, 72). Third, complex formation between $^{1-5}$F1 and TickFnBR is unusually slow, such that equilibrium had not been reached between subsequent injections of TickFnBR into pNTD in the ITC experiments. The third

explanation is most likely because it is supported by SEC MALLS experiments (Section 5.4.3).
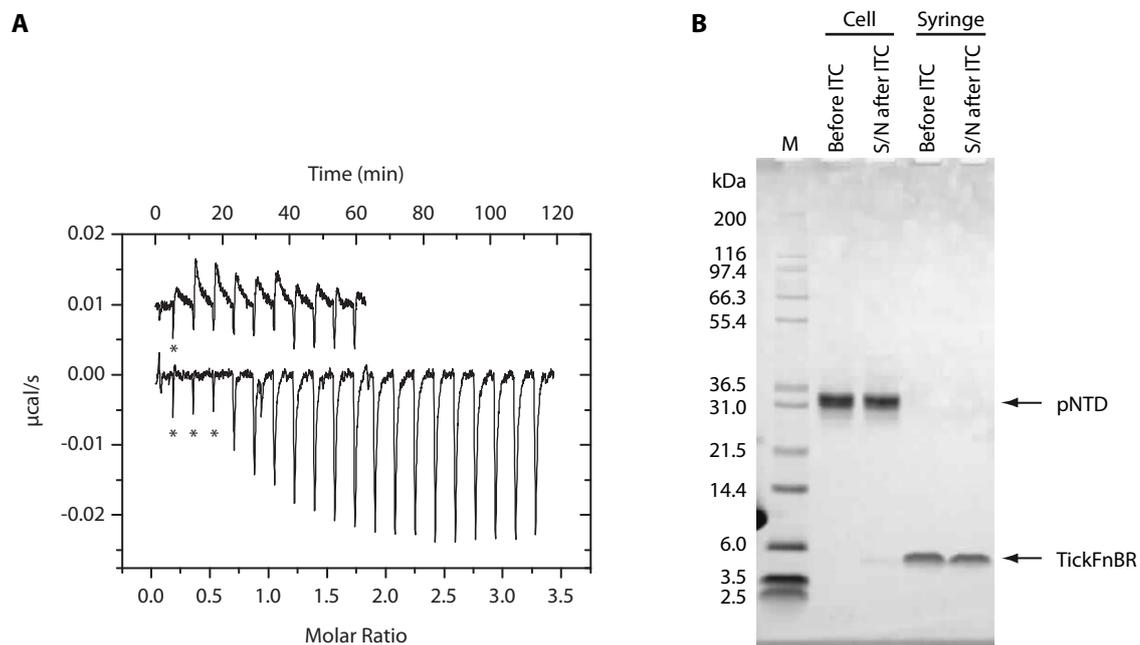
Figure 75. Attempted ITC experiment of TickFnBR binding to pNTD in PBS at 25°C. A. ITC experimental trace as TickFnBR was titrated into pNTD. The trace from the control experiment of TickFnBR titrated into PBS is shown offset relative to the binding titration by 0.01 μcal/s. All injections were 15 μL except for the first injection in both titrations, which was 2 μL, and the injections marked by an asterisk (*), which were 6 μL (since the heats were small the injection volume was adjusted part-way through the experiment). **B.** SDS PAGE analysis of samples from the ITC of TickFnBR with pNTD. Lane contents are indicated, where: 'before ITC′ is the cell and syringe solutions before the ITC experiment, they contained pNTD and TickFnBR, respectively; S/N after ITC, the supernatant (S/N) from the cell solution and remaining syringe solution after centrifugation (5 min, 13000 rpm).

5.4.3 TickFnBR forms soluble aggregates but forms a heterodimeric complex with $^{1-5}$F1

SEC MALLS experiments (Section 2.5.3) were performed to confirm that TickFnBR was able to bind to $^{1-5}$F1 in the absence of GST, and also to determine the stoichiometry of the $^{1-5}$F1:TickFnBR complex. SEC MALLS experiments of TickFnBR and $^{1-5}$F1:TickFnBR samples indicate that by itself TickFnBR primarily forms soluble aggregates in PBS (Figure 76A). However, in the presence of an approximately

equimolar concentration of [1-5]F1, TickFnBR forms a 1:1 complex[19] with [1-5]F1 (Figure 76A; Table 41). That is, the SEC MALLS results indicate that an equilibrium exists between monomeric and aggregated TickFnBR. On addition of [1-5]F1, monomeric TickFnBR binds to [1-5]F1, shifting the equilibrium towards monomeric TickFnBR, until most TickFnBR is bound to [1-5]F1 (Figure 76B). It is plausible that reaching equilibrium between [1-5]F1, aggregated TickFnBR and [1-5]F1:TickFnBR is fairly slow. Therefore, lack of measurable binding by ITC could suggest equilibrium had not been reached between TickFnBR injections in the ITC experiment. Most importantly, however, these experiments show that TickFnBR interacts with [1-5]F1 in a 1:1 complex, confirming the results of the pull-down experiments.
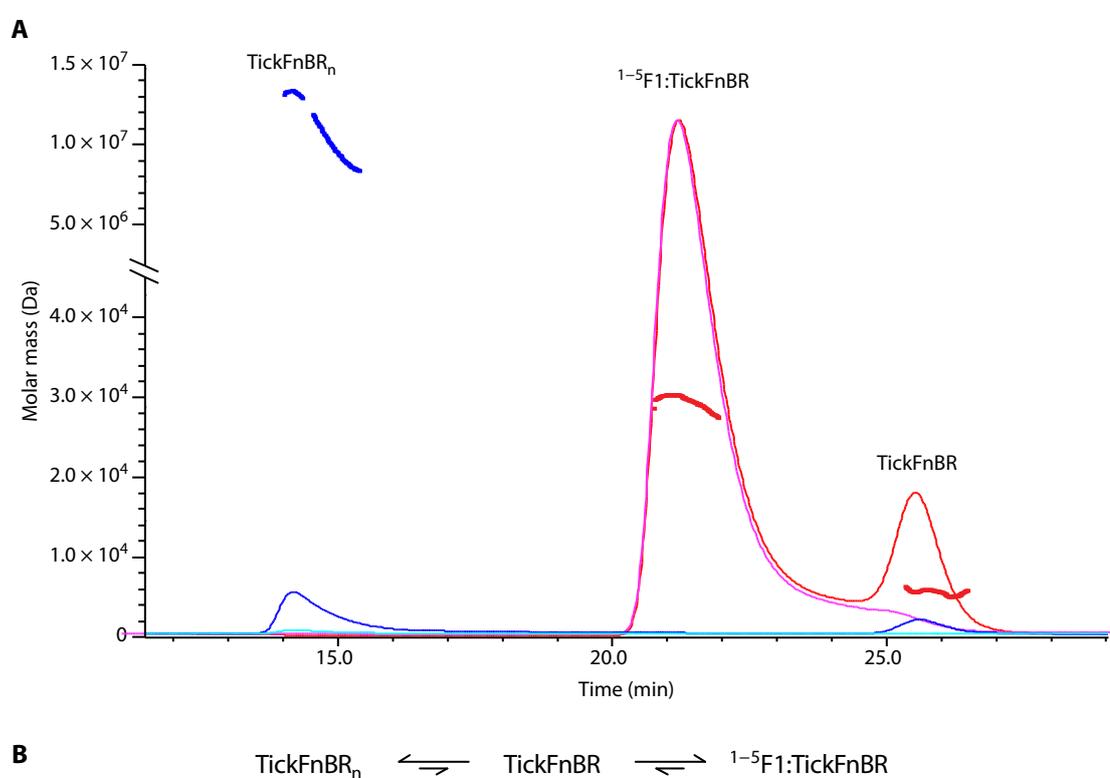
Figure 76. SEC MALLS of TickFnBR both in the presence and absence of [1-5]F1 suggests that TickFnBR reversibly forms soluble aggregates. A. Samples of ~100 μM TickFnBR either with or without ~40 μM [1-5]F1 were separated on a Superdex 75 column while the A_{280}, refractive index and MALLS were monitored. The data from the latter two were used to calculate the molar mass. The thin blue line, thin cyan line and bold blue line show the refractive index, A_{280} and calculated molar mass, respectively, for the TickFnBR sample. The thin red line, the thin magenta line and the bold red line show the refractive index, A_{280} and calculated molar mass, respectively for the [1-5]F1:TickFnBR sample. The

---

[19] It is also plausible that the [1-5]F1:TickFnBR complex has a molar ratio of 1:2 and not 1:1. However, a 1:2 complex is considered unlikely because the crystal structure of [2]F1[3]F1 in complex with a TickFnBR peptide, IxTT1, is a 1:1 heterodimer (Figure 79, Section 5.4.5).

peaks are labelled by protein species, as determined by comparison between theoretical MW and those determined experimentally from the MALLS data (Table 41), where TickFnBR$_n$ indicates large soluble aggregates of TickFnBR. **B.** The proposed equilibrium between TickFnBR, TickFnBR$_n$ and $^{1-5}$F1.

Table 40. Comparison between the theoretical and experimental MWs from SEC MALLS for TickFnBR alone and in complex with $^{1-5}$F1. The error was calculated as |Experimental MW – Theoretical MW|. Note that the experimental MW of the TickFnBR species in the $^{1-5}$F1:TickFnBR sample may be inaccurate, because MW determination from this peak was probably affected by the tail of the peak corresponding to the complex (Figure 76). Although the same TickFnBR species is present in the TickFnBR sample (Figure 76), the MALLS signal for this peak was not strong enough to determine the MW.

Species	MW (kDa)		Error (kDa)
	Theoretical	Experimental	
TickFnBR	3.9	5.8	1.8
$^{1-5}$F1:TickFnBR	28.7	29.6	1.0

5.4.4   TickFnBR peptide binds to ^{2}F1^{3}F1

Given the problems with solubility and aggregation encountered when working with TickFnBR, synthetic peptides were designed and purchased of the predicted ^{2}F1^{3}F1- and ^{4}F1^{5}F1-binding regions in TickFnBR, called IxTT1 and IxFF1, respectively. IxFF1 could not be purified from contaminants by the manufacturer, and has currently not been pursued further. In contrast, IxTT1 is soluble to about 5 mM in water (pH 6.5), and its binding to ^{2}F1^{3}F1 has been demonstrated using NMR spectroscopy (Section 5.4.4.1) and ITC (Section 5.4.4.2).

5.4.4.1   Chemical shift perturbation assay of ^{15}N-labelled ^{2}F1^{3}F1 with IxTT1

In an NMR chemical shift perturbation assay (defined in Section 5.4.2), a series of HSQC experiments were recorded of ^{15}N-labelled ^{2}F1^{3}F1 (0.13 mM, pH 6.5) with increasing concentrations of IxTT1 (Figure 77). The HSQC spectra show binding is occurring and that there is slow exchange on the NMR timescale between unbound ^{2}F1^{3}F1 and ^{2}F1^{3}F1 bound to IxTT1 (Figure 77; Section 1.6.3). That is, at intermediate molar ratios of ^{2}F1^{3}F1 to IxTT1 (1:0.3 to 1:0.6), the peaks for two ^{2}F1^{3}F1 populations can be seen for most residues in ^{2}F1^{3}F1 (*e.g.,* black spectrum in the boxed region, Figure 77). These residues are in different chemical environments in free and bound ^{2}F1^{3}F1 (*e.g.,* D68 and I151, Figure 77). At a higher molar ratio of ^{2}F1^{3}F1 to IxTT1

(1:1.2 and above) only peaks corresponding to ^{2}F1^{3}F1 bound to IxTT1 are seen (*e.g.,* D68 and I151, Figure 77).

Although the HSQC spectra of unbound ^{15}N-labelled ^{2}F1^{3}F1 has been assigned previously by Rudiño-Piñera *et al.* (2007), the bound spectra could not be assigned from these assignments because slow exchange means that it is unclear which bound peaks correspond to which unbound peaks. Sometimes peaks that do not move can be used to define the surface of a protein that is not involved in binding. However, in these experiments, the chemical shifts of all but three residues changed to some extent (not including sidechains). Previously, in a chemical shift perturbation assay of ^{15}N-labelled ^{2}F1^{3}F1 with PyTT5, only 15 chemical shifts of the assigned residues did not change or only changed slightly (Schwarz-Linek *et al.*, 2004b). Thus, this chemical shift perturbation assay showed that IxTT1 binds to ^{2}F1^{3}F1 but slow exchange between unbound and bound ^{2}F1^{3}F1 meant that the binding surface on ^{2}F1^{3}F1 could not be readily identified. It may have been possible to change the exchange on the NMR timescale by changing the temperature or by using a spectrometer of a different frequency. If necessary, the binding surface could be determined by reassigning the bound spectrum. However, this was not necessary because a high resolution crystal structure of IxTT1 bound to ^{2}F1^{3}F1 was solved (Section 5.4.5).

