STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF SURFACE PROTEINS FROM PATHOGENIC BACTERIA

Maria Antonietta Gentile

Dottorato in Scienze Biotecnologiche – XXIV ciclo
Indirizzo Biotecnologie Molecolari e Industriali
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XXIV ciclo
Indirizzo Biotecnologie Molecolari e Industriali
Università di Napoli Federico II

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF SURFACE PROTEINS FROM PATHOGENIC BACTERIA

Maria Antonietta Gentile

Dottoranda:	Maria Antonietta Gentile
Relatore:	Prof. Ettore Benedetti
Co-relatori:	Dott. Marcello Merola Dott. Maria Scarselli
Coordinatore:	Prof. Giovanni Sannia

...A mio zio Nicola per avermi insegnato
che la vita va vissuta giorno per giorno
e per avermi seguito da "lontano"
Al mio grande amore Diego
per avermi restituito il sorriso e
per avermi insegnato il significato dell'amore...

INDICE

...Sono convinto del fatto che l'unica cosa che mi ha consentito di proseguire sia stato l'amore che provavo per quello che facevo. dovete trovare ciò che amate. E' questo è tanto vero per il vostro lavoro quanto per chi vi ama. Il lavoro riempirà gran parte della vostra vita e l'unico modo per essere veramente soddisfatti e quello di fare quello che pensate sia il lavoro migliore. E l'unico modo..... per fare il lavoro migliore e quello di amare quello che fate. Se non lo avete ancora trovato, continuate a cercare. Non vi fermate. Come tutti gli affari di cuore, lo saprete quando lo troverete. E, come nelle migliori relazioni, diventerà sempre migliore al passare degli anni. Quindi, continuate a cercarlo fino a quando non l'avrete trovato. Non fermatevi...

Steve Jobs

RIASSUNTO

Gli approcci per lo sviluppo di vaccini hanno avuto notevoli progressi durante l'ultimo secolo. Tutti i vaccini commerciali in questo momento disponibili sono stati realizzati seguendo l'approccio tradizionale e si basano, quindi, sull'utilizzo di patogeni morti o attenuati, di subunità purificate dai patogeni come tossine detossificate mediante trattamenti chimici o genetici, di antigeni purificati o di polisaccaridi coniugati a proteine. L'approccio tradizionale, tuttavia, non ha avuto successo per la produzione di vaccini contro i patogeni che non possono essere coltivati *in vitro* e per i quali non esistono ovvi antigeni protettivi immunodominanti. Nell'ultimo decennio la genomica ha rivoluzionato il mondo della ricerca dei vaccini grazie allo sviluppo di nuove tecnologie per il sequenziamento del DNA.

L'approccio genomico per la produzione di vaccini elabora le informazioni ottenute dal sequenziamento dell'intero genoma del patogeno d'interesse. Le migliaia di geni presenti in tutta la sequenza del DNA sono identificate in seguito ad analisi computerizzata che consente di predire con buona accuratezza sia la localizzazione cellulare di nuovi open reading frame (ORFs) che la funzione delle risultanti proteine e le loro omologie con fattori di virulenza noti prodotti da altri patogeni (analisi *in silico*). I risultati di questo tipo di analisi includono non solo tutti gli antigeni in precedenza identificati mediante approcci convenzionali ma anche la scoperta di nuovi antigeni. Proteine secrete o espresse sulla superficie esterna del patogeno, cioè facilmente "riconoscibili" dal sistema immunitario, saranno selezionate quali candidati per lo sviluppo di un vaccino. La loro capacità di provocare una risposta immunologica efficace sarà valutata in un lungo percorso di saggi *in vivo* o *in vitro* prima di poter attraversare tutto l'iter che porta all'utilizzo nell'uomo. L'approccio genomico per sviluppare vaccini partendo dall'informazione genetica del patogeno è definito "Reverse Vaccinology". La Reverse Vaccinology è stata utilizzata per lo sviluppo di vaccini contro numerosi batteri patogeni come gli Streptococchi di gruppo A e di gruppo B, *Neisseria meningitidis*, *Escherichia coli* extraintestinale e *Streptococcus pneumoniae*.

Gli studi strutturali rientrano in una nuova fase della vaccinologia che è definita Structural Vaccinology. La Structural Vaccinology consiste nella caratterizzazione strutturale dei fattori di virulenza di patogeni ad alta priorità mediante cristallografia ai raggi X o spettroscopia di risonanza magnetica nucleare (NMR) e nella mappatura degli epitopi sulle strutture risolte. La comprensione degli aspetti strutturali alla base dell'interazione anticorpo/epitopo stabilisce, in secondo momento, la base per il design razionale dei vaccini, in modo da ottimizzare la produzione industriale su larga scala, l'omogeneità conformazionale e la stabilità dei nuovi antigeni.

Il progetto di dottorato rientra nell'ambito della Structural Vaccinology. Il lavoro di dottorato è stato focalizzato su un antigene del patogeno *Neisseria meningitidis* ed un antigene del patogeno *Streptococcus pneumoniae*, entrambi identificati mediante la "Reverse Vaccinology".

Neisseria meningitidis rappresenta la principale causa dell'insorgere e della diffusione della meningite tra i bambini e gli adolescenti. *Neisseria meningitidis* è un batterio diplocco Gram-negativo dotato di capsula polisaccaridica che colonizza il tratto superiore della via naso-faringea negli uomini. Il principale fattore di virulenza è la capsula polisaccaridica. Il vaccino meningococcico in uso è un vaccino tetravalente coniugato contro i sierogruppi A, C, Y e W135; tuttavia il vaccino non è

efficace contro il sierogruppo B. Una delle strategie usate per la realizzazione di un vaccino contro il sierogruppo B prevede la realizzazione di vaccini basati su antigeni, una soluzione che è efficace per la protezione contro l'infezione dal ceppo omologo all'antigene utilizzato, ma che risulta incapace di proteggere contro l'infezione da batteri eterologhi a causa dell'elevata variabilità antigenica dei patogeni. L'approccio genomico ha consentito di superare tutti i problemi legati ai metodi tradizionali, infatti la Reverse Vaccinology applicata a *Neisseria meningitidis* ha portato allo sviluppo di un vaccino pentavalente costituito da cinque nuovi antigeni: factor H binding protein (fHbp), NadA, GNA2132, GNA1030 e GNA2091. Durante il primo e il terzo anno di dottorato il lavoro è stato focalizzato su fHbp.

fHbp è una lipoproteina esposta in superficie di circa 28 kDa costituita da 255 residui amminoacidici; oltre ad essere uno dei componenti del vaccino pentavalente contro il sierogruppo B di *Neisseria meningitidis*, è essenziale per la patogenesi perché consente al batterio di sopravvivere e crescere nel sangue umano attraverso il legame con il fattore H, un regolatore negativo del pathway alternativo del complemento. Il rivestimento della superficie batterica con il fattore H consente al patogeno di mimare i tessuti umani e di sfuggire alla lisi mediata dall'attivazione del complemento. fHbp può essere classificata in tre varianti caratterizzate da diversa sequenza amminoacidica. Questa diversità ha un forte impatto sulle proprietà immunologiche della proteina, poiché ogni variante induce una forte immunità solo contro i ceppi di meningococco che presentano alleli omologhi.

Inizialmente studi immunologici sono stati eseguiti con la variante 1 di fHbp; sieri policlonali ed un anticorpo monoclonale (Mab502) sono stati utilizzati per identificare gli epitopi lineari e conformazionali e sulla base dei risultati ottenuti la proteina è stata suddivisa in tre regioni: la regione A (1-101), la regione B (101-164) e la regione C (164-255). Gli epitopi battericidi sono localizzati prevalentemente nel dominio BC della proteina e i residui chiave per il riconoscimento antigene-anticorpo sono il residuo Arg204 e la regione 146-149.

Per comprendere la conformazione degli epitopi protettivi, studi strutturali in soluzione mediante spettroscopia di risonanza magnetica nucleare (NMR) sono stati compiuti sulla variante 1 di fHbp. La proteina è costituita da due β-barrel indipendenti tenuti insieme da un corto linker; saggi immunologici hanno mostrato che gli epitopi battericidi sono localizzati sul dominio C-terminale della proteina. La disponibilità della struttura in soluzione ha consentito di effettuare la mappatura degli epitopi anche mediante esperimenti NMR confermando i dati precedentemente ottenuti medianti saggi immunologici.

Durante il lavoro di dottorato è stata studiata l'interazione tra fHbp e il fattore H mediante esperimenti NMR bidimensionali. Recentemente è stata risolta la struttura cristallografica del complesso fHbp-fH (domini 6-7). I dati rivelano che i residui chiave di fHbp coinvolti nell'interazione sono i residui E218 e E239 e che il complesso è tenuto insieme da numerose interazioni di tipo elettrostatico tra fHbp e il dominio 6 del fattore H. Gli esperimenti NMR in soluzione consentono di studiare l'interazione in un contesto molto più simile a quello *in vivo*; inoltre, gli esperimenti sono stati condotti usando due costrutti diversi del fattore H, il costrutto formato dai domini 67 (CCP67) ed il costrutto formato dai domini 567 (CCP567). In un primo momento si è studiata l'interazione tra fHbp ed il costrutto CCP567 attraverso esperimenti NMR bidimensionali analizzando le perturbazioni dei chemical shift sugli spettri $^{1}H–^{15}N$ TROSY-HSQC della proteina marcata con l'isotopo radioattivo ^{15}N in seguito all'aggiunta di quantità crescenti del fattore H.

La maggior parte dei residui perturbati nella struttura cristallografica che costituiscono l'area di contatto tra le due proteine risultano perturbati anche dai dati NMR, due eccezioni importanti sono rappresentate dai residui E219 e E239 che vengono perturbati solo marginalmente dalla presenza del fattore H. La concordanza dei dati NMR e X-ray rivela che anche in soluzione è mantenuta l'interazione tra la proteina ed i domini 6 e 7 del fH. Tuttavia il numero totale dei residui esposti in superficie perturbati nelle misure NMR definisce un'ampia area di contatto rispetto a quella definita nella struttura cristallografica. Un primo gruppo di residui è localizzato nel linker che connette i due domini della proteina e molti residui localizzati all'interfaccia tra i due domini. La sensibilità di tali residui alla presenza del fattore H suggerisce che durante la formazione del complesso fHbp-CCP567 potrebbero avvenire riarrangiamenti molecolari della proteina. Una seconda regione è formata da residui la cui localizzazione sulla superficie proteica è compatibile con eventuali contatti con il dominio 5 del fH. Per confermare quest'ipotesi gli stessi esperimenti sono stati eseguiti utilizzando il costrutto CCP67 del fH. L'analisi rivela che il numero di residui perturbati rimane invariato, escludendo la possibilità che il dominio 5 possa stabilire contatti specifici con la proteina e suggerendo che l'interazione con il fH è limitata ai domini 67. I dati NMR ottenuti sono in accordo con gli studi precedenti di mappatura degli epitopi eseguiti su fHbp. Beernink e colleghi hanno dimostrato che Gly121 e Lys122, due residui esposti in superficie e localizzati nel loop che congiunge gli strand 7 e 8 del dominio N-terminale della proteina, sono riconosciuti da anticorpi monoclonali come Jar3 e Jar5, che inibiscono il legame al fH. In accordo, questi residui risultano perturbati nella titolazione mediante spettroscopia NMR con i costrutti del fH. Inoltre, è stato dimostrato attraverso esperimenti FACS che l'interazione con il fH non è influenzata dall'interazione della proteina con l'anticorpo MAb502 (11). I residui coinvolti nell'interazione con tale anticorpo risultano infatti non perturbati durante gli esperimenti NMR, confermando che non esiste sovrapposizione tra i siti di legame del fH e dell'anticorpo. I dati NMR possono essere utilizzati come punto di partenza per uno studio di mutagenesi sito-diretta dei residui perturbati allo scopo di identificare altri residui chiave coinvolti nell'interazione tra il fH e fHbp.

Durante il terzo anno del dottorato l'attività di ricerca è stata focalizzata sullo studio del possibile ruolo di fHbp nei processi di assorbimenti del ferro dall'ambiente durante la patogenesi del microorganismo. L'analisi strutturale della proteina evidenzia una forte similitudine con le lipocaline, proteine trovate in piante, funghi, batteri, vertebrati e invertebrati che sono coinvolte in numerosi processi biologici incluso il legame a molecole coinvolte nei processi di reclutamento del ferro dall'ambiente chiamate siderofori. I siderofori sono piccole molecole sintetizzate dai batteri e dai funghi in condizioni di stress in presenza di poco ferro; essi fungono da agenti chelanti specifici per tale metallo. Il ruolo dei siderofori durante la patogenesi è di rendere disponibile il ferro alla cellula batterica catturando il metallo dall'ambiente, per questo motivo queste piccole molecole rappresentano un possibile bersaglio per il sistema immunitario della cellula ospite. Molti batteri hanno evoluto la capacità di utilizzare siderofori prodotti da altri microorganismi, ad esempio *Neisseria meningitidis*, pur non producendo siderofori esprime una proteina di superficie, FetA, nota essere un recettore per l'enterobactina, un sideroforo prodotto dai batteri enterici. Sulla base di queste osservazioni è stato studiato il possibile ruolo di fHbp come siderophore binding protein mediante spettroscopia NMR, in particolare è stata studiata la possibile interazione con tre siderofori: enterobactina, yersiniabactina e la salmochelina. Lo studio, come nel caso della mappatura del sito di binding con il

fattore H, è stato eseguito analizzando le perturbazioni sullo spettro $^{1}H–^{15}N$- HSQC della proteina marcata con l'isotopo radioattivo ^{15}N in seguito all'aggiunta di quantità crescente di sideroforo. Sono state eseguite diverse titolazioni in modo da investigare la possibile interazione con diversi siderofori, per gli esperimenti NMR sono stati usati enterobactina, Fe-enterobactina, Fe-Yersiniabactina e Fe-Salmochelina. I dati NMR rivelano un'affinità solo per la Fe-enterobactina, l'aggiunta di tale sideroforo alla proteina risulta in un progressivo incremento della larghezza di riga dei segnali di alcuni residui sugli spettri, che scompaiono al rapporto molare proteina:sideroforo pari a 1:1.5. Tutti i residui perturbati sono localizzati sul dominio C-terminale della proteina. I dati NMR sono stati, in seguito, usati per generare un modello del complesso fHbp-enterobactina usando il programma HADDOCK2.0. I dati ottenuti, inoltre, rivelano che la regione che interagisce con l'enterobactina non è sovrapposta alla regione contenente il sito di legame per il fH, questo suggerisce un possibile ruolo di fHbp nei processi di acquisizione del ferro del meningococco e nella sopravvivenza del microorganismo nella cellula ospite durante la patogenesi.

Durante il secondo anno di dottorato il lavoro è stato focalizzato su un antigene del patogeno *Streptococcus pneumoniae*, RrgB.

Streptococcus pneumoniae è un batterio diplococco Gram-postitivo e rappresenta la principale causa dell'insorgere delle malattie del tratto respiratorio fra cui la polmonite ed è causa minoritaria dell'insorgere di malattie invasive fra cui la meningite. Il principale fattore di virulenza per questo microrganismo è rappresentato dalla capsula polisaccaridica, la cui differente composizione chimica determina la divisione in vari sierogruppi. Il vaccino disponibile in commercio è un vaccino polivalente coniugato ma non universale. Recentemente, sulla superficie batterica di *Streptococcus pneumoniae* sono stati identificati pili, i risultati ottenuti dagli studi effettuati evidenziano un ruolo chiave per queste strutture superficiali nei processi di adesione del patogeno. La produzione del pilo di *Streptococcus pneumoniae* è regolata dall'operone rlrA costituito da sette geni: un gene codificante un regolatore trascrizionale positivo; i geni srtB, srtC e srtD, codificanti enzimi richiesti per l'assemblaggio del pilo sulla parete cellulare; e i geni rrgA, rrgB e rrgC, codificanti le tre proteine strutturali del pilo. L'analisi strutturale del pilo mediante microscopia crio-elettronica mostra che i pili sono organizzati in strutture coiled-coil costituite da almeno due protofilamenti, in cui RrgB costituisce il componente principale e forma lo scheletro del pilo, mentre RrgA e RrgC sono ancorate allo scheletro proteico del pilo e coinvolte nei fenomeni di adesione alle cellule ospiti. Studi immunologici hanno rivelato che le tre subunità del pilo conferiscono protezione e quindi sono considerati come potenziali candidati per vaccini di nuova generazione basati su proteine. Da un punto di vista biochimico le proteine dei pili dei batteri Gram-positivi sono caratterizzate dalla presenza di legami isopeptidici intramolecolari e intermolecolari. I legami intramolecolari si formano tra le catene laterali di un residuo di Asn e di un residuo di Lys, mentre i legami intermolecolari sono fondamentali per la polimerizzazione del pilo e si formano tra le catene laterali di un residuo di Thr presente nel motivo LPXTG all'estremità C-terminale di una subunità RrgB e di un residuo di Lys localizzato all'estremità N-terminale della subunità successiva.

RrgB è una proteina di superficie batterica di 665 amminoacidi composta di diversi domini: il peptide leader (1-30), il dominio D1 (30-184), il dominio D2 (185-327), il dominio D3 (328-446), il dominio D4 (447-627) e il dominio LPxTG, contenente il motivo riconosciuto dalle sortasi (628-665). Recentemente è stata risolta la struttura cristallografica dei domini D2, D3, D4 (187-627); tutti i domini presentano un folding IgG-like e sono caratterizzati dalla presenza di un legame

isopeptidico intramolecolare. Saggi immunologici hanno inoltre consentito di identificare il dominio D1 come la regione che conferisce maggiore protezione, tuttavia non esistono dati strutturali sul dominio D1 e quindi non è possibile approfondire gli aspetti strutturali alla base del riconoscimento antigene-anticorpo.

Durante il lavoro di dottorato si è lavorato sulla caratterizzazione strutturale del dominio D1 di RrgB mediante esperimenti NMR in collaborazione con il CERM di Firenze allo scopo di comprendere gli aspetti molecolari alla base della maggiore immunogenicità di questo dominio. Sono stati effettuati esperimenti NMR bidimensionali e tridimensionali su campioni marcati con gli isotopi radioattivi ^{15}N e $^{13}C/^{15}N$ utilizzando spettrometri 500, 600, 700, 800 e 900 MHz, gli spettri sono stati analizzati mediante il software Cara. I dati ottenuti dall'analisi e dall'interpretazione degli spettri sono stati utilizzati per la determinazione della struttura tridimensionale della proteina attraverso l'uso del programma CYANA2.1. La struttura è stata raffinata mediante il programma AMBER10.0 utilizzato per la minimizzazione energetica. Il dominio D1 presenta un folding β sandwich IgG-like. Il core proteico è caratterizzato da sette β strand paralleli e antiparalleli raggruppati in due β sheet disposti l'uno contro l'altro e circondati da due lunghi segmenti; un α-elica è localizzata all'interno del primo segmento e due β strand sono localizzati all'interno del secondo segmento. Una delle caratteristiche che differenzia il dominio D1 dai domini D2, D3 e D4 è l'assenza di legami isopeptidici intramolecolari e la presenza di molte regioni flessibili. La flessibilità del D1 potrebbe essere fondamentale nei processi di riconoscimento specifico antigene-anticorpo. Infatti, la plasticità strutturale della proteina potrebbe essere correlata all'abilità del dominio D1 di andare incontro a cambi conformazionali locali, di adattare la struttura in modo da ottimizzare l'interazione con gli anticorpi e d'incrementare l'affinità e la specificità dei processi di riconoscimento antigene-anticorpo.

Una ricerca per le proteine strutturali omologhe, effettuata usando il server DALI, ha portato all'identificazione di tre proteine con folding simile: il dominio C-terminale CNA3 di BcpA che rappresenta la pilina principale di *Bacillus cereus*, il dominio N-terminale della pilina SpaA di *Corynebacterium diphtheriae* e il dominio N1 della pilina minore di *Streptococcus agalactiae* GBS52. Tutte queste proteine non presentano legami isopeptidici intramolecolari, tuttavia solo nel caso della SpaA il dominio che corrisponde al D1 è localizzato all'estremità N-terminale così come il D1 nella proteina intera RrgB. La sovrapposizione delle due strutture rivela, inoltre, la stessa orientazione spaziale delle catene laterali dei residui di Lys localizzati nel pilin motif YPKN e coinvolte nella formazione del legame isopeptidico intermolecolare durante la polimerizzazione del pilo; entrambe le lisine non sono completamente esposte, in particolare nel dominio D1 la catena laterale della lisina è proiettata in una cavità formata tra il core della proteina ed un loop flessibile. Questo dato suggerisce che anche durante la polimerizzazione del pilo siano necessari dei cambi conformazionali del dominio D1. Il ruolo del residuo di lisina, così identificato, nel processo di polimerizzazione del pilo è stato confermato mediante esperimenti di mutagenesi sito-diretta.

In conclusione, questo studio fornisce informazioni addizionali riguardanti le caratteristiche delle proteine dei pili e pone le basi per il design razionale di nuovi antigeni per implementare il vaccino proteico contro le malattie da pneumococco. Le nuove informazioni strutturali e dinamiche acquisite su RrgB suggeriscono che la flessibilità conformazionale del dominio D1 sia fondamentale per i processi di riconoscimento antigene-anticorpo.

Tutte le informazioni ottenute durante questi anni hanno permesso di elucidare alcuni aspetti riguardanti le funzioni biologiche di alcuni antigeni. Tutti i dati ottenuti potranno essere utilizzati in futuro per il design razionale di nuovi antigeni allo scopo di creare proteine con maggiore immunogenicità e stabilità in modo da ottimizzare la produzione industriale su larga scala. Questo lavoro di ricerca fornisce, quindi, un contributo nell'ambito della Structural Vaccinology, che sfrutta le tecniche di biologia strutturale per incrementare, tramite ingegneria biomolecolare, l'efficacia e la stabilità dei vaccini attualmente disponibili sul mercato o in fase di sviluppo clinico.

CHAPTER 1

Introduction

1.1. Historical background

Despite advances in the treatment of infectious diseases, pathogenic microorganisms are still the most important threat to health worldwide. Remarkable progresses have been made in vaccine development in the last 200 years and vaccination has prevented illness and death for millions of individuals. However, efficacious treatments against many infectious disease agents or reemerging infections still need to be discovered. For these reasons, novel vaccines together with new ways to discover and approach them are needed. The oldest vaccines currently available are based on killed or live-attenuated microorganisms and on toxins detoxified by chemical treatments. The knowledge of the pathogenesis of many microorganisms, the identification of the main virulence factors, and the characterization of the immune response after infection have been fundamental for the design of new vaccines based on highly purified antigenic components, on genetically detoxified toxins, and on polysaccharides or oligosaccharides conjugated to proteins (1).

The first important innovation in the vaccine field was the introduction of molecular biology and genetic techniques which allowed the development of two efficacious recombinant vaccines: the hepatitis B vaccine, which is based on a highly purified capsid protein (2), and the vaccine against *Bordetella pertussis*, based on three highly purified proteins, including a genetically detoxified toxin (3, 4). A second revolution in vaccine design started with the genomic era. The “shotgun sequencing” strategy was initially developed and applied to several genome projects allowing in 1995 the completion of the first sequence for a free-living organism (*Haemophilus influenzae*) by The Institute for Genomic Research (TIGR) (5). Nowadays, new sequencing technologies are emerging (6). Thanks to these technical advances, in the last few years the number of available genomes has grown considerably with 725 bacterial sequences completed, and more than 2000 other microorganisms being sequenced in various laboratories around the world (GOLD Genomes OnLine database at http://www.genomesonline.org/). This panel of bacterial genomes already covers most of the pathogens impacting heavily on human health. By the vaccinology point of view, the complete genome of a bacterium represents a large reservoir of genes encoding for potential antigens that can be selected, screened and tested as vaccine candidates. Therefore, potentially surface-exposed proteins can be identified in a reverse manner, starting from the genome rather than from the microorganism. This novel approach has been termed “reverse vaccinology” (7) (Fig. 1). In addition, the availability of complete annotated genome sequences, coupled with bioinformatics, has spawned the new scientific discipline of “comparative genomics” that can be applied to different strains, species and genera and represents a powerful approach for studying differences in phenotype, host range, and molecular evolution of virulence.

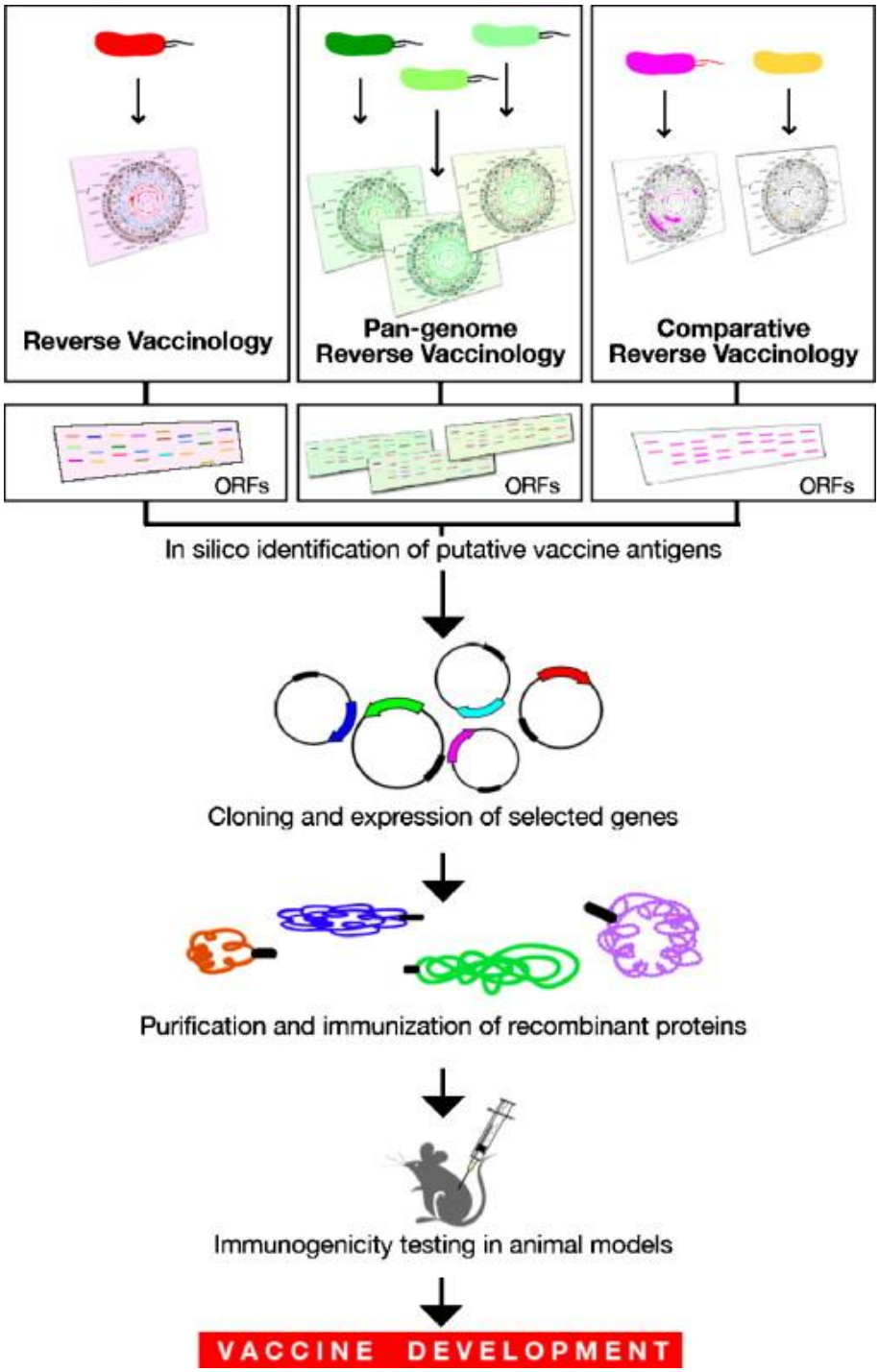


Figure 1. Reverse vaccinology approaches. The reverse vaccinology approach consists in mining a genome sequence of a single bacterial isolates. The pan-genome reverse vaccinology approach represents an advance, since it highlights the potential of looking at more than one genome for the same bacterial species (i.e. sequencing more than one isolates) to overcome the problems represented by gene presence and variability. The comparative reverse vaccinology approach consists in comparing pathogenic (pink bacteria) and non-pathogenic (yellow bacteria) genome sequences in order to select those antigens that are specific for pathogens and that might be involved in pathogenesis. Once potential vaccine candidates are identified *in silico*, the selected ORFs are amplified, cloned in expression vectors, purified and used to immunize mice. The immune sera are used in *in vitro* assays to measure the ability of specific antibodies to kill the pathogen (e.g. the bactericidal assay, which was used for MenB); the immunized animals can be also used for challenging experiments with a lethal dose of the pathogen to measure the protection rate. The best candidates selected by this screening process will enter the final vaccine development step.

1.2. The application of Reverse Vaccinology*: Neisseria meningitidis*

The concept of reverse vaccinology was applied for the first time to serogroup B *Neisseria meningitidis* (MenB). *N. meningitidis* is the major cause of meningitis and sepsis, two devastating diseases that can kill children and young adults within hours, despite the availability of effective antibiotics. It is a Gram-negative bacterium that colonizes asymptomatically the upper nasopharinx tract of about 5–15% of human population. However, in a significant number of cases, the bacterium can traverse the epithelium and reach the bloodstream causing septicemia. From the blood meningococcus is able to cross the blood brain barrier and infect the meninges, causing meningitis (8)(Fig. 2).