5.4.4.2   IxTT1 binds ^{2}F1^{3}F1 by ITC

An ITC experiment was performed of ^{2}F1^{3}F1 binding to IxTT1 at 25°C in 10 mM sodium/phosphate buffer, pH 7.4. This experiment showed IxTT1 bound to ^{2}F1^{3}F1 with a K_d of 0.76 μM (Figure 78). In the same conditions, ^{2}F1^{3}F1 bound to PyTT4 from SfbI-4 with a K_d of 32 μM and to PyTT5 from SfbI-5 with a K_d of 3.6 μM (Schwarz-Linek *et al.*, 2004b). These are the only previous measurements for FnBR peptides binding to ^{2}F1^{3}F1, thus, IxTT1 is the tightest binder of ^{2}F1^{3}F1 measured to date.
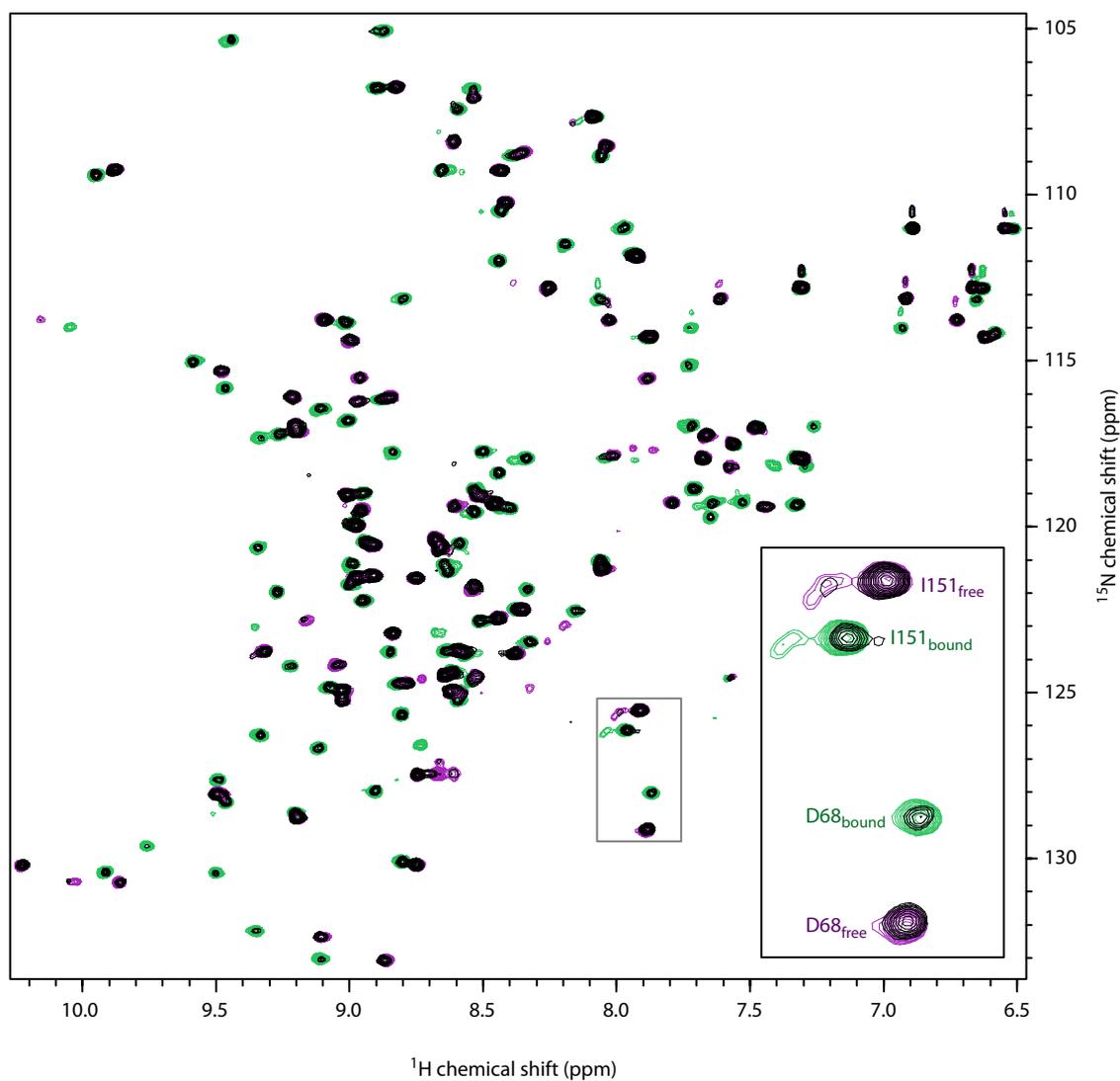
Figure 77. Series of HSQC spectra of ^{15}N-labelled ^{2}F1^{3}F1 (0.13 mM, pH 6.5) with 0 (purple), 0.3 (black) and 2.4 (green) molar equivalents of IxTT1. Spectra were recorded on a Bruker 700 MHz spectrometer at 298 K. These are the first, third and last of six experiments. Inset is an enlargement of the grey boxed region of the spectra. The peaks from free ^{2}F1^{3}F1(purple spectrum) in this region have been assigned two to ^{2}F1^{3}F1 residues based on the published chemical shifts of ^{2}F1^{3}F1 (Rudiño-Piñera *et al.*, 2007). In the same region, two peaks from bound ^{2}F1^{3}F1 (green spectrum) have been tentatively assigned to these two residues by assuming that the two peaks nearest to each other in bound and free ^{2}F1^{3}F1 are from the same residue. A subscript of bound and free indicates the peak corresponding to that residue either free from or bound to IxTT1, respectively.
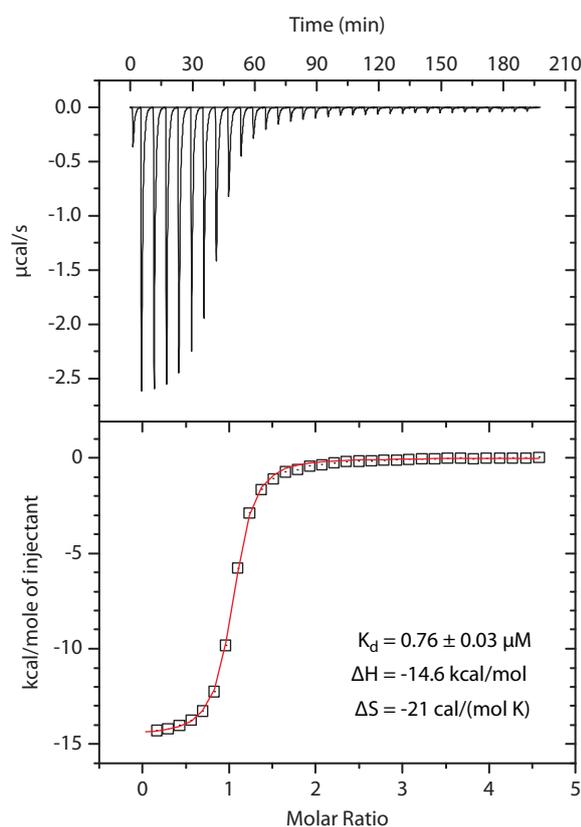
Figure 78. ITC experiment for IxTT1 binding to ^{2}F1^{3}F1. ITC experiment for IxTT1 (~1.2 mM) binding to ^{2}F1^{3}F1 (52.1 µM) performed in 10 mM sodium/potassium phosphate buffer, pH 7.4 at 25°C. Top panel shows the experimental trace as IxTT1 was titrated into ^{2}F1^{3}F1. The last five data points were used to control for heat of dilution effects. The bottom panel shows the processed molar heats of binding normalized to the ratio between IxTT1 to ^{2}F1^{3}F1. The curve from non-linear least squares regression analysis using a single-site binding model is shown in red. The concentration of IxTT1, originally estimated from the weight of lyophilized peptide, was increased by 10% so that stoichiometry of the interaction was 1:1.

5.4.5 Crystallization of ^{2}F1^{3}F1:IxTT1

Crystallization of the ^{2}F1^{3}F1:IxTT1 complex was successful. Crystals grew within two weeks in Hampton Research Crystal Screen condition 6 (200 mM MgCl$_2$, 100 mM Tris, pH 8.5, 30% (w/v) PEG 4000). Data was kindly collected from the crystal shown in Figure 79 by Mr Sam Hart (YSBL, Department of Chemistry, University of York). The data was processed with MOSFLM (Leslie, 2006), scaled in SCALA (Evans, 2006) and solved by molecular replacement with Phaser (McCoy *et al.*, 2007) using ^{2}F1 and ^{3}F1 as search models. This solution showed clear unmodelled density over the E-strands of both ^{2}F1 and ^{3}F1. Subsequent rounds of model-building in Coot (Emsley & Cowtan, 2004) and refinement with Refmac 5 (Murshudov *et al.*, 1997) revealed that IxTT1 was bound in an anti-parallel orientation across the E-strands of both F1-modules, that is, *via*

a tandem β-zipper (Figure 80). This solution was refined to give an R_{work} of 16.6% and an R_{free} of 20.8% (Table 41).

The ^{2}F1^{3}F1:IxTT1 structure was validated by the MolProbity server (Davis *et al.*, 2007), which gave the structure a MolProbity score of 1.52 and a clashscore of 5.72, which placed it in the 92nd and 93rd percentile, respectively, for structures in the resolution range 1.5–2.0 Å, where the best structures are in the 100th percentile (further statistics in Table 41; more information on MolProbity in Section 4.2.2). In conclusion, the crystal structure of ^{2}F1^{3}F1:IxTT1 shows that this peptide from *I. scapularis* binds to ^{2}F1^{3}F1 *via* a tandem β-zipper. This structure strongly supports the hypothesis that TickFnBR is an FnBR that binds $^{2-5}$F1 *via* a tandem β-zipper.
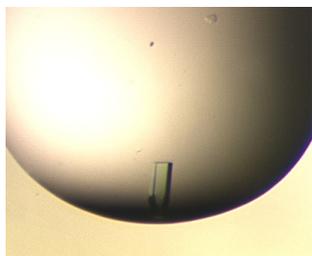
Figure 79. Photo of ^{2}F1^{3}F1:IxTT1 crystal. The starting drop condition was a 1:1 mixture of well solution (200 mM MgCl2, 100 mM Tris, pH 8.5, 30% (w/v) PEG 4000) and ^{2}F1^{3}F1:IxTT1 solution (400 µM ^{2}F1^{3}F1, ~4.2 mM IxTT1, pH 6.0).
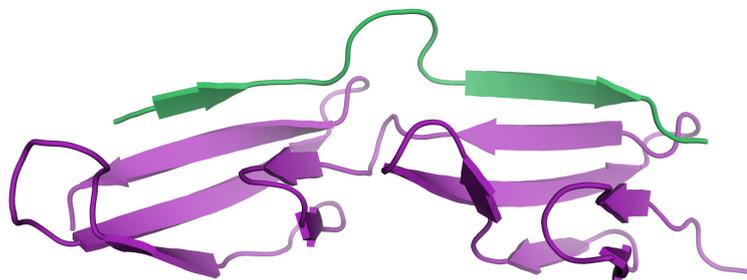
Figure 80. Ribbon representation of the asymmetric unit of the ^{2}F1^{3}F1:IxTT1 structure. ^{2}F1^{3}F1 and IxTT1 are shown in purple and green, respectively.

Table 41. Statistics for ^{2}F1^{3}F1:IxTT1 structure. Highest resolution shell data is shown in brackets where applicable.

Data collection			
Wavelength (Å)	1.542	Number of images	400
Exposure time (s)	120	Crystal to detector distance (mm)	110.0
Oscillation angle (°)	400	Mean I/σ(I)	11.6 (3.3)
Space group	P2$_1$	Number of unique reflections	8693 (1261)
Cell dimensions	a = 20.70 Å, b = 68.05 Å,	R$_{sym}$	0.063 (0.316)
	c = 31.32 Å,	Completeness (%)	100.0 (100.0)
	α = γ = 90°, β=98.74°	Redundancy	3.9 (3.8)
Resolution (Å)	30.95–1.75 (1.84–1.75)	Wilson B factor	21.2
Refinement		*R.M.S. deviations*	
Resolution (Å)	28.2–1.75 (1.795–1.750)	Bond lengths (Å)	0.011
R$_{work}$	0.166 (0.366)	Bond angles (°)	1.235
R$_{free}$	0.208 (0.411)		
Number of atoms		*B factors*	
^{2}F1^{3}F1	718	^{2}F1^{3}F1	15.53
IxTT1	123	IxTT1	16.83
Water	125	Water	25.89

5.5 Discussion

5.5.1 The four putative FnBPs are plausible based on three criteria

There are three criteria that need to be met for a sequence to be a plausible FnBP: i) the putative FnBP can specifically bind Fn; ii) the putative FnBP has access to Fn, either the originating organism's or a host's; and iii) an FnBP's usefulness can be justified based on the biology of the originating organisms (Section 5.1). The four putative FnBPs from the deer tick *I. scapularis* (Figure 64) are highly plausible since they are likely to meet all these criteria, as addressed below.

5.5.1.1 Criterion 1: The putative FnBPs can specifically bind Fn.

It has been demonstrated that the predicted FnBR in one of the four putative tick FnBPs, called 'TickFnBR', interacts with $^{1-5}$F1 from Fn using GST pull-down experiments (Section 5.3.1). A competition GST pull-down experiment showed that TickFnBR

cannot bind $^{1-5}$F1 in the presence of SfbI-5, indicating TickFnBR interacts with $^{1-5}$F1 at the same or a similar site as SfbI-5 (Section 5.3.2). Furthermore, TickFnBR was able to pull-down a similar amount of Fn from plasma as SfbI-5, indicating TickFnBR interacts with full-length Fn both specifically and with an affinity similar to another FnBR (Section 5.3.3).

A MALLS experiment confirmed the interaction between TickFnBR and $^{1-5}$F1, showing the stoichiometry was 1:1 and that, in the absence of $^{1-5}$F1, TickFnBR forms soluble aggregates (Section 5.4.3). A slow equilibrium between monomer and aggregates of TickFnBR is probably the reason an attempt to measure the binding between pNTD and TickFnBR by ITC was unsuccessful (Section 5.4.2). It is currently unclear if the reversible formation of soluble aggregates by TickFnBR is functionally relevant, or an artefact of studying the peptide *in vitro*. TickFnBR does not contain the C-terminus of the full-length FnBP from which it is derived (Figure 64). The C-terminus is rich in glycine and proline, and thus it is possible the full-length FnBP may have different solubility properties (Figure 64). Thus, TickFnBR has been shown to bind $^{1-5}$F1 by both GST pull-down experiments and MALLS.

IxTT1, the predicted ^{2}F1^{3}F1-binding peptide from TickFnBR, was shown to bind ^{15}N-labelled ^{2}F1^{3}F1 in a chemical shift perturbation assay (Section 5.4.4.1). However, slow exchange on the NMR timescale meant the binding surface could not be identified from the previously assigned ^{2}F1^{3}F1 resonances. An ITC experiment confirmed the interaction, and the measured K_d of 0.76 μM makes this peptide the tightest ^{2}F1^{3}F1-binding peptide identified thus far (Section 5.4.4.2). Finally, and most importantly, the structure of ^{2}F1^{3}F1:IxTT1 shows that IxTT1 binds to ^{2}F1^{3}F1 *via* a tandem β-zipper (Section 5.4.5). This interaction will be discussed further in Section 5.5.3. IxTT1 is the first eukaryotic peptide shown to bind F1 modules in this way.

In conclusion, there are multiple levels of evidence to support the hypothesis that TickFnBR is an FnBR that binds to $^{1-5}$F1 in Fn *via* a tandem β-zipper. Given the sequence homology between TickFnBR and the other three putative FnBRs (Figure 64), it is likely that the other putative FnBPs also specifically bind Fn. Therefore, the four tick putative FnBPs meet the first criterion to be plausible FnBPs.

5.5.1.2 Criterion 2: The putative FnBPs have access to host Fn

The second criterion can be addressed by considering the sequences of the four putative FnBPs and also how these sequences were identified. All four FnBPs are strongly predicted to contain a eukaryotic signal sequence by the Signal P server (Figure 64), which indicates these proteins are secreted. All putative FnBPs were identified in an analysis of Expressed Sequence Tags (ESTs) from the salivary glands of adult female ticks (Ribeiro *et al.*, 2006). ESTs arise from sequencing the 5′ and/or 3′ end of mRNA that has been processed for translation, and thus, represent genes that are probably being translated into protein in the tissue sample at the time it was collected. In the initial study the ESTs were sequenced from six salivary gland libraries: i) nymphs fed for 2 days; ii) nymphs fed for 2 days, previously fed as larvae on mice infected with *B. burgdorferi*; iii) unfed adult females (iii); and (iv–vi) adult females after attachment to a rabbit for, respectively, 6–12 hours, 18–24 hours or 3–4 days (Ribeiro *et al.*, 2006). Sixteen ESTs were clustered to give rise to the four putative FnBP sequences. One was from library (iii), the rest were from library (iv) (Table 38; Ribeiro *et al.*, 2006). Therefore, given the expression pattern and signal sequences of the four putative FnBPs, it is likely that these proteins, regardless of their function, are being expressed in the tick's salivary gland and secreted into saliva during the initial stages of feeding where they will come into contact with the host's Fn in both the ECM and the blood.

5.5.1.3 Criterion 3: The putative FnBPs could be functionally useful to *I. scapularis*

There are at least two plausible functions for secreted FnBPs: as an attachment aid for the tick during feeding; and to interfere with role of Fn in wound healing and clot formation. These two functions will now be discussed in turn. Ixodid ticks feed on blood by attaching to a host for 4–15 days (Kaufman, 1989). Within 5–30 minutes after initial attachment *via* its mouth parts, an ixodid tick secretes a substance into the wound that subsequently hardens to a 'latex-like' material called attachment 'cement', which helps the tick stay attached to the host during feeding[20] (Kaufman, 1989; Sonenshine, 1991). The cement is comprised of proteins, lipids, lipoproteins and glycoproteins, but it

---

[20] Nearly all references found regarding attachment cement mention that all ixodid ticks excrete cement except for a few *Ixodes* species, but no reference mentioned which species these are. However, since Ribeiro *et al.* (2006) predict some proteins from *I. scapularis* are involved in cement formation we assume it is likely *I. scapularis* is one of the *Ixodes* species that does secrete cement.

is currently not known what causes hardening (Kemp *et al.*, 1982; reviewed by Sauer *et al.*, 1995). The putative FnBPs may be a part of this cement, aiding attachment to the host *via* Fn in the ECM. In fact, Ribeiro *et al.* (2006) proposed a role in attachment cement for the putative FnBPs, but for a different reason. The four putative FnBPs are grouped by these researchers with other peptides that are similarly rich in proline and glycine, 'Group 4'. Ribeiro *et al.* (2006) note that Group 4 peptides contain at least one (L/P/A/E)PG motif (Figure 64), the proline of which can be post-translationally hydroxylated by prolyl hydroxylase. They also note that this modification is plausible given that they have identified an EST that is a putative alpha-subunit of this enzyme. This leads them to the hypothesis that Group 4 proteins are collagen-like and thus, may function as part of the attachment cement. Interestingly, a known cement protein from the related ixodid tick *Rhipicephalus appendiculatus*, RIM36, is also proline and glycine rich (Bishop *et al.*, 2002), although this protein does not contain any prolyl hydroxylase motifs. Therefore, it is plausible that the putative FnBPs aid attachment of the tick by being part of the attachment cement.

During the prolonged feeding period, the tick manipulates the host's haemostatic, inflammatory and cell-mediated immune responses *via* components in its saliva in order to maintain a blood pool and remain undetected by the host (Section 1.8.3.1; reviewed in Valenzuela, 2004). It is plausible that the putative FnBPs are involved in interfering with wound healing. Fn is involved in many steps of wound healing, including the initial step of platelet spreading and probably also platelet aggregation (reviewed in Cho & Mosher, 2006a; Grinnell, 1984). Additionally, it has been shown that Fn enhances thrombus (blood clot) formation, and this enhancement is blocked by the 'functional upstream domain' from *S. pyogenes*, which is the UR and first FnBR from SfbI (Figure 3; Cho & Mosher, 2006a; 2006b). Thus, it is plausible that the *I. scapularis* could use an FnBP to interfere with wound healing.

In conclusion, the putative *I. scapularis* FnBPs appear to meet all three criteria for being plausible FnBPs. In the work reported here, it has been shown experimentally that an FnBR from one of the FnBPs binds $^{1-5}$F1 in Fn; it is highly likely the other putative FnBRs do so too. Since these FnBPs are expressed in the saliva of ticks during feeding, it is likely that the FnBPs can access host Fn while *I. scapularis* feeds on host blood. Finally, it is likely putative FnBPs could either aid attachment of the tick to its host during feeding, or interfere with wound healing. Furthermore, as discussed in Section

4.4.3, it is very unlikely that these putative FnBPs are a contaminant from *B. burgdorferi* for which *I. scapularis* is the disease vector, rather than from *I. scapularis*. Therefore, it appears likely that the four putative FnBPs from *I. scapularis* are the first eukaryotic examples of FnBPs.

5.5.2 Analysis of the putative FnBPs expression patterns based on their ESTs

Along with sequence information, EST analysis also indicates where and when a sequence is expressed. As well as the initial study that identified the mRNA of the four putative FnBP sequences in the salivary glands of female ticks, two additional studies of *I. scapularis* have identified ESTs that cluster with the FnBP sequences (Table 38). One of these studies was of a pooled library of ESTs from fed and unfed ticks at various life stages (Hill *et al.*, 2007). Therefore, this library does not give any information about where or when the sequences are expressed, so will not be discussed further. The other study was of ESTs isolated from the synganglia of feeding adult female ticks (Huang *et al.*, 2007). The synganglion of a tick is its entire central nervous system; it is the fusion of the cerebral ganglia and the abdominal nerve cord in to a condensed 'nerve mass' (Sonenshine, 1991). The presence of ESTs corresponding to two of the four putative FnBPs in tick synganglia is not readily explained, but indicates these peptides may be multi-functional and have another role as secreted neurotransmitters. Unfortunately, while the EST sequences identified from synganglia are available in the UniGene database, this study has not been published in a journal, so it is difficult to analyse these sequences in the context of other synganglia ESTs.

All but one of the 20 FnBP ESTs from studies at discrete time points are from an adult feeding tick (Table 38), indicating these sequences are important during feeding, including the synganglia ESTs. It does not appear synganglia ESTs have been sequenced from unfed adult females, so it is not known whether these ESTs are differentially expressed during feeding. In contrast, it appears the salivary gland ESTs are important in the initial stages of feeding of adult female ticks. As previously mentioned, Ribeiro *et al.* (2006) examined six salivary gland EST libraries, including glands from nymphs, unfed adults and fed adults harvested at different time points (Section 5.5.1.2). That 15 of the putative FnBP ESTs were found in fed adults 6–12 hours after host attachment and none were found in fed adults 18–24 hours or 3–4 days after attachment (Table 38) indicates that the putative FnBPs are useful for initial

attachment and feeding, and not later on. This expression pattern might indicate these peptides are used in attachment rather than wound healing since initial cement is secreted in the first 5–30 minutes after feeding, where as anti-haemostatic compounds would be needed throughout feeding. However, to combat the immune response, the tick varies the proteins it expresses, so it is still possible these peptides interfere with wound healing.

The number of ESTs in a cluster from a single study can also give an indication of the level of expression, because the number of transcripts is likely to be proportional to amount of protein expressed. Interestingly, the UniGene database highlights Q4PN03 as unusually abundant, since with 12 other ESTs it is the equal fourth most identified EST from the initial study (Ribeiro *et al.*, 2006). This result and the presence of this peptide in the other two studies indicate that it is functionally important to feeding *I. scapularis*.

In conclusion, analysis of the putative FnBP ESTs indicate a role for these peptides in feeding adult ticks, both in the salivary glands in the initial stages of feeding and also, surprisingly, in the tick's central nervous system. Furthermore, the number of ESTs identified for, in particular, putative FnBP Q4PN03, indicate these peptides are functionally important to the tick.

5.5.3 Analysis of the ^{2}F1^{3}F1:IxTT1 structure

The interaction between ^{2}F1^{3}F1 and IxTT1 shares many similarities with the streptococcal and staphylococcal ^{2}F1^{3}F1:FnBR peptide structures. Like the other interactions between ^{2}F1^{3}F1 and FnBR peptides, IxTT1 forms extensive polar and van der Waals contacts with ^{2}F1^{3}F1 (Figure 81). These interactions will now be discussed with respect to the other ^{2}F1^{3}F1:FnBR peptide structures, with particular reference to ^{2}F1^{3}F1:PyTT5.

5.5.3.1 Important mainchain-mainchain interactions between ^{2}F1^{3}F1 and IxTT1

IxTT1 binds across the E-strand of both F1-modules, extending each β-sheet by an additional anti-parallel β-strand (Figure 80). IxTT1 interacts *via* seven and five mainchain-mainchain hydrogen bonds with ^{2}F1 and ^{3}F1, respectively; with residues 37–39 and 46–50 adopting φ and ψ angles typical of β-strand (Figures 80, 81). Therefore, IxTT1 interacts with ^{2}F1^{3}F1 *via* a tandem β-zipper.