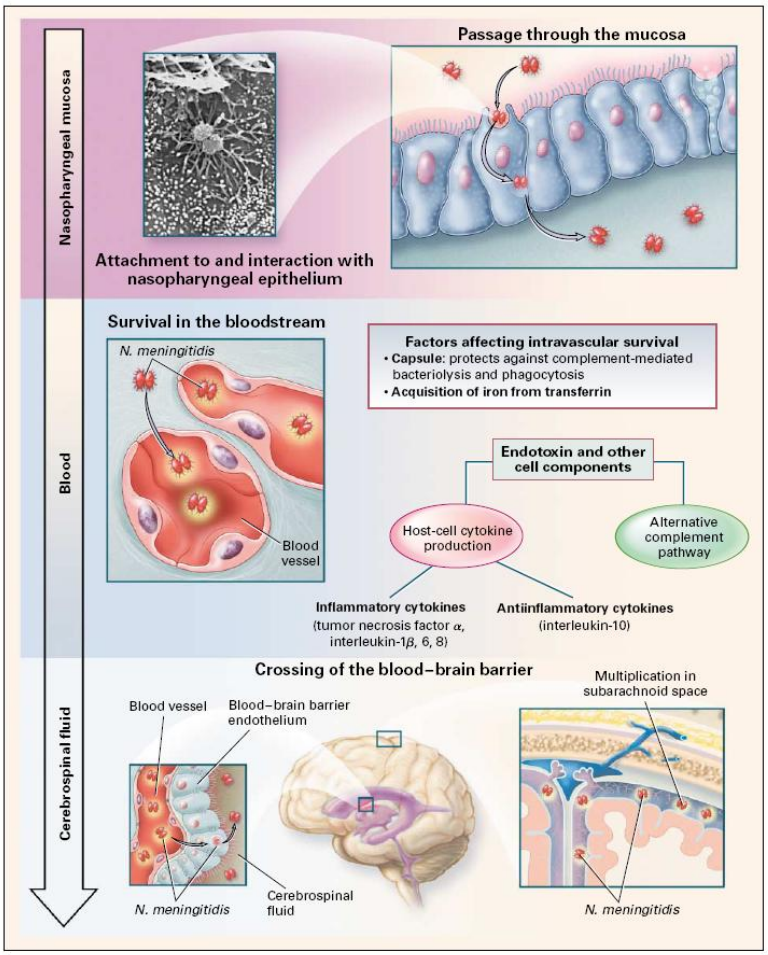


Figure 2. Colonization of nasopharinx by *N. meningitidis* and entry into the bloodstream and cerebrospinal fluid.

N. meningitidis can be classified in 13 serogroups on the basis of the chemical composition of the capsule polysaccharide. However, more than 95% of total cases of invasive disease are caused by five major serogroups: A, B, C, Y and W135. Vaccines against serogroups A, C, Y and W135 were developed in the 1960s by using the purified capsular polysaccharide as antigen. Second generation, conjugated vaccines are now being introduced. The chemical composition of the polysaccharide of serogroup B, which resembles a molecule present in human tissues, makes a polysaccharide-based vaccine poorly immunogenic and a possible cause of auto-immunity. In the last 40 years a lot of efforts have been directed to the identification of meningococcus B antigens as basis of new vaccines. However, the high variability of these proteins among the different MenB strains represents a

serious obstacle to the production of a globally effective anti-MenB vaccine (9). As a consequence there are no effective vaccines available for the prevention of MenB disease, which is responsible of one third of meningococcal disease in the United States of America, and up to 80% of cases in Europe. To identify novel vaccine candidates against MenB, it was applied the reverse vaccinology approach based on the whole-genome sequencing of a serogroup B strain (10)(Fig. 3).

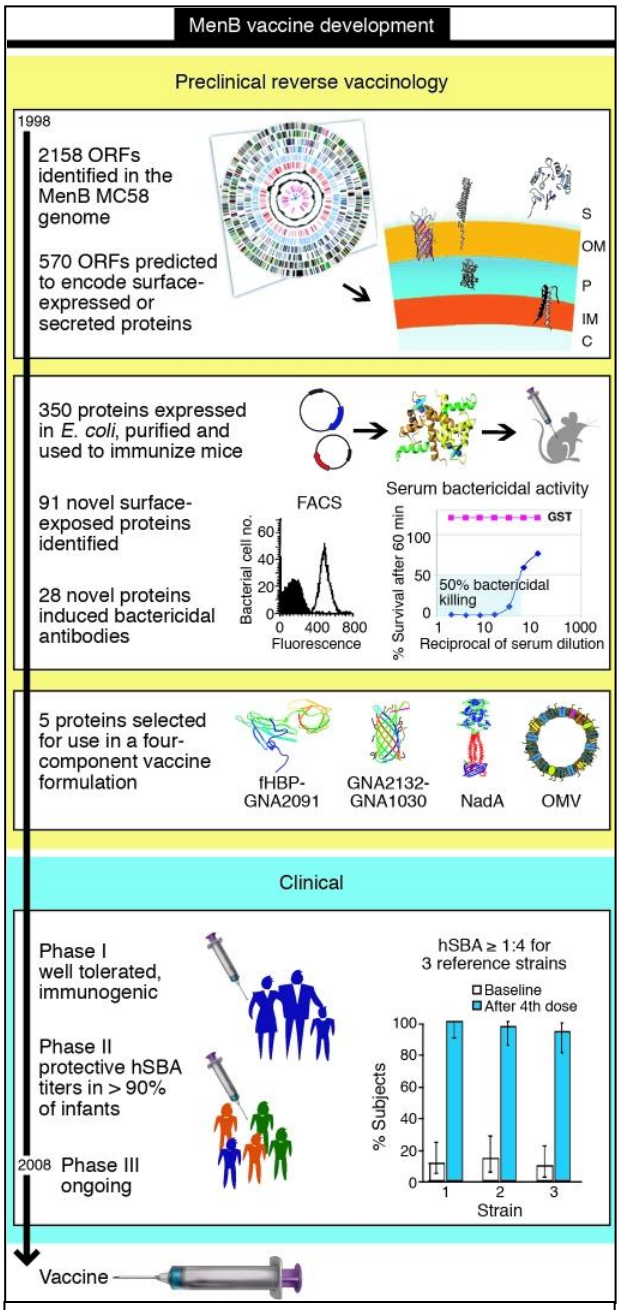


Figure 3. Flow chart of MenB vaccine development.

Briefly, the sequence of the virulent MC58 strain was determined by the shotgun strategy (11). Based on the concept that surface-exposed antigens are susceptible to antibody recognition and are therefore the most suitable vaccine candidates, the complete MenB genome was screened using bioinformatic algorithms in order to select open reading frames (ORFs) coding for putative surface-exposed or secreted proteins. Genome mining allowed the prediction of approximately 600 potential antigens, 350 of which were successfully expressed in *Escherichia coli*, purified and used to immunize mice. The antisera obtained were tested by FACS and ELISA to evaluate the surface localization of the antigens. In addition, the antisera were tested for bactericidal antibodies (BCA), a property known to correlate with protection in humans. Ninety-one novel surface-exposed antigens were identified, 28 of which induced bactericidal antibodies. Several of the antigens previously identified using conventional approaches showed strain variability or were only expressed in some strains, and most of them were effective only against the homologous strains. Therefore, the potential vaccine candidates identified were evaluated for degree of sequence variability among multiple isolates and serogroups of *N. meningitidis*. The vaccine antigens selected by reverse vaccinology were prioritized based on their ability to induce broad strain coverage as inferred by bactericidal activity or observed in passive protection in the infant rat model.

Five antigens selected by this approach were combined into one multi-component vaccine, to increase the breadth of the vaccine coverage and avoid selection of escape mutants. Moreover, to facilitate large-scale production and to further increase their immunogenicity, four of the five antigens were expressed as

fusion proteins, GNA2132 with GNA1030, GNA2091 with GNA1870. Neisseria Adhesin A (NadA) was included as single antigen.

In order to properly test whether the MenB vaccine formulation is able to induce protection against diverse MenB strains, a very large collection of clinical isolates representing the MenB global population diversity has been assembled and characterized and 85 strains selected for the bactericidal assay. The vaccine formulated with alum hydroxide induces bactericidal antibodies in mice against 78% of the 85 genetically diverse strains; coverage can be increased to 90% using different adjuvants (12). The MenB vaccine combined with OMV from the New Zealand strain have been tested in human clinical trials with very promising results, suggesting that a vaccine able to prevent meningococcus B can become a reality. While the vaccine is in clinical development, research efforts are focused on the functional and immunological characterization of the main vaccine antigens.

Neisseria Adhesin A belongs to the Oca family (Oligomeric coiled-coil adhesins) a group of autotransporters adhesins characterized by the presence of a coiled-coil motif that possibly mediates the oligomerization process. Consistently to the predicted structural features, NadA forms high-molecular weight oligomers, stable in denaturing conditions. The role of NadA in mediating the binding to human epithelial cells was initially reported by using the recombinant purified protein (13).
Further characterization revealed that *E. coli* strains expressing the oligomeric form of NadA are able to interact with Chang conjunctiva epithelial cells. The use of a meningococcus *nadA* isogenic knockout mutant confirmed the role of this protein in the intimate association of the bacteria with host cells. Interestingly, the NadA gene is not ubiquitous in Neisseria strains, being present only in 50% of the strains analyzed, but it is always present in strains of hyper-virulent clonal complexes ET5, ET37 and A4, whereas it is absent from Lineage III, in *N. gonorrhoeae, N. cinerea* and *N. lactamica*. The gene clusters in three well conserved alleles, whose overall identity ranges from 96% to 99%; bactericidal activity is not influenced by allele diversity (13). NadA is recognized by children convalescent sera suggesting that it is expressed *in vivo* and immunogenic (14). These evidences support the hypothesis that NadA is an important virulence factor in the pathogenesis of *Neisseria* highlighting the additional role that NadA antibodies could have in preventing meningococcus colonization.

GNA1870 (also named fHBP, factor H Binding Protein or LP2086) is a 26.9 kDa lipoprotein exposed on the surface of meningococcus. The protein is expressed at different levels in the different strains that have been therefore classified as high, intermediate and low expressors. Sequencing of the gene in several strains, representative of the genetic and geographic diversity of the *N. meningitidis* population, showed that the protein is present as three variants. Conservation within each variant ranges between 91.6% and 100%, while conservation among the three variants can be as low as 62.8%. Bactericidal activity is variant-specific and antibodies against a recombinant form of the protein elicit complement-mediated killing of strains that carry the same variant, whereas the activity against strains carrying a different variant is very low or absent (15). Immunological characterization indicated that bactericidal antibodies were directed against conformational epitopes mainly located in the C-terminal domain of the protein, including the residues 100–255 (16). Nuclear Magnetic Resonance (NMR) spectroscopy was applied to solve its solution structure. The protein is composed of two independent barrels connected by a short linker (17). Sequence alignment of the three representative variants together with the analysis of the GNA1870 structure shows that, as opposed to integral outer membrane proteins, regions of extensive variability of the protein are not associated

with mobile loops but instead are spread on the whole molecule, suggesting that the entire protein is exposed to immune recognition. GNA1870 is able to bind the human complement factor H (fH) and for this reason has been renamed fHbp. Binding to fH allows the bacterium to escape the alternative pathway of complement activation and survive and grow in human blood. Level of fH binding correlates with the expression level of fHbp and when factor H is bound complement activation is reduced, enabling bacterial survival. Antibodies to GNA1870 can induce bacterial killing through a dual mechanism: directly, by activation of the classical complement pathway and indirectly by diverting fH away from the bacterial surface (18).

GNA2132 protein is predicted to be a novel surface-exposed lipoprotein. Serum antibody from mice immunized with recombinant GNA2132 are able to bind on the surface of diverse *N. meningitidis* strains and elicited complement-mediated bactericidal activity (10, 15). Moreover, anti-GNA2132 antibody elicited deposition of human C3b on the bacterial surface and passively protected infant rats against meningococcal bacteremia (19). GNA2132 is thus a promising vaccine candidate for prevention of meningococcal disease.

1.3. A pilus-based vaccine against *Streptococcus pneumoniae*

The reverse vaccinology approach was also applied to *Streptococcus pneumoniae* to identify vaccines candidates. *S. pneumoniae* is one of the most important human pathogens, accounting for considerable morbidity and mortality rates worldwide. Pneumococcus is the most common etiologic agent of community-acquired pneumonia, as well as bacterial meningitis, otitis media, and sepsis. It is a Gram-positive bacterium that colonizes the nasopharynx of children and healthy adults. *S. pneumoniae* produces a range of different colonization and virulence factors including the polysaccharide capsule, surface proteins and enzymes, and the pneumolysin toxin, that allow the bacterium to escape the host immune system and cause disease (Fig. 4).

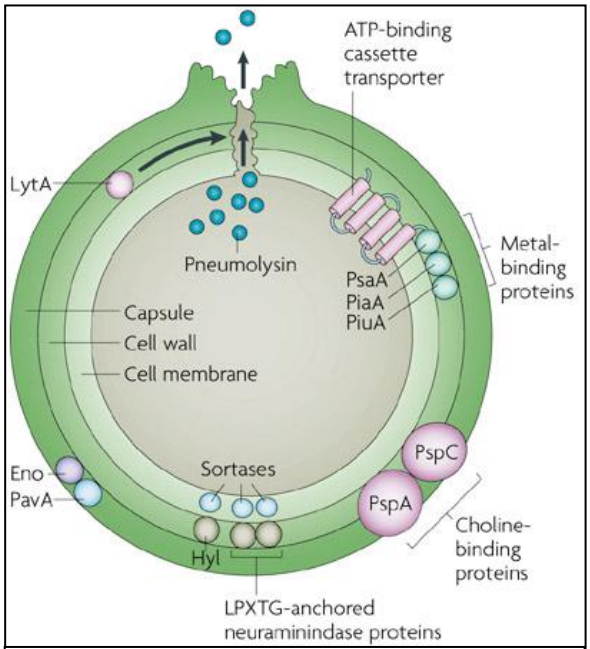


Figure 4. *Streptococcus pneumoniae* virulence is multi-faceted.

The polysaccharide capsule is probably the most important pneumococcal virulence factor, as it protects the bacteria from phagocytosis; it is crucial for colonization, prevents mechanical removal by mucus and can also restrict autolysis and reduce exposure to antibiotics. Capsule polysaccharides are highly heterogeneous and, thus far, at least 93 different capsular serotypes have been described, according to their unique serological profiles and chemical structures. Conjugated polysaccharide-based vaccines are available and extremely effective against serotypes included in the formulation (20). However, the vaccine was primarily developed for US and Europe, whereas in developing countries where non-vaccine types cause most of the disease, the coverage is low. Furthermore, few years after the introduction of Prevenar glycoconjugated vaccine,

emergence of non-vaccine serotypes has been reported, thus posing some concern on the long-term effectiveness of these types of vaccines (21). To broaden the protection, the development of a protein-based pneumococcal vaccine could be a feasible and preferable alternative. However, so far the identification of protein antigens suitable for the formulation of a universal pneumococcal vaccine has remained elusive. In order to fill this gap “reverse vaccinology” were used to identify all possible surface-exposed and immunogenic molecules encoded by *S. pneumoniae*. The whole genome sequence was scanned and 130 potential ORFs with significant homology to surface proteins and virulence factors of other bacteria were selected (22). From this set, 108 ORFs were expressed and purified for mice immunization and evaluated as vaccine candidates. A subset of six novel antigens was able to induce protective antibodies against pneumococcal challenge in a mouse sepsis model. Furthermore, the six predictive targets showed a high degree of cross-reactivity against the majority of capsular antigens that are expressed *in vivo* and are immunogenic during human infection.

Interestingly, this methodology resulted in the identification of a genomic locus in the pathogenic strain TIGR4 that possesses all the typical features of a pilus-encoding island. This locus, named *rlrA* pathogenicity islet (or pilus 1, PI-1) contains seven genes: *rlrA*, encoding a RofA-like transcriptional regulator, *rrgA*, *rrgB* and *rrgC*, which encode the LPxTG cell-wall anchored pilin subunits, and *srtC-1*, *srtC-2* and *srtC-3* coding for the three sortases involved in the linkage and assembly of the pilus structure on the bacterial surface.

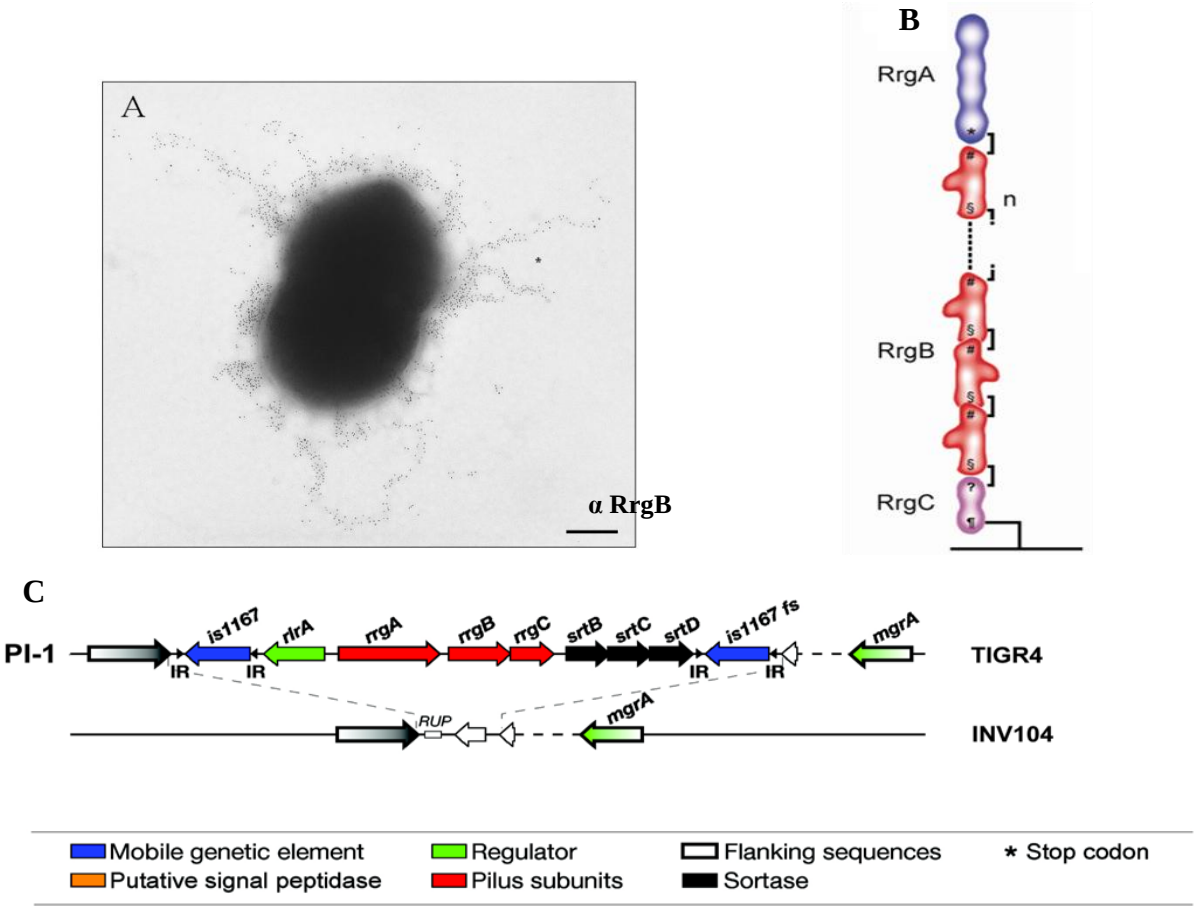


Figure 5: Pilus 1. A) Immunogold localization of RrgB in pilus-1 of *S. pneumoniae* TIGR4 using α-RrgB antibobies. B) Schematic *representation* of pilus-1 subunits arrangement. C) Schematic representation of PI-1 islet.

It has been demonstrated that the major pilin subunit, forming the backbone of the pilus, is RrgB, while RrgA and RrgC are the two ancillary pilins, probably localized at the opposite extremes of the pilus; RrgA is the pilus adhesin, while the role of RrgC is still under investigation (23, 24) (Fig. 5).

The deletion of all three pilus-associated sortase genes, *srtC-1*, *srtC-2* and *srtC-3*, completely prevent pilus biogenesis, and expression of SrtC-1 alone is sufficient to covalently associate RrgB subunits to one another as well as linking the RrgA adhesion and RrgC to the polymer (25). In particular, Falker's studies suggested that, SrtC-1 and SrtC-2 act as redundant pilus subunit polymerases, with SrtC-1 processing all three pilus subunits proteins, while SrtC-2 only RrgA and RrgB. In contrast, SrtC-3 seems to have no pilus polymerase activity, but appears to be required for wild type focal presentation of the pili on the bacterial surface.

Antibodies raised against the three Rrg subunits were tested in immuno-electron microscopy experiments allowing for the first time the visualization of filamentous structures on the surface of the pneumococcus (23). Pneumococcal pili appear as long, thin filaments, protruding from the surface of the bacterium in a hair-like fashion (Fig. 5A). These peculiar features of being abundant and extraordinarily well exposed onto the cell wall and capsular polysaccharide of the bacterium – much more than any other surface-exposed molecule –make pilus subunits exceptional candidates for vaccine development. As expected, both active and passive immunization studies confirmed that the three pilus subunits are able to confer highly significant protection against lethal challenge with *S. pneumoniae* TIGR4 strain (26).
From a functional viewpoint, pneumococcal PI-1 pili were shown to be involved in mucosal colonization and virulence in animal models, as well as to play a role in the host inflammatory response (23). However, a major limitation of pilus subunits as vaccine candidates is represented by the fact that the *rlrA* islet is not widespread in pneumococcal isolates. In order to identify additional pilus types in pneumococcus that could be used to complement pilus 1 subunits for the development of a universal pilus-based vaccine, the reverse vaccinology approach was extended to the whole collection of publicly available *S. pneumoniae* genomes. This global genomic approach eventually led to the identification of a second pilus island (pilus islet 2, PI-2) in a serotype 1 strain (27). Interestingly, as revealed by genome analyses, some strains contain both pilus islet 1 and pilus islet 2, and in these strains both pili are surface exposed and independently assembled. Recent data provide evidence that *in vitro* the pilus encoded by PI-2 of *S. pneumoniae* is involved in adherence to lung alveolar cells, while *in vivo* the PI-2 subunits elicit protection against challenge with PI-2 positive strains (Ruggiero et al., personal communication). Since pneumococcal PI-1 covers between 33% and 50% (in the antibiotic resistant subset) of the strains, and *PI-2* is contained in approximately 16% of the strains, the combination of the two pili antigens could cover about 49–66% of currently circulating pneumococcal isolates.

Although additional widely conserved vaccine candidates still need to be identified to offer a universal protection against pneumococcal infections, the role of pili in pneumococcal pathogenesis, in addition to strain coverage, represents an eligible criterion for these antigens to be considered as valuable components of the next generation of protein-based vaccines.

1.4. The need for “Structural Vaccinology”

The use of native antigens has proven remarkably successful against a wide range of pathogens, but not against all agents for which vaccines are needed. The pathogens for which we have successful vaccines share certain characteristics. In general they are agents against which humoral immunity is protective and their surface antigens do not vary more quickly than new variants can be incorporated into updated vaccines. Examples include diphtheria, poliovirus and hepatitis B virus, which have not undergone significant antigenic change over many decades. Despite great progress, we still face some very tough pathogens that so far have eluded vaccine design efforts. These pathogens usually have the common property of showing to the immune system antigens of high variability, which allows escape from immune recognition. A typical example is HIV, which changes its antigenic surface daily, making it impossible to design a conventional vaccine. These pathogens usually have two types of surface epitopes: those that are highly visible by the immune system, conventionally named *immunodominant epitopes* (which have the property of being highly variable), and those that do not mutate and are common to all strains. The latter are usually completely invisible or poorly visible to the immune system, and the immunity against them cannot be elicited by conventional vaccinology (28). Moreover, for some complex pathogens, in the impoverished settings where new vaccines are urgently needed, the cost of production and distribution of vaccines is a major barrier to use. Even in the developed world, where expense is less of an issue, the number of immunizations a child receives is becoming increasingly burdensome. Optimizing vaccine antigens by combining the tools of genetic engineering with the insights provided by high resolution structural analysis may allow us to overcome these barriers.

Structural biology is increasingly being applied to vaccine development focusing on determining and understanding the structural basis of immunodominant and immunosilent antigens, to enable the rational design of peptide mimetics of bactericidal epitopes (29, 30). This new vaccinology field is named “Structural Vaccinology”. The explosion of genome and proteome data, as well as improved protein expression, purification, and structural determination technologies, has led to the rapid development of the field of structural vaccinology (31)(Fig. 6). Approximately 30,000 high-resolution protein structures are available in public databases (predominantly for soluble proteins), and several initiatives have been established to pursue high-throughput characterization of protein structures on a genome-wide scale (32). The structure-based design of antiviral therapeutics is a well-established approach that has led to the development of drugs directed toward the active sites of the HIV-1 protease (33) and influenza neuraminidase (34).

One important goal of structural vaccinology is to selectively present the conserved determinants of complex and variable antigens. Distinguishing structural components that elicit protective and disease enhancing immunity could avoid vaccine-mediated exacerbation of disease. Using high resolution structures, antigens can be designed to be more efficiently produced and stably stored than native molecules, thereby lowering costs and eliminating barriers to distribution. By engineering antigens amenable for use in combination vaccines, immunization regimens can be simplified. By knowing which parts of an antigen must be retained to preserve basic characteristics and which can be altered, vaccine antigens can be modified more rapidly in response to changing epidemiology.

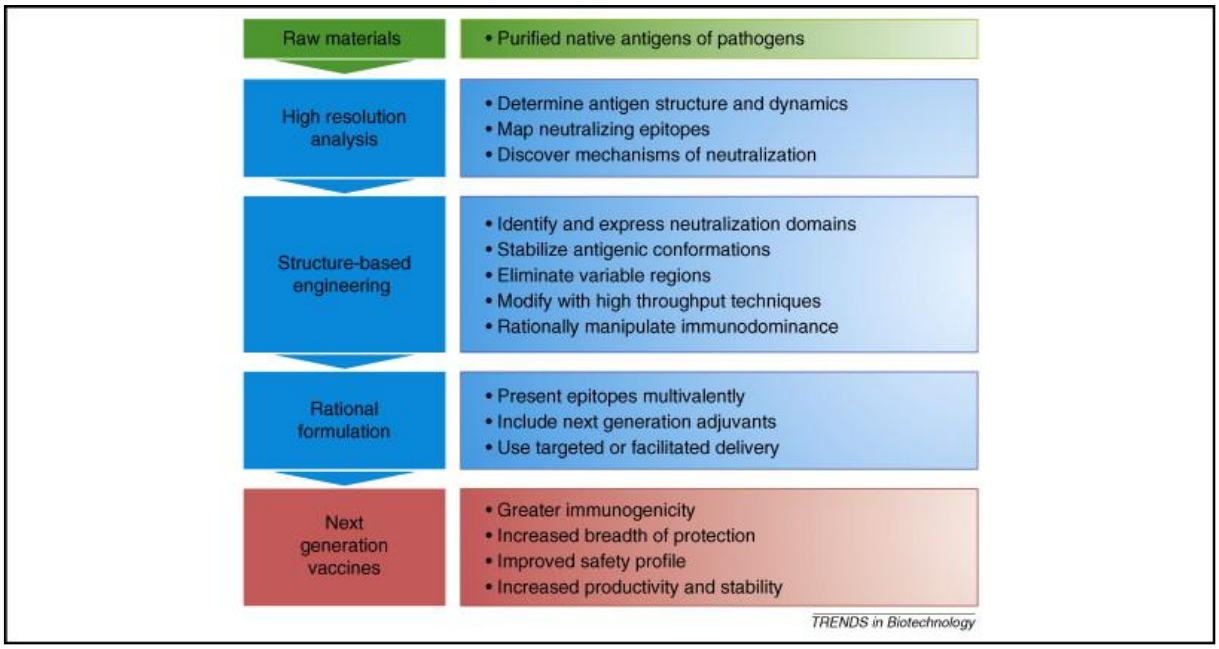


Figura 6. Flow chart of structural vaccinology approach

The biophysical, biochemical, and genetic engineering technologies needed for structural vaccinology are sophisticated, but the concept is simple: first, study the native molecular architecture of neutralization determinants of antigens by NMR spectroscopy and X-ray crystallography; and then, use this knowledge to modify the molecules and engineer immunogens that are optimally designed for inclusion in protective vaccines. Success of structural vaccinology will require the application of high throughput methods to generate a large number of antigenic structures and assays to predict their effectiveness as immunogens. With current knowledge, it is already possible to optimize protein antigens for efficient production, conformational homogeneity, and stable storage. However, fully realizing the potential of structure based design of vaccine antigens will require longer-term basic research into the structural determinants of immunodominance and immunogenicity. This research must take into account the diversity of the antibody repertoire generated not only in a single individual, but also in those of different genetic backgrounds and histories of antigen exposure (30). Moreover, matching the rational design of antigens to advances in the rational design of adjuvants can enable a new generation of vaccines that are more safe, broadly protective, practical, and affordable than those available today.

An example of structural vaccinology approach was the resolution structure of the fHbp, one of the MenB antigens identified by reverse vaccinology to elicit protective human immune responses. The structure of fHbp has enabled the identification of epitopes recognized by bactericidal fHbp-specific antibodies raised against fHBP variants, and these epitopes have been used as the basis for engineering a chimeric fHbp vaccine candidate (30, 35).

1.5. Aim of research

In the “Structural Vaccinology” contest the objective of the PhD project was the functional and structural characterization of surface proteins from pathogenic bacteria for the improvement of vaccines against these agents. During the PhD the

work was focused on two antigens from different pathogens: factor H binding protein (fHbp) from *N. Meningitidis* and RrgB from *S. Pneumoniae*.

In the first case I focused my research on the study of the interaction between fHbp and factor H by using NMR spectroscopy in order to identify key residues engaged in the interaction. Moreover, I pointed the attention also on a new potential biological function of this antigen; in particular, I investigated about a possible role of the protein in the iron recruitments from the environments during the pathogenesis. Based on the work already published by Scarselli M. about the rational design of chimeric fHbp, all the data collected could be used in order to engineer new vaccine candidate.