5.5.3.2 Important interactions with $^2F1^3F1$ involving PyTT5 sidechains

IxTT1 has $^3F1$-binding, linker and $^2F1$-binding regions of residue ranges 36–40, 41–45 and 46–52, respectively. As for the analysis of $^2F1^3F1$:PyTT5, the important interactions involving the sidechains of IxTT1 in each of these regions will now be discussed in turn.

Like PyTT5, the sidechains of the IxTT1 $^3F1$-binding region interact with $^3F1$ primarily *via* van der Waals contacts: I37 contacts R125 and W146 in $^3F1$, and H40 stacks against the aliphatic sidechain carbons of K143 (Figures 81, 82A). Furthermore, as for all bacterial $^2F1^3F1$:FnBP structures, a glycine is positioned over W146 to avoid sterically unfavourable interactions (Chapter 4; Bingham *et al.*, 2008).

The IxTT1 linker region loops away from $^2F1^3F1$ similarly to the $^2F1^3F1$:STATT1 and STATT5 structures but unlike the $^2F1^3F1$:PyTT5 structure (compare Figures 5B, 36F and 80). The linker regions in IxTT1, STATT1, STATT5 and PyTT5 are five, seven, four and four residues, respectively. Therefore, IxTT1 may loop out partly because it has a longer linker. The structure of this region appears to be primarily caused by intra- and intermolecular hydrogen bonds. The 'looping out' of IxTT1 away from $^3F1$ appears to be as a result of the orientation of N41, which in turn is caused by its many favourable interactions with both $^2F1$ and $^3F1$. This residue forms five hydrogen bonds: one each with R109, H111 and K143 from $^3F1$, one with the I106 in $^2F1$, and one with a bridging water that in turn forms hydrogen bonds with H111 and G144 in $^3F1$. Furthermore, the sidechain of N41 stacks against the sidechain of I106 in $^3F1$. Q45 appears to stabilize the loop and also appears to cause the linker region to 'loop in' again and bind $^2F1$. The sidechain of Q45 stabilizes the 'hairpin' structure of the linker region by the formation of three hydrogen bonds *via* its amide group with the backbone carbonyl groups of N41, Q42 and G43, and also a hydrogen bond with the sidechain of N41 *via* its carbonyl group. Furthermore, the backbone carbonyl of Q45 forms hydrogen bonds with both I106 and T105 in $^2F1$. That the most N-terminal and C-terminal residues in the linker region, N41 and Q45 respectively, form hydrogen bonds with I106 in $^2F1$ appears to contribute significantly to the structure of the linker.

Figure 81. Schematic of the interaction of IxTT1 with ^{2}F1^{3}F1. Residues in ^{2}F1^{3}F1 (purple) which form van der Waals contacts and/or hydrogen bonds with IxTT1 (green) are shown, as are bridging water molecules (blue circles). Hydrogen bonds between mainchain atoms on both PyTT5 and ^{2}F1^{3}F1 are black, all others are grey. Diagram created with LIGPLOT (Wallace *et al.*, 1995)
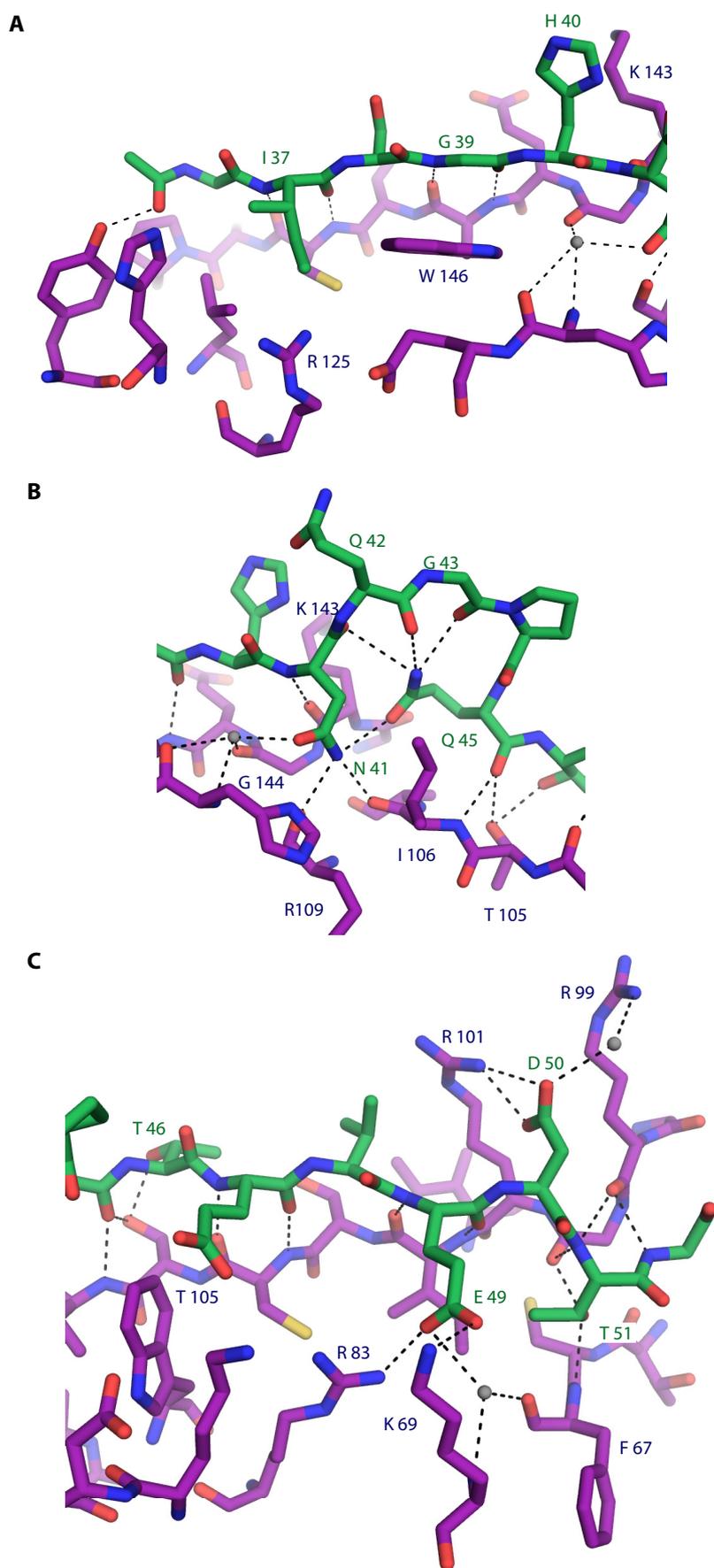
Figure 82. The interactions between ²F1³F1 and IxTT1. ²F1³F1 and PyTT5 interface residues are shown in stick representation in purple and green, respectively. Hydrogen bonds are black dashes and

important bridging water molecules are shown as grey spheres. Important residues mentioned in the text are labelled. **A.** The interaction between ^{3}F1 and the ^{3}F1-binding region of IxTT1. **B.** The interaction between ^{2}F1^{3}F1 and the linker region of IxTT1. **C.** The interaction between ^{2}F1 and the ^{2}F1-binding region of IxTT1.

The IxTT1 ^{2}F1-binding region binds to ^{2}F1 *via* both van der Waals and hydrogen bond interactions. The two key van der Waals contacts are the I571 and D574 sidechains interacting with the sidechains of W90 and R99, respectively. T46 forms a hydrogen bond with the sidechain of T105 in ^{2}F1, forming a favourable staggered confirmation, which has been shown to be a determinant of strand register (Hutchinson *et al.*, 1998). Equivalent residues are forming similar contacts in the ^{2}F1^{3}F1:PyTT5 structure. The sidechain of E49 exists in a different conformation than the glutamate in an equivalent position in the other three ^{2}F1^{3}F1:FnBP structures. Rather than forming end-on-end hydrogen bonds with R83, it forms one hydrogen bond with R83 and one with K69. Possibly stabilizing this conformation is a water molecule that, through hydrogen bonds, connects E49 with K69 and F67. Similarly to PyTT5 D574, D50 forms hydrogen bonds with both R101 and R99, although unlike PyTT5 the latter hydrogen bond is *via* a bridging water molecule. T51 forms hydrogen bonds to the backbone of G100 and F67, as do the equivalent residues in all ^{2}F1^{3}F1:FnBR structures.

In conclusion, the key interactions between IxTT1 and ^{2}F1^{3}F1 are highly similar to the interactions between bacterial FnBR peptides and ^{2}F1^{3}F1. This is perhaps not surprising given that the sequence was identified based on its similarity to the other known FnBR sequences. However, it is remarkable that a sequence from a eukaryote appears to have independently evolved the same mechanism for binding to ^{2}F1^{3}F1.

6 Conclusions and future directions

The proposed tandem β-zipper interaction between SfbI-5 and $^{1-5}$F1 was investigated in this thesis. That is, it was proposed that SfbI-5 is intrinsically disordered, undergoes a disorder-to-order transition and binds to $^{1-5}$F1 by extending each F1-module by an additional, anti-parallel β-strand.

The first part of the tandem β-zipper model was tested in Chapter 3, that is, whether SfbI-5 is intrinsically disordered. *In silico* sequence predictors, chemical shift analysis of assigned NMR spectra and analysis by SEC all strongly suggest that SfbI-5 is an IDP with no stable secondary or tertiary structure. This result is in agreement with existing studies on homologous proteins (House-Pompeo *et al.*, 1996; Penkett *et al.*, 1998). Thus, objective 1 of the Aims was met (Section 1.9). Additionally, chemical shift analysis of SfbI-5 nuclei indicated that the C-terminal third of SfbI-5, which includes the predicted ^{1}F1-binding region, contained some propensity for adopting β-strand conformations (objective 2, Section 1.9). Therefore, this region may be a preformed contact site (PCS), a region of SfbI-5 that already has some propensity for its bound structure, and so might increase the 'on' rate by acting as the first contact site for binding to pNTD (Fuxreiter *et al.*, 2004). The presence of a PCS in the ^{1}F1-binding region is interesting because SfbI-5 is the only FnBR in SfbI predicted to contain a ^{1}F1-binding region. Whether the ^{1}F1-binding region is a PCS could be tested by limited proteolysis of SfbI-5. This technique has been used in studies of other IDPs to show that a PCS exists; because these sites are relatively exposed they are cleaved preferentially to the rest of an IDP (Csizmok *et al.*, 2005). Furthermore, kinetic studies with techniques such as stopped-flow or surface plasmon resonance could be used to test whether the ^{1}F1-binding region has a significantly faster association rate than the other F1-binding regions, which would be predicted if this region is a PCS.

A preliminary HSQC spectrum of ^{15}N-labelled SfbI-5 in complex with pNTD also indicated that the complex could be detected with NMR spectroscopy, and that the ^{1}F1-binding region was the most perturbed by binding to pNTD. The latter is particularly interesting given that this region might be a PCS. Future work will be to optimize the NMR conditions and assign the spectrum of SfbI-5 when it is bound to pNTD to obtain insight into which residues interact with pNTD (objective 3, Section 1.9).

The crystal structure of an SfbI-5 peptide, PyTT5, in complex with ^{2}F1^{3}F1 at a resolution of 1.7 Å is reported in Chapter 4, which shows that the peptide binds to ^{2}F1^{3}F1 *via* a tandem β-zipper. This structure provides the first structural evidence that a streptococcal FnBR binds to ^{3}F1 *via* a β-zipper (objective 4, Section 1.9). Comparison of this structure to those of FnBPA peptides in complex with ^{2}F1^{3}F1 revealed that residues that were conserved between these peptides formed similar interactions with ^{2}F1^{3}F1 (Bingham *et al.*, 2008). This suggests that sequence similarity to structurally characterized FnBRs is a good indication that a sequence will bind FnBRs *via* a tandem β-zipper. Given the sequence similarity between the other F1-binding regions of SfbI with FnBR peptides which have been solved in complex with ^{4}F1^{5}F1 and ^{1}F1^{2}F1 (Bingham *et al.*, 2008; Schwarz-Linek *et al.*, 2003), it is highly likely that SfbI-5 will bind to ^{1}F1 and ^{4}F1^{5}F1 *via* a tandem β-zipper. Support for this hypothesis could be obtained from the aforementioned chemical shift analysis of SfbI-5 in complex with pNTD, or from X-ray crystallography studies with SfbI-5 peptides and either ^{1}F1^{2}F1 or ^{4}F1^{5}F1. Therefore, there is currently strong structural evidence in support of SfbI-5 binding to $^{1-5}$F1 *via* a tandem β-zipper.

Further analysis of the interface between PyTT5 and ^{2}F1^{3}F1 revealed that it is both extended and large, with 1950 Å becoming buried on complex formation. Furthermore, PyTT5 binds ^{2}F1^{3}F1 very efficiently: it contributes more to the buried surface area with fewer residues than ^{2}F1^{3}F1, with at least one atom from every residue contributing to the interface. These features of a large interface created by a relatively small peptide are predicted to apply to the interaction between $^{1-5}$F1 and SfbI-5, making SfbI-5 extremely efficient for its MW. Furthermore, the efficiency of PyTT5 and, by extension SfbI-5, is a direct result of the disordered state of the free peptide. The lack of a globular fold means that the FnBR does not need to include residues to maintain a fold, and so, can have a much higher proportion of residues involved in the interface than a globular protein of the same MW (Gunasekaran *et al.*, 2003). Therefore, it appears that SfbI-5 is efficient at binding to $^{1-5}$F1 because it is intrinsically disordered.

The role of conserved FnBR residues was investigated in Chapter 4. The main function of SfbI-5 appears to be its interaction with $^{1-5}$F1. Therefore, it was hypothesized that FnBR residues are under selective pressure because they contribute significantly to the binding affinity of SfbI-5 for $^{1-5}$F1. To test this hypothesis, the effect of individually mutating the conserved residues in SfbI-5 on the ΔΔG° for SfbI-5 binding to pNTD was

determined by ITC (objectives 5.a, b, Section 1.9). No residue tested had a $\Delta\Delta G°$ greater than 2 kcal/mol, suggesting SfbI-5 is rather tolerant to the mutation of conserved FnBR residues. This is possibly because most SfbI-5 atoms predicted to be buried at the interface between SfbI-5 and pNTD are mainchain or sidechain atoms that were not affected by the mutation. Furthermore, the linear, extended nature of the tandem β-zipper interaction that is proposed to occur between SfbI-5 and pNTD predicts that no residue sidechain is completely occluded from solvent on binding. Thus, water might be able to partially compensate for the loss of any favourable interactions made by the atoms that are deleted in the mutant, making SfbI-5 tolerant to mutation. Indeed, analysis of the relative effect of mutating SfbI-5 residues in the context of F1 module:FnBR peptide structures revealed that those residues that on mutation are predicted to lose either a buried interface atom and/or the ability to form hydrogen bonds usually have a more unfavourable effect on binding to pNTD than residues that are predicted to lose only accessible interface atoms (objective 5.c, Section 1.9). In summary, binding of SfbI-5 to $^{1-5}$F1 appears to be relatively tolerant to mutation because of the large, extended intermolecular interface. It is postulated that interactions of other IDPs might also be tolerant to mutations because of an extended interface with their globular protein partner. This hypothesis awaits quantitative mutational studies of other IDPs.

It was hypothesized that the conserved FnBR residues must be under some selective pressure not directly tested in the ITC experiments. A series of GST pull-down experiments was used to investigate whether mutation in SfbI-5 had a larger effect on binding full-length Fn (compared to binding pNTD) in the more biologically relevant context of human plasma. These experiments showed that, within high experimental error, SfbI-5 mutants did not bind significantly less Fn in plasma than would be predicted from the ITC results of SfbI-5 mutants binding pNTD (objective 5.b, Section 1.9). It is proposed that SfbI-5 might act as a mechanical tether between the bacteria and Fn while under applied forces such as blood flow. Then, the dissociation rate, rather than the equilibrium constant would be under selective pressure. This hypothesis could be tested by measuring the dissociation rates of SfbI-5 and its mutants from pNTD or Fn with either AFM or surface plasmon resonance.

The HMM searches identified four putative FnBR-containing FnBPs from the deer tick *Ixodes scapularis,* also in Chapter 4. Analysis of the DNA sequences available for these

proteins revealed that they were highly likely to be eukaryotic and not prokaryotic in origin. One of the putative FnBRs, TickFnBR, was investigated in Chapter 5 to determine if it was a functional FnBR. That is, whether TickFnBR is disordered and binds $^{1-5}$F1 *via* a tandem β-zipper. A one-dimensional ^{1}H spectrum indicated that TickFnBR is probably disordered. GST pull-downs showed that TickFnBR bound to $^{1-5}$F1 at the same or a similar site as SfbI-5, and that it was able to specifically interact with human Fn in the context of plasma (objectives 6, 7, Section 1.9). A SEC MALLS experiment confirmed the formation of a complex between TickFnBR and $^{1-5}$F1 and demonstrated the stoichiometry of the complex was 1:1 and that, in the absence of $^{1-5}$F1, TickFnBR forms soluble aggregates. It was shown that a TickFnBR peptide, IxTT1, bound ^{2}F1^{3}F1 in both a chemical shift perturbation assay and by ITC. Finally, and most importantly, the structure of IxTT1 with ^{2}F1^{3}F1 was solved to a resolution of 1.75 Å, revealing that IxTT1 binds ^{2}F1^{3}F1 *via* a tandem β-zipper (objective 8, Section 1.9). This is the first eukaryotic sequence that has been shown to bind to F1-modules in this way. Therefore, there is strong evidence to support the hypothesis that TickFnBR is a functional FnBR that binds to $^{1-5}$F1 *via* a tandem β-zipper. The potential role of TickFnBR in the interaction of tick saliva with human tissue remains to be investigated.

The structures of ^{2}F1^{3}F1 in complex with PyTT5 and IxTT1 have been compared both to each other and also to the structures of ^{2}F1^{3}F1 bound to FnBPA peptides (Chapters 4 and 5, respectively). This analysis revealed that all the peptides contain a conserved glycine and EDT-motif that, in the absence of crystal contacts, form highly similar interactions with ^{2}F1^{3}F1. The importance of these conserved motifs has been noted by Bingham *et al.* (2008) in their analysis of the F1-module:FnBPA peptide structures. Therefore, the streptococcal, staphylococcal and *I. scapularis* peptides all interact with ^{2}F1^{3}F1 *via* a tandem β-zipper interaction that includes highly similar side-chain interactions. This is striking because it indicates that, not only have two related staphylococcal and streptococcal species evolved the same mechanism for binding ^{2}F1^{3}F1, but that also, the eukaryote *I. scapularis* appears to have independently evolved the same mechanism. This suggests that a tandem β-zipper interaction with the E-strands of $^{1-5}$F1 in Fn is a useful way for harmful organisms to 'hijack' Fn and exploit Fn's native functions for the organisms' own uses.

The properties of FnBRs that facilitate the tandem β-zipper interaction between FnBRs and F1 modules have been the focus of this thesis. Given how efficient this type of

interaction is, it is likely that other examples of tandem β-zipper interactions exist. Future work could include an analysis of the properties of the E-strands of F1 modules that allow the tandem β-zipper interaction to occur, so that other modular β-sheet-containing proteins that might form such interactions with IDPs could be predicted. In this way, it might be possible to determine whether the tandem β-zipper interaction is more widespread in the interactions of IDPs with modular proteins.

Abbreviations

A_{280}	Absorbance at 280 nm
$A_{280-320}$	Absorbance at 280 nm minus the absorbance at 320 nm
AEBSF	4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride
AFM	Atomic force microscopy
Amp	Ampicillin, sodium salt
ASA	Accessible surface area
AU	Asymmetric unit
B3T	Synthetic peptide that binds to $^1F1^2F1$, corresponds to residues 1042–1066 from *S. dysgalactiae* protein FnbB
Bis-tris propane	1,3-bis (tris (hydroxymethyl) methylamino) propane
bp	basepairs
BSA	Buried surface area
CD	Circular dichroism
cDNA	Complementary DNA
CSI	Chemical shift index
CV	Column volumes
$\Delta\delta$	Secondary chemical shift
dNTPs	Deoxynucleotide triphosphates
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
DTT	Dithiothreitol
E	Eluate fraction (individual or pooled) gel sample
ECM	Extracellular matrix
EDA	Extradomain A
EDB	Extradomain B
EDTA	Ethylenediaminetetraacetic acid
ESMS	Electrospray mass spectrometry
ESRF	European synchrotron radiation facility
EST	Expressed sequence tag
E-value	Expectation value
F	Flow-through gel sample
F1	Fn type I module
F2	Fn type II module
F3	Fn type III module
Fn	Fibronectin

FnBP	Fn-binding protein
FnBR	Fn-binding repeat
GAS	Group A streptococcus
GBF	Gelatin binding fragment
GSH	Glutathione
GST	Glutathione *S*-transferase
HMM	Hidden Markov model
HPW	High purity water
HSQC	heteronuclear single quantum correlation
+I	Post-induction gel sample
–I	Pre-induction gel sample
IDP	Intrinsically disordered protein
IPTG	Isopropyl β-*D*-thiogalactopyranoside
ITC	Isothermal titration calorimetry
IxTT1	Synthetic peptide that binds to $^2F1^3F1$, corresponds to residues 36–53 of a putative *I. scapularis* FnBP with UniProt ID Q4PN68
K_d	Dissociation constant
LDS	Lithium dodecyl sulfate
LTA	Lipoteichoic acid
M	Mark 12 protein molecular weight marker for SDS PAGE
MALLS	Multi-angle laser light scattering
MES	2-(*N*-morpholino) ethanesulfonic acid
MOPS	3-(*N*-morpholino) propanesulfonic acid
MoRF	Molecular recognition feature
MSA	Multiple sequence alignment
MSCRAMM	Microbial surface components recognizing adhesive matrix molecules
MW	Molecular weight
MWCO	Nominal MW cut-off
$^nF1$	The nth F1 module of Fn
NMR	Nuclear magnetic resonance
NTD	N-terminal domain of Fn, contains $^{1-5}F1$.
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

PCS	Primary contact site
pNTD	29 kDa proteolytic N-terminal domain of Fn from tryptic digest. Corresponds to residues 1–259 of mature human Fn (Figure 1).
PSSI	Probability-based secondary structure identification
PyTT5	Synthetic peptide that binds to ^{2}F1^{3}F1, corresponds to residues 560–577 of SfbI, UniProt ID Q01924
rpHPLC	Reversed-phase high-performance liquid chromatography
rpm	Rotations per minute
R$_s$	Stokes radius
S	Soluble cellular protein gel sample
SDS	Sodium dodecyl sulfate
SDS PAGE	SDS polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEM	Scanning electron microscopy
SfbI	Streptococcus Fn-binding protein I
SfbI-5	The fifth FnBR of SfbI, corresponds to residues 541–591 of UniProt ID Q01924
SOF	Serum opacity factor
SSP	Secondary structural propensity
STAFF1	Synthetic peptide that binds to ^{4}F1^{5}F1, corresponds to residues 508–530 of FnBPA, UniProt ID P14738
STAFF5	Synthetic peptide that binds to ^{4}F1^{5}F1, corresponds to residues 638–655 of FnBPA
STATT1	Synthetic peptide that binds to ^{2}F1^{3}F1, corresponds to residues 529–549 of FnBPA
STATT5	Synthetic peptide that binds to ^{2}F1^{3}F1, corresponds to residues 655–672 of FnBPA
T	Total cellular protein gel sample
TBS	Tris buffered saline
TFA	Trifluoroacetic acid
TickFnBR	Putative FnBR from the deer tick Ixodes Scapularis, corresponds to residues 36–53 of UniProt ID ZZ.
Tris	Tris (hydroxymethyl) aminomethane
TROSY	Transverse relaxation-optimized spectroscopy
UniProt ID	UniProt accession number, which is a unique number for each protein sequence available in the UniProt database (www.uniprot.org)
UR	Upstream Fn-binding region
UV	Ultraviolet
WJ2 dataset	Random coil chemical shift dataset from Wang and Jardetzky (2002b)
WS dataset	Random coil chemical shift dataset from Wishart *et al.* (1995)

YNB   Yeast nitrogen base without amino acids or ammonium sulfate

YSBL   York structural biology laboratory

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology,* **215:** 403–410.

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

Amann, E., Brosius, J., & Ptashne, M. (1983). Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli. Gene,* **25:** 167-178.

Anderson, J.F. (2002). The natural history of ticks. *Medical Clinics of North America,* **86:** 205–218.

Bahadur, R.P., Chakrabarti, P., Rodier, F., & Janin, J. (2004). A dissection of specific and non-specific protein-protein interfaces. *Journal of Molecular Biology,* **336:** 943–955.

Bailey, T.L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings, International Conference on Intelligent Systems for Molecular Biology,* **2:** 28–36.

Bendtsen, J.D., Nielsen, H., von Heijne, G., & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology,* **340:** 783–795.

Berkelman, T. (2006). Sensitivity and protein-to-protein consistency of Flamingo fluorescent gel stain compared to other fluorescent gel stains, *Technical note 5705*: Bio-Rad Laboratories.