My research focused, on another antigen from *S.pneumoniae*, RrgB. The work proceeded with the solution structural characterization by NMR spectroscopy of the RrgB D1 domain. The solution structure of the D1 domain could improve the knowledge about the distribution of protective epitopes on the protein surface in order to understand the molecular basis for the immunogenity of the protein. In addition, the new acquired structural information on the RrgB molecule pave the way to future rational design attempts to generate new RrgB-based molecules to implement a protein-based vaccine against pneumococcal disease.

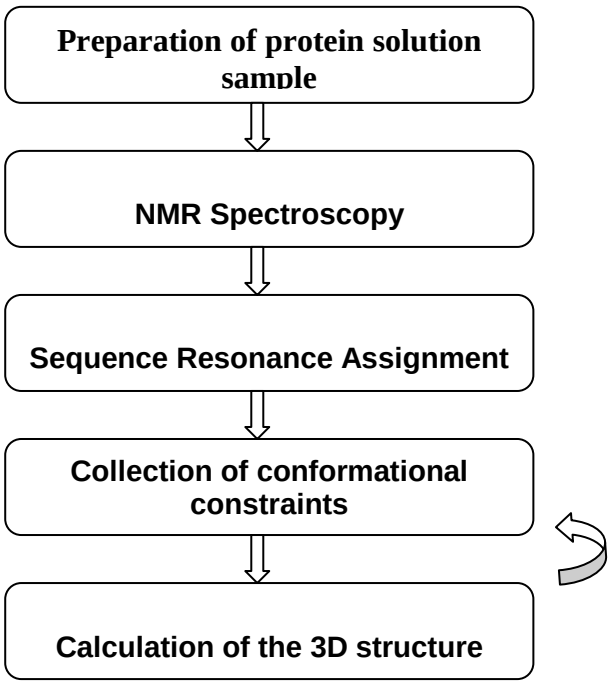
CHAPTER 2

Methodologies Aspects

2.1 NMR and high-resolution structure determination

X-ray crystallography and NMR spectroscopy are the two main techniques that can provide structures of macromolecules at atomic resolution. Both techniques are well established and play already a key role in structural biology as a basis for a detailed understanding of molecular functions. Whereas X-ray crystallography requires single crystals, NMR measurements are carried out in solution under conditions that can be as close as possible to the physiological state. Furthermore, the analysis through NMR spectroscopy easily allows the characterization under several, different experimental conditions, such as different ionic strength and pH. NMR measurements not only provide structural data but also can supply information on the internal mobility of proteins on various time scales, on protein folding and on intra-, as well as, intermolecular interactions.

The principles of NMR structure determination can be summarized as follows:



The power of NMR technique respect the other spectroscopic techniques, results from the fact that each NMR active nucleus gives rise to an individual signal in the spectrum that can be resolved by multi-dimensional NMR techniques. This becomes more difficult for larger molecular structures (more than 50,000 Da) and puts a practical limit to the molecular size that can be studied in detail by NMR.

2.2. Preparation of the protein solution sample

The first step to solve the three dimensional structure of biological macromolecules is the preparation of the protein solution, since a highly purified protein preparation is required. An inhomogeneous preparation and/or aggregation of the protein as well as low molecular weight protonated impurities may severely hard the structure determination. The first step in every protein NMR study therefore involves optimization of the measurement conditions as pH, ionic strength, and temperature that can often be adjusted to mimic physiological conditions. The macromolecule under study should be stable in the chosen conditions for many

weeks. Proteins with a molecular weight larger than 10 kDa must be isotope enriched in ^{15}N and ^{13}C for an efficient structure determination. ^{15}N and ^{13}C are used because the most abundant carbon isotope (^{12}C) does not give a NMR signal and the most abundant nitrogen isotope (^{14}N) has undesired NMR properties.

2.3. NMR spectroscopy

NMR spectra of biological macromolecules contain hundreds or even thousands of resonance lines which cannot be resolved in a conventional one-dimensional spectrum (1D). Further, the interpretation of NMR data requires correlations between different nuclei, which are implicitly contained in 1D spectrum but often difficult to extract. Multidimensional NMR spectra provide both, increased resolution and correlations which are easy to analyse.

The crucial step in increasing the dimensionality of NMR experiments lies in the extension from one to two dimensions. A higher dimensional NMR experiment consists of a combination of two-dimensional (2D) experiments. All 2D NMR experiments use the same basic scheme which consists of the four following, consecutive time periods:

excitation - evolution - mixing - detection

During the excitation period the spins are prepared in the desired state from which the chemical shifts of the individual nuclei are observed during the evolution period t1. In the mixing period the spins are correlated with each other and the information on the chemical shift of one nucleus ends up on another nucleus of which the frequency is measured during the detection period t2. A resonance in the 2D spectrum, such as a cross peak, represents a pair of nuclei that suitably interact during the mixing time.

The extension from a 2D to a n-dimensional (nD) NMR experiment consists in the combination of (n-1) two-dimensional experiments which contains only one excitation and one detection period but repeats the evolution and mixing times (n-1) times. A typical nD NMR experiment thus follows the scheme:

excitation - (evolution - mixing)$_{n-1}$ - detection

where the bracket repeats (n–1) times. Only during the detection period the signal is physically measured and this period is often referred to as the direct dimension in contrast to the evolution periods which are referred to as indirect dimensions.

The NMR multidimensional measurements almost always use protons (^{1}H) and depending on the isotope labelling, ^{13}C and/or ^{15}N nuclei. A 3D spectrum can, for example, be obtained by correlating the amide groups with the α-carbon nuclei attached to ^{15}N. The chemical shifts of these carbon nuclei are used to spread the resonances from the 2D plane into a third dimension. The sensitivity obtainable with these types of nuclei greatly varies even if the sample is fully isotope labelled with ^{13}C or ^{15}N. The proton offers the best sensitivity and for this reason constitutes the preferred nucleus for detection of the NMR signal. The other nuclei are usually measured during evolution periods of multidimensional NMR experiments and their information is transferred to protons for detection. For small proteins (less than 10kDa), it is not required to label the sample with ^{13}C or ^{15}N. In this case the assignment strategy makes use of a combination of 2D homonuclear ^{1}H NMR experiments such as COSY/TOCSY, and NOESY/ROESY spectra. COSY- and

TOCSY-type experiments, where COSY stands for *CO*rrelation *S*pectroscop*Y* and TOCSY for *To*tal *C*orrelation *S*pectrosco*py*, correlate different nuclei *via J* coupling (36, 37) In proteins which are isotope labelled with ^{15}N and ^{13}C *J* couplings between ^{1}H, ^{15}N and ^{13}C allow *through-bond* correlations across the peptide bond.

Through-space correlations are instead measured via the nuclear Overhauser effect (NOE) and provide the basis for geometric information required to determine the structure of a macromolecule (38). The NMR method for proteins structure determination relies on a dense network of distance constraints derived from NOEs between nearby hydrogen atoms in the protein (39). NOEs connect pairs of hydrogen atoms separated by less than about 5 Å. In contrast to COSY-type experiments the nuclei involved in the NOE correlation can belong to amino acid residues that may be far apart along the protein sequence but close in space. For molecules with a molecular weight of more than 5 kDa the intensity of an NOE is approximately proportional to r^{-6} and to the molecular weight, where *r* is the distance between the two interacting spins. NMR experiments which measure the NOE are often referred to as NOESY experiments where NOESY stands for *NOE S*pectroscop*Y* (40).

2.4. Sequence Resonance Assignment

For a detailed analysis of the information content of NMR spectra nearly complete assignments of signals in the spectra to individual atoms in the molecule are a requirement. The application of multidimensional NMR spectroscopy allowed the development of general strategies for the assignment of signals in proteins. All procedures use the known protein sequence to connect nuclei of amino acid residues which are neighbours in the sequence. As mentioned before, for unlabelled proteins smaller than 10 kDa the combination of the [^{1}H,^{1}H]-COSY or TOCSY, used for the sequential assignment, with the [^{1}H, ^{1}H]-NOESY spectrum allows the assignment of most proton NMR signals to individual protons (39).

For larger proteins extensive signal overlap prevents complete assignments of all ^{1}H signals in proton spectra. This barrier can be overcome with 3D NMR technique and uniformly ^{13}C and ^{15}N labelled proteins. The resonance assignment of single (^{15}N or ^{13}C) labelled proteins using 3D experiments is basically an extension of Wüthrich's strategy which exclusively relies on homonuclear ^{1}H NMR experiments. With these methods, systems with molecular weights up to approximately 30 kDa can be studied. In ^{13}C, ^{15}N-labeled proteins a sequential assignment strategy is based on through-bond correlations across the peptide-bond between sequential amino acids. This procedure circumvents the use of NOESY spectra already in the assignment step. Most of these correlation experiments use the three types of nuclei ^{1}H, ^{15}N, ^{13}C and are referred to as triple resonance experiments.

The 3D triple resonance experiments exclusively correlate the resonances of the peptide backbone (HN(i), N(i), Cα(i), Hα(i), Cα(i-1), CO(i) and CO(i-1)). Figure 1 shows the spin system of the peptide backbone and indicates the size of the coupling constants used for magnetization transfer in double ^{13}C-, ^{15}N-labelled proteins. The 3D experiments used to identify the backbone resonances are, usually, HNCA or HNCACB, HN(CO)CA or HN(CO)CACB, HNCO, HN(CA)CO and HNHA (41). The HNCACB for example, correlates each H-^{15}N group with both the intra- and the neighbouring inter-residue Cα and and Cβ. These four types of connectivity are discriminated using the HN(CO)CACB experiment, in which only the inter-residue HN-Cα and Cβ couplings are observed.

Similar strategy can be used to assign the other resonances in the other triple resonance spectra.

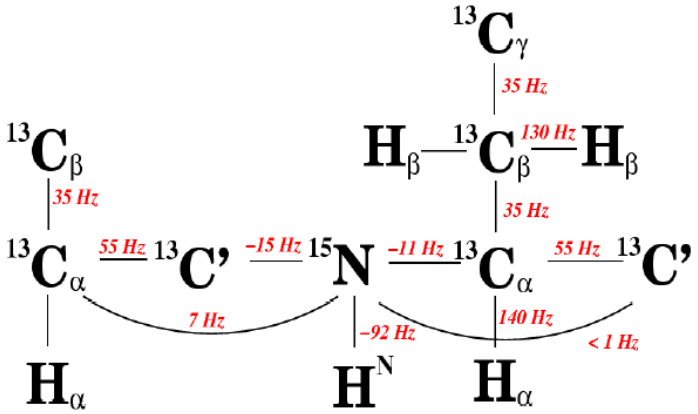


Figure 1. Spin system of the peptide backbone and the size of the ^{1}J and ^{2}J coupling constants that are used for magnetization transfer in ^{13}C-, ^{15}N-labelled proteins.

In the case of proteins with a molecular weight larger than 30 kDa the use of TROSY-type experiments (42) is necessarily. TROSY experiment can reduce the signal loss, which is the direct consequence of the slower correlation tumbling of large molecules which results in faster relaxation and consequently broader lines in the NMR spectrum. TROSY uses constructive interference between different relaxation mechanisms and works best at the highest available magnetic field strengths in the range of 700 to 900 MHz proton resonance frequency. With TROSY the molecular size of proteins accessible for detailed NMR investigations has been extended several folds. The TROSY technique benefits a variety of triple resonance NMR experiments as the 3D HNCA and HNCOCA (43) and the TROSY-based NOESY experiments for the collection of structural constraints are also available (44). CRINEPT-TROSY is another recently developed technique for even larger proteins. It can yield a further significant gain in sensitivity for molecular sizes above 200kDa. The best magnetic field strengths for CRINEPT-TROSY are in the range 900-1000MHz proton resonance frequencies. Since the Hα and Cα/β chemical shifts have been assigned, 3D H(C)CH-TOCSY and (H)CCH-TOCSY (45) experiments are then used to link the side chain spin systems to the backbone assignments. These two experiments provide information for the assignment of the side chain protons and of the side chain carbons, respectively.

A complete set of backbone chemical shifts for all Hα, Cα, Cβ and CO resonances can be used to predict the secondary structure of the protein (46). One technique in particular, the Chemical Shift Index (CSI) (47), has been widely used for the quantitative identification and location of secondary structure in proteins. The method relies on the fact that the chemical shifts of the different nuclei in the protein backbone are related both to the type of amino acid and to the nature of the secondary structure they are located in. By comparing the actual chemical shift for a nucleus in a specific amino acid with a reference value, it is possible to predict in what secondary structure element the nucleus resides. The reference value that you compare with is the random coil chemical shift for that same nucleus in the same amino acid. Another system for the secondary structure prediction is TALOS+ (Torsion Angle Likelihood Obtained from Shift and sequence similarity) which allowed

to predict phi and psi backbone torsion angles using a combination of five kinds (HA, CA, CB, CO, N) of chemical shift assignments for a given protein sequence (48). The method relies on the fact that the phi and psi of the different nuclei in the protein backbone are related both to the type of amino acid and to the nature of the secondary structure they are located in.

2.5. Collection of conformational constraints

For use in structure calculation, geometric conformational information in the form of distances and/or torsion angles has to be derived from the NMR data. The latter have to be supplemented by information about the covalent structure of the protein, such as the amino acid sequence, bond lengths, bond angles, chiralities, and planar groups, as well as by steric repulsion between non-bonded atom pairs. Although a variety of NMR parameters contain structural information, the crucial information comes from NOE measurements which provide distance information between pairs of protons. Supplementary constraints can be derived from through bond correlations in the form of dihedral angles (39). Further, CSI data, provides, as before mentioned, information on the type of secondary structure. Such information can be included in a structure calculation by restricting the local conformation of a residue to the α–helical or β-sheet regions of the Ramachandran plot through torsion angle restraints. Furthermore, hydrogen bonds can be experimentally detected via through-bond interactions (49) and they can be useful during structure calculations of larger proteins when not enough NOE data are available yet. Few years ago a new class of conformational restraints has been introduced that originates from residual dipolar couplings (RDC) in partially aligned or paramagnetic molecules and gives information on angles between covalent bonds and globally defined axes in the molecule, namely those of the magnetic susceptibility tensor (50, 51).

The intensity of a NOE, i.e. the volume *V* of the corresponding cross peak in a NOESY spectrum (17,18)(52), is related to the distance *r* between the two interacting spins by:

$$V=<r^{-6}>f(\tau_c) \qquad (1)$$

NOEs are usually treated as upper inter-atomic distance rather than as precise distance restraints. Adding the information on the distance between pairs of protons and their position in the polypeptide sequence allows construction of possible arrangements in which all distance constraints are fulfilled.

Of particularly utility for the NMR structure determination are scalar three-bond $^3J_{HX}$ coupling constants. The magnitude of these constants is related to dihedral angles subtended by the covalent bonds that connect the coupled nuclei H and X, where X can be ^{1}H, ^{15}N or ^{13}C. According to Karplus (53), the dependence of the three-bond J coupling constant on the dihedral angle θ subtended by the three successive covalent bonds that connect the coupled nuclei is embodied in the relation:

$$^3J(\theta) = A\cos^2(\theta) + B\cos(\theta) + C \qquad (2)$$

where θ = torsion angle (ϕ, φ, χ etc); *A*, *B* and *C* are Karplus constants depending on the type of the torsion angle. $^3J_{HNH\alpha}$ coupling constants (ϕ torsion angle) are obtained from the ratio between the intensity of the diagonal peak and that of the cross-peak of the HNHA map. Through the analysis of the HNHB spectrum the $^3J_{HNH\beta}$ coupling constants (χ torsion angles) can be derived. The process of defining useful

conformational constraints on the basis of J coupling constants usually requires the translation of these quantities into molecular-geometry related dihedral-angle values.

2.6. Structure calculation, refinement and validation

Using NMR constrains, calculation programs fold a random generated 3D structure, in order to maximize the agreement between the structure and the structural constraints. A NMR structure is represented by a family of conformers which are in good agreement with the structural constraints imposed. The precision of the structure is measured by the *root-mean-square-deviation* (RMSD) of the coordinates of the protein atoms for each conformer of the family from the mean structure and the accuracy of the structure is measured by an average *target function* of the family; the latter measures the agreement between the structure and the given set of constraints.

Starting from chemical shifts and NOESY spectra all the calculation works have been done with the software UNIO based on the programs ATNOS-CANDID (54, 55) coupled with CYANA-2.1 (56) by using torsion angle dynamics algorithm. In this algorithm, the molecular dynamics simulation uses torsion angles as degree of freedom, while bond lengths, angles, and backbone peptide plane angles are fixed. In particular, ATNOS/CANDID program, based on the backbone and side-chains assignment, allows obtaining the automated NOESY peak picking (ATNOS) and the automated NOE assignment (CANDID). During the integration time steps, the best conformations were searched by minimized the deviation between the constraints and obtained conformations. Compared with other algorithms, torsion angle dynamics provides at present the most efficient way to calculate NMR structures.

At the end of the structure calculation, an important question is if the structure calculation was “successful”. There are two important features, i) whether the calculated structures fulfil the given restraints, ii) whether the calculations are converged. The first feature could be described by statistical report of the target function. In each calculation, 20 or 30 conformers, among 200 or 350 conformers, with lowest final target functions were chosen to represent the solution structures. The average values of target function of the represent structures should be smaller than 1$Å^2$ and each single violation should be smaller than 0.3 $Å^2$. The second feature was judged by RMSD between these models. Only the family structure with backbone RMSD values close or less than 1 Å will be considered as “good”. The “good” structures obtained from torsion angle dynamics finally will be refined in AMBER program (57). In the final refinement, force field parameters are also considered. The refinement in explicit water box after the refinement *in vacuum* make the structure qualities further improved.

The solved structures could be flawed or present a wrong fold, even if RMSD and target function values are accepted. The defect could be induced by only a few wrong assigned NOEs or other local constraints.

The validation package provides a tool to understand the quality of the structures. In this thesis, the software iCING based on the programs PROCHEK_NMR, WHATIF and QUEEN (58-60) was used for judging the structure quality. The solution structures will be accepted if more than 90% residues fall into the allowed region of Ramachandran plot and less than 1% residues in the disallowed region.

PROCHEK_NMR provides a detailed check on the stereochemistry of a protein structure, while WHATIF try to assess the quality of a structure primarily by

checking whether a number of different parameters are in agreement with their values in databases derived from high-resolution X-ray structures.

QUEEN (*QU*antitative *E*valuation of *E*xperimental *N*mr restraints) is a method which describes the quantitative evaluation of experimental NMR restraints. This method successfully identifies the crucial restraints in a structure determination: those restraints that are both important and unique.

2.7. ^{15}N relaxation in proteins

Protein mobility has an increasing relevance to our understanding of biological and chemical phenomena for the recognized importance of protein internal motions in biological processes. Macromolecular functions are often associated to energetic transition, which are intimately connected with structural changes and molecular flexibility. The measurement of ^{15}N relaxation rates in isotopically enriched proteins is particularly useful for obtaining dynamics information since the relaxation of this nucleus is governed predominantly by the dipolar interaction with directly bound proton and by the chemical shift anisotropy (CSA) mechanism.

In a ^{15}N relaxation experiment, one creates a non-equilibrium spin order and records how this relaxes back to equilibrium. At equilibrium, the ^{15}N magnetization is aligned along the external field, and this alignment can be changed by radio frequency pulses. The magnetization will relax back to equilibrium along the direction of the magnetic filed with a constant time called longitudinal relaxation time T_1. When outside equilibrium the magnetization can have also a component perpendicular to the external magnetic field. The time constant for this spin component to return to equilibrium is called transverse relaxation time, T_2. A third source of relaxation parameter is the heteronuclear NOE. This is measured by saturating the proton (^{1}H) signal and observing changes in the ^{15}N signal intensities.

The relaxation parameters are related to the spectral density function, of the ^{1}H-^{15}N bond vector by the following equations (61, 62):

$$T_1^{-1} = R_1 = (d^2/4)\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + c^2 J(\omega_N) \quad (3)$$

$$T_2^{-1} = R_2 = (d^2/8)\left[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\right] + (c^2/6)\left[4J(0) + 3J(\omega_N)\right] + R_{ex} \quad (4)$$

$$NOE = 1 + (d^2/4R_1)(\gamma_H/\gamma_N)\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right] \quad (5)$$

in which $d=(\mu_0 h\gamma_N\gamma_H/8\pi r^3_{NH})$ and $c=\omega_N\Delta_N/\sqrt{3}$. R_{ex} is a term introduced to account for microsecond to millisecond conformational exchange contributions to R_2.

The dynamic information contained in the relaxation rates are represented by the values of the spectral density function, $J(\omega)$, at several frequencies. The descriptions of the dynamics require therefore the identification of a suitable model

for the spectral density function, which must be consistent with the experimental relaxation rates. The spectral density is often expressed in terms of global tumbling parameters, and of local motional parameters.

2.8. Data Analysis

2.8.1. Spectra Density Mapping

The spectral density mapping approach was developed by Peng and Wagner in 1992 (63). It makes use of six different relaxation parameters, which are used to map the spectral density function. These parameters are: 1) the longitudinal ^{15}N relaxation rate, R_1, 2) the transverse ^{15}N relaxation rate, R_2, 3) the $^{1}H \rightarrow ^{15}N$ heteronuclear NOE, 4) the relaxation rate of longitudinal two-spin order $R_{NH}(2H_zN_z)$, 5) the relaxation rate of anti-phase ^{15}N coherence $R_{NH}(2H_zN_{x,y})$ and 6) longitudinal relaxation of the amide protons. The method of spectral density mapping does not require any assumption regarding the form of the spectral density functions. Values of the spectral density functions of ^{1}H-^{15}N vectors are directly sampled at several relevant frequencies (e.g. 0, ω_H-ω_N, ω_H+ω_N, ω_H, ω_N). Drawbacks in this method reside in the fact that the three measurable relaxation parameters, R_1, R_2 and heteronuclear NOE, are insufficient to determine uniquely the values of the spectral density function at the five frequencies in equations (1) to (3) and that anomalous behaviour can be expected for the spectral densities at the three highest frequencies (63). These problems can be overcome by a more recent approach called reduced spectral density mapping, in which the values of the spectral density function at ω_H-ω_N, ω_H+ω_N and ω_H frequencies can be combined in a average spectral density $<J(\omega_H)>$ (64, 65). A comparison of $J(\omega)$ measured at high and low frequencies provides a quantitative measure of the breadth of the frequency distribution accessed through the spatial fluctuations of the bond and the overall tumbling of the molecule.

2.8.2. Model Free Approach

A different method, called Model Free approach, has been introduced by Lipari and Szabo (66). This model assumes that the overall rotation of the molecule can be described by a single correlation time (isotropic motion) and that this overall motion and the internal motions are independent. Then the total correlation function can be factored as:

$$C(t) = C_O(t) \times C_I(t) \qquad (6)$$

where C(t) is the total correlation function, $C_O(t)$ is the correlation function characterizing the overall rotation and $C_I(t)$ is the correlation function characterizing the internal motions.

As a consequence, the global spectral density function can be expressed as a weighted sum of Lorentzian functions. This is correct rigorously for isotropic rotational diffusion and approximately for anisotropic rotational diffusion.

The Lipari and Szabo formalism employs a minimum number of parameters to describe the overall isotropic tumbling motion of a macromolecule and the internal motions of the ^{15}N-^{1}H bond vector. The central equation in the Model Free Approach is:

$$J(\omega) = \left[\frac{S^2 \tau_m}{1 + (\omega \tau_m)^2} + \frac{(1 - S^2)\tau'}{1 + (\omega \tau')^2} \right] \quad (7)$$

where τ_m is the correlation time as a result of the isotropic tumbling motion of the entire molecule. The effective correlation time resulting from internal motions is described by τ_e, where $\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$. The order parameter S^2 describes the degree of spatial restriction of the internal motion of the ^1H-^{15}N bond vector. It satisfies the inequality $0 \leq S \geq 1$ and lower values indicate larger amplitudes of internal motions. As a consequence, for a nucleus rotating as a whole with the molecule, all contributions to the spectral density function derive from the overall tumbling; alternatively, extra contributions will be described by other motions with correlation times faster than the overall tumbling. An extended form of the model-free spectral density function has been developed by Clore and co-workers (67) to describe internal motions that take place on two distinct time scales, differing by at least an order of magnitude.

2.9. The contribution to relaxation of exchange processes

The presence of exchange processes occurring in the micro-millisecond time scale produces dephasing of magnetization and contributes to make the transverse relaxation time shorter. A method to obtain a detailed analysis of exchange contribution is the measurement of R_2 rates as a function of the refocusing times $\tau_{CPMG}(=1/(2\nu_{CPMG})$, where ν_{CPMG} is the frequency of repetition of 180° pulses during the Carr-Purcell-Meiboom-Gill (CPMG) sequence (68). The contribution of the exchange processes (R_{ex}) to the transverse relaxation rate can be expressed as follow (69):

$$R_{ex} = \frac{k_{ex}}{2} - 2\nu_{eff} \sinh^{-1}\left(\frac{k_{ex}}{\xi} \sinh \frac{\xi}{4\nu_{eff}} \right) \quad (6)$$

where $\xi = (k_{ex}^2 - 4p_A p_B \delta\omega^2)^{1/2}$, $k_{ex} = 1/\tau_{ex}$ and $\nu_{eff}(s^{-1}) = \frac{1}{2(T_\pi + \tau_{CPMG})}$

p_A and p_B are the populations of the sites A and B in a two-site exchange process, $\delta\omega$ is the difference of Larmor frequencies between the sites and τ_{ex} is the time constant for the exchange process.

CHAPTER 3

Structural and Functional Characterization of the *Streptococcus pneumoniae* RrgB Pilus Backbone D1 Domain

Maria Antonietta Gentile[1], Sara Melchiorre[1], Carla Emolo[1], Monica Moschioni[1], Claudia Gianfaldoni[1], Laura Pancotto[1], Ilaria Ferlenghi[1], Maria Scarselli[1], Werner Pansegrau[1], Daniele Veggi[1], Marcello Merola[1,2], Francesca Cantini[3], Paolo Ruggiero[1], Lucia Banci[3], and Vega Masignani[1]

[1]Novartis Vaccines and Diagnostics Research Center, Via Fiorentina 1, Siena 53100, Italy,
[2]Universita` di Napoli Federico II, Monte Sant'Angelo, Via Cintia, Napoli 80126, Italy
[3]Magnetic Resonance Center, Department of Chemistry, University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

3.1. Introduction

Streptococcus pneumoniae is an important human pathogen responsible for diseases such as otitis media, pneumonia, sepsis meningitis (70-75). However, *S. pneumoniae* is also a common inhabitant of the respiratory tract of children and healthy adults. This carriage state could represent a risk factor for the development of respiratory diseases but also the source for pneumococcal transmission to other individuals (76-78). Like most streptococci, *S. pneumoniae* decorates its surface with long filaments known as pili (23, 79-82). Pneumococcal pili have previously been associated with virulence and the capability of the microorganism to adhere better to epithelial cells and to colonize the nasopharynx (23, 24, 83).

The pneumococcal pilus is a multimeric structure consisting of three proteins (RrgA, RrgB, and RrgC) polymerized by three sortase enzymes (SrtC1, SrtC2, and SrtC3) through the formation of covalent intermolecular isopeptide bonds (25, 84-87). In particular, multiple copies of RrgB are polymerized to form the scaffold of the pilus, whereas the major adhesin, RrgA, and the putative anchor, RrgC, are localized at the tip and at the base of the pilus, respectively (24, 88, 89). Recently, the structure of a major portion of RrgB (residues 184–627) was solved at a 1.6 Å resolution (90) and revealed an organization into three independently folded IgG-like domains (D2, D3, and D4, residues 184–326, 326–446, and 446–627, respectively). On the contrary, the structure of the RrgB N-terminal region (D1, residues 1–184), likely constituting a fourth independently folded domain, remained unsolved due to the failure to obtain the crystals of the full-length (FL) RrgB (90). Interestingly, each of the D2, D3, and D4 domains is stabilized by one intramolecular isopeptide bond. These covalent linkages, formed between Lys and Asn residues, have been found in other pilus proteins (86, 91-94) and are thought to play a role similar to that of disulfide bonds;

they confer in fact a rigid molecular architecture to the pili and make them less susceptible to proteolytic cleavage (Fig. 1).

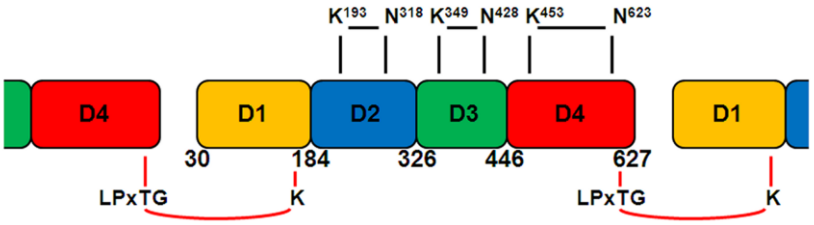


Figure 1. Schematic representation of pilus backbone protein RrgB. Pilus scaffold is composed by multiples copies of RrgB protein in a head-to-tail arrangement. Pilus polymerization occurs through intermolecular isopeptide bonds (*red*), whereas each RrgB protein is stabilized by intramolecular isopeptide bonds (*black*). Lys-183 as a residue involved in the intermolecular bond has been identified in the present work.

In the pilus backbone assembly RrgB molecules are linked together by sortases through intermolecular isopeptide bond formation between a Thr within the C-terminal LP*X*TG motif of a molecule and a Lys located at the N terminus of the following molecule (85, 94, 95) (Fig. 1). In *Corynebacterium diphtheriae*, where the general principles of pilus assembly were first established, the functional Lys is located within a conserved YPKN “pilin” motif (85, 92, 96). Nevertheless, this sequence is not absolutely required for polymerization as demonstrated by studies on the Spy0128 pilin of *Streptococcus pyogenes*, where the lysine forming the intermolecular isopeptide bond and responsible for pilus polymerization is located into ^{159}GSKVPI^{164} motif even though the YPKN pilin motif is also present (94, 97).