Bingham, R.J., Rudiño-Piñera, E., Meenan, N.A.G., Schwarz-Linek, U., Turkenburg, J.P., Höök, M., Garman, E.F., & Potts, J.R. (2008). Crystal structures of fibronectin-binding sites from *Staphylococcus aureus* FnBPA in complex with fibronectin domains. *Proceedings of the National Academy of Sciences of the United States of America,* **105:** 12254–12258.

Bishop, R., Lambson, B., Wells, C., Pandit, P., Osaso, J., Nkonge, C., Morzaria, S., Musoke, A., & Nene, V. (2002). A cement protein of the tick *Rhipicephalus appendiculatus*, located in the secretory e cell granules of the type III salivary gland acini, induces strong antibody responses in cattle. *International Journal for Parasitology,* **32:** 833–842.

Bisno, A.L., Brito, M.O., & Collins, C.M. (2003). Molecular basis of group A streptococcal virulence. *Lancet Infectious Diseases,* **3:** 191–200.

Blow, D. (2002). *Outline of crystallography for biologists*. Oxford: Oxford University Press.

Bogan, A.A., & Thorn, K.S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology,* **280:** 1–9.

Brandt, C.M., Allerberger, F., Spellerberg, B., Holland, R., Lutticken, R., & Haase, G. (2001). Characterization of consecutive *Streptococcus pyogenes* isolates from patients with pharyngitis and bacteriological treatment failure: special reference to *prtF1* and *sic / drs*. *Journal of Infectious Diseases,* **183:** 670–674.

Bright, J.R., Pickford, A.R., Potts, J.R., & Campbell, I.D. (2000). Preparation of isotopically labeled recombinant fragments of fibronectin for functional and

structural study by heteronuclear nuclear magnetic resonance spectroscopy. *Methods in Molecular Biology,* **139:** 59–69.

Brossard, M., & Wikel, S.K. (2004). Tick immunobiology. *Parasitology,* **129 Suppl:** S161–176.

Brouillette, E., Talbot, B.G., & Malouin, F. (2003). The fibronectin-binding proteins of *Staphylococcus aureus* may promote mammary gland colonization in a lactating mouse model of mastitis. *Infection and Immunity,* **71:** 2292–2295.

Carapetis, J.R., Steer, A.C., Mulholland, E.K., & Weber, M. (2005). The global burden of group A streptococcal diseases. *Lancet Infectious Diseases,* **5:** 685–694.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G., & Skelton, N.J. (1995). *Protein NMR spectroscopy: Principles and practice* (1 ed.). San Diego: Academic Press.

CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica. Section D, Biological Crystallography,* **50:** 760–763.

Chakrabarti, P., & Janin, J. (2002). Dissecting protein-protein recognition sites. *Proteins,* **47:** 334–343.

Chilkoti, A., Boland, T., Ratner, B.D., & Stayton, P.S. (1995). The relationship between ligand-binding thermodynamics and protein-ligand interaction forces measured by atomic force microscopy. *Biophysical Journal,* **69:** 2125–2130.

Cho, J., & Mosher, D.F. (2006a). Impact of fibronectin assembly on platelet thrombus formation in response to type I collagen and von Willebrand factor. *Blood,* **108:** 2229–2236.

Cho, J., & Mosher, D.F. (2006b). Enhancement of thrombogenesis by plasma fibronectin cross-linked to fibrin and assembled in platelet thrombi. *Blood,* **107:** 3555–3563.

Clamp, M., Cuff, J., Searle, S.M., & Barton, G.J. (2004). The Jalview Java alignment editor. *Bioinformatics,* **20:** 426–427.

Colombatti, A., Bonaldo, P., & Doliana, R. (1993). Type A modules: interacting domains found in several non-fibrillar collagens and in other extracellular matrix proteins. *Matrix,* **13:** 297–306.

Courtney, H.S., Li, Y., Dale, J.B., & Hasty, D.L. (1994). Cloning, sequencing, and expression of a fibronectin/fibrinogen-binding protein from group A streptococci. *Infection and Immunity,* **62:** 3937–3946.

Courtney, H.S., Hasty, D.L., & Dale, J.B. (2002). Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci. *Annals of Medicine,* **34:** 77–87.

Courtney, H.S., Zhang, Y.M., Frank, M.W., & Rock, C.O. (2006). Serum opacity factor, a streptococcal virulence factor that binds to apolipoproteins A-I and A-II and disrupts high density lipoprotein structure. *Journal of Biological Chemistry,* **281:** 5515–5521.

Csizmok, V., Bokor, M., Banki, P., Klement, E., Medzihradszky, K.F., Friedrich, P., Tompa, K., & Tompa, P. (2005). Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry,* **44:** 3955–3964.

Cunningham, M.W. (2000). Pathogenesis of group A streptococcal infections. *Clinical Microbiology Reviews,* **13:** 470–511.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., & Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research,* **35:** W375–383.

Deane, J.E., Ryan, D.P., Sunde, M., Maher, M.J., Guss, J.M., Visvader, J.E., & Matthews, J.M. (2004). Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex. *EMBO Journal,* **23:** 3589–3598.

Deivanayagam, C.C., Rich, R.L., Carson, M., Owens, R.T., Danthuluri, S., Bice, T., Höök, M., & Narayana, S.V. (2000). Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure,* **8:** 67–78.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR,* **6:** 277–293.

DeSimone, D.W., Norton, P.A., & Hynes, R.O. (1992). Identification and characterization of alternatively spliced fibronectin mRNAs expressed in early *Xenopus* embryos. *Developmental Biology,* **149:** 357–369.

Dickinson, C.D., Veerapandian, B., Dai, X.P., Hamlin, R.C., Xuong, N.H., Ruoslahti, E., & Ely, K.R. (1994). Crystal structure of the tenth type III cell adhesion module of human fibronectin. *Journal of Molecular Biology,* **236:** 1079–1092.

Dinkla, K., Rohde, M., Jansen, W.T., Carapetis, J.R., Chhatwal, G.S., & Talay, S.R. (2003). *Streptococcus pyogenes* recruits collagen *via* surface-bound fibronectin: a novel colonization and immune evasion mechanism. *Molecular Microbiology,* **47:** 861–869.

Dosztanyi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005a). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics,* **21:** 3433–3434.

Dosztanyi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005b). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of Molecular Biology,* **347:** 827–839.

Dosztanyi, Z., & Tompa, P. (2008). Prediction of protein disorder. *Methods in Molecular Biology,* **426:** 103–115.

Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., & Uversky, V.N. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics,* **9 Suppl 2:** S1.

Dyson, H.J., & Wright, P.E. (2001). Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods in Enzymology,* **339:** 258–270.

Dyson, H.J., & Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews. Molecular Cell Biology,* **6:** 197–208.

Eddy, S.R. (1996). Hidden Markov models. *Current Opinion in Structural Biology,* **6:** 361–365.

Eddy, S.R. (1998). Profile hidden Markov models. *Bioinformatics,* **14:** 755–763.

Eddy, S.R. (2003a). HMMER2 (version 2.3.2) Retrieved 23-01-2009, from http://hmmer.janelia.org

Eddy, S.R. (2003b). HMMER user's guide, Biological sequence analysis using profile hidden Markov models. Available from ftp://selab.janelia.org/pub/software/hmmer/CURRENT/Userguide.pdf

Eddy, S.R. (2004). What is a hidden Markov model? *Nature Biotechnology,* **22:** 1315–1316.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography,* **60:** 2126–2132.

Espinoza-Fonseca, L.M. (2009). Reconciling binding mechanisms of intrinsically disordered proteins. *Biochemical and Biophysical Research Communications,* **382:** 479–482.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica. Section D, Biological Crystallography,* **62:** 72–82.

Facinelli, B., Spinaci, C., Magi, G., Giovanetti, E., & P, E.V. (2001). Association between erythromycin resistance and ability to enter human respiratory cells in group A streptococci. *Lancet,* **358:** 30–33.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., & Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Research,* **36:** D281–D288.

Fraser, J.D., & Proft, T. (2008). The bacterial superantigen and superantigen-like proteins. *Immunological Reviews,* **225:** 226–243.

Fuxreiter, M., Simon, I., Friedrich, P., & Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of Molecular Biology,* **338:** 1015–1026.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (1st ed., pp. 571–607). Totowa, N.J: Humana Press.

Geiger, B., Bershadsky, A., Pankov, R., & Yamada, K.M. (2001). Transmembrane crosstalk between the extracellular matrix–cytoskeleton crosstalk. *Nature Reviews. Molecular Cell Biology,* **2:** 793–805.

George, E.L., Georges-Labouesse, E.N., Patel-King, R.S., Rayburn, H., & Hynes, R.O. (1993). Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. *Development,* **119:** 1079–1091.

Grinnell, F. (1984). Fibronectin and wound healing. *Journal of Cellular Biochemistry,* **26:** 107–116.

Gunasekaran, K., Tsai, C.J., Kumar, S., Zanuy, D., & Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. *Trends in Biochemical Sciences,* **28:** 81–85.

Guzman, C.A., Talay, S.R., Molinari, G., Medina, E., & Chhatwal, G.S. (1999). Protective immune response against *Streptococcus pyogenes* in mice after intranasal vaccination with the fibronectin-binding protein SfbI. *Journal of Infectious Diseases,* **179:** 901–906.

Hanski, E., & Caparon, M. (1992). Protein F, a fibronectin-binding protein, is an adhesin of the group A streptococcus *Streptococcus pyogenes. Proceedings of the National Academy of Sciences of the United States of America,* **89:** 6172–6176.

Harris, R.K., Becker, E.D., De Menezes, S.M., Granger, P., Hoffman, R.E., & Zilm, K.W. (2008). Further conventions for NMR shielding and chemical shifts (IUPAC Recommendations 2008). *Magn Reson Chem,* **46:** 582–598.

Hill, C.A., Wikel, S.K., Alarcon-Chaidez, F., Nene, V.M., Kodira, C.H., Devon, K., Nusbaum, C., Galagan, J., & Birren, B. (2007). *Ixodes scapularis* EST sequencing.

Hore, P.J. (1995). *Nuclear magnetic resonance*. Oxford: Oxford University Press.

House-Pompeo, K., Xu, Y., Joh, D., Speziale, P., & Höök, M. (1996). Conformational changes in the fibronectin binding MSCRAMMs are induced by ligand binding. *Journal of Biological Chemistry,* **271:** 1379–1384.

Huang, J., Sun, J., Alarcon-Chaidez, F.J., & Wikel, S.K. (2007). Transcriptome analysis of *Ixodes scapularis* (Acari:Ixodidae) synganglia.

Hutchinson, E.G., Sessions, R.B., Thornton, J.M., & Woolfson, D.N. (1998). Determinants of strand register in antiparallel β-sheets of proteins. *Protein Science,* **7:** 2287–2300.

Hynes, W. (2004). Virulence factors of the group A streptococci and genes that regulate their expression. *Frontiers in Bioscience,* **9:** 3399–3433.

Jaffe, J., Natanson-Yaron, S., Caparon, M.G., & Hanski, E. (1996). Protein F2, a novel fibronectin-binding protein from *Streptococcus pyogenes*, possesses two binding domains. *Molecular Microbiology,* **21:** 373–384.

Jancarik, J., Pufan, R., Hong, C., Kim, S.H., & Kim, R. (2004). Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. *Acta Crystallographica. Section D, Biological Crystallography,* **60:** 1670–1673.

Janin, J., Bahadur, R.P., & Chakrabarti, P. (2008). Protein-protein interaction and quaternary structure. *Quarterly Reviews of Biophysics,* **41:** 133–180.

Joh, D., Speziale, P., Gurusiddappa, S., Manor, J., & Höök, M. (1998). Multiple specificities of the staphylococcal and streptococcal fibronectin-binding microbial surface components recognizing adhesive matrix molecules. *European Journal of Biochemistry,* **258:** 897–905.

Joh, D., Wann, E.R., Kreikemeyer, B., Speziale, P., & Höök, M. (1999). Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells. *Matrix Biology,* **18:** 211–223.

Johnson, B.A., & Blevins, R.A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR,* **4:** 603–614.

Kanelis, V., Forman-Kay, J.D., & Kay, L.E. (2001). Multidimensional NMR methods for protein structure determination. *IUBMB Life,* **52:** 291–302.

Katerov, V., Andreev, A., Schalen, C., & Totolian, A.A. (1998). Protein F, a fibronectin-binding protein of *Streptococcus pyogenes*, also binds human fibrinogen: isolation of the protein and mapping of the binding region. *Microbiology,* **144 (Pt 1):** 119–126.

Katerov, V., Lindgren, P.E., Totolian, A.A., & Schalen, C. (2000). Streptococcal opacity factor: a family of bifunctional proteins with lipoproteinase and fibronectin-binding activities. *Current Microbiology,* **40:** 149–156.

Kaufman, W.R. (1989). Tick-host interaction: a synthesis of current concepts. *Parasitol Today,* **5:** 47–56.

Kemp, D.H., Stone, B.F., & Binnington, K.C. (1982). Tick attachment and feeding: role of the mouthparts, feeding apparatus, salivary gland secretions and the host response. In F. D. Obenchain & G. R. (Eds.), *Physiology of Ticks* (pp. 119–168). Oxford: Pergamon.

Kim, J.H., Singvall, J., Schwarz-Linek, U., Johnson, B.J.B., Potts, J.R., & Höök, M. (2004). BBK32, a fibronectin binding MSCRAMM from *Borrelia burgdorferi*, contains a disordered region that undergoes a conformational change on ligand binding. *Journal of Biological Chemistry,* **279:** 41706–41714.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene,* **234:** 187–208.

Kreikemeyer, B., McIver, K.S., & Podbielski, A. (2003). Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends in Microbiology,* **11:** 224–232.

Krogh, A., Brown, M., Mian, I.S., Sjolander, K., & Haussler, D. (1994). Hidden Markov models in computational biology. Applications to protein modeling. *Journal of Molecular Biology,* **235:** 1501–1531.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., & Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics,* **23:** 2947–2948.

Leiss, M., Beckmann, K., Giros, A., Costell, M., & Fassler, R. (2008). The role of integrin binding sites in fibronectin matrix assembly *in vivo. Current Opinion in Cell Biology,* **20:** 502–507.

Lembke, C., Podbielski, A., Hidalgo-Grass, C., Jonas, L., Hanski, E., & Kreikemeyer, B. (2006). Characterization of biofilm formation by clinically relevant serotypes of group A streptococci. *Applied and Environmental Microbiology,* **72:** 2864–2875.

Leslie, A.G. (2006). The integration of macromolecular diffraction data. *Acta Crystallographica. Section D, Biological Crystallography,* **62:** 48–57.

Letunic, I., Doerks, T., & Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Research,* **37:** D229–232.

Lindmark, H., Nilsson, M., & Guss, B. (2001). Comparison of the fibronectin-binding protein FNE from *Streptococcus equi* subspecies *equi* with FNZ from *S. equi* subspecies *zooepidemicus* reveals a major and conserved difference. *Infection and Immunity,* **69:** 3159–3163.

Lo Conte, L., Chothia, C., & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology,* **285:** 2177–2198.

Long, F., Vagin, A.A., Young, P., & Murshudov, G.N. (2008). BALBES: a molecular-replacement pipeline. *Acta Crystallographica. Section D, Biological Crystallography,* **64:** 125–132.

Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S., & Canard, B. (2003). The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-

terminal moiety of the phosphoprotein. *Journal of Biological Chemistry,* **278:** 18638–18648.

Loughman, A., Sweeney, T., Keane, F.M., Pietrocola, G., Speziale, P., & Foster, T.J. (2008). Sequence diversity in the A domain of *Staphylococcus aureus* fibronectin-binding protein A. *BMC Microbiology,* **8:** 74.

Ma, X., Kikuta, H., Ishiguro, N., Yoshioka, M., Ebihara, T., Murai, T., Kobayashi, I., & Kobayashi, K. (2002). Association of the *prtF1* gene (encoding fibronectin-binding protein F1) and the *sic* gene (encoding the streptococcal inhibitor of complement) with *emm* types of group A streptococci isolated from Japanese children with pharyngitis. *Journal of Clinical Microbiology,* **40:** 3835–3837.

Magnusson, M.K., & Mosher, D.F. (1998). Fibronectin: structure, assembly, and cardiovascular implications. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **18:** 1363–1370.

Mao, Y., & Schwarzbauer, J.E. (2005). Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biology,* **24:** 389–399.

Marsh, J.A., Singh, V.K., Jia, Z., & Forman-Kay, J.D. (2006). Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: implications for fibrillation. *Protein Science,* **15:** 2795–2804.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., & Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr,* **40:** 658–674.

Meenan, N.A.G., Visai, L., Valtulina, V., Schwarz-Linek, U., Norris, N.C., Gurusidappa, S., Höök, M., Speziale, P., & Potts, J.R. (2007). The tandem β-zipper model defines high affinity fibronectin-binding repeats within *Staphylococcus aureus* FnBPA. *Journal of Biological Chemistry,* **282:** 25893–25902.

Mészáros, B., Tompa, P., Simon, I., & Dosztányi, Z. (2007). Molecular principles of the interactions of disordered proteins. *Journal of Molecular Biology,* **372:** 549–561.

MicroCal (2004). ITC data analysis in Origin 7.0. 22 Industrial Drive East, Northampton, MA 01060 USA.

Mitchell, G., Lamontagne, C.A., Brouillette, E., Grondin, G., Talbot, B.G., Grandbois, M., & Malouin, F. (2008). *Staphylococcus aureus* SigB activity promotes a strong fibronectin-bacterium interaction which may sustain host tissue colonization by small-colony variants isolated from cystic fibrosis patients. *Molecular Microbiology,* **70:** 1540–1555.

Mittag, T., & Forman-Kay, J.D. (2007). Atomic-level characterization of disordered protein ensembles. *Current Opinion in Structural Biology,* **17:** 3–14.

Mohan, A., Oldfield, C.J., Radivojac, P., Vacic, V., Cortese, M.S., Dunker, A.K., & Uversky, V.N. (2006). Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology,* **362:** 1043–1059.

Molinari, G., Talay, S.R., Valentin-Weigand, P., Rohde, M., & Chhatwal, G.S. (1997). The fibronectin-binding protein of *Streptococcus pyogenes*, SfbI, is involved in the internalization of group A streptococci by epithelial cells. *Infection and Immunity,* **65:** 1357–1363.

Molinari, G., & Chhatwal, G.S. (1999). Streptococcal invasion. *Current Opinion in Microbiology,* **2:** 56–61.

Moretti, F.A., Chauhan, A.K., Iaconcig, A., Porro, F., Baralle, F.E., & Muro, A.F. (2007). A major fraction of fibronectin present in the extracellular matrix of tissues is plasma-derived. *Journal of Biological Chemistry,* **282:** 28057–28062.

Mosher, D.F., & Proctor, R.A. (1980). Binding and Factor XIIIa-mediated cross-linking of a 27-kilodalton fragment of fibronectin to *Staphylococcus aureus. Science,* **209:** 927–929.

Mosher, D.F. (2006). Plasma fibronectin concentration: a risk factor for arterial thrombosis? *Arteriosclerosis, Thrombosis, and Vascular Biology,* **26:** 1193–1195.

Murdoch, D.A. (1998). Gram-positive anaerobic cocci. *Clinical Microbiology Reviews,* **11:** 81–120.

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

Natanson, S., Sela, S., Moses, A.E., Musser, J.M., Caparon, M.G., & Hanski, E. (1995). Distribution of fibronectin-binding proteins among group A streptococci of different M types. *Journal of Infectious Diseases,* **171:** 871–878.

Neeman, R., Keller, N., Barzilai, A., Korenman, Z., & Sela, S. (1998). Prevalence of internalisation-associated gene, *prtF1*, among persisting group-A streptococcus strains isolated from asymptomatic carriers. *Lancet,* **352:** 1974–1977.

Nitsche-Schmitz, D.P., Rohde, M., & Chhatwal, G.S. (2007). Invasion mechanisms of Gram-positive pathogenic cocci. *Thrombosis and Haemostasis,* **98:** 488–496.

Nuttall, P.A., & Labuda, M. (2004). Tick-host interactions: saliva-activated transmission. *Parasitology,* **129 Suppl:** S177–189.

Nyberg, P., Sakai, T., Cho, K.H., Caparon, M.G., Fassler, R., & Bjorck, L. (2004). Interactions with fibronectin attenuate the virulence of *Streptococcus pyogenes. EMBO Journal,* **23:** 2166–2174.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., & Dunker, A.K. (2005). Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins,* **61 Suppl 7:** 176–182.

Oliver, J.H., Jr., Owsley, M.R., Hutcheson, H.J., James, A.M., Chen, C., Irby, W.S., Dotson, E.M., & McLain, D.K. (1993). Conspecificity of the ticks *Ixodes scapularis* and *I. dammini* (Acari: Ixodidae). *Journal of Medical Entomology,* **30:** 54–63.

Österlund, A., & Engstrand, L. (1997a). An intracellular sanctuary for *Streptococcus pyogenes* in human tonsillar epithelium — studies of asymptomatic carriers and *in vitro* cultured biopsies. *Acta Otolaryngol,* **117:** 883–888.

Österlund, A., Popa, R., Nikkila, T., Scheynius, A., & Engstrand, L. (1997b). Intracellular reservoir of *Streptococcus pyogenes in vivo*: a possible explanation for recurrent pharyngotonsillitis. *Laryngoscope,* **107:** 640–647.

Ozeri, V., Tovi, A., Burstein, I., Natanson-Yaron, S., Caparon, M.G., Yamada, K.M., Akiyama, S.K., Vlodavsky, I., & Hanski, E. (1996). A two-domain mechanism for group A streptococcal adherence through protein F to the extracellular matrix. *EMBO Journal,* **15:** 989–998.

Ozeri, V., Rosenshine, I., Mosher, D.F., Fässler, R., & Hanski, E. (1998). Roles of integrins and fibronectin in the entry of *Streptococcus pyogenes* into cells *via* protein F1. *Molecular Microbiology,* **30:** 625–637.

Ozeri, V., Rosenshine, I., Ben-Ze'Ev, A., Bokoch, G.M., Jou, T.S., & Hanski, E. (2001). *De novo* formation of focal complex-like structures in host cells by invading Streptococci. *Molecular Microbiology,* **41:** 561–573.

Pankov, R., & Yamada, K.M. (2002). Fibronectin at a glance. *Journal of Cell Science,* **115:** 3861–3863.

Patti, J.M., Allen, B.L., McGavin, M.J., & Höök, M. (1994). MSCRAMM-mediated adherence of microorganisms to host tissues. *Annual Review of Microbiology,* **48:** 585–617.

Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., & Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics,* **7:** 208.

Penkett, C.J., Redfield, C., Dodd, I., Hubbard, J., McBay, D.L., Mossakowska, D.E., Smith, R.A., Dobson, C.M., & Smith, L.J. (1997). NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. *Journal of Molecular Biology,* **274:** 152–159.

Penkett, C.J., Redfield, C., Jones, J.A., Dodd, I., Hubbard, J., Smith, R.A., Smith, L.J., & Dobson, C.M. (1998). Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from *Staphylococcus aureus*. *Biochemistry,* **37:** 17054–17067.

Pickford, A.R., Smith, S.P., Staunton, D., Boyd, J., & Campbell, I.D. (2001). The hairpin structure of the ^{6}F1^{1}F2^{2}F2 fragment from human fibronectin enhances gelatin binding. *EMBO Journal,* **20:** 1519–1529.

Pope, B., & Kent, H.M. (1996). High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Research,* **24:** 536–537.

Potterton, E., Briggs, P., Turkenburg, M., & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallographica. Section D, Biological Crystallography,* **59:** 1131–1137.

Potts, J.R., & Campbell, I.D. (1996). Structure and function of fibronectin modules. *Matrix Biology,* **15:** 313–321.

Potts, J.R., Bright, J.R., Bolton, D., Pickford, A.R., & Campbell, I.D. (1999). Solution structure of the N-terminal F1 module pair from human fibronectin. *Biochemistry,* **38:** 8304–8312.

Ptitsyn, O.B., & Finkelstein, A.V. (1983). Theory of protein secondary structure and algorithm of its prediction. *Biopolymers,* **22:** 15–25.

Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE,* **77:** 257–286.

Raibaud, S., Schwarz-Linek, U., Kim, J.H., Jenkins, H.T., Baines, E.R., Gurusiddappa, S., Höök, M., & Potts, J.R. (2005). *Borrelia burgdorferi* binds fibronectin through a tandem β-zipper, a common mechanism of fibronectin binding in staphylococci, streptococci, and spirochetes. *Journal of Biological Chemistry,* **280:** 18803–18809.

Ramachandran, V., McArthur, J.D., Behm, C.E., Gutzeit, C., Dowton, M., Fagan, P.K., Towers, R., Currie, B., Sriprakash, K.S., & Walker, M.J. (2004). Two distinct genotypes of *prtF2*, encoding a fibronectin binding protein, and evolution of the

gene family in *Streptococcus pyogenes. Journal of Bacteriology,* **186:** 7601–7609.

Receveur-Brechot, V., Bourhis, J.M., Uversky, V.N., Canard, B., & Longhi, S. (2006). Assessing protein disorder and induced folding. *Proteins,* **62:** 24–45.

Rehm, T., Huber, R., & Holak, T.A. (2002). Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure,* **10:** 1613–1618.