RrgB, along with the other two pilus proteins RrgA and RrgC, was previously shown to confer protection in mouse models of infection and therefore is regarded as a potential candidate for a new generation of protein-based vaccines (26, 98). We have investigated the protective ability of the single recombinant D1, D2, D3, and D4 domains of RrgB in a mouse model of sepsis, and here we provide evidence that D1 is the most protective, followed by D4.

Furthermore, we present the solution structure of the recombinant D1 obtained by NMR spectroscopy and show that Lys-183 of D1 is engaged in the intermolecular isopeptide bond formation during pilus polymerization. Finally, we propose a possible model of the entire RrgB molecule and show the positions of two linear epitopes possibly involved in the protection mechanism.

3.2. Material and Methods

Bacterial Strains and Cultures- For the animal experiments, the *S. pneumoniae* TIGR4 (serotype 4) strain was used. Bacteria were grown, frozen in aliquots at −80 °C, and titrated as already reported (26). Immediately prior to challenge, frozen aliquots were thawed and diluted in PBS to reach the working concentration.

Cloning and Protein Expression and Purification- Standard recombinant DNA techniques were used to construct the expression plasmids (pET21b^{+}; Novagen) and to express and purify the recombinant C-terminal His-tagged proteins. The affinity-purified proteins were subsequently used to immunize CD1 mice or rabbits for antibody generation (Charles River Laboratory) and BALB/c mice to evaluate the protective efficacy.

Complementation Plasmids- Wild-type or mutant *rrgB* genes were amplified from chromosomal DNA of TIGR4 strain by PCR. Point mutations were introduced by overlap extension PCR by using specific primers. PCR products were then cloned into the complementation plasmid pMU1328 between the BamHI and SalI restriction sites (26). Expression of RrgB or RrgB mutated forms was under control of the erythromycin constitutive promoter (Pc) which was amplified with the specific and cloned immediately upstream *rrgB* (EcoRI, BamHI). All plasmids were confirmed by sequencing.

Generation of rrgB Deletion Mutants and rrgB Complementants- A TIGR4 Δ*rrgB* isogenic mutant was generated by allelic exchange. Fragments of ~500 bp upstream and downstream the target gene were amplified by and spliced into a kanamycin resistance cassette by using overlap extension PCR; the PCR fragments were then cloned into pGEMt (Promega) and transformed in *S. pneumoniae* with conventional methods (99). To select the bacteria in which the target gene was replaced with the resistance cassette, bacteria were plated on blood-agar plates with kanamycin (500 µg/ml). The presence of the isogenic mutation was confirmed by PCR and Western blot (WB) analysis. To obtain RrgB-complemented mutants, pMU1328 plasmids containing WT *rrgB* or *rrgB* mutated forms were transformed into TIGR4 Δ*rrgB* with conventional methods. Transformants selection was performed by supplementing media with kanamycin (500 µg/ml) and erythromycin (1 µg/ml). The complemented mutants were then analyzed by PCR; expression of FL WT RrgB or RrgB mutants was detected by WB analysis of whole cell lysates.

SDS-PAGE and Western Blot Analysis- SDS-PAGE analysis was performed using NuPAGE TM 4–12% BisTris gradient gels (Invitrogen) according to the manufacturer's instructions. Hi-MarkTM prestained HMW (Invitrogen) served as protein standard. Gels were processed for WB analysis by using standard protocols. Mouse polyclonal antibodies raised against recombinant His-tagged RrgB were used at 1/3000 dilution. Secondary goat anti-mouse IgG alkaline phosphatase-conjugated antibodies (Promega) were used at 1/5000, and signals were developed by using Western Blue Stabilized Substrate for Alkaline Phosphatase (Promega).

Animal Experiments- Animal studies were done in compliance with the current law approved by the local Animal Ethics Committee and authorized by the Italian Ministry of Health. Female, 6-week-old, specific pathogen-free BALB/c mice (Charles River) received three intraperitoneal immunizations, 2 weeks apart. Each dose was composed of 20 µg of either the single RrgB domains or the FL RrgB, or of a combination of the four RrgB domains (D1+D2+D3+D4), 10 µg each, along with 400µg of aluminum hydroxide as an adjuvant, in a final volume of 200 µl of PBS. Control animals received the same course of saline plus adjuvant. Ten days after the third immunization, samples of sera were obtained from each animal and pooled according to the immunization group to be used in passive serum transfer experiment. Two weeks after the third immunization, each mouse was challenged intraperitoneally with a mean dose of 1.6x102 cfu of TIGR4. Bacteremia was evaluated 24 h after challenge, and mortality course monitored for 10 days after challenge as already reported (26). The animals were euthanized when they exhibited defined humane end points that had been pre-established for the study in agreement with Novartis Animal Welfare Policies. For passive protection experiments, 8-week-old mice were used. Fifteen minutes before TIGR4 intraperitoneal challenge (102 cfu/mouse), each mouse received intraperitoneally

50µl of pooled mouse sera against recombinant D1 or D4, or of control sera obtained immunizing with adjuvant plus saline.

Statistical Analysis- Bacteremia and mortality course were analyzed by the Mann-Whitney *U* test. Survival rates were analyzed by Fisher's exact test. One-tailed or two-tailed tests were used to compare immunized groups with the control group or each other, respectively. Values of $p \leq 0.05$ were considered and referred to as significant. Values of $p \leq 0.1$ were referred to as a trend.

Flow Cytometry on Entire Bacteria- TIGR4 were grown in Todd-Hewitt yeast extract broth to an exponential phase (A_{600} = 0.25), fixed with 2% formaldehyde, and then stained with pooled mouse antisera raised against FL RrgB or RrgB domains at 1:400 dilution. Mouse IgG were detected with FITC-conjugated goat anti-mouse IgG (Jackson Laboratories) at 1:100 dilution, and bacterial staining was analyzed by using a FACS-Calibur cytometer (Becton Dickinson). Sera from mice immunized with PBS plus adjuvant were used as negative control.

ELISA- 96-well MaxiSorpTM flat-bottom plates (Nunc) were coated with 0.2 µg/well FL RrgB overnight at 4 °C. Plates were then washed three times with PBS/0.05% Tween 20 and saturated for 1 h at 37 °C with PBS/1% BSA. Following three washing steps with PBS/0.05% Tween 20, the plates were incubated for 2 h at 37 °C with serial dilutions of the pooled mouse sera. After another three washing steps, bound antigen-specific mouse IgGs were revealed with alkaline phosphatase-conjugated goat anti-mouse IgG (Sigma), followed by the phosphatase alkaline substrate *p*-nitrophenyl phosphate (Sigma). The intensity of color was quantified with an ELISA plate reader at A_{405}. The antibody titer was expressed as the log10 of the reciprocal of the serum dilution that gave A_{405} = 1.

PepScan Analysis- Arrayed peptides were synthesized *in situ* on glass fiber membranes. Membranes were conditioned by wetting with ethanol and washing three times for 5 min in TTBS (50 mM Tris-HCl, pH 7.0, 137 mM NaCl, 2.7 mM KCl, 0.05% Tween 20). After overnight blocking at 4 °C in MBS (2% dry milk in TTBS), membranes were incubated for 1.5 h at 37 °C with polyclonal antisera (1:3000 in MBS) followed by secondary goat anti-mouse IgG alkaline phosphatase-conjugated antibodies (1:5000 in MBS; Promega), and signals were developed by using Western Blue Stabilized Substrate for Alkaline Phosphatase. For image processing, membranes were scanned using an Epson V750 Pro scanner at 800 dpi, 48-bit color depth and with gamma 1.0 full linear response.

NMR Characterization of RrgB D1 Domain- NMR spectra were acquired at 298 K on Avance 900, 800, 700, and 500 MHz Bruker spectrometers, all equipped with a triple resonance cryoprobe. The NMR experiments, used for the backbone and the aliphatic side chain resonances assignment recorded on $^{13}C/^{15}N$ and ^{15}N enriched samples or on unlabeled D1 samples, are summarized in Table 1. Backbone dihedral angle constraints were derived from 15N, 13C', 13Cα, 13Cα, and Hα chemical shifts, using TALOS+ (100). Distance constraints for structure determination were obtained from ^{15}N-edited and ^{13}C-edited three-dimensional NOESY-HSQC. 3131 meaningful proton-proton distance restraints (Table 1), with 114 φ and 120 ψ backbone dihedral angles restraints were included in structure calculations.

Experiments [b]	Dimension of acquired data (nucleus)			Spectral width (ppm)			n[a]
	t_1	t_2	t_3	F_1	F_2	F_3	
[^{1}H-^{1}H]-NOESY [c]	1024(^{1}H)	2048(^{1}H)		16	16		32
[^{1}H-^{1}H]-TOCSY [c]	1024(^{1}H)	2048(^{1}H)		16	16		32
^{1}H-^{15}N-HSQC	128(^{15}N)	1024(^{1}H)		36	14		8
^{1}H-^{13}C-HSQC	128(^{13}C)	2048(^{1}H)		80	14		8
trhncocacb [d]	96(^{13}C)	56(^{15}N)	2048(^{1}H)	80	36	14	8
trhncacb [d]	96(^{13}C)	56(^{15}N)	2048(^{1}H)	80	36	14	8
trhnco [d]	96(^{13}C)	56(^{15}N)	1024(^{1}H)	36	36	14	16
trhncaco [d]	96(^{13}C)	56(^{15}N)	1024(^{1}H)	36	36	14	24
trhnca [d]	96(^{13}C)	56(^{15}N)	2048(^{1}H)	42	36	14	8
trhncoca [d]	96(^{13}C)	56(^{15}N)	2048(^{1}H)	42	36	14	16
Hbhaconh	88(^{1}H)	52(^{15}N)	1024(^{1}H)	12	36	12	8
(H)CCH-TOCSY	224(^{13}C)	56(^{13}C)	2048(^{1}H)	14	36	14	16
^{15}N-edited [^{1}H-^{1}H]-NOESY	256(^{1}H)	56(^{15}N)	1024(^{1}H)	14	36	14	16
^{13}C-edited [^{1}H-^{1}H]-NOESY	248(^{1}H)	104(^{13}C)	1024(^{1}H)	14	24	14	8
^{2}J-HSQC [e]	128(^{1}H)	1024(^{1}H)		170	20		48
^{15}N R_1 [f]	160(^{15}N)	1024(^{1}H)		38	14		16
^{15}N R_2 [f]	160(^{15}N)	1024(^{1}H)		38	14		16
^{1}H-^{15}N NOE [f]	320(^{15}N)	1024(^{1}H)		38	14		32

[a] number of acquired scans. [b] All the experiments were acquired on a 500, 700, 800 and 900 MHz spectrometer equipped with a triple resonance cryoprobe at 298 K. All the triple resonance (TXI 5-mm) probes used were equipped with Pulsed Field Gradients along the z-axis. [c] The experiment was acquired in D_2O for the assignment of the aromatic protons, the spectra analysis provided the assignment of the side chain of aromatic residues. [d] The experiments were acquired on ^{13}C/^{15}N enriched D1 sample. [e] Histidine coordination mode was determined through ^{1}H-^{15}N heteronuclear experiments to detect the ^{2}J ^{1}H-^{15}N coupling between the imidazole nitrogens and non exchangeable imidazole protons.

[f] The ^{15}N longitudinal relaxation rates, R_1, were measured using delays in the pulse sequence of 2.5, 35, 75, 125, 200, 350, 500, 750, 1000, 1500, 2000 and 3000 ms for all samples. The ^{15}N transverse relaxation rates, R_2, were measured using the CPMG sequence. The relaxation delays used were 16.96, 33.92, 50.88, 84.80, 118.72, 152.64, 186.56, 220.48, 254.40 and 288.32 ms. The interscan delay (d1) was 3 s for R_1 and R_2 and 5 s for ^{1}H-^{15}N NOE respectively. R_1 and R_2 relaxation rates were obtained by fitting the cross peak volumes (I), measured as a function of the relaxation delay, to a single exponential decay as described in the literature. Heteronuclear NOE values were calculated as the ratio of peak volumes in spectra recorded with and without saturation. In all experiments the water signal was suppressed with the 'water flipback' scheme. All 3D and 2D spectra were processed using the standard Bruker software TOPSPIN and analyzed through CARA and XEASY programs.

Table 1. Acquisition parameters for NMR experiments performed on D1 domain

The exchangeability of the backbone amide hydrogen nuclei with solvent protons was investigated through a ^{1}H-^{15}N HSQC experiment performed on a protein sample dialyzed against deuterated buffer for 3 days. Hydrogen bond constraints for the slowly deuterium-exchanging amide protons of the β-strands were introduced at later steps of the structure calculations. Structure calculations were performed through iterative cycles of CYANA-2.1 (55) followed by restrained energy minimization with the AMBER 10.0 Package in explicit water solvent (57). The quality of the structures was evaluated by using the iCING validation program (for details, see Table 2). The program MOLMOL was subsequently used for structure analysis (64, 101).The final bundle of 20 conformers of D1 has an average target function of 1.36 ± 0.13 (CYANA units). The average backbone r.m.s.d. value (over residues 28–183) is 0.71 ± 0.19 Å, and the all-heavy atoms r.m.s.d. value is 0.96 ± 0.16 Å. Per-residue r.m.s.d. values are shown in Fig. 2.

Heteronuclear Relaxation Data- The dynamic properties of D1 have been characterized experimentally through ^{15}N relaxation measurements. ^{15}N longitudinal and traverse relaxation rates (64) and $^{15}N\{^{1}H\}$ NOEs (102) were recorded at 298 K at 500 MHz, using a protein concentration of 0.8 mM. The average backbone ^{15}N-longitudinal R_1 and transversal R_2 relaxation rates and $^{15}N\{1H\}$ NOE values are 1.45 ± 0.1 s^{-1}, 16.18 ± 1.5 s^{-1}, and 0.71 ± 0.04, respectively (Fig. 3). They are essentially homogeneous along the entire polypeptide sequence, with the exception of residues located at the C- and N-termini and three loop regions (56–69, 148–162, and 173–177). The correlation time for the molecule tumbling (τ_c), as estimated from the R_2/R_1 ratio, is 11.49 ± 1.9 ns, consistent with the molecular mass of the monomeric protein. The relaxation data were analyzed according to the model-free approach of Lipari and Szabo (66, 103) using TENSOR2 (104)(Fig. 3).

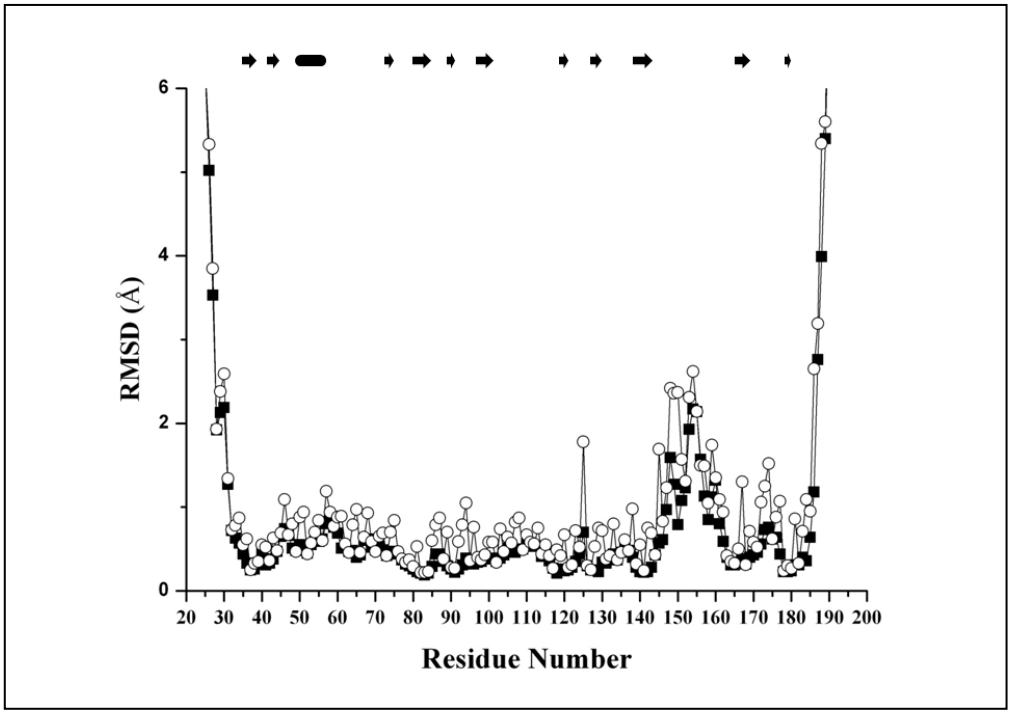


Figure 2. RMSD values per residue to the mean structure for the backbone (filled squares) and all heavy atoms (open circles) of the family of 20 conformers of D1 domain after energy minimization. The secondary structure elements are also reported.

Table 2. Statistical analysis of the energy minimized family of D1 domain conformers.

	RrgB D1 domain (20 Conformers)
Total number of meaningful NOE upper distance constraints	3131
Intra-protein NOEs[b]	
Intra-residue	676
Inter-residue	
Sequential (|i-j| = 1)	801
Medium-range (|i-j| < 4)	450
Long-range (|i-j > 5)	1204
Total meaningful dihedral angle restraints[b]	234
Phi	114
Psi	120
RMS violations per meaningful distance constraint (Å):	
Intraresidue	0.8738±0.3661
Sequential	0.0071±0.2822
Medium range	0.0110±0.0011
Long range	0.0076±0.0008
RMS violations per meaningful dihedral angle constraints (°):	
Phi	5.9314±0.7973
Psi	6.7689±1.6990
Violations	
< 0.1 Å	3
> 0.1 Å	47
> 0.3 Å	0
> 0.5 Å	0
Average RMSD to the mean (Å)	
Residue range 28-183 (backbone atoms)	0.73±0.12
Residue range 28-183 (all heavy atoms)	0.97±0.07
Secondary structure elements (backbone atoms)	0.46±0.11
Secondary structure elements (all heavy atoms)	0.83±0.11
residual CYANA Target Function ($Å^2$)	1.36±0.13
Structural analysis	
% of residues in most favorable regions	83.80%
% of residues in allowed regions	15.30%

% of residues in generously allowed regions	0.60%
% of residues in disallowed regions	0.30%
WHAT IF structure Z-scores[d]	
1st generation packing quality	-3.29
2st generation packing quality	-2.43
Ramachandran plot appearance	-2.85
χ_1/χ_2 rotamer normality	-4.94
Backbone conformation	-1.1
WHAT IF RMS Z-scores[e]	
Bond lengths	1.49
Bond angles	0.82
Omega angle restraints	1.24
Side chain planarity	1.38
Improper dihedral distribution	1.04
Inside/Outside distribution	1.06
Overall NOE completeness within 4 Å[f]	55.4
QUEEN[f]	
I_{uni}/I_{total} (%) < 0.001	2481 (72.7%)
Average/SD	0.0017 ± 0.0055
Max	0.12

[a] Structure calculations were performed with the program CYANA 2.1 (55,56). A total of 400 random conformers were subjected to 8000 steps of a simulated annealing process. Each member of the family was submitted to restrained energy minimization (REM in explicit solvent) with the Amber-10 package (57). Values of 50 kcal mol-1 $Å^{-2}$ and 32 kcal $mol^{-1}rad^{-2}$ were used as force constants for the NOE and torsion angle restraints, respectively. The data are calculated over the 20 conformers representing the NMR structure and on the energy minimized mean structure. The mean value and the standard deviation are given

[b] Number of meaningful constraints for each class. Backbone dihedral angle constraints were derived from ^{15}N,$^{13}C'$,$^{13}C\alpha$,$^{13}C\beta$, and Ha chemical shifts, using TALOS and added as restrains in the structure calculations as well.

[c] As it results from the Ramachandran plot analysis performed with CinG

[d] Values calculated on secondary structure elements. A Z-score is defined as the deviation from the average value for this indicator observed in a database of high-resolution crystal structures, expressed in units of the standard deviation of this database-derived average.

[e] Values calculated on all residues with iCinG

[f] As it results from the analysis performed with iCinG

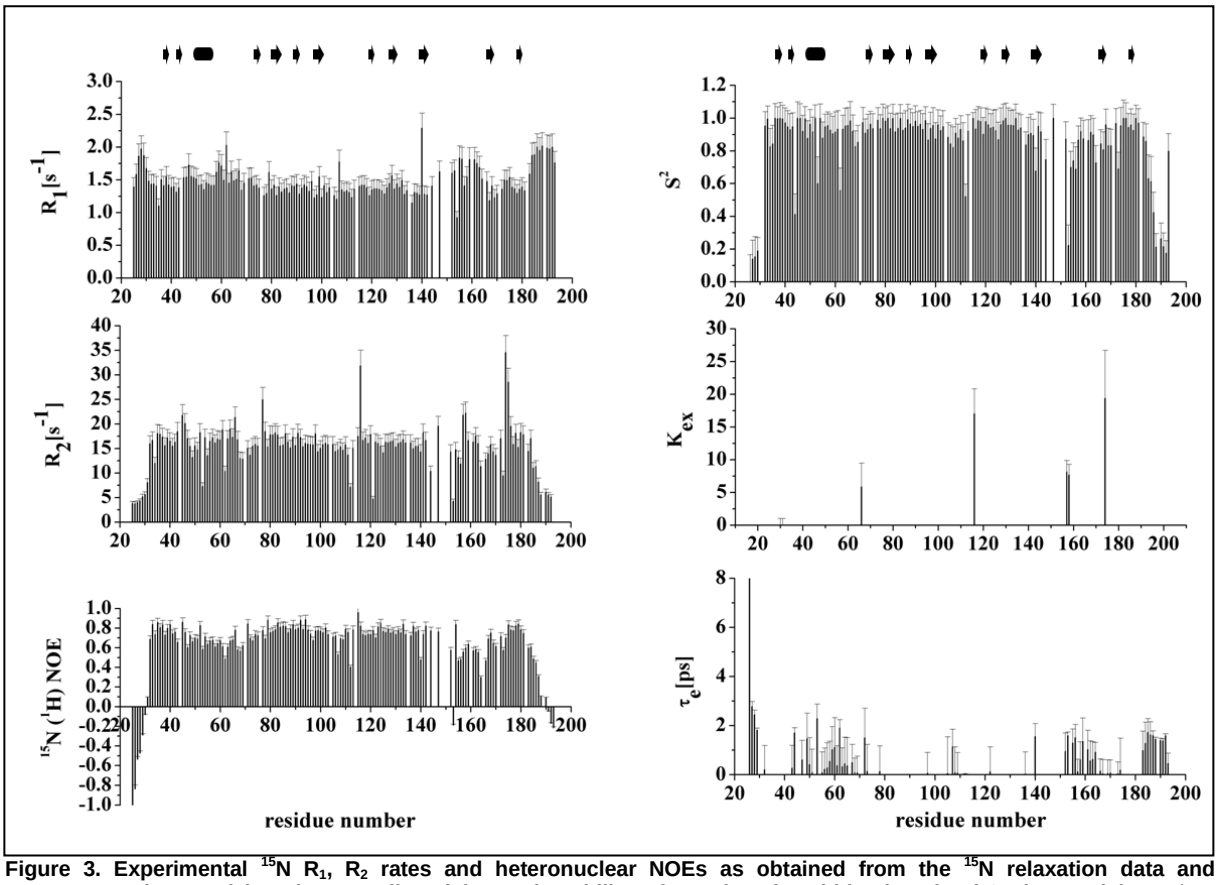


Figure 3. Experimental ^{15}N R_1, R_2 rates and heteronuclear NOEs as obtained from the ^{15}N relaxation data and parameters characterizing the overall and internal mobility of D1 domain within the Lipari-Szabo model. Nuclear relaxation rates are expressed in terms of spectral density functions J(ω) which depend on the overall rotational correlation time, τ_c, the generalized order parameter, S^2 which describe the amplitude of local motions of backbone NHs occurring at rates faster than the τ_c, and the correlation time for internal motions, τ_e, faster than τ_c. The secondary structure elements are reported at the top. The backbone amide order parameters (S^2), the effective correlation time for motions faster than the overall tumbling rate (τ_e), and the conformational exchange contribution to R_2 are shown.

Rigid Body Fitting— The procedure used to accommodate the NMR D1 structure into the EM map of the whole pilus previously generated (90) followed the same procedures used for the fitting of the RrgBD2-D4 x-ray crystal structure and the D1 computer model into the EM pilus map (88). A preliminary rigid body fitting of the D2–D4 crystal fragment was performed by using CHIMERA (105-107) followed by a rigid body fitting of the D1 NMR coordinates into the leftover N-terminal apical volume of the pilus. The D1 NMR coordinates were first fitted manually using CHIMERA by placing as much of the atomic structure as possible into the EM density map, approximately in the position thought to be correct. This step was then followed by a rigid-body fitting using the "fit model in map" tool from CHIMERA. This tool calculates, for the selected atoms, a position that maximizes locally the sum of the densities. The evaluation of the correlation coefficient values, the resulting average map value at the fit atoms, the number of fits atoms outside the lowest value contour surface displayed were the parameters used to assess the fit of the D1 molecule.

3.3. Results

Distinct Domains of RrgB Provide Protection in Active Immunization Experiments- The protective efficacy of each of the four RrgB domains or of a mixture of them was assessed in a mouse model of intraperitoneal challenge (TIGR4) in two distinct experiments, performed under the same conditions, which

were combined to reach *n*= 13–16 mice/group. All RrgB domains except D3 afforded significant protection against bacteremia (Fig. 3A), giving a reduction of the cfu geometric mean by 1–2 logs with respect to the controls. These values were similar to those afforded by the FL RrgB and by the combination of the four domains D1+D2+D3+D4. The results of mortality are reported in Fig. D1 and D4 conferred significant increase of survival time, similar to FL RrgB and the combination D1+D2+D3+D4. In particular, the median survival for the D1 group was 2.5 days higher respect to that of the control group (4 *versus* 1.5 days). At the end of the mortality observation, a significant survival rate was found for D1 (44% survival), D4 (27%), FL RrgB (44%), and D1+D2+D3+D4 (31%) groups.

N-terminal D1 Domain Is the Most Protective in Vivo— Among the single RrgB domains, D1 and D4 showed the most significant protective efficacy and were therefore analyzed further in a larger group of mice. Four different experiments, carried out under the same conditions, were combined to reach *n* = 31 animals/group. The results are shown in Fig. 3, *C* and *D*. In terms of bacteremia (Fig. 3*C*), both D1 and D4 afforded highly significant protection, with a cfu geometric mean by 2.6 and 1.5 logs lower, respectively, than that of the control group, and 8 animals from the D1 group in which cfu were undetectable. The reduction of bacteremia was significantly superior in D1 than in the D4 group. In terms of mortality course (Fig. 3*D*), both D1 and D4 conferred significant protection. The increase of survival time afforded by D1 showed a better trend than that of D4.

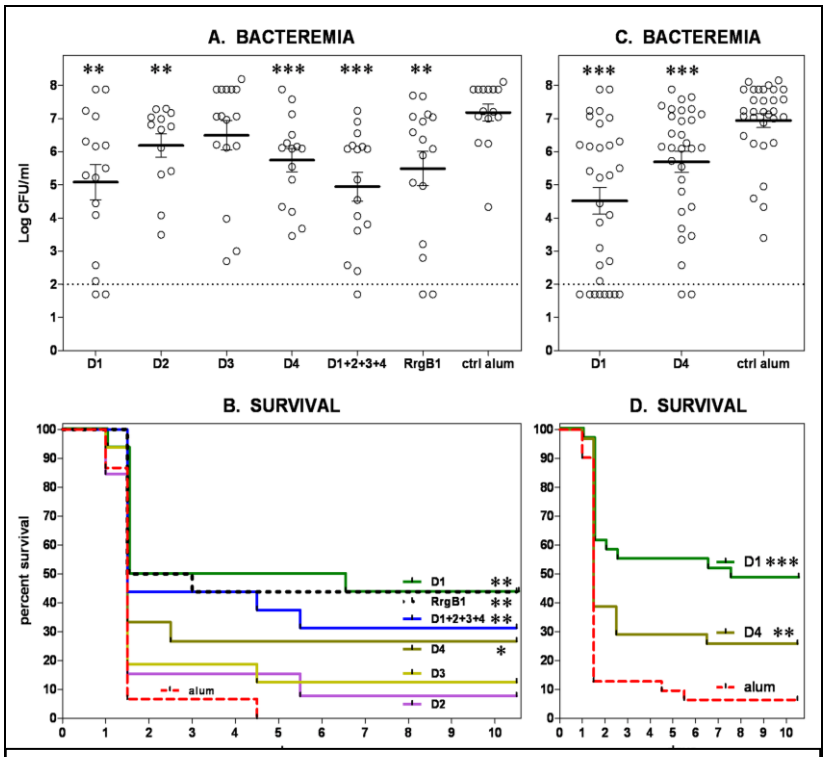


Figura 3. RrgB domains are protective in active immunization experiments. Bacteremia (A and C): circles represent the Log CFU per ml of blood for individual animals; horizontal bars represent the mean value of the Log CFU/ml ± SEM for the group; the dotted line marks the detection limit (values under the dotted line correspond to animals in which no CFU were detected). The treatment groups are indicated on the bottom. Survival (B and D): the survival course for each group is represented. The treatment groups are indicated close to each of the corresponding survival lines. A and B: *N* = 13 for D2 group, 15 for D4 and control (alum ctrl) groups, 16 for the remaining groups; C and D: *N* = 31 for each of the groups. *** *P* < 0.001;** *P* < 0.01; * *P* < 0.05.

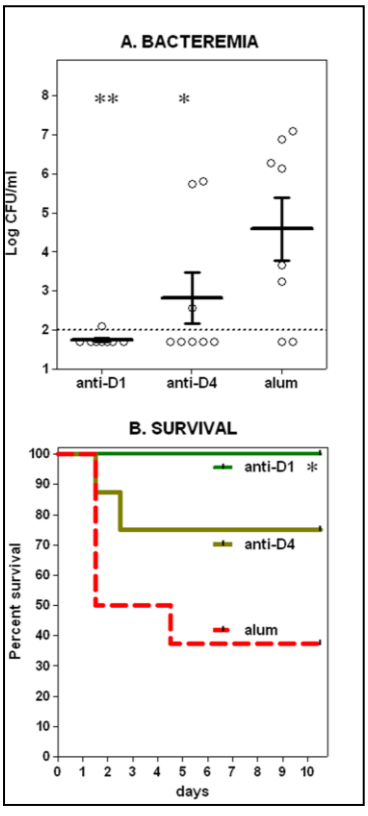


Figura 4. Anti-RrgB D1 and D4 sera are protective in passive serum transfer experiments. Symbols are described in the Figure 2 caption. *N* = 8 for each of the groups.

In particular, only for the D1 group was the median survival time higher than that of the control group (7.5 *versus* 1.5 days). At the end of the mortality observation, the D1 group showed the highest survival rate, *i.e.* 45% *versus* 21% observed for D4. An evident difference of survival rates between D1 and D4 groups was observed. The possible relevance of antibodies in the protection elicited by D1 and D4 was investigated by a passive serum transfer experiment, with groups of 8 mice.
The results are shown in Fig. 4. Both anti-D1 and anti-D4 sera elicited significant protection against bacteremia (Fig. 4*A*), with cfu geometric means lower by 2.6 and 1.6 logs, respectively, than that of the control group. Only anti-D1 serum afforded significant protection against mortality, giving 100% survival rate, whereas santi-D4 gave a protective trend, with 60% survival rate. These results indicate that immunization with D1 domain and at a lower extent D4 domain, elicits functional antibodies that may play a role in the protection.

D1 and D4 Antisera Recognize the Native Pilus and Linea Epitopes within RrgB—To investigate the differences in protective efficacy exerted by the isolated RrgB domains with respect to the FL RrgB in the *in vivo* assays, mouse sera were tested for their capability of recognizing the native pilus and the recombinant FL RrgB. Sera were probed against entire TIGR4 bacteria by FACS analysis. Sera raised against D1, D4, and D1+D2+D3+D4 gave a fluorescence intensity comparable with that obtained with anti-FL RrgB, whereas D2- and D3-specific antisera recognized the native pilus less effectively. To explain the lower recognition efficacy observed with D2 and D3, the same sera were titrated in ELISA against the recombinant FL RrgB. Antibody titers elicited by D2 and D3 immunization toward FL RrgB were about 10 times lower than those obtained by immunization with D1, D4, D1+D2+D3+D4, and FL RrgB, consistent with the results obtained by FACS analysis on the polymerized
native RrgB. To gain more insights on the epitopes recognized by the protective D1 and D4 polyclonal antibodies, a PepScan approach, suitable for identifying linear epitopes, was applied. Overlapping 15-mer peptides with an offset of 5 residues were synthesized *in situ* on a glass fiber membrane. The library of peptides tested covered residues 25–190 of D1 and 444–628 of D4. Incubation with the anti-D1 serum (previously used in passive protection experiments) revealed a single linear epitope covering residues 40–59 (D1-1) (Fig. 5*B*), whereas anti-D4 serum detected a unique linear epitope (D4-1) spanning residues 494– 508 (Fig. 5*C*).

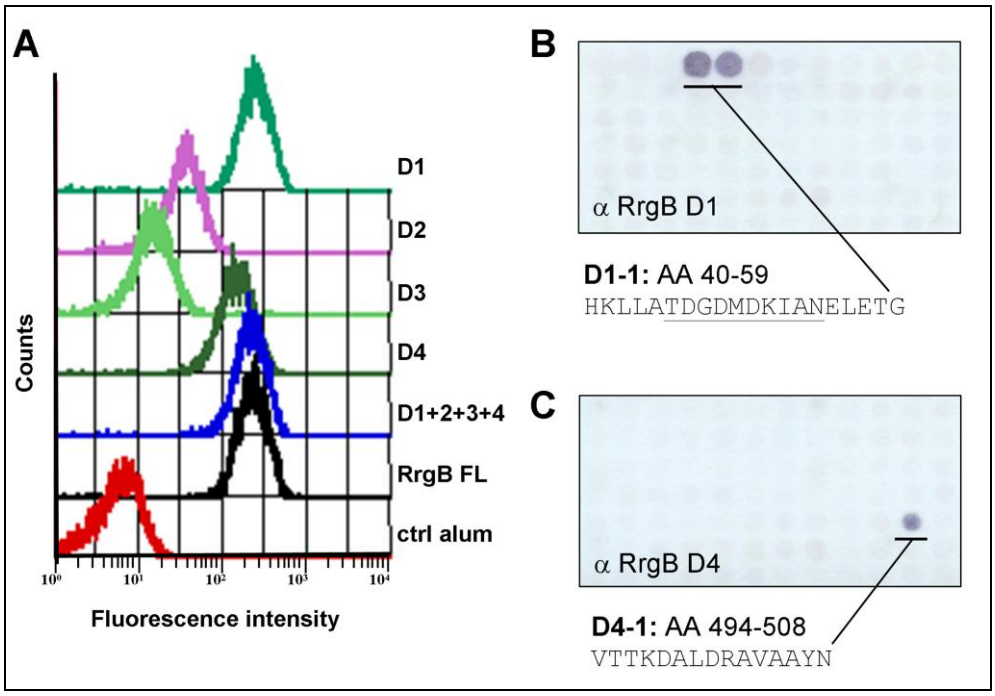


Figura 5. Polyclonal antibodies raised against D1 and D4 recognize efficiently the native pilus and linear epitopes within RrgB. A) TIGR4 bacteria were incubated with mouse primary antibodies directed against the specified recombinant proteins (1:400 dilution), followed by FITC-conjugated goat anti mouse IgG (1:100 dilution). Bacterial staining was analyzed by flow cytometry (FACS-Calibur). Mouse control sera (immunized with PBS plus alum) were used as negative control. **B-C)** Glass fiber membranes with arrayed peptides synthesized *in situ* covering residues 25 to 190 (D1) and 444 to 628 (D4) of RrgB were incubated with anti-D1 (A) or -D4 (B) polyclonal mouse antibodies (1:3000), and then with goat anti-mouse IgG alkaline phosphatase-conjugated antibodies (1:5000). Linear epitopes corresponding to peptide sequences recognized by the antibodies are reported. Underlining marks common residues present in peptides adjacent in the PepScan

Solution Structure of the D1 Domain—The solution structure of D1 was investigated by NMR spectroscopy. Its ^{1}H-$^{15}NHSQC$ spectrum showed well dispersed resonances indicative of an overall well folded protein (Fig. 6). D1 showed a common IgG-like β sandwich fold (41 Å x 48 Å x 30 Å) and a topology of secondary structure elements drawn in Fig. 7. The core of the structure was formed by seven parallel and antiparallel β-strands: β_1(36–39), β_4(80–85), β_7(119–121), β_8(127–130), β_9(138–143), β_{10}(166–169), and β_{11}(178–180). These β-strands were arranged in two sheets (comprising β_1, β_8, β_{11} andβ_4, β_7, β_9, β_{10}, respectively) packed against each other and flanked by two long segments (40–78, 87–117) located between strands β_1 and β_4 and strands β_4 and β_7, respectively. An α-helix (49–57), flanked by two short β strands β_2 (42–44) and β_3 (73–75), was inserted within the first segment. Two additional β-strands β_5 (89–91) and β_6 (97–101) formed α-hairpin structure inserted within the second segment. In 50% of 20 conformers of the D1 family an additional β-sheet was formed by two short hydrogen-bonded β-strands (stretches 161–163 and 184–187).

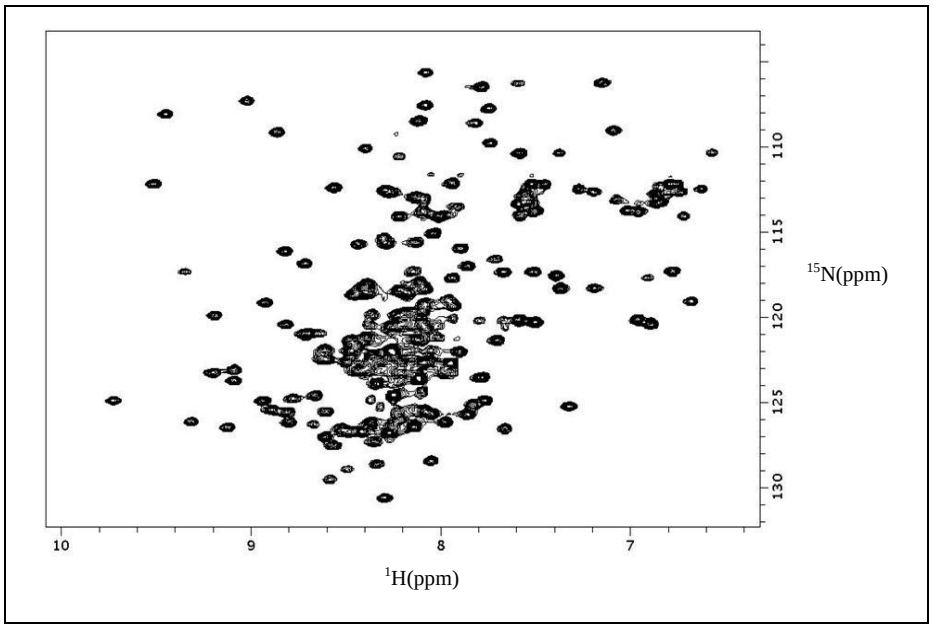


Figure 6. 2D 15N-1H HSQC spectrum (900MHz, 298K) of D1 domain. All the amide protons of the D1 domain were assigned, with the exceptions of the first 7 residues of the N-terminal end and the last 6 residues of the C-terminal end and the residues Ile76, Glu143, His145, Ser146, Ser148,Thr149, Tyr150,Val152 and Gly160.

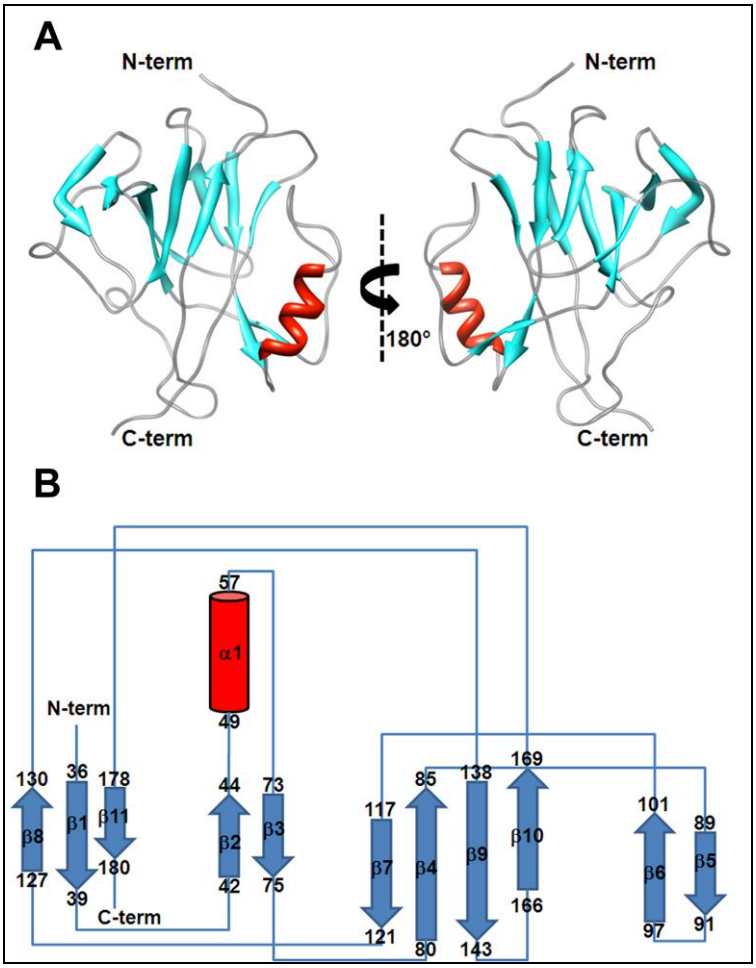


Figura 7. Solution structure of the D1 domain. A) Ribbon diagram with the secondary structure elements: β-strands are shown in cyan, the α-helix in red. **B)** Topology diagram. The α-helix is represented by a red cylinder and the β-strands are cyan arrows.

The structure was well defined with the exception of three long loops corresponding to the stretches 56–69, 148–162, and 173–177 (Fig. 2). Heteronuclear relaxation

measurements revealed that residues in the first two loops had heteronuclear NOE and longitudinal R_1 values lower and higher than average, respectively (Fig. 3). This behavior was a consequence of local internal motions occurring on a faster time scale with respect to the overall reorientational correlation time (τ_r) of the molecule and accordingly, a correlation time (τ_e) for these fast motions can be fitted for the previous mentioned loops (Fig. 3). In addition, conformational exchange processes, occurring on the millisecond- microsecond scale, affected some residues located in the region 155–164, as monitored by transverse R_2 relaxation rates higher than the average (Fig. 3). These data indicated that these loops experience higher flexibility, showing accordingly a low number of long ranges ^{1}H-^{1}H NOEs. For a few residues, the assignment of the backbone resonances was also not achieved (Glu-143, His-145, Ser- 146, Ser-148, Thr-149, Tyr-150, Val-152, and Gly-160), likely as a consequence of an increased local mobility. The loop between strands β_9 and β_{10} (residues 148–162) was the most disordered, with a r.m.s.d. value of about 2 Å (Fig. 2). Conformational exchange processes on the millisecond-microsecond time scale are observed also for the loop between strands β_{10} and β_{11} comprising residues 174–178. The protein core is characterized by hydrophobic interactions between residues located on the first (strands β_1, β_8, β_{11}) and second (strands β_4, β_7, β_9, β_{10}) sheets. A salt bridge between two complementarily charged side chains of residues Lys-41 and Glu-143 was also present. The aliphatic side chains of residues Met-48, Ile-52, Ala-53, and Leu-56, all located on one side of the α-helix, formed hydrophobic interactions with aliphatic residues of β_2 and β_3 strands; these interactions determined the position of the helix with respect to the rest of the protein. A search for related protein structures performed through the DALI server (108) retrieved the N-terminal domain of the SpaA pilus backbone protein of *C. diphtheriae* (Protein Data Bank (PDB) code 3HTL; r.m.s.d. 2.1 Å), the C-terminal CNA3 domain of the major pilin protein of *Bacillus cereus* BcpA (PDB code 3KPT; r.m.s.d. 2.4 Å), and the N1 domain of the *Streptococcus agalactiae* minor pilin GBS52 (PDB code 2PZ4; r.m.s.d. 3.4 Å). Like D1, none of these domains contained intramolecular isopeptide bonds. However, only the SpaA domain is located at theNterminus of pilus backbone protein as D1; their overlay is presented in Fig. 6*A*. In an attempt to determine the orientation of D1 with respect to the D2–D4 RrgB portion, a rigid body fitting of all the domains into the shape of the native pilus obtained from crio-electron microscopy (cryo-EM) was attempted (90). Initial rigid body fitting of the D2–D4 crystal structure into the cryo-EM density map of the native pilus left an apical unoccupied volume, likely due to the absence of D1. However, when the simultaneous fitting of the D1 and the D2–D4 domains was carried out, some portions of D1 could not be accommodated into the apical empty volume. To model the FL RrgB, we merged the D1 and D2–D4 structures into a single molecule by overlapping residues 188–191, shared by the C terminus of D1 and the N terminus of D2–D4. The relative orientation of D1 with respect to the others domains was then varied to best fit into the cryo-EM map (Fig. 8). The final model, which represents only one of the possible orientations of D1 with respect to the rest of the protein, does not present steric clashes between D1 and the D2–D4 domains. The two linear epitopes, previously identified to be located within the D1 and D4 domains by peptide hybridization with specific protective polyclonal antibodies, when mapped onto the FL model of RrgB, confirmed their superficial localization (Fig. 8).

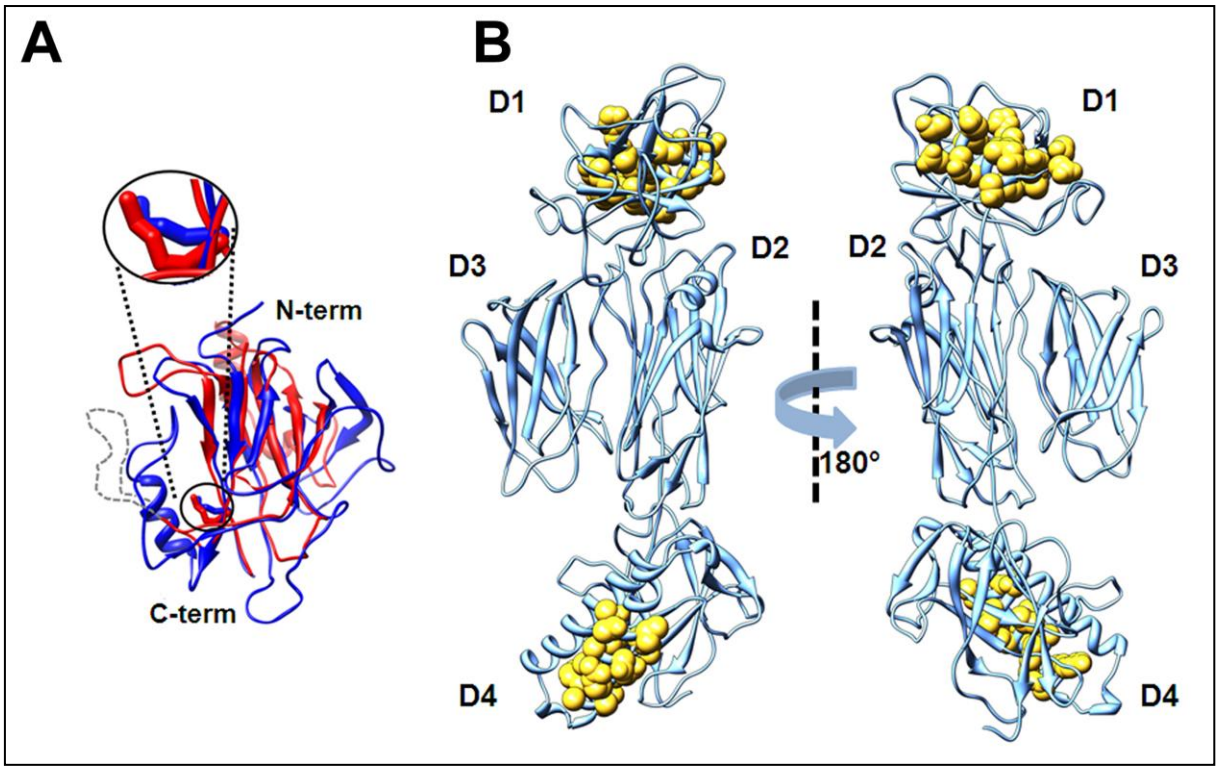


Fig. 8. Analysis of RrgB linker flexibility through superimposition and modeling. **A)** Superimposition of the D1 domain (blue) and the *C. diphtheriae* SpaA N-terminal domain (red). The position of Lys190 13 residue, forming the inter-molecular isopeptide bond between two neighboring SpaA molecules is shown along with residue Lys183 of RrgB, which occupies a similar position, and is indicated in a black circle. The missing loop in the SpaA crystal structure is indicated as grey dotted line. **B)** Modeled conformation of the full length RrgB molecule obtained by combining the D1 NMR coordinates with the RrgB D2-D4 X-ray coordinates. D1_1 and D4_1 epitopes (residues 40-59 and 494-508 respectively) are rendered as yellow spheres.

Lysine 183 of D1 Is Required for Intermolecular Isopeptide Bond Formation and Pilus Polymerization—Pili of Gram-positive bacteria are polymerized by means of intermolecular isopeptide bonds occurring between the Thr of the C-terminal LP*X*TG motif of a RrgB molecule and a Lys located at the N-terminus of the following molecule (85, 95, 96).

To identify the Lys implicated in the intermolecular isopeptide bond formation between two consecutive RrgB monomers, a TIGR4 RrgB deletion mutant (no longer able to assemble a pilus on its surface) was created. Subsequently, RrgB expression and pilus polymerization were restored in TIGR4_RrgB by transforming the mutant with plasmids expressing either wild-type RrgB or RrgB mutated in single Lys residues (Lys3Ala substitutions). Sequence analysis revealed that the D1 contains the canonical 181YPKN184 pilin motif, with Lys-183 nicely superimposing onto the functional SpaA Lys-190, as shown in Fig. 8.

Noteworthy, D1 also presents the sequence fragment 160GSKAVP165, similar to the motif containing the Lys residue functional for the pilus assembly in *S. pyogenes* Spy0128 (97). Lys-183 and Lys-162, along with two additional Lys residues (Lys-138 and Lys-309), located on D1 and D3 domains, respectively, were selected and mutated. WB analysis performed on whole cell bacterial lysates using rabbit polyclonal antisera raised against RrgA, RrgB, or RrgC revealed that all of the pilus proteins were expressed in all of the complemented mutants. However, the typical pilus-associated high molecular weight ladder was revealed only in the case of TIGR4_RrgB complemented with RrgB wild type or RrgB mutated in Lys-162, Lys-138, or Lys-309. In contrast, in the mutant complemented with K183A, RrgB was detectable only as a monomer, clearly indicating that Lys-183 is the residue implicated in intermolecular isopeptide bond formation. Moreover, in the presence of

the RrgB K183A substitution only hetero-oligomers composed of RrgA-RrgC (about 160 kDa) or RrgB-RrgC (about 110 kDa) could be detected by WB analysis. The absence of RrgARrgB heterodimers suggests that RrgA and RrgB are linked exclusively through intermolecular isopeptide bonds involving the D1 residue Lys-183 of RrgB.

3.4. Discussion

Ever since their initial discovery, pili of Gram-positive bacteria have received considerable attention because they are associated with a number of different virulence mechanisms (24, 83, 98), and elicit protection in animal models (26). In particular, the *S. pneumoniae* pilus was found to be implicated in the initial attachment to epithelial cells (24, 83, 98), and the pilus components, being protective in mouse models of infection, are regarded as potential vaccine candidates (26, 98).

The major pilin RrgB is organized into four Ig-like domains (D1– D4), as shown by combined structural approaches approaches (90) and this work). We have tested the four RrgB domains in animal model experiments and have shown that the protective efficacy exerted by the combination of the four RrgB domains D1+D2+D3+D4 is comparable with that afforded by the FL RrgB. Among the single domains, D1 is the most protective, and D4 retained an important part of the protective efficacy of the FL protein. The lower protection achieved by D2 and D3 compared with the FL protein, as well as with the D1 and D4, is probably because the antibodies elicited by the former two domains recognize the FL protein less efficiently, in both its native and recombinant forms. This may be the result of smaller exposed surface areas experienced by D2 and D3 in the FL RrgB compared with the D2 and D3 isolated domains and with D1 and D4. It is possible, in fact, that the antibodies generated against the isolated domains are recognizing areas of D2 and D3 that are buried in the FL protein. Alternatively, D2 and D3 could assume a slightly different conformation when expressed as single domains, thus generating nonfunctional antibodies. On the other hand, the two linear epitopes (Fig. 8) identified within D1 and D4 by PepScan analysis performed with protective polyclonal antibodies raised against the two domains (conformational epitopes are not detectable with this method) are well exposed on the surface of the RrgB molecule and could contribute to the protective activity exerted by D1 and D4. Further experiments are needed to understand to what extent these linear epitopes contribute to the protective activity exerted by the two domains. Taken together, these results suggest that RrgB contains multiple protective epitopes, thus confirming the potential of this vaccine candidate. Furthermore, although the existence of possible conformational epitopes involving residues from different domains cannot be excluded, their contribution to the overall protective efficacy might not be essential.

To obtain more information about the structural role of D1 and to try to correlate it with the protection data discussed above, we solved the solution structure of this domain by using NMR spectroscopy. D1 shows an Ig-like fold, does not contain any intramolecular isopeptide bond, and has many flexible regions. The observed D1 flexibility, indeed, could play a fundamental role in the specific antigen-antibody recognition process (109), thus accounting for D1 enhanced protection capability with respect to the more rigid D2–D4 domains, each one containing an intramolecular isopeptide bond. In fact, the protein structural plasticity could be related to the ability of D1 to undergo local conformational changes and to adapt its structure to optimize the interactions with the antibodies and increase the affinity and the specificity of the antigen-antibody recognition process. The dynamics of D1 could

therefore strongly contribute to the interface adaptation for molecular recognition such that the antibody can select an optimal conformer from a wide distribution of possible D1 conformations. The rigid structure of the D2–D4 region prevents such an effective conformational selection for these domains. The above described phenomenon has been observed for other protein-protein or protein-DNA interaction processes (109, 110). To shed light on the molecular mechanism driving pilus polymerization in *S. pneumoniae*, we investigated which of the lysine residues of D1 was engaged in the intermolecular isopeptide bond formation. Site-directed mutagenesis followed by complementation identified Lys-183 as crucial for the pilus assembly. This result is consistent with the observation that the spatial position of RrgB Lys-183 can be superimposed onto Lys- 190 of the *C. diphtheriae* pilus backbone subunit SpaA, known to be involved in the intermolecular isopeptide linkage (96). Interestingly, as shown in Fig. 8, both lysines are not fully available to form an external bond. In particular, the average relative solvent accessibility of the Lys-183 over the D1 family of conformers is 27.2% ± 4.8, as the Lys side chain projects into a cleft between the main body of the protein and the segment 40–78 containing the mobile loop 56–69 (92). This suggests that pilus backbone proteins, to be polymerized, might undergo conformational changes, probably involving not only the Lys residue (Lys-183) but also the flexible regions spatially close to it (49–69, 152–167, and 183–193), to allow the formation of the covalent intermolecular isopeptide bond. NMR mobility data indicate that the C terminus of the D1, where Lys-183 is located, and other loops are highly mobile and that such dynamics could be relevant for the intermolecular isopeptide bond formation (Figura 3). Consistently, the absence of stabilizing intramolecular isopeptide bonds renders D1, unlike the other domains, less rigid and prone to conformational rearrangements.

Furthermore, the occurrence of a structural rearrangement of D1 within the native pilus structure is also in line with the partial fitting of its NMR structure onto the molecular shape of the native pilus determined by cryo-EM. Mutagenesis and complementation data were used to analyze the covalent links established among the three pilus proteins either in the absence of RrgB or in the presence of the RrgB K183A mutant, in an attempt to provide further insights into the possible organization of the native pilus. In the presence of the non polymerizing RrgB K183A mutant, the lack of RrgA-RrgB heterodimers provides evidence that RrgA and RrgB are unidirectionally linked only through the Lys-183 of RrgB and the C-terminal Thr of RrgA. This arrangement is in accordance with the model proposed by Hilleringmann *et al.*, positioning RrgA at the pilus terminus, thus ruling out the alternative possibility of RrgA being incorporated along the pilus shaft (88). Conversely, an RrgA-RrgC multimer is detectable either when rrgB is not expressed, as already reported by Falker *et al.* and Le Mieux *et al.* (25, 111), or in the presence of nonpolymerizing RrgB K183A. In this case, the mutated RrgB is competing with RrgA for the linkage to RrgC, as demonstrated by the presence of RrgB-RrgC hetero-oligomers. Concomitant detection of RrgA-RrgC hetero-oligomers even under these conditions further strengthens the idea that in wild-type bacteria, although not detectable by electron microscopy analysis of purified pili, a fraction of RrgA and RrgC might be directly linked to each other.

In conclusion, this study provides additional information elucidating pilus proteins features and also paves the way to the rational design of new RrgB-based molecules to implement a protein-based vaccine against pneumococcal disease. Moreover, the newly acquired structural and dynamic information on the RrgB molecule provided by this study suggests that the conformational flexibility of D1 is

pivotal for the protein-antibody recognition process. These findings together with the new functional information could be used to better understand pilus functions and its role in pathogenesis.

CHAPTER 4
Investigation on the interaction between factor H binding protein and factor H

Paper in preparation

4.1. Introduction

The ability to evade the surveillance of the immune system is a key determinant for the survival of pathogens within their hosts. In particular, the complement cascade provides the first line of defence against external pathogenic agents, and it is therefore quite expected that co-evolution of humans and microorganisms resulted in a variety of microbial molecules able to protect the microbe from the complement attack. Different molecular mechanisms have been set up by bacterial pathogens to escape human complement, including recruitment of complement inhibitors on the surface and degradation or inactivation of key effectors (112). Bacterial molecules able to sequester complement inhibitors on bacterial surfaces are in principle good targets of vaccine-induced antibodies and may be used to design vaccines focused on the generation of complement dependent bactericidal and opsonopahgocytic antibodies activities (113).

The factor H binding protein (fHbp) of *Neisseria meningitidis* is a 27kDa membrane-anchored lipoprotein (10, 15) and represents an example of protein able to contribute to the bacterial survival by binding a negative regulator of the complement cascade. It can binds human factor H (hfH) (18). The bacterium is able to cover its surface with factor H allowing mimicking a human tissue and avoiding complement-mediated lysis. Factor H is a single polypeptide chain plasma glycoprotein (155 kDa), composed of 20 units of 60 amino acids (114), named short consensus repeats (SCR) or complement control protein modules (CCP). The 3D structure of several SCRs have been determined, showing a globular structure with six-stranded antiparallel β-sheets connected with loops and turns. fH plays a central role in inhibiting the alternative complement pathway by contrasting the assembly of an active C3 convertase (115-117).

Recently, important advances have been done on the knowledge of structural basis of the interaction between fHbp and human factor H. The crystal structure of the complex formed by fHbp and the sixth and seventh domains of human factor H (CCP_{6-7}) solved by Schneider et al. (118) evidenced that the factor H binds fHbp at the same site devoted to recognition of host glucosamino-glycans (GAGs). On the other side, fHbp contacts factor H with the groove formed between its N- and C-domains (figure 1).

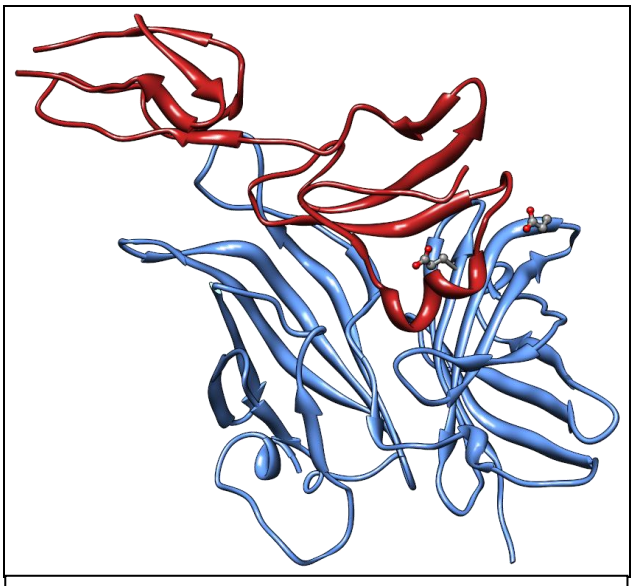


Figure 1. Crystal structure of the complex fHbp (blue) and CCP67 (res). PDB ID 2W80

The centre of the interaction site is occupied by two residues, namely E218 and E239, which have been proposed to mimic the poly-anionic structure of GAGs that normally drives the interaction of fH with human cell surfaces (119).

In the present study we characterized the interaction of fHbp with fH in aqueous solution by nuclear magnetic resonance (NMR). Chemical shift perturbation analysis performed in presence of recombinant CCP6-7 and CCP5-7 domains evidenced that in both cases the fHbp binding sites to fH were virtually identical and largely overlapping to that already identified by X-ray crystallography in CCP6-7, excluding that CCP5 could interact in a direct and stable way with a specific portion of the fHbp surface. Prompted by signal perturbations observed on amino acids located at periphery of fH binding site, it will be possible to design in the future a panel of fHbp in order to explore the influence of such residues on the fHbp-fH interaction. This approach will allow us to identify additional amino acids contributing to the fHbp ability to bind fH, providing evidences that other mechanisms in addition to molecular mimicry can contribute to the fHbp affinity for fH.

4.2. MATERIALS AND METHODS

NMR sample production and NMR measurements - Recombinant fHbp (residues Met-7–Gln-255) was expressed in *Escherichia coli* as already described. The interaction of CCP5-6 and CCP5-7 with ^{15}N-labeled fHbp was investigated by NMR spectroscopy with ^{1}H–^{15}N TROSY-HSQC experiments. All the NMR experiments were performed at 298 K on a Bruker Advance 900 spectrometer, working at 900.13 MHz frequency and equipped with a cryogenically cooled probe. Titrations with the fH constructs were performed on 0.2 mM ^{15}N-labelled fHbp protein samples with the unlabelled CCP5-6 or CCP5-7 up to an fHbp CCP molar ratio of 1:1. ^{1}H and ^{15}N resonances assignment for the fHbp protein was already available (17). All the samples used were in 20 mM PBS (90% H_2O and 10% D_2O) at pH 7.0. Solvent accessibility for fHbp residues was calculated with the program NACCESS (120); only residues with a GLOBAL solvent accessibility larger than 40% were considered for the mapping.

4.3. Results and discussion

NMR mapping of the fH-fHbp interaction -The interaction between fHbp and the recombinant CCP5-7 was investigated through NMR spectroscopy, by analyzing the perturbations caused in the $^{1}H–^{15}N$ TROSY-HSQC spectra of ^{15}N-labeled fHbp upon addition of unlabelled CCP5-7 up to a fHbp:CCP5-7 ratio of 1:1. Chemical shift variation of NH cross peaks observed in the $^{1}H–^{15}N$ TROSY-HSQC spectra upon addition of CCP5-7 indicated that the two proteins are interacting. The fHbp-CCP5-7 complex exchanged with the free proteins at rates slower than the resonance frequency differences between the two forms (i.e., in the range of milliseconds). Residues of fHbp experiencing different chemical shift upon addition of CCP5-7, are shown in Figure 2 and listed in Table 1. They comprise an extensive region which involves both the N- and C-terminal domains of fHbp, largely overlapping to the fH binding site described in the crystal structure of the fHBP-CCP6-7 complex (118). The total number of perturbed, surface-exposed residues in solution define a larger contact area compared to that found in the crystal, but still contains all the residues located at the interface between fHbp and CCP6-7, suggesting that also in solution and in the presence of a longer construct of fH, the fHbp interaction with domains 6 and 7 is essentially maintained. Moreover, additional residues, distinct from those forming the CCP6-7 binding site, were perturbed by the addition of CCP5-7. A first group consisted of solvent-accessible amino acids (Figure2A) located in the linker connecting the N- and C-domains of fHbp (Thr139, Phe141, Asp142 and Lys143) and several buried residues located at the domain-domain interface of fHbp (Gln97, Tyr99, Gln101, His103, Phe129, Gly132, Ala135, Ile226, Gly236, Ser237, His248, Ile249, Gly250 and Leu251, Figure2B). Their sensitivity to the presence of CCP5-7 suggested that a molecular rearrangement could occur in fHbp during the formation of the fHbp-CCP5-7 complex. A second region was formed by residues whose localization on the protein surface was theoretically compatible with contacts to CCP5. To test this hypothesis we performed the same NMR measurements in presence of CCP6-7. However the pattern of perturbed residues remained essentially unchanged, ruling out the possibility that CCP5 engage specific and stable contacts with the fHbp surface and suggesting that the interaction of fHbp with fH was limited to domains 6 and 7. Our NMR data are in line with previous epitope mapping studies on fHbp. Beernink and colleagues have shown that Gly121 and Lys122, two residues solvent-exposed located in the loop between strands 7 and 8 of the N-terminal domain of the molecule, are recognized by monoclonal antibodies as JAR3 and JAR5 (121) that inhibit fH binding. Accordingly, spectral perturbations were observed for their NH cross peaks upon addition of CCP5-7 and CCP67 to fHbp. On the other side, we demonstrated by FACS experiments that the fHbp interaction with fH could not be prevented by the monoclonal antibody MAb502 binding to fHbp (122). Residues involved in the interaction with MAb502 are Arg204, Glu146-Arg149, Ala174, Lys230 and Glu233. As expected, the latter residues are indeed not affected by CCP5-7 and CCP67 binding confirming that there is not overlapping between fH and MAb502 binding sites. Based on the NMR data it will be possible to investigate by mutagenesis of single residues about the role of additional residues in the interaction between fH and fHbp.

RESIDUES	
BURIED	Leu31,Gln32,Leu36,Ser39,Val40,Glu44,Tyr57,Asn64,Ser74,Phe76,Phe78, Arg80,Gln81,Ile82,Ile89,Ser93,Phe96,Gln97,Val98,Tyr99,Gln101,His103, Thr107,Phe109,Gln110,Thr111,Ile114,Asp116,Phe129,Gly132,Ala135,Tyr152, Lys165Gly177,Gly179,Ala196,Ile198,Ala206,Val207,Gly210,Val212,Gly220, Leu224,Gly225,Ile226,Gly236,Ser237,Ala238,Thr242,His248,Ile249,Gly250, Leu251
EXPOSED	Asp37,Gln38,Arg41,Lys42,Asn43,Lys45,Thr56,Glu83,Gly86,Leu91,Glu95, Glu112,His119,Gly121,Lys122,Val124,Lys126,Arg127,Gln128,Arg130,Thr139, Ser140,Phe141,Asp142,Lys143,Asp160,Gly163,Glu188,Lys199,Ser211, Leu213,Lys219,Ser221,Ser223,Lys241

Table 1

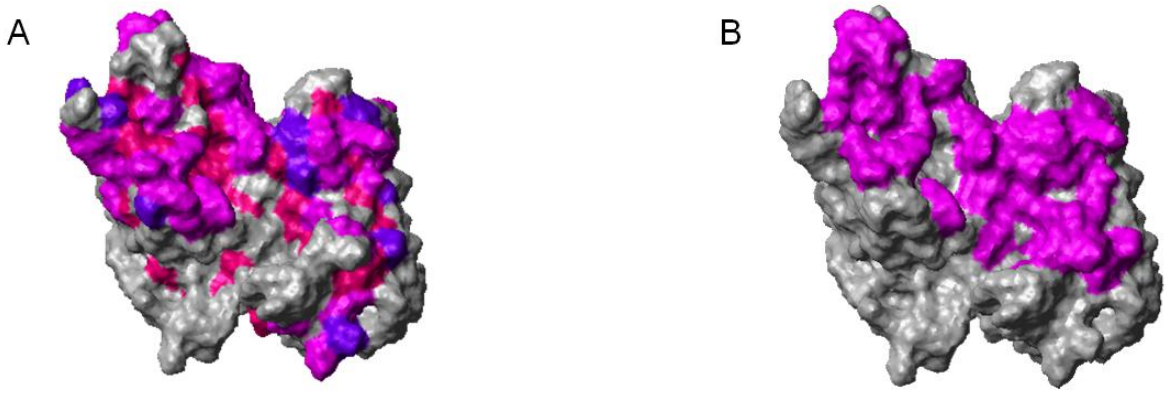


Figure 2. A) Residues experiencing chemical shift differences mapped on the molecular surface of the free fHbp. Surface-exposed and buried residues, whose amide NMR signals were perturbated (> 0.1 ppm) upon addition of hfH567, are shown in magenta and red, respectively. Residues whose amide signals were not perturbated upon addition of hfH567, prolines, and residues that could not be accurately analyzed due to spectral overlap, are in gray. **B)** Residues mapped on the molecular surface of the free fHbp (PDBID 2KC0) and located at the interface of the fHbp-hfH67 complex in the crystal structure are shown in magenta

CHAPTER 5

Neisseria meningitidis fHbp interacts with xenosiderophores *in vitro*

Paper in preparation

5.1. Introduction

Neisseria meningitidis, a Gram-negative bacterium that colonizes the upper respiratory tract of 10% of healthy human population, is adapted to grow only in humans. With a frequency of 1 in 100,000 populations, the bacterium invades the bloodstream and becomes a severe pathogen, causing sepsis and meningitis. Vaccination with capsular polysaccharides induces serogroup specific protective antibodies. Meningococcal capsular polysaccharide vaccines are available against serogroups A, C, Y, and W135 (123-125). On the contrary, the development of a vaccine against serogroup B, still responsible for a significant percentage of invasive diseases, has been protracted due to the immunologic cross-reactivity of B polysaccharide with human tissues. Recently, new perspectives to meningococcal B (menB) prevention have been opened by the identification of suitable protein-based vaccine antigens identified by mining the bacterial genome (12). One of the most promising antigens is the factor H-binding protein (fHbp), whose solution structure was recently determined by NMR spectroscopy (17). This is a 27kDa membrane-anchored lipoprotein (10, 15) present on the surface of all strains of *N. meningitidis* and elicits protective bactericidal antibodies. fHbp binds human factor H (hfH), a negative regulator of the alternative complement activation (18). The bacterium is able to cover its surface with factor H allowing mimicking a human tissue and avoiding complement-mediated lysis. However, little else was known about the in vivo functions of this protein prior to vaccine development. NMR analysis of the C-terminal structure of fHbp (17) revealed similarities to lipocalins, a diverse group of proteins found in plants, fungi, bacteria, vertebrates and invertebrates that participate in many biological functions including binding to molecules devoted to iron recruitment from the environment collectively called siderophores (126). Siderophores (from the Greek: “iron carriers”) are defined as relatively low molecular weight, ferric ion specific chelating agents elaborated by bacteria and fungi growing under low iron stress Even when bacteria colonize multicellular organisms, whether they are saprophytic or pathogenic microbes, they do not find freely available iron, the role of the siderophores is to scavenge iron from the environment and to make the mineral, which is almost always essential, available to the microbial cell (127) (Figure 1). Siderophores are frequent targets of the innate defense system (128) and many bacteria evolved the ability to use siderophores produced by other microorganisms (xenosiderophores) (129, 130). Although no siderophores produced by *N.meningitidi*s have been identified so far, a meningococcal surface protein, FetA, is a known receptor for enterobactin (131), a siderophore produced by many enteric bacteria.

Prompted by its structural similarity to lipocalins, we investigated on the possible activity of fHbp as siderophore binding protein. In particular, we analyzed the ability by the protein to bind prototypic siderophores such as enterobactin, salmochelin and yersiniabactin. Our data revealed affinity for enterobactin and the binding site has been mapped on the protein surface by NMR spectroscopy.
This study provides the first evidences that, additionally to bind factor H, this protein can be involved also in additional mechanisms for the survival in the host.

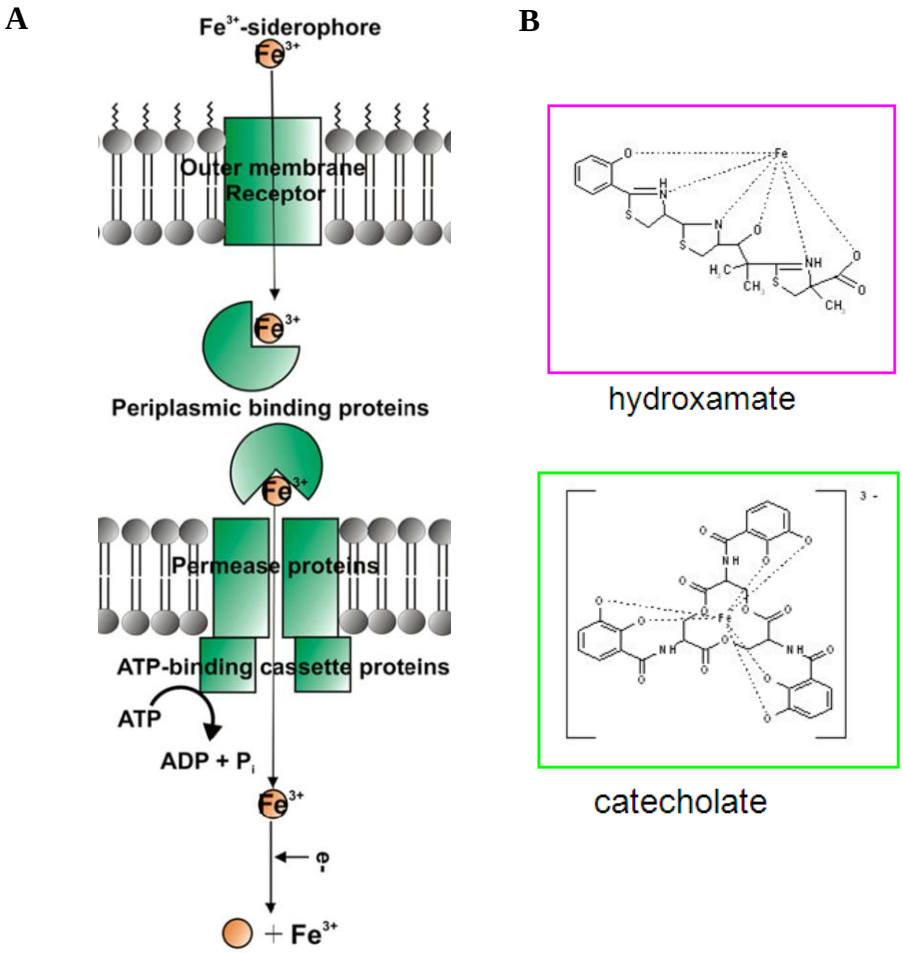


Figure 1.**A) Schematic representation of iron uptake mediated by siderophores in Gram-negative bacteria**. The pathway requires an outer membrane receptor, a PBP, and an inner-membrane ABC transporter. Transport through the outer membrane receptor requires the action of the TonB system (TonB, ExbB, ExbD). **B)** Two different class of siderophores.

5.2. Materials and methods

rfHbp-His Expression and purification- The *fhbp* gene was amplified by PCR on the chromosomal DNA of *N. meningitidis* strain MC58 and cloned into pET21b+ vector (Novagen) under the control of the T7 promoter. fHbp was expressed in the *Escherichia coli* BL21 (DE3) strain as a C-terminal histidine fusion. Protein expression was induced by Isopropyl1-thio-β-D-galactopyranoside (IPTG), (Sigma), then rfhbp was purified through a nickel chelating affinity chromatography using a HisTrap HP column (GE Healthcare) followed by cationic exchange chromatography. The purified protein was dialyzed in PBS buffer. For NMR analysis, the clone BL21 (DE3) *E.coli* was grown in ISOGRO^{TM} ^{15}N (Sigma). After protein purification, the final

product was dialyzed in 20 mM sodium phosphate, pH 7.2, a specific requirement for NMR.

fhbp enterobactin interaction - Native Page Gel-100 micrograms of purified full-length rfHBP-his or rpK1 0405B fragment from ExpEc (control), were mixed and incubated for 1h or 24 h with 3, 6 or 9 µg of ferric enterobactin (EMC). Samples were mixed with Tris-glycine native sample buffer and loaded on Novex Tris-glycine or Novex Tris-acetate native gels (Invitrogen). Gels were run at 80 V constantly for 2.0 h and stained with SimplyBlue Safe Stain (Invitrogen).

fHbp enterobactin interaction studies by NMR spectroscopy-The interaction of ferric–enterobactin ($FeIII(Ent)^{3-}$), -salmochelin ($FeIIISalm)^{3-}$) and -Yersiniabactin ($FeIII(Yers)^{3-}$) with ^{15}N-labeled fHbp was investigated by NMR spectroscopy performed at 298 K on a Bruker Avance 900 and 800 spectrometers, at proton nominal frequencies of 899.20 and 800.13 MHz, respectively, using triple-resonance (TXI 5-mm) probes equipped with pulsed-field gradients along the z-axis. The interaction between fHbp and enterobactin iron free was also investigated. Titrations were performed on 0.5 mM ^{15}N-labeled fHbp protein samples with enterobactin up to a fHbp–enterobactin molar ratio of 1:1.5. ^{1}H and ^{15}N resonance assignments for the fHbp protein were available (17). All samples were in 20 mM phosphate buffer (90% H2O and 10% D2O) at pH 7.2. All 2D spectra were processed using TOPSPIN software.
Docking calculations between fHbp and $FeIII(Ent)^{3-}$ were performed with HADDOCK2.0 (132, 133). Twenty conformers of the fHbp solution structure (PDB ID 2KC0) and the PDB coordinates of the small molecule extracted from the crystal structure on the human siderocalin NGAL complexed with $FeIII(Ent)^{3-}$ (PDBID 3CMP) were used as input. The topologies and parameter files for $FeIII(Ent)^{3-}$, which are necessary to run HADDOCK2.0, were generated by PRODRG (http://davapc1.bioch.dundee.ac.uk/prodrg/) (134) using the PDB coordinates extracted from the 3CMP crystal structure.
The docking calculations in HADDOCK were driven by 9 active and 9 passive residues (Table 1), defined following the HADDOCK protocol (132, 133). Solvent accessibility of fHbp was calculated with NACCESS (120). Flexible regions of the proteins were defined based on the active and passive residues plus two preceding and following residues. In the initial rigid-body docking calculation phase, 1000 structures of the complex were generated; the best 100 in terms of total intermolecular energy were further submitted to semiflexible simulated annealing and final refinement in water. The initial temperature for the second Torsion Angle Dynamic (TAD) cooling step with flexible side-chain at the interface was set up to 500 K. The number of MD steps for rigid body high temperature TAD and during the first rigid body cooling stage were set up to zero while the number of MD steps during the second and the third cooling stages with flexible side-chains at interface were set up to 500. The final 100 structures were then clustered using a cutoff of 2.0 Å of RMSD among any structure of a cluster. Clusters containing at least 5 structures were analyzed.

Active residues	**Gly148, Arg149, Ala150, Thr151, Arg153, Lys230, Gln232, Lys254, Gln255**
Passive residues	**Asp71, His103, Thr139, Gly147, Thr167, Thr169, Ala173, Gly229, Gln232**

Table 1

5.3. Results

fHbp-siderophores interaction - Native Page Gel-The ability of recombinant fHbp to bind native enterobactin was initially investigated by native PAGE. Incubation with and without ferric enterobactin resulted in different patterns of migration.

NMR mapping of fHbp-siderophores interaction-To address the molecular aspects of the interaction between FeIII(Ent)$^{3-}$ and fHbp, the perturbations caused in the $^{1}H–^{15}N$- HSQC spectrum upon addition of Enterobactin iron free, FeIII(Ent)$^{3-}$, FeIII(Salm)$^{3-}$ and FeIII(Yers)$^{3-}$ were analyzed and the fHbp residues involved in the protein-small molecule interaction were mapped onto the NMR structure (PDB ID 2KC0). Addition of increasing amount of paramagnetic FeIII(Ent)$^{3-}$ to ^{15}N-labeled fHbp resulted in progressive increase of the line-widths of some NH cross-peaks in the $^{1}H–^{15}N$HSQC spectra (table 1), which disappeared at the 1:1.5. The protein residues involved in the protein-small molecule interaction were mapped onto the NMR structure (figure 2).

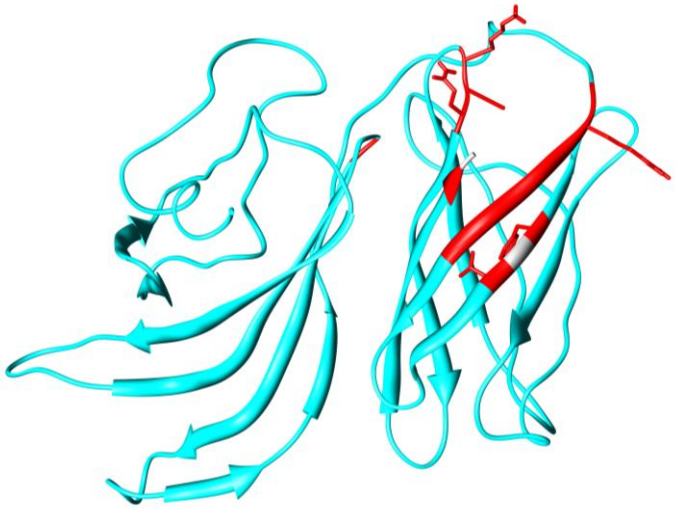


Figure 2. Solution structure of fHbp (PDB ID 2KC0) with the mapping of residues engaged in the interaction with FeIII(Ent)$^{3-}$.

All residues, whose NH cross peaks disappear upon addition of FeIII(Ent)$^{3-}$, were located on the C-terminal domain with exception of Ser102. No interaction was observed between protein and enterobactin iron free, salmochelin and yersiniabactin. Enterobactin iron free is less negatively charged at pH 7.2 than FeIII(Ent)$^{3-}$ and this could be one of the reason why we did not observed any interaction. We also added a solution of $FeCl_3$ to the protein/apo-enterobactin 1:1.5 mixture, and the same signals broaden beyond detection indicating that both Coulombic attraction and cation-p interaction of the cathecole moiety could be important for binding. The importance of the presence of cathecole moieties is confirmed by the absence of interaction of fHbp with the hydromaxate siderophores yersiniabactin. Any changes were indeed observed in the the $^{1}H–^{15}N$ HSQC fHbp spectrum upon addition of yersiniabactin up to a 1:5 protein:small molecule ratio (Figure 3). Finally, the best model of the molecular docking shows the orientation of the FeIII(Ent)$^{3-}$ with respect to fHbp (figure 4).

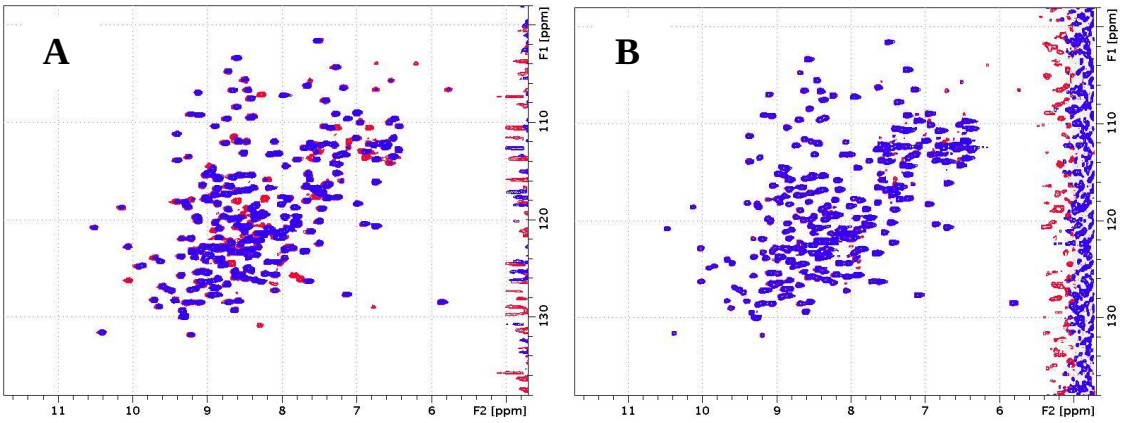


Figure 3. **A)** Titration of fHbp with FeIII(Ent)$^{3-}$. Superposition of the spectra on the protein alone (red) and the spectra at the end of titration with a molar ratio fHbp:FeIII(Ent)$^{3-}$ 1:2. **B)** Titration of fHbp with FeIII(Yers)$^{3-}$. Superposition of the spectra on the protein alone (red) and the spectra at the end of titration with a molar ratio fHbp:FeIII(Yers)$^{3-}$ 1:5.

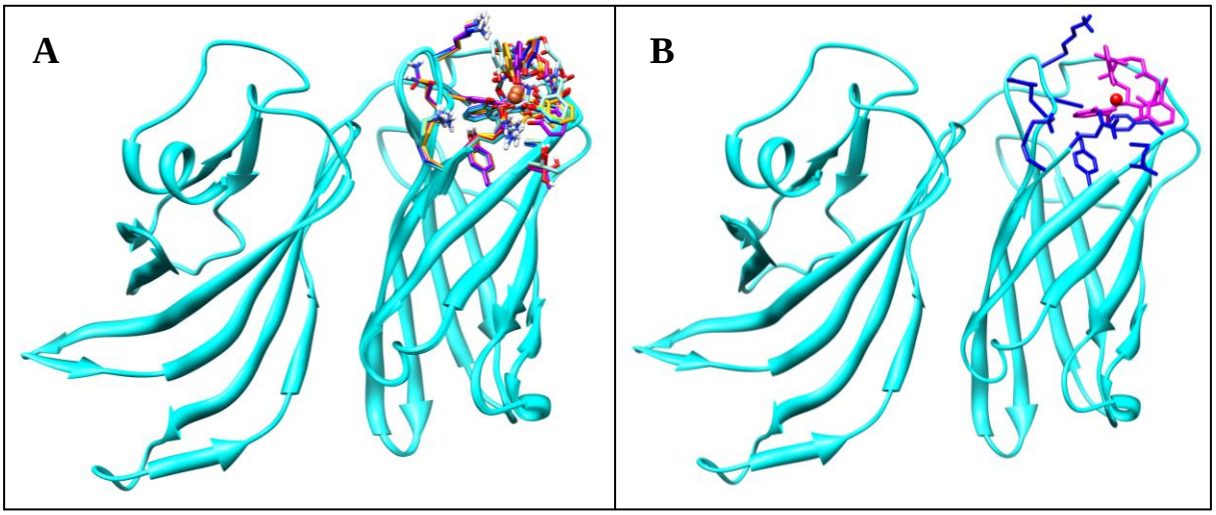


Figure 4. Complex between fHbp and FeIII(Ent)$^{3-}$ obtained with HADDOCK2.0. A) Superposition of the best ten complex structure obtained by HADDOCK2.0. **B)** Focalizing on one complex structure.

Discussion

fHbp, a subcapsular membrane bound protein, has gained importance as a meningococcal surface antigen because it is a component of several different protein-based candidate vaccines against meningococcal serogroup B strains (135, 136). The ability to bind human factor H (hfH) without affecting its function, allows fHbp to play a crucial role in the evasion of the host defence mediated by the alternative complement pathway. NMR and X-ray resolution of the three-dimensional structure of meningococcal fHbp revealed a lipocalin-like fold in the Carboxyl-terminal domain, which might suggest a role in iron uptake. We found that fHbp interacts with enterobactin, a xenosiderophore produced by *E.coli*, *in vitro.* Our data identified the region interacting with enterobactin not in overlap with the fH contact area. This finding, which was confirmed by multiple experiments, strongly suggests that fHbp plays an important role in iron acquisition for the meningococcus and in the survival of the microorganism in the host during pathogenesis.

Reference List

1. Del Giudice G, Rappuoli R (1999) Genetically derived toxoids for use as vaccines and adjuvants. *Vaccine* 17 Suppl 2:S44-S52.

2. Andre FE (1990) Overview of a 5-year clinical experience with a yeast-derived hepatitis B vaccine. *Vaccine* 8 Suppl:S74-S78.

3. Pizza M *et al.* (1989) Mutants of pertussis toxin suitable for vaccine development. *Science* 246:497-500.

4. Nencioni L *et al.* (1990) Characterization of genetically inactivated pertussis toxin mutants: candidates for a new vaccine against whooping cough. *Infect. Immun.* 58:1308-1315.

5. Fleischmann RD *et al.* (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 269:496-512.

6. Medini D *et al.* (2008) Microbiology in the post-genomic era. *Nat. Rev. Microbiol.* 6:419-430.

7. Rappuoli R (2001) Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 19:2688-2691.

8. Stephens DS, Greenwood B, Brandtzaeg P (2007) Epidemic meningitis, meningococcaemia, and Neisseria meningitidis. *Lancet* 369:2196-2210.

9. Stephens DS (2007) Conquering the meningococcus. *FEMS Microbiol. Rev.* 31:3-14.

10. Pizza M *et al.* (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287:1816-1820.

11. Tettelin H *et al.* (2000) Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. *Science* 287:1809-1815.

12. Giuliani MM *et al.* (2006) A universal vaccine for serogroup B meningococcus. *Proc. Natl. Acad. Sci. U. S. A* 103:10834-10839.

13. Comanducci M *et al.* (2002) NadA, a novel vaccine candidate of Neisseria meningitidis. *J. Exp. Med.* 195:1445-1454.

14. Litt DJ *et al.* (2004) Putative vaccine antigens from Neisseria meningitidis recognized by serum antibodies of young children convalescing after meningococcal disease. *J. Infect. Dis.* 190:1488-1497.

15. Masignani V *et al.* (2003) Vaccination against Neisseria meningitidis using three variants of the lipoprotein GNA1870. *J. Exp. Med.* 197:789-799.

16. Giuliani MM *et al.* (2005) The region comprising amino acids 100 to 255 of Neisseria meningitidis lipoprotein GNA 1870 elicits bactericidal antibodies. *Infect. Immun.* 73:1151-1160

17. Cantini F *et al.* (2009) Solution structure of the factor H binding protein, a survival factor and protective antigen of neisseria meningitidis. *J. Biol. Chem.* 284:9022-9026.

18. Madico G *et al.* (2006) The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance. *J. Immunol.* 177:501-510.

19. Welsch JA *et al.* (2003) Antibody to genome-derived neisserial antigen 2132, a Neisseria meningitidis candidate vaccine, confers protection against bacteremia in the absence of complement-mediated bactericidal activity. *J. Infect. Dis.* 188:1730-1740.

20. Shields B (2001) Prevnar (heptavalent pneumococcal conjugate vaccine): disease prevention in infants and children. *J. Pediatr. Health Care* 15:203-208.

21. Beall B *et al.* (2006) Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. *J. Clin. Microbiol.* 44:999-1017.

22. Wizemann TM *et al.* (2001) Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. *Infect. Immun.* 69:1593-1598.

23. Barocchi MA *et al.* (2006) A pneumococcal pilus influences virulence and host inflammatory responses. *Proc. Natl. Acad. Sci. U. S. A* 103:2857-2862.

24. Hilleringmann M *et al.* (2008) Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A. *PLoS. Pathog.* 4:e1000026.

25. Falker S *et al.* (2008) Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. *Mol. Microbiol.* 70:595-607.

26. Gianfaldoni C *et al.* (2007) Streptococcus pneumoniae pilus subunits protect mice against lethal challenge. *Infect. Immun.* 75:1059-1062.

27. Bagnoli F *et al.* (2008) A second pilus type in Streptococcus pneumoniae is prevalent in emerging serotypes and mediates adhesion to host cells. *J. Bacteriol.* 190:5480-5492.

28. Rappuoli R (2007) Bridging the knowledge gaps in vaccine design. *Nat. Biotechnol.* 25:1361-1366.

29. Serruto D, Rappuoli R (2006) Post-genomic vaccine development. *FEBS Lett.* 580:2985-2992.

30. Dormitzer PR, Ulmer JB, Rappuoli R (2008) Structure-based antigen design: a strategy for next generation vaccines. *Trends Biotechnol.* 26:659-667.

31. Lundstrom K (2007) Structural genomics and drug discovery. *J. Cell Mol. Med.* 11:224-238.

32. Todd AE, Marsden RL, Thornton JM, Orengo CA (2005) Progress of structural genomics initiatives: an analysis of solved target structures. *J. Mol. Biol.* 348:1235-1260.

33. Kim CU *et al.* (1997) Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* 119:681-690.

34. Kaldor SW *et al.* (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* 40:3979-3985.

35. Scarselli M *et al.* (2011) Rational design of a meningococcal antigen inducing broad protective immunity. *Sci. Transl. Med.* 3:91ra62.

36. Aue WP, Bartholdi E, Ernst RR (1976) Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* 64:2229-2235.

37. Wider G *et al.* (1984) Homonuclear Two-Dimensional 1H NMR of Proteins. Experimental Procedures. *J. Magn. Reson.* 56:207-234.

38. Wider G (1998) Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. *Progr. NMR Spectrosc.* 32:193-275.

39. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).

40. Kumar A, Ernst RR, Wüthrich K (1980) *Biochem. Biophys. Res. Commun.* 95:1104.

41. Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* 89:496-514.

42. Pervushin K (2000) Impact of tranverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q. Rev. Biophys.* 33:161-197.

43. Salzmann M *et al.* (1999) TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* 121:844-848.

44. Pervushin KV, Wider G, Riek R, Wüthrich K (1999) The 3D NOESY-[(1)H,(15)N,(1)H]-ZQ-TROSY NMR experiment with diagonal peak suppression. *Proc. Natl. Acad. Sci. USA* 96:9607-9612.

45. Kay LE *et al.* (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chains ^{1}H and ^{13}C correlations in H_2O samples of proteins. *J. Magn. Reson. Ser. B* 101:333-337.

46. Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* 222:311-333.

47. Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31:1647-1651.

48. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* 44:213-223.

49. Cordier F, Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue 3hJNC' scalar couplings. *J. Am. Chem. Soc.* 121:1601-1602.

50. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci. USA* 92:9279-9283.

51. Tjandra N, Grzesiek S, Bax A (1996) Magnetic field dependence of nitrogen-proton J splittings in 15N-enriched human Ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc.* 118:6264-6272.

52. Macura S, Ernst RR (1980) Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy. *Mol. Phys.* 41:95.

53. Karplus M (1963) Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* 85:2870-2871.

54. Herrmann T, Guntert P, Wuthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* 24:171-189.

55. Herrmann T, Guntert P, Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 319:209-227.

56. Lopez-Mendez B, Guntert P (2006) Automated protein structure determination from NMR spectra. *J. Am. Chem. Soc.* 128:13112-13122.

57. Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A. *et al.* AMBER 10. (8.0). 2008. San Francisco, CA, University of California.
Ref Type: Computer Program

58. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8:52-6, 29.

59. Laskowski RA *et al.* (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8:477-486.

60. Nabuurs SB *et al.* (2003) Quantitative evaluation of experimental NMR restraints. *J. Am. Chem. Soc.* 125:12026-12034.

61. Abragam, A. (1961) *The Principles of Nuclear Magnetism* (Oxford University Press, Oxford).

62. Wagner G (1993) NMR relaxation and protein mobility. *Curr. Opin. Struct. Biol.* 3:748-754.

63. Peng JW, Wagner G (1992) Mapping of spectral density function using heteronuclear NMR relaxation measurements. *J. Magn. Reson.* 98:308-332.

64. Farrow NA *et al.* (1995) Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR* 6:153-162.

65. Ishima R, Nagayama K (1995) Protein backbone dynamics revealed by quasi spectral density function analysis af amide N-15 nuclei. *Biochemistry* 34:3162-3171.

66. Lipari G, Szabo A (1982) Model-Free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546-4559.

67. Clore GM *et al.* (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112:4989-4991.

68. Mulder FA, Van Tilborg PJ, Kaptein R, Boelens R (1999) Microsecond time scale dynamics in the RXR DNA-binding domain from a combination of spin-echo and off-resonance rotating frame relaxation measurements. *J. Biomol. NMR* 13:275-288.

69. Palmer AG, III, Williams J, McDermott A (1996) Nuclear Magnetic Resonance Studies of Biopolymer Dynamics. *J. Phys. Chem.* 100:13293-13310.

70. Fletcher MA, Fritzell B (2007) Brief review of the clinical effectiveness of PREVENAR against otitis media. *Vaccine* 25:2507-2512.

71. Kim KS (2010) Acute bacterial meningitis in infants and children. *Lancet Infect. Dis.* 10:32-42.

72. O'Brien KL *et al.* (2009) Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet* 374:893-902.

73. Pelton SI, Leibovitz E (2009) Recent advances in otitis media. *Pediatr. Infect. Dis. J.* 28:S133-S137.

74. Ryan MW, Antonelli PJ (2000) Pneumococcal antibiotic resistance and rates of meningitis in children. *Laryngoscope* 110:961-964.

75. van der PT, Opal SM (2009) Pathogenesis, treatment, and prevention of pneumococcal pneumonia. *Lancet* 374:1543-1556.

76. Melegaro A, Gay NJ, Medley GF (2004) Estimating the transmission parameters of pneumococcal carriage in households. *Epidemiol. Infect.* 132:433-441.

77. Melegaro A *et al.* (2010) Dynamic models of pneumococcal carriage and the impact of the Heptavalent Pneumococcal Conjugate Vaccine on invasive pneumococcal disease. *BMC. Infect. Dis.* 10:90.

78. Millar EV *et al.* (2009) Nasopharyngeal carriage of Streptococcus pneumoniae in Navajo and White Mountain Apache children before the introduction of pneumococcal conjugate vaccine. *Pediatr. Infect. Dis. J.* 28:711-716.

79. Dramsi S *et al.* (2006) Assembly and role of pili in group B streptococci. *Mol. Microbiol.* 60:1401-1413.

80. Lauer P *et al.* (2005) Genome analysis reveals pili in Group B Streptococcus. *Science* 309:105.

81. LeMieux J, Hava DL, Basset A, Camilli A (2006) RrgA and RrgB are components of a multisubunit pilus encoded by the Streptococcus pneumoniae rlrA pathogenicity islet. *Infect. Immun.* 74:2453-2456.

82. Mora M *et al.* (2005) Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc. Natl. Acad. Sci. U. S. A* 102:15641-15646.

83. Nelson AL *et al.* (2007) RrgA is a pilus-associated adhesin in Streptococcus pneumoniae. *Mol. Microbiol.* 66:329-340.

84. Telford JL *et al.* (2006) Pili in gram-positive pathogens. *Nat. Rev. Microbiol.* 4:509-519.

85. Ton-That H, Schneewind O (2004) Assembly of pili in Gram-positive bacteria. *Trends Microbiol.* 12:228-234.

86. El Mortaji L *et al.* (2010) Stability and assembly of pilus subunits of Streptococcus pneumoniae. *J. Biol. Chem.* 285:12405-12415.

87. Manzano C *et al.* (2008) Sortase-mediated pilus fiber biogenesis in Streptococcus pneumoniae. *Structure.* 16:1838-1848.

88. Hilleringmann M *et al.* (2009) Molecular architecture of Streptococcus pneumoniae TIGR4 pili. *EMBO J.* 28:3921-3930.

89. Izore T *et al.* (2010) Structural basis of host cell recognition by the pilus adhesin from Streptococcus pneumoniae. *Structure.* 18:106-115.

90. Spraggon G *et al.* (2010) Supramolecular organization of the repetitive backbone unit of the Streptococcus pneumoniae pilus. *PLoS. ONE.* 5:e10919.

91. Budzik JM *et al.* (2009) Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U. S. A* 106:19992-19997.

92. Kang HJ *et al.* (2009) The Corynebacterium diphtheriae shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc. Natl. Acad. Sci. U. S. A* 106:16967-16971.

93. Alegre-Cebollada J, Badilla CL, Fernandez JM (2010) Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. *J. Biol. Chem.* 285:11235-11242.

94. Kang HJ, Baker EN (2009) Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of Streptococcus pyogenes. *J. Biol. Chem.* 284:20729-20737.

95. Guttilla IK *et al.* (2009) Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria. *J. Bacteriol.* 191:5603-5612.

96. Ton-That H, Schneewind O (2003) Assembly of pili on the surface of Corynebacterium diphtheriae. *Mol. Microbiol.* 50:1429-1438.

97. Kang HJ *et al.* (2007) Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science* 318:1625-1628.

98. Moschioni M *et al.* (2010) The two variants of the Streptococcus pneumoniae pilus 1 RrgA adhesin retain the same function and elicit cross-protection in vivo. *Infect. Immun.* 78:5033-5042.

99. Alloing G, Martin B, Granadel C, Claverys JP (1998) Development of competence in Streptococcus pneumonaie: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. *Mol. Microbiol.* 29:75-83.

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

101. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14:51-32.

102. Grzesiek S, Bax A (1993) The importance of not saturating H2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* 115:12593-12594.

103. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559-4570.

104. Dosset P, Hus JC, Blackledge M, Marion D (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* 16:23-28.

105. Goddard TD, Huang CC, Ferrin TE (2007) Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* 157:281-287.

106. Meng EC *et al.* (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC. Bioinformatics.* 7:339.

107. Goddard TD, Huang CC, Ferrin TE (2005) Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. *Structure.* 13:473-482.

108. Holm L, Sander C (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* 20:478-480.

109. Mittag T *et al.* (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure.* 18:494-506.

110. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197-208.

111. LeMieux J, Woody S, Camilli A (2008) Roles of the sortases of Streptococcus pneumoniae in assembly of the RlrA pilus. *J. Bacteriol.* 190:6002-6013.

112. Lambris JD, Ricklin D, Geisbrecht BV (2008) Complement evasion by human pathogens. *Nat. Rev. Microbiol.* 6:132-142.

113. Meri S, Jordens M, Jarva H (2008) Microbial complement inhibitors as vaccines. *Vaccine* 26 Suppl 8:I113-I117.

114. Ripoche J *et al.* (1988) Two populations of complement factor H differ in their ability to bind to cell surfaces. *Biochem. J.* 253:475-480.

115. Perkins SJ, Haris PI, Sim RB, Chapman D (1988) A study of the structure of human complement component factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods. *Biochemistry* 27:4004-4012.

116. Barlow PN *et al.* (1993) Solution structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol.* 232:268-284.

117. Barlow PN *et al.* (1992) Solution structure of the fifth repeat of factor H: a second example of the complement control protein module. *Biochemistry* 31:3626-3634.

118. Schneider MC *et al.* (2009) Neisseria meningitidis recruits factor H using protein mimicry of host carbohydrates. *Nature* 458:890-893.

119. Prosser BE *et al.* (2007) Structural basis for complement factor H linked age-related macular degeneration. *J. Exp. Med.* 204:2277-2283.

120. Hubbard, S. J. & Thornton, J. M. NACCESS. 1993. Department Biochemistry and Molecular Biology, University College, London.
Ref Type: Computer Program

121. Beernink PT *et al.* (2009) A region of the N-terminal domain of meningococcal factor H-binding protein that elicits bactericidal antibody across antigenic variant groups. *Mol. Immunol.* 46:1647-1653.

122. Scarselli M *et al.* (2009) Epitope mapping of a bactericidal monoclonal antibody against the Factor H Binding Protein of Neisseria meningitidis. *J. Mol. Biol.* 386:97-108.

123. Borrow R, Andrews N, Goldblatt D, Miller E (2001) Serological basis for use of meningococcal serogroup C conjugate vaccines in the United Kingdom: reevaluation of correlates of protection. *Infect. Immun.* 69:1568-1573.

124. Girard MP, Preziosi MP, Aguado MT, Kieny MP (2006) A review of vaccine research and development: meningococcal disease. *Vaccine* 24:4692-4700.

125. Snape MD *et al.* (2008) Immunogenicity of a tetravalent meningococcal glycoconjugate vaccine in infants: a randomized controlled trial. *JAMA* 299:173-184.

126. Flower DR (1996) The lipocalin protein family: structure and function. *Biochem. J.* 318 (Pt 1):1-14.

127. Neilands JB (1995) Siderophores: structure and function of microbial iron transport compounds. *J. Biol. Chem.* 270:26723-26726.

128. Fohn MJ *et al.* (1987) Human immunoglobulin G antibody response to the major gonococcal iron-regulated protein. *Infect. Immun.* 55:3065-3069.

129. Cornelis P (2010) Iron uptake and metabolism in pseudomonads. *Appl. Microbiol. Biotechnol.* 86:1637-1645.

130. Zawadzka AM *et al.* (2009) Characterization of a Bacillus subtilis transporter for petrobactin, an anthrax stealth siderophore. *Proc. Natl. Acad. Sci. U. S. A* 106:21854-21859.

131. Chen CJ *et al.* (1996) Identification and purification of a hemoglobin-binding outer membrane protein from Neisseria gonorrhoeae. *Infect. Immun.* 64:5008-5014.

132. Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125:1731-1737.

133. de Vries SJ *et al.* (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* 69:726-733.

134. Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.* 60:1355-1363.

135. Sadarangani M, Pollard AJ (2010) Serogroup B meningococcal vaccines-an unfinished story. *Lancet Infect. Dis.* 10:112-124.

136. Koeberling O, Giuntini S, Seubert A, Granoff DM (2009) Meningococcal outer membrane vesicle vaccines derived from mutant strains engineered to express factor H binding proteins from antigenic variant groups 1 and 2. *Clin. Vaccine Immunol.* 16:156-162.

RINGRAZIAMENTI

Finalmente dopo tanti anni di studio riesco ad arrivare a questo traguardo, quello per me più importante che segna la fine di un lungo percorso e l'inizio di un nuovo capitolo della mia vita. In questo giorno particolare vorrei ringraziare in primo luogo i miei genitori, mia sorella e mio fratello per il supporto e per tutta la pazienza avuta in questi anni, per aver subito i miei sfoghi, i miei momenti no e i miei lunghi silenzi mentre mi richiudevo in camera a studiare. Ringrazio i miei genitori per aver cresciuto così come sono e per avermi dato la libertà di scegliere e di seguire i miei sogni pur sapendo quanto difficile sarebbe stato il mio percorso e quanto ancora dovrà esserlo.
Ringrazio di cuore il mio amore Diego per l'infinita pazienza e per il suo modo di rendere ogni cosa, dalla più bella alla più brutta, un buon motivo per sorridere. Ringrazio la sua mamma sprint "Evelin" per tutto l'affetto e per il sostegno regalatomi nell'ultimo anno, per avermi accolto come una figlia nella sua casa.

Con affetto ringrazio la mia Prof. Delia Picone per avermi aiutato a trovare questo percorso e per la fiducia dimostratami in questi anni.

Ringrazio infine per il supporto scientifico e "burocratico" Matthew Bottomley, capo dell'Unità di Biologia Strutturale della Novartis Vaccines and Diagnostics ed i miei co-relatori Dott. Maria Scarselli e Dott. Marcello Merola.

Infine vorrei ringraziare tutti quelli che in qualche modo, sia nel bene che nel male, hanno contribuito al conseguimento di questo obiettivo.

PUBLICATIONS

"Structural and Functional Characterization of the *Streptococcus pneumoniae* RrgB Pilus Backbone D1 Domain" Gentile MA et al., 2011, *THE JOURNAL OF BIOLOGICAL CHEMISTRY,* 286:14588–14597

"Investigation on the interaction between factor H binding protein and factor H" *To be submitted*

"Neisseria meningitidis* fHbp interacts with xenosiderophores *in vitro" *To be submitted*

CONGRESS AND COURSES

29 July – 5 August 2011
BNMRZ, Department Chemie, TU München; Garching (Germany)
"Structure, dynamics and function of biomacromolecules by solution NMR"
EMBO Practical Course

8 September 2010
Novartis Pharma Basel (Switzerland)
Talk title "Protein NMR and its application to epitope mapping in the optimization of novel vaccine"

4-9 July 20010
"World Wide Magnetic Resonance" Florence (Italy)
EUROMAR 2010 and 17th ISMAR Conference
Poster: "Solution structure of the protective D1 domain of *Streptococcus pneumoniae* RrgB pilus subunit"

31 August – 4 September 2009
Foundation for Biotechnologies Turin (Italy)
"Nuclear Magnetic Resonance School" Advanced Course
Theoretical and Practical courses

External experience:
1 March 2011 - 31 May 2011
Bijvoet Center for Biomolecular Research (Netherlands)
Structural characterization of the *Neisseria meningitides* antigen NMB1030

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 16, pp. 14588–14597, April 22, 2011

Structural and Functional Characterization of the *Streptococcus pneumoniae* RrgB Pilus Backbone D1 Domain*[S]

Received for publication, November 11, 2010, and in revised form, February 16, 2011 Published, JBC Papers in Press, March 2, 2011, DOI 10.1074/jbc.M110.202739

Maria Antonietta Gentile[‡], Sara Melchiorre[‡], Carla Emolo[‡], Monica Moschioni[‡], Claudia Gianfaldoni[‡], Laura Pancotto[‡], Ilaria Ferlenghi[‡], Maria Scarselli[‡], Werner Pansegrau[‡], Daniele Veggi[‡], Marcello Merola[‡§], Francesca Cantini[¶], Paolo Ruggiero[‡], Lucia Banci[¶1], and Vega Masignani[‡2]

From the [‡]Novartis Vaccines and Diagnostics Research Center, Via Fiorentina 1, Siena 53100, Italy, the [¶]Magnetic Resonance Center, Department of Chemistry, University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, and the [§]Università di Napoli Federico II, Monte Sant'Angelo, Via Cintia, Napoli 80126, Italy

Streptococcus pneumoniae **expresses on its surface adhesive pili, involved in bacterial attachment to epithelial cells and virulence. The pneumococcal pilus is composed of three proteins, RrgA, RrgB, and RrgC, each stabilized by intramolecular isopeptide bonds and covalently polymerized by means of intermolecular isopeptide bonds to form an extended fiber. RrgB is the pilus scaffold subunit and is protective *in vivo* in mouse models of sepsis and pneumonia, thus representing a potential vaccine candidate. The crystal structure of a major RrgB C-terminal portion featured an organization into three independently folded protein domains (D2–D4), whereas the N-terminal D1 domain (D1) remained unsolved. We have tested the four single recombinant RrgB domains in active and passive immunization studies and show that D1 is the most effective, providing a level of protection comparable with that of the full-length protein. To elucidate the structural features of D1, we solved the solution structure of the recombinant domain by NMR spectroscopy. The spectra analysis revealed that D1 has many flexible regions, does not contain any intramolecular isopeptide bond, and shares with the other domains an Ig-like fold. In addition, we demonstrated, by site-directed mutagenesis and complementation in *S. pneumoniae*, that the D1 domain contains the Lys residue (Lys-183) involved in the formation of the intermolecular isopeptide bonds and pilus polymerization. Finally, we present a model of the RrgB protein architecture along with the mapping of two surface-exposed linear epitopes recognized by protective antisera.**

Streptococcus pneumoniae is an important human pathogen responsible for diseases such as otitis media, pneumonia, sepsis, and meningitis (1–6). However, *S. pneumoniae* is also a common inhabitant of the respiratory tract of children and healthy adults. This carriage state could represent a risk factor for the development of respiratory diseases but also the source for pneumococcal transmission to other individuals (7–9). Like most streptococci, *S. pneumoniae* decorates its surface with long filaments known as pili (10–14). Pneumococcal pili have previously been associated with virulence and the capability of the microorganism to adhere better to epithelial cells and to colonize the nasopharynx (10, 15, 16). The pneumococcal pilus is a multimeric structure consisting of three proteins (RrgA, RrgB, and RrgC) polymerized by three sortase enzymes (SrtC1, SrtC2, and SrtC3) through the formation of covalent intermolecular isopeptide bonds (17–21). In particular, multiple copies of RrgB are polymerized to form the scaffold of the pilus, whereas the major adhesin, RrgA, and the putative anchor, RrgC, are localized at the tip and at the base of the pilus, respectively (15, 22, 23).

Recently, the structure of a major portion of RrgB (residues 184–627) was solved at a 1.6 Å resolution (24) and revealed an organization into three independently folded IgG-like domains (D2, D3, and D4, residues 184–326, 326–446, and 446–627, respectively). On the contrary, the structure of the RrgB N-terminal region (D1, residues 1–184), likely constituting a fourth independently folded domain, remained unsolved due to the failure to obtain the crystals of the full-length (FL)[3] RrgB (24). Interestingly, each of the D2, D3, and D4 domains is stabilized by one intramolecular isopeptide bond. These covalent linkages, formed between Lys and Asn residues, have been found in other pilus proteins (19, 25–28) and are thought to play a role similar to that of disulfide bonds; they confer in fact a rigid molecular architecture to the pili and make them less susceptible to proteolytic cleavage (Fig. 1).

In the pilus backbone assembly RrgB molecules are linked together by sortases through intermolecular isopeptide bond formation between a Thr within the C-terminal LP*X*TG motif of a molecule and a Lys located at the N terminus of the following molecule (18, 26, 29) (Fig. 1). In *Corynebacterium diphtheriae*, where the general principles of pilus assembly were first

* This work was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (Fondi per gli Investimenti della Ricerca di Base-Proteomica RBRN07BMCT) and Integrated Structural Biology Infrastructure for Europe Contract 211252. Because some of the authors are employees of Novartis Vaccines, there are competing financial interests.

[S] The on-line version of this article (available at http://www.jbc.org) contains Materials and Methods, supplemental Tables S1–S6, and supplemental Figs. S1–S6.

The atomic coordinates and structure factors (code 2L40) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

The NMR chemical shifts have been deposited in the BioMagResBank (accession no. 17246).

[1] To whom correspondence may be addressed. Tel.: 39-055-4574263; Fax: 39-055-4574253; E-mail: banci@cerm.unifi.it.

[2] To whom correspondence may be addressed. Tel.: 39-0577-243319; Fax: 39-0577-243564; E-mail: vega.masignani@novartis.com.

[3] The abbreviations used are: FL, full-length; BisTris, bis(2-hydroxyethyl)iminotris(hydroxymethyl)methane; cryo-EM, cryoelectron microscopy; cfu, colony-forming units; HSQC, heteronuclear single quantum coherence; PDB, Protein Data Bank; r.m.s.d., root mean square deviation; WB, Western blot.

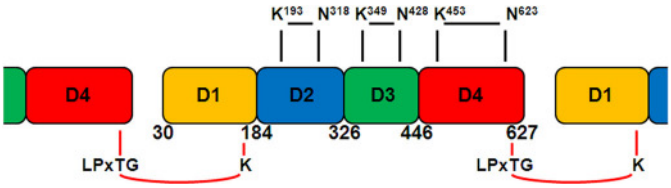


FIGURE 1. **Schematic representation of pilus backbone protein RrgB.** Pilus scaffold is composed by multiples copies of RrgB protein in a head-to-tail arrangement. Pilus polymerization occurs through intermolecular isopeptide bonds (*red*), whereas each RrgB protein is stabilized by intramolecular isopeptide bonds (*black*). Lys-183 as a residue involved in the intermolecular bond has been identified in the present work.

established, the functional Lys is located within a conserved YPKN "pilin" motif (18, 27, 30). Nevertheless, this sequence is not absolutely required for polymerization as demonstrated by studies on the Spy0128 pilin of *Streptococcus pyogenes*, where the lysine forming the intermolecular isopeptide bond and responsible for pilus polymerization is located into ^{159}GSKVPI^{164} motif even though the YPKN pilin motif is also present (26, 31).

RrgB, along with the other two pilus proteins RrgA and RrgC, was previously shown to confer protection in mouse models of infection and therefore is regarded as a potential candidate for a new generation of protein-based vaccines (32, 33). We have investigated the protective ability of the single recombinant D1, D2, D3, and D4 domains of RrgB in a mouse model of sepsis, and here we provide evidence that D1 is the most protective, followed by D4. Furthermore, we present the solution structure of the recombinant D1 obtained by NMR spectroscopy and show that Lys-183 of D1 is engaged in the intermolecular isopeptide bond formation during pilus polymerization. Finally, we propose a possible model of the entire RrgB molecule and show the positions of two linear epitopes possibly involved in the protection mechanism.

MATERIALS AND METHODS

Bacterial Strains and Cultures—For the animal experiments, the *S. pneumoniae* TIGR4 (serotype 4) strain was used. Bacteria were grown, frozen in aliquots at −80 °C, and titrated as already reported (32). Immediately prior to challenge, frozen aliquots were thawed and diluted in PBS to reach the working concentration.

Cloning and Protein Expression and Purification—Standard recombinant DNA techniques were used to construct the expression plasmids (pET21b$^+$; Novagen) and to express and purify the recombinant C-terminal His-tagged proteins (for details, see supplemental Materials and Methods, and the primers used are listed in supplemental Table S1). The affinity-purified proteins were subsequently used to immunize CD1 mice or rabbits for antibody generation (Charles River Laboratory) and BALB/c mice to evaluate the protective efficacy.

Complementation Plasmids—Wild-type or mutant *rrgB* genes were amplified from chromosomal DNA of TIGR4 strain by PCR by using the primers listed in supplemental Table S2; point mutations were introduced by overlap extension PCR by using specific primers (supplemental Table S2). PCR products were then cloned into the complementation plasmid pMU1328 between the BamHI and SalI restriction sites (34). Expression of RrgB or RrgB mutated forms was under control of the erythromycin constitutive promoter (Pc) which was amplified with the primers listed in supplemental Table S2 and cloned immediately upstream *rrgB* (EcoRI, BamHI). All plasmids were confirmed by sequencing.

Generation of rrgB Deletion Mutants and rrgB Complementants—A TIGR4 Δ*rrgB* isogenic mutant was generated by allelic exchange. Fragments of ~500 bp upstream and downstream the target gene were amplified by PCR (oligonucleotides are listed in supplemental Table S2) and spliced into a kanamycin resistance cassette by using overlap extension PCR; the PCR fragments were then cloned into pGEMt (Promega) and transformed in *S. pneumoniae* with conventional methods (35). To select the bacteria in which the target gene was replaced with the resistance cassette, bacteria were plated on blood-agar plates with kanamycin (500 μg/ml). The presence of the isogenic mutation was confirmed by PCR and Western blot (WB) analysis. To obtain RrgB-complemented mutants, pMU1328 plasmids containing WT *rrgB* or *rrgB* mutated forms were transformed into TIGR4 Δ*rrgB* with conventional methods. Transformants selection was performed by supplementing media with kanamycin (500 μg/ml) and erythromycin (1 μg/ml). The complemented mutants were then analyzed by PCR; expression of FL WT RrgB or RrgB mutants was detected by WB analysis of whole cell lysates.

SDS-PAGE and Western Blot Analysis—SDS-PAGE analysis was performed using NuPAGE™ 4–12% BisTris gradient gels (Invitrogen) according to the manufacturer's instructions. Hi-Mark™ prestained HMW (Invitrogen) served as protein standard. Gels were processed for WB analysis by using standard protocols. Mouse polyclonal antibodies raised against recombinant His-tagged RrgB were used at 1/3000 dilution. Secondary goat anti-mouse IgG alkaline phosphatase-conjugated antibodies (Promega) were used at 1/5000, and signals were developed by using Western Blue Stabilized Substrate for Alkaline Phosphatase (Promega).

Animal Experiments—Animal studies were done in compliance with the current law approved by the local Animal Ethics Committee and authorized by the Italian Ministry of Health. Female, 6-week-old, specific pathogen-free BALB/c mice (Charles River) received three intraperitoneal immunizations, 2 weeks apart. Each dose was composed of 20 μg of either the single RrgB domains or the FL RrgB, or of a combination of the four RrgB domains (D1+D2+D3+D4), 10 μg each, along with 400 μg of aluminum hydroxide as an adjuvant, in a final volume of 200 μl of PBS. Control animals received the same course of saline plus adjuvant. Ten days after the third immunization, samples of sera were obtained from each animal and pooled according to the immunization group to be used in passive serum transfer experiment. Two weeks after the third immunization, each mouse was challenged intraperitoneally with a mean dose of 1.6×10^2 cfu of TIGR4. Bacteremia was evaluated 24 h after challenge, and mortality course monitored for 10 days after challenge as already reported (32). The animals were euthanized when they exhibited defined humane end points that had been preestablished for the study in agreement with Novartis Animal Welfare Policies.

For passive protection experiments, 8-week-old mice were used. Fifteen minutes before TIGR4 intraperitoneal challenge (10^2 cfu/mouse), each mouse received intraperitoneally 50 μl of pooled mouse sera against recombinant D1 or D4, or of control sera obtained immunizing with adjuvant plus saline.

Statistical Analysis—Bacteremia and mortality course were analyzed by the Mann-Whitney *U* test. Survival rates were analyzed by Fisher's exact test. One-tailed or two-tailed tests were used to compare immunized groups with the control group or each other, respectively. Values of $p \leq 0.05$ were considered and referred to as significant. Values of $p \leq 0.1$ were referred to as a trend.

Flow Cytometry on Entire Bacteria—TIGR4 were grown in Todd-Hewitt yeast extract broth to an exponential phase (A_{600} = 0.25), fixed with 2% formaldehyde, and then stained with pooled mouse antisera raised against FL RrgB or RrgB domains at 1:400 dilution. Mouse IgG were detected with FITC-conjugated goat anti-mouse IgG (Jackson Laboratories) at 1:100 dilution, and bacterial staining was analyzed by using a FACS-Calibur cytometer (Becton Dickinson). Sera from mice immunized with PBS plus adjuvant were used as negative control.

ELISA—96-well MaxiSorp^{TM} flat-bottom plates (Nunc) were coated with 0.2 μg/well FL RrgB overnight at 4 °C. Plates were then washed three times with PBS/0.05% Tween 20 and saturated for 1 h at 37 °C with PBS/1% BSA. Following three washing steps with PBS/0,05% Tween 20, the plates were incubated for 2 h at 37 °C with serial dilutions of the pooled mouse sera. After another three washing steps, bound antigen-specific mouse IgGs were revealed with alkaline phosphatase-conjugated goat anti-mouse IgG (Sigma), followed by the phosphatase alkaline substrate *p*-nitrophenyl phosphate (Sigma). The intensity of color was quantified with an ELISA plate reader at A_{405}. The antibody titer was expressed as the log10 of the reciprocal of the serum dilution that gave A_{405} = 1.

PepScan Analysis—Arrayed peptides were synthesized *in situ* on glass fiber membranes. Membranes were conditioned by wetting with ethanol and washing three times for 5 min in TTBS (50 mM Tris-HCl, pH 7.0, 137 mM NaCl, 2.7 mM KCl, 0.05% Tween 20). After overnight blocking at 4 °C in MBS (2% dry milk in TTBS), membranes were incubated for 1.5 h at 37 °C with polyclonal antisera (1:3000 in MBS) followed by secondary goat anti-mouse IgG alkaline phosphatase-conjugated antibodies (1:5000 in MBS; Promega), and signals were developed by using Western Blue Stabilized Substrate for Alkaline Phosphatase. For image processing, membranes were scanned using an Epson V750 Pro scanner at 800 dpi, 48-bit color depth and with gamma 1.0 full linear response.

NMR Characterization of RrgB D1 Domain—Expression and purification of labeled D1 were carried out as described in the supplemental Materials and Methods. NMR spectra were acquired at 298 K on Avance 900, 800, 700, and 500 MHz Bruker spectrometers, all equipped with a triple resonance cryoprobe. The NMR experiments, used for the backbone and the aliphatic side chain resonances assignment recorded on $^{13}C/^{15}N$ and ^{15}N enriched samples or on unlabeled D1 samples, are summarized in supplemental Table S3. Backbone dihedral angle constraints were derived from ^{15}N, $^{13}C'$, $^{13}C\alpha$, $^{13}C\beta$, and Ha chemical shifts, using TALOS+ (36). Distance constraints for structure determination were obtained from ^{15}N-edited and ^{13}C-edited three-dimensional NOESY-HSQC. 3131 meaningful proton-proton distance restraints (supplemental Table S4), with 114 φ and 120 ψ backbone dihedral angles restraints were included in structure calculations. The exchangeability of the backbone amide hydrogen nuclei with solvent protons was investigated through a ^{1}H-^{15}N HSQC experiment performed on a protein sample dialyzed against deuterated buffer for 3 days. Hydrogen bond constraints for the slowly deuterium-exchanging amide protons of the β-strands were introduced at later steps of the structure calculations.

Structure calculations were performed through iterative cycles of CYANA-2.1 (37) followed by restrained energy minimization with the AMBER 10.0 Package in explicit water solvent (38). The quality of the structures was evaluated by using the iCING validation program (for details, see supplemental Table S4). The program MOLMOL was subsequently used for structure analysis (39).

The final bundle of 20 conformers of D1 has an average target function of 1.36 ± 0.13 (CYANA units). The average backbone r.m.s.d. value (over residues 28–183) is 0.71 ± 0.19Å, and the all-heavy atoms r.m.s.d. value is 0.96 ± 0.16. Per-residue r.m.s.d. values are shown in supplemental Fig. S1.

Heteronuclear Relaxation Data—The dynamic properties of D1 have been characterized experimentally through ^{15}N relaxation measurements. ^{15}N longitudinal and traverse relaxation rates (40) and $^{15}N\{^{1}H\}$ NOEs (41) were recorded at 298 K at 500 MHz, using a protein concentration of 0.8 mM.

The average backbone ^{15}N longitudinal R_1 and transversal R_2 relaxation rates and $^{15}N\{^{1}H\}$ NOE values are 1.45 ± 0.1 s^{-1}, 16.18 ± 1.5 s^{-1}, and 0.71 ± 0.04, respectively (supplemental Fig. S2). They are essentially homogeneous along the entire polypeptide sequence, with the exception of residues located at the C and N termini and three loop regions (56–69, 148–162, and 173–177). The correlation time for the molecule tumbling (τ_c), as estimated from the R_2/R_1 ratio, is 11.49 ± 1.9 ns, consistent with the molecular mass of the monomeric protein. The relaxation data were analyzed according to the model-free approach of Lipari and Szabo (42, 43) using TENSOR2 (44) (supplemental Fig. S2).

Rigid Body Fitting—The procedure used to accommodate the NMR D1 structure into the EM map of the whole pilus previously generated (24) followed the same procedures used for the fitting of the RrgBD2-D4 x-ray crystal structure and the D1 computer model into the EM pilus map (22). A preliminary rigid body fitting of the D2–D4 crystal fragment was performed by using CHIMERA (45–47) followed by a rigid body fitting of the D1 NMR coordinates into the leftover N-terminal apical volume of the pilus. The D1 NMR coordinates were first fitted manually using CHIMERA by placing as much of the atomic structure as possible into the EM density map, approximately in the position thought to be correct. This step was then followed by a rigid-body fitting using the "fit model in map" tool from CHIMERA. This tool calculates, for the selected atoms, a position that maximizes locally the sum of the densities. The evaluation of the correlation coefficient values, the resulting average map value at the fit atoms, the number of fits atoms outside

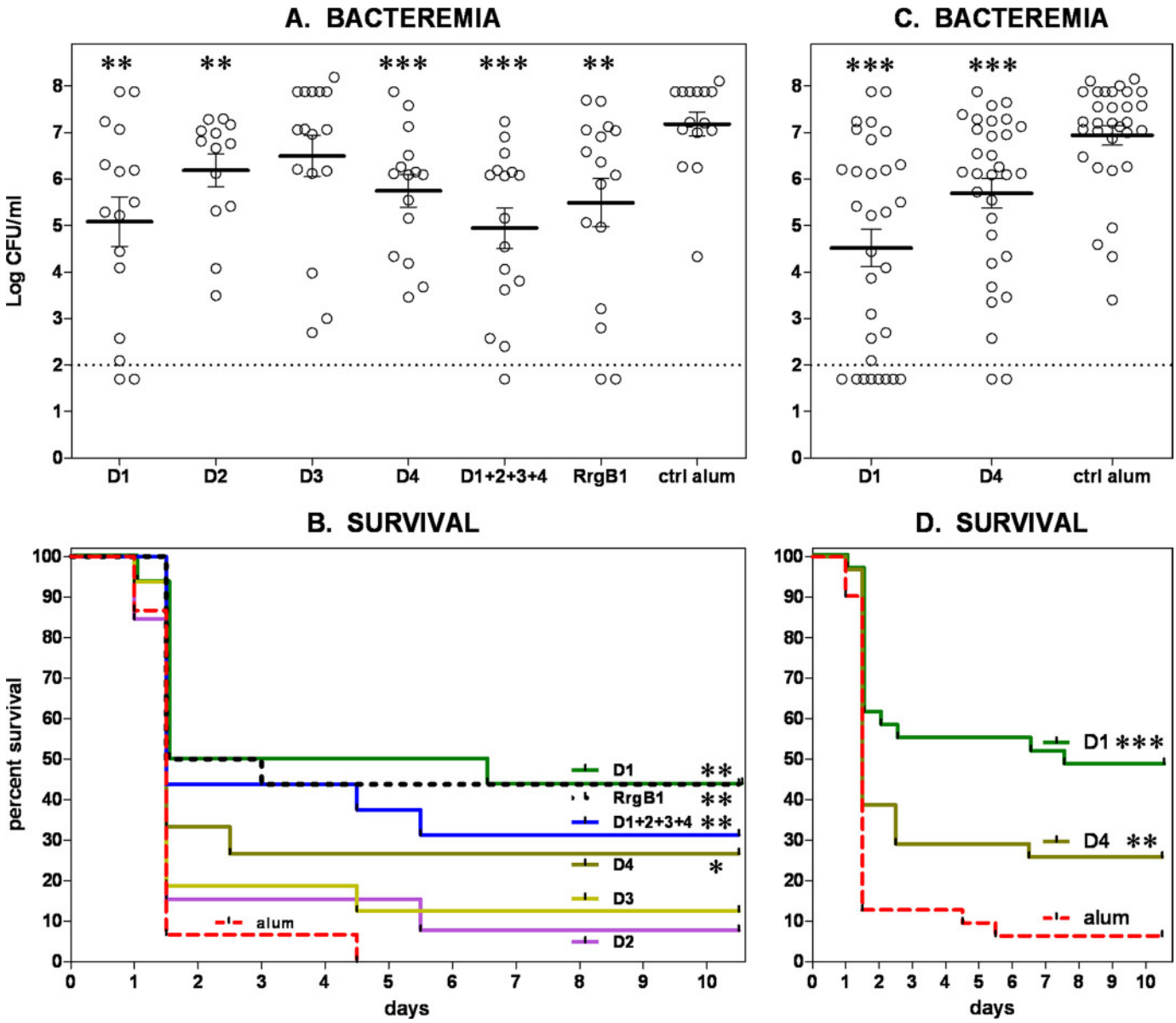


FIGURE 2. **RrgB domains are protective in active immunization experiments.** *A* and *C*, bacteremia. *Circles* represent the log cfu/ml of blood for individual animals; *horizontal bars* represent the mean value of the log cfu/ml ± S.E. for the group; the *dotted line* marks the detection limit (values under the *dotted line* correspond to animals in which no cfu were detected). The treatment groups are indicated on the *bottom*. *B* and *D*, survival. The survival course for each group is represented. The treatment groups are indicated close to each of the corresponding survival lines. *A* and *B*, $n = 13$ for D2 group, 15 for D4 and control (*alum ctrl*) groups, 16 for the remaining groups. *C* and *D*, $n = 31$ for each of the groups. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$.

the lowest value contour surface displayed were the parameters used to assess the fit of the D1 molecule.

RESULTS

Distinct Domains of RrgB Provide Protection in Active Immunization Experiments—The protective efficacy of each of the four RrgB domains or of a mixture of them was assessed in a mouse model of intraperitoneal challenge (TIGR4) in two distinct experiments, performed under the same conditions, which were combined to reach $n = 13–16$ mice/group. The results are summarized in Fig. 2, *A* and *B*, and detailed statistical analysis is provided in supplemental Table S5*A*.

All RrgB domains except D3 afforded significant protection against bacteremia (Fig. 2*A*), giving a reduction of the cfu geometric mean by 1–2 logs with respect to the controls. These values were similar to those afforded by the FL RrgB and by the combination of the four domains D1+D2+D3+D4. The results of mortality are reported in Fig. 2*B* and supplemental Table S5*A*. D1 and D4 conferred significant increase of survival time, similar to FL RrgB and the combination D1+D2+D3+D4. In particular, the median survival for the D1 group was 2.5 days higher respect to that of the control group (4 *versus* 1.5 days). At the end of the mortality observation, a significant survival rate was found for D1 (44% survival), D4 (27%), FL RrgB (44%), and D1+D2+D3+D4 (31%) groups.

N-terminal D1 Domain Is the Most Protective in Vivo—Among the single RrgB domains, D1 and D4 showed the most significant protective efficacy and were therefore analyzed further in a larger group of mice. Four different experiments, carried out under the same conditions, were combined to

reach $n = 31$ animals/group. The results are shown in Fig. 2, *C* and *D*, and detailed statistical analysis is provided in supplemental Table S5*B*. In terms of bacteremia (Fig. 2*C*), both D1 and D4 afforded highly significant protection, with a cfu geometric mean by 2.6 and 1.5 logs lower, respectively, than that of the control group, and 8 animals from the D1 group in which cfu were undetectable. The reduction of bacteremia was significantly superior in D1 than in the D4 group.

In terms of mortality course (Fig. 2*D*), both D1 and D4 conferred significant protection. The increase of survival time afforded by D1 showed a better trend than that of D4. In particular, only for the D1 group was the median survival time higher than that of the control group (7.5 *versus* 1.5 days). At the end of the mortality observation, the D1 group showed the highest survival rate, *i.e.* 45% *versus* 21% observed for D4. An evident difference of survival rates between D1 and D4 groups was observed.

The possible relevance of antibodies in the protection elicited by D1 and D4 was investigated by a passive serum transfer experiment, with groups of 8 mice. The results are shown in Fig. 3, and detailed statistical analysis is provided in supplemental Table S5*C*. Both anti-D1 and anti-D4 sera elicited significant protection against bacteremia (Fig. 3*A*), with cfu geometric means lower by 2.6 and 1.6 logs, respectively, than that of the control group. Only anti-D1 serum afforded significant protection against mortality, giving 100% survival rate, whereas santi-D4 gave a protective trend, with 60% survival rate. These results indicate that immunization with D1 domain and at a lower extent D4 domain, elicits functional antibodies that may play a role in the protection.

D1 and D4 Antisera Recognize the Native Pilus and Linear Epitopes within RrgB—To investigate the differences in protective efficacy exerted by the isolated RrgB domains with respect to the FL RrgB in the *in vivo* assays, mouse sera were tested for their capability of recognizing the native pilus and the recombinant FL RrgB. Sera were probed against entire TIGR4 bacteria by FACS analysis. Sera raised against D1, D4, and D1+D2+D3+D4 gave a fluorescence intensity comparable with that obtained with anti-FL RrgB, whereas D2- and D3-specific antisera recognized the native pilus less effectively. To explain the lower recognition efficacy observed with D2 and D3, the same sera were titrated in ELISA against the recombinant FL RrgB. As shown in supplemental Fig. S3, antibody titers elicited by D2 and D3 immunization toward FL RrgB were about 10 times lower than those obtained by immunization with D1, D4, D1+D2+D3+D4, and FL RrgB, consistent with the results obtained by FACS analysis on the polymerized native RrgB.

To gain more insights on the epitopes recognized by the protective D1 and D4 polyclonal antibodies, a PepScan approach, suitable for identifying linear epitopes, was applied. Overlapping 15-mer peptides with an offset of 5 residues were synthesized *in situ* on a glass fiber membrane. The library of peptides tested covered residues 25–190 of D1 and 444–628 of D4. Incubation with the anti-D1 serum (previously used in passive protection experiments) revealed a single linear epitope covering residues 40–59 (D1-1) (Fig. 4*B*), whereas anti-D4 serum

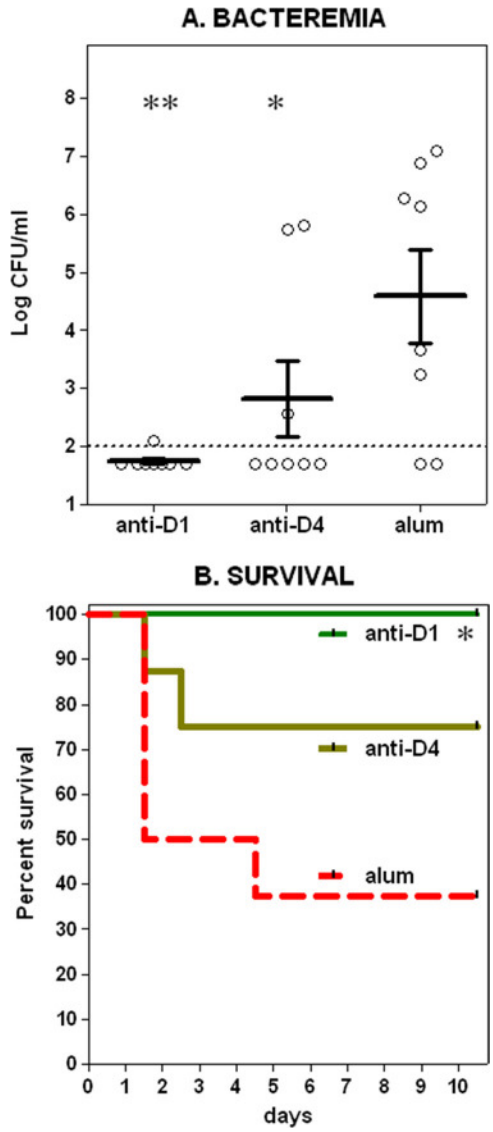


FIGURE 3. **Anti-RrgB D1 and D4 sera are protective in passive serum transfer experiments.** *Symbols* are described in the Fig. 2 legend. $n = 8$ for each of the groups.

detected a unique linear epitope (D4-1) spanning residues 494–508 (Fig. 4*C*).

Solution Structure of the D1 Domain—The solution structure of D1 was investigated by NMR spectroscopy. Its ^{1}H-^{15}N HSQC spectrum showed well dispersed resonances indicative of an overall well folded protein (supplemental Fig. S4). D1 showed a common IgG-like β sandwich fold (41 Å × 48 Å × 30 Å) and a topology of secondary structure elements drawn in Fig. 5. The core of the structure was formed by seven parallel and antiparallel β-strands: β_1(36–39), β_4(80–85), β_7(119–121), β_8(127–130), β_9(138–143), β_{10}(166–169), and β_{11}(178–180). These β-strands were arranged in two sheets (comprising β_1, β_8, β_{11} and β_4, β_7, β_9, β_{10}, respectively) packed against each other and flanked by two long segments (40–78, 87–117) located between strands β_1 and β_4 and strands β_4 and β_7, respectively. An α-helix (49–57), flanked by two short β-strands β_2 (42–44) and β_3 (73–75), was inserted within the first segment. Two additional β-strands β_5 (89–91) and β_6 (97–101) formed a β-hairpin structure inserted within the sec-

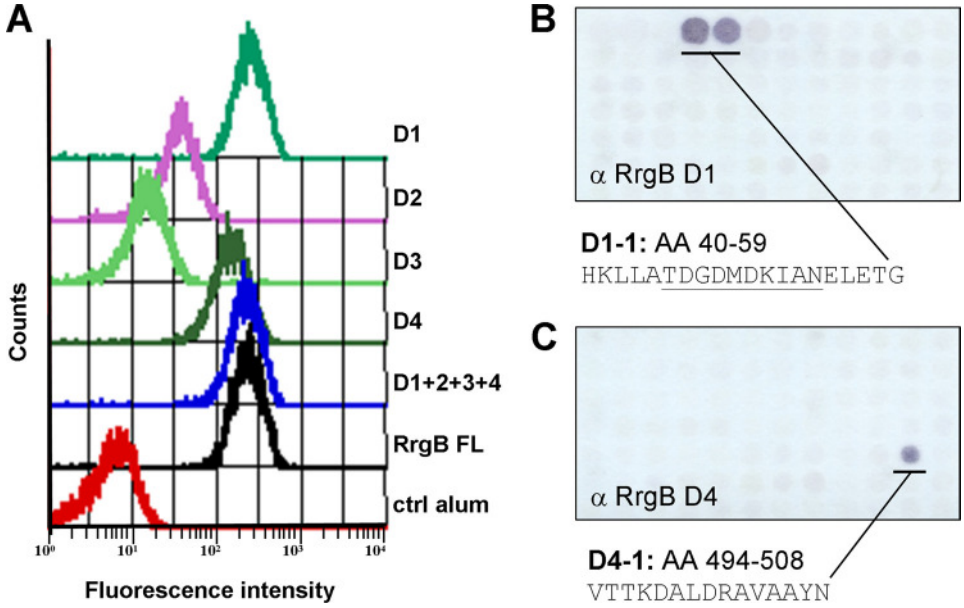


FIGURE 4. **Polyclonal antibodies raised against D1 and D4 recognize the native pilus and linear epitopes within RrgB efficiently.** *A*, TIGR4 bacteria were incubated with mouse primary antibodies directed against the specified recombinant proteins (1:400 dilution) followed by FITC-conjugated goat anti-mouse IgG (1:100 dilution). Bacterial staining was analyzed by flow cytometry (FACS-Calibur). Mouse control sera (immunized with PBS plus alum) were used as negative control. *B* and *C*, glass fiber membranes with arrayed peptides synthesized *in situ* covering residues 25–190 (D1) and 444–628 (D4) of RrgB were incubated with anti-D1 (*A*) or anti-D4 (*B*) polyclonal mouse antibodies (1:3000) and then with goat anti-mouse IgG alkaline phosphatase-conjugated antibodies (1:5000). Linear epitopes corresponding to peptide sequences recognized by the antibodies are reported. *Underlining* marks common residues present in adjacent peptides in the PepScan.

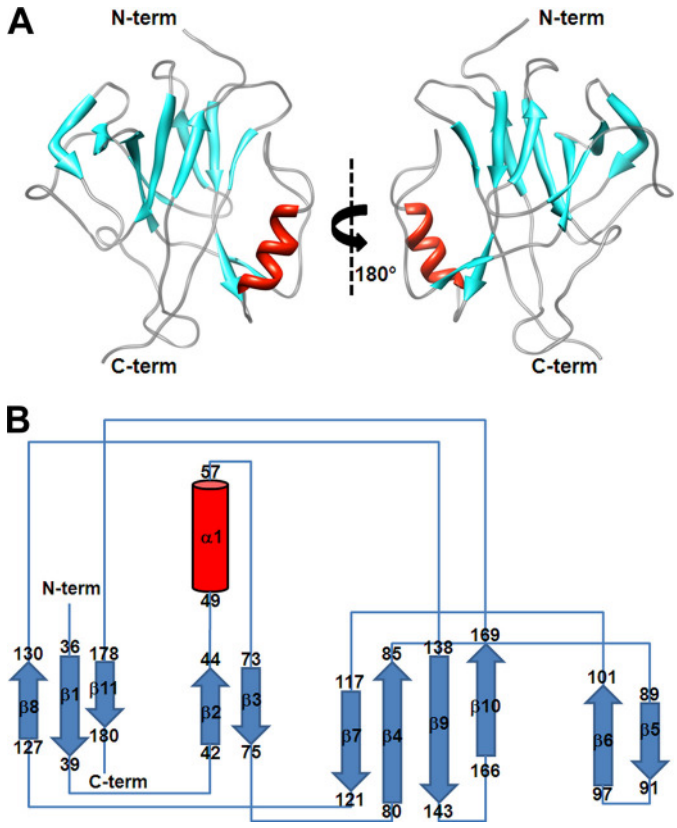


FIGURE 5. **Solution structure of the D1 domain.** *A*, ribbon diagram with the secondary structure elements. β-Strands are shown in *cyan*, the α-helix in *red*. *B*, topology diagram. The α-helix is represented by a *red cylinder*, and the β-strands are *cyan arrows*.

ond segment. In 50% of 20 conformers of the D1 family an additional β-sheet was formed by two short hydrogen-bonded β-strands (stretches 161–163 and 184–187).

The structure was well defined with the exception of three long loops corresponding to the stretches 56–69, 148–162, and 173–177 (supplemental Fig. S1). Heteronuclear relaxation measurements revealed that residues in the first two loops had heteronuclear NOE and longitudinal R_1 values lower and higher than average, respectively (supplemental Fig. S2). This behavior was a consequence of local internal motions occurring on a faster time scale with respect to the overall reorientational correlation time (τ_r) of the molecule and accordingly, a correlation time (τ_e) for these fast motions can be fitted for the previous mentioned loops (supplemental Fig. S2). In addition, conformational exchange processes, occurring on the millisecond-microsecond scale, affected some residues located in the region 155–164, as monitored by transverse R_2 relaxation rates higher than the average (supplemental Fig. S2). These data indicated that these loops experience higher flexibility, showing accordingly a low number of long ranges ^{1}H-^{1}H NOEs (supplemental Fig. S5). For a few residues, the assignment of the backbone resonances was also not achieved (Glu-143, His-145, Ser-146, Ser-148, Thr-149, Tyr-150, Val-152, and Gly-160), likely as a consequence of an increased local mobility. The loop between strands β_9 and β_{10} (residues 148–162) was the most disordered, with a r.m.s.d. value of about 2 Å (supplemental Fig. S1). Conformational exchange processes on the millisecond-microsecond time scale are observed also for the loop between strands β_{10} and β_{11} comprising residues 174–178.

The protein core is characterized by hydrophobic interactions between residues located on the first (strands β_1, β_8, β_{11}) and second (strands β_4, β_7, β_9, β_{10}) sheets. A salt bridge between two complementarily charged side chains of residues Lys-41 and Glu-143 was also present. The aliphatic side chains of residues Met-48, Ile-52, Ala-53, and Leu-56, all located on one side of the α-helix, formed hydrophobic interactions with

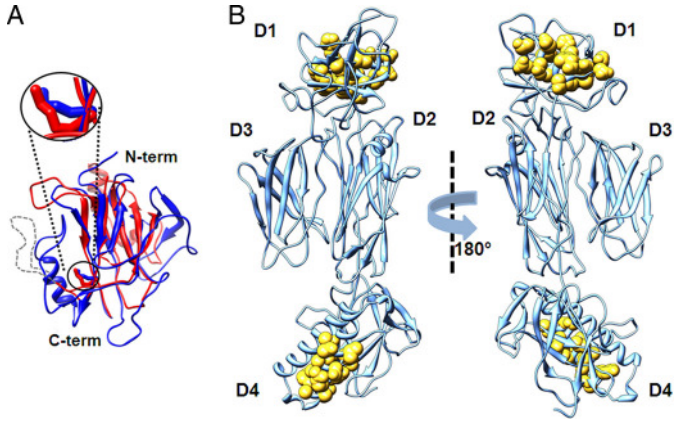


FIGURE 6. **Analysis of RrgB linker flexibility through superimposition and modeling.** *A*, superimposition of the D1 domain (*blue*) and the *C. diphtheriae* SpaA N-terminal domain (*red*). The position of Lys-190 residue, forming the intermolecular isopeptide bond between two neighboring SpaA molecules, is shown along with residue Lys-183 of RrgB, which occupies a similar position, and is indicated in a *black circle*. The missing loop in the SpaA crystal structure is indicated as a *gray dotted line*. *B*, modeled conformation of the full-length RrgB molecule obtained by combining the D1 NMR coordinates with the RrgB D2-D4 x-ray coordinates. D1-1 and D4-1 epitopes (residues 40–59 and 494–508 respectively) are rendered as *yellow spheres*.

aliphatic residues of β_2 and β_3 strands; these interactions determined the position of the helix with respect to the rest of the protein.

A search for related protein structures performed through the DALI Server (48) retrieved the N-terminal domain of the SpaA pilus backbone protein of *C. diphtheriae* (Protein Data Bank (PDB) code 3HTL; r.m.s.d. 2.1 Å), the C-terminal CNA3 domain of the major pilin protein of *Bacillus cereus* BcpA (PDB code 3KPT; r.m.s.d. 2.4 Å), and the N1 domain of the *Streptococcus agalactiae* minor pilin GBS52 (PDB code 2PZ4; r.m.s.d. 3.4 Å). Like D1, none of these domains contained intramolecular isopeptide bonds. However, only the SpaA domain is located at the N terminus of pilus backbone protein as D1; their overlay is presented in Fig. 6*A*.

In an attempt to determine the orientation of D1 with respect to the D2–D4 RrgB portion, a rigid body fitting of all the domains into the shape of the native pilus obtained from cryo-electron microscopy (cryo-EM) was attempted (24). Initial rigid body fitting of the D2–D4 crystal structure into the cryo-EM density map of the native pilus left an apical unoccupied volume, likely due to the absence of D1. However, when the simultaneous fitting of the D1 and the D2–D4 domains was carried out, some portions of D1 could not be accommodated into the apical