Ribeiro, J.M., Alarcon-Chaidez, F., Francischetti, I.M., Mans, B.J., Mather, T.N., Valenzuela, J.G., & Wikel, S.K. (2006). An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochemistry and Molecular Biology,* **36:** 111–129.

Robert, P., Benoliel, A.M., Pierres, A., & Bongrand, P. (2007). What is the biological relevance of the specific bond properties revealed by single-molecule studies? *Journal of Molecular Recognition,* **20:** 432–447.

Rohde, M., Muller, E., Chhatwal, G.S., & Talay, S.R. (2003). Host cell caveolae act as an entry-port for group A streptococci. *Cellular Microbiology,* **5:** 323–342.

Romberger, D.J. (1997). Fibronectin. *International Journal of Biochemistry and Cell Biology,* **29:** 939–943.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., & Dunker, A.K. (2001). Sequence complexity of disordered protein. *Proteins,* **42:** 38–48.

Rost, B., & Sander, C. (1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proceedings of the National Academy of Sciences of the United States of America,* **90:** 7558–7562.

Rudiño-Piñera, E., Ravelli, R.B.G., Sheldrick, G.M., Nanao, M.H., Korostelev, V.V., Werner, J.M., Schwarz-Linek, U., Potts, J.R., & Garman, E.F. (2007). The solution and crystal structures of a module pair from the *Staphylococcus aureus*-binding site of human fibronectin — a tale with a twist. *Journal of Molecular Biology,* **368:** 833–844.

Saha, R.P., Bahadur, R.P., Pal, A., Mandal, S., & Chakrabarti, P. (2006). ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. *BMC Structural Biology,* **6:** 11.

Sali, A., & Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology,* **234:** 779–815.

Sambrook, J., & Russell, D.W. (2001). *Molecular cloning: A laboratory manual* (3rd ed.). Cold Spring Harbor, N.Y: Cold Spring Harbour Laboratory Press.

Sauer, J.R., McSwain, J.L., Bowman, A.S., & Essenberg, R.C. (1995). Tick salivary gland physiology. *Annual Review of Entomology,* **40:** 245–267.

Schulze, K., Medina, E., Talay, S.R., Towers, R.J., Chhatwal, G.S., & Guzman, C.A. (2001). Characterization of the domain of fibronectin-binding protein I of *Streptococcus pyogenes* responsible for elicitation of a protective immune response. *Infection and Immunity,* **69:** 622–625.

Schulze, K., Medina, E., Chhatwal, G.S., & Guzman, C.A. (2003). Identification of B- and T-cell epitopes within the fibronectin-binding domain of the SfbI protein of *Streptococcus pyogenes. Infection and Immunity,* **71:** 7197–7201.

Schuster-Bockler, B., Schultz, J., & Rahmann, S. (2004). HMM Logos for visualization of protein families. *BMC Bioinformatics,* **5:** 7.

Schwarz-Linek, U., Werner, J.M., Pickford, A.R., Gurusiddappa, S., Kim, J.H., Pilka, E.S., Briggs, J.A.G., Gough, T.S., Höök, M., Campbell, I.D., & Potts, J.R. (2003). Pathogenic bacteria attach to human fibronectin through a tandem β-zipper. *Nature,* **423:** 177–181.

Schwarz-Linek, U., Höök, M., & Potts, J.R. (2004a). The molecular basis of fibronectin-mediated bacterial adherence to host cells. *Molecular Microbiology,* **52:** 631–641.

Schwarz-Linek, U., Pilka, E.S., Pickford, A.R., Kim, J.H., Höök, M., Campbell, I.D., & Potts, J.R. (2004b). High affinity streptococcal binding to human fibronectin requires specific recognition of sequential F1 modules. *Journal of Biological Chemistry,* **279:** 39017–39025.

Schwarz-Linek, U., Höök, M., & Potts, J.R. (2006). Fibronectin-binding proteins of gram-positive cocci. *Microbes and Infection,* **8:** 2291–2298.

Schwarzbauer, J.E. (1991). Identification of the fibronectin sequences required for assembly of a fibrillar matrix. *Journal of Cell Biology,* **113:** 1463–1473.

Schwarzinger, S., Kroon, G.J., Foss, T.R., Wright, P.E., & Dyson, H.J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *Journal of Biomolecular NMR,* **18:** 43–48.

Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E., & Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *Journal of the American Chemical Society,* **123:** 2970–2978.

Sela, S., Aviv, A., Tovi, A., Burstein, I., Caparon, M.G., & Hanski, E. (1993). Protein F: an adhesin of *Streptococcus pyogenes* binds fibronectin *via* two distinct domains. *Molecular Microbiology,* **10:** 1049–1055.

Sela, S., Neeman, R., Keller, N., & Barzilai, A. (2000). Relationship between asymptomatic carriage of *Streptococcus pyogenes* and the ability of the strains to adhere to and be internalised by cultured epithelial cells. *Journal of Medical Microbiology,* **49:** 499–502.

Sharp, K. (2001). Entropy-enthalpy compensation: fact or artifact? *Protein Science,* **10:** 661–667.

Signäs, C., Raucci, G., Jonsson, K., Lindgren, P.E., Anantharamaiah, G.M., Höök, M., & Lindberg, M. (1989). Nucleotide sequence of the gene for a fibronectin-binding protein from *Staphylococcus aureus:* use of this peptide sequence in the synthesis of biologically active peptides. *Proceedings of the National Academy of Sciences of the United States of America,* **86:** 699–703.

Smith, D.B., & Johnson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene,* **67:** 31-40.

Sonenshine, D.E. (1991). *The biology of ticks* (1 ed. Vol. 1). Oxford: Oxford University Press.

Sonenshine, D.E. (1993). *The biology of ticks* (1 ed. Vol. 2). Oxford: Oxford University Press.

Sun, P.D., Foster, C.E., & Boyington, J.C. (2004). Overview of protein structural and functional folds. *Curr Protoc Protein Sci,* **Chapter 17:** Unit 17 11.

Symersky, J., Patti, J.M., Carson, M., House-Pompeo, K., Teale, M., Moore, D., Jin, L., Schneider, A., DeLucas, L.J., Höök, M., & Narayana, S.V. (1997). Structure of the collagen-binding domain from a *Staphylococcus aureus* adhesin. *Nature Structural Biology,* **4:** 833–838.

Talay, S.R., Ehrenfeld, E., Chhatwal, G.S., & Timmis, K.N. (1991). Expression of the fibronectin-binding components of *Streptococcus pyogenes* in *Escherichia coli* demonstrates that they are proteins. *Molecular Microbiology,* **5:** 1727–1734.

Talay, S.R., Valentin-Weigand, P., Jerlstrom, P.G., Timmis, K.N., & Chhatwal, G.S. (1992). Fibronectin-binding protein of *Streptococcus pyogenes:* sequence of the binding domain involved in adherence of streptococci to epithelial cells. *Infection and Immunity,* **60:** 3837–3844.

Talay, S.R., Valentin-Weigand, P., Timmis, K.N., & Chhatwal, G.S. (1994). Domain structure and conserved epitopes of Sfb protein, the fibronectin-binding adhesin of *Streptococcus pyogenes. Molecular Microbiology,* **13:** 531–539.

Talay, S.R., Zock, A., Rohde, M., Molinari, G., Oggioni, M., Pozzi, G., Guzman, C.A., & Chhatwal, G.S. (2000). Co-operative binding of human fibronectin to Sfbl protein triggers streptococcal invasion into respiratory epithelial cells. *Cellular Microbiology,* **2:** 521–535.

Tart, A.H., Walker, M.J., & Musser, J.M. (2007). New understanding of the group A Streptococcus pathogenesis cycle. *Trends in Microbiology,* **15:** 318–325.

Terao, Y., Kawabata, S., Kunitomo, E., Murakami, J., Nakagawa, I., & Hamada, S. (2001). Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the *fba* gene is positively transcribed under the Mga regulator. *Molecular Microbiology,* **42:** 75–86.

Tindall, B.J., & Euzeby, J.P. (2006). Proposal of *Parvimonas* gen. nov. and *Quatrionicoccus* gen. nov. as replacements for the illegitimate, prokaryotic, generic names *Micromonas* Murdoch and Shah 2000 and *Quadricoccus* Maszenan *et al.* 2002, respectively. *International Journal of Systematic and Evolutionary Microbiology,* **56:** 2711–2713.

Tompa, P. (2003). Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays,* **25:** 847–855.

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters,* **579:** 3346–3354.

Tompa, P., Szasz, C., & Buday, L. (2005). Structural disorder throws new light on moonlighting. *Trends in Biochemical Sciences,* **30:** 484–489.

Toolan, H.W. (1954). Transplantable human neoplasms maintained in cortisone-treated laboratory animals: H.S. No. 1; H.Ep. No. 1; H.Ep. No. 2; H.Ep. No. 3; and H.Emb.Rh. No. 1. *Cancer Research,* **14:** 660–666.

Towers, R.J., Fagan, P.K., Talay, S.R., Currie, B.J., Sriprakash, K.S., Walker, M.J., & Chhatwal, G.S. (2003). Evolution of *sfbI* encoding streptococcal fibronectin-binding protein I: horizontal genetic transfer and gene mosaic structure. *Journal of Clinical Microbiology,* **41:** 5398–5406.

Uversky, V.N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry,* **32:** 13288–13298.

Uversky, V.N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Science,* **11:** 739–756.

Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N., & Dunker, A.K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. *Journal of Proteome Research,* **6:** 2351–2366.

Valdar, W.S., & Thornton, J.M. (2001). Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins,* **42:** 108–124.

Valdar, W.S. (2002). Scoring residue conservation. *Proteins,* **48:** 227–241.

Valenzuela, J.G. (2004). Exploring tick saliva: from biochemistry to 'sialomes' and functional genomics. *Parasitology,* **129 Suppl:** S83–94.

Velazquez-Campoy, A., Leavitt, S.A., & Freire, E. (2004). Characterization of protein-protein interactions by isothermal titration calorimetry. *Methods in Molecular Biology,* **261:** 35–54.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., & Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins,* **59:** 687–696.

Wallace, A.C., Laskowski, R.A., & Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering,* **8:** 127–134.

Wang, Y., & Jardetzky, O. (2002a). Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Science,* **11:** 852–861.

Wang, Y., & Jardetzky, O. (2002b). Investigation of the neighboring residue effects on protein chemical shifts. *Journal of the American Chemical Society,* **124:** 14075–14084.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., & Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology,* **337:** 635–645.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., & Barton, G.J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. doi:10.1093/bioinformatics/btp033

Wierzbicka-Patynowski, I., & Schwarzbauer, J.E. (2003). The ins and outs of fibronectin matrix assembly. *Journal of Cell Science,* **116:** 3269–3276.

Williams, M.J., Phan, I., Harvey, T.S., Rostagno, A., Gold, L.I., & Campbell, I.D. (1994). Solution structure of a pair of fibronectin type 1 modules with fibrin binding activity. *Journal of Molecular Biology,* **235:** 1302–1311.

Wishart, D.S., Sykes, B.D., & Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* **31:** 1647–1651.

Wishart, D.S., & Sykes, B.D. (1994a). Chemical shifts as a tool for structure determination. *Methods in Enzymology,* **239:** 363–392.

Wishart, D.S., & Sykes, B.D. (1994b). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *Journal of Biomolecular NMR,* **4:** 171–180.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., & Sykes, B.D. (1995). [1]H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR,* **5:** 67–81.

Wishart, D.S., & Nip, A.M. (1998). Protein chemical shift analysis: a practical guide. *Biochemistry and Cell Biology,* **76:** 153–163.

Wishart, D.S., & Case, D.A. (2001). Use of chemical shifts in macromolecular structure determination. *Methods in Enzymology,* **338:** 3–34.

Wright, P.E., & Dyson, H.J. (2009). Linking folding and binding. *Current Opinion in Structural Biology,* **19:** 31–38.

Yao, J., Dyson, H.J., & Wright, P.E. (1997). Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Letters,* **419:** 285–289.

Zeidner, N.S., Schneider, B.S., Nuncio, M.S., Gern, L., & Piesman, J. (2002). Coinoculation of *Borrelia* spp. with tick salivary gland lysate enhances spirochete load in mice and is tick species-specific. *Journal of Parasitology,* **88:** 1276–1278.

Zhang, D., Cupp, M.S., & Cupp, E.W. (2002). Thrombostasin: purification, molecular cloning and expression of a novel anti-thrombin protein from horn fly saliva. *Insect Biochemistry and Molecular Biology,* **32:** 321–330.

Zhang, D., Cupp, M.S., & Cupp, E.W. (2004). Processing of pro-thrombostasin by a recombinant subtilisin-like proprotein convertase derived from the salivary glands of horn flies (*Haematobia irritans*). *Insect Biochemistry and Molecular Biology,* **34:** 1289–1295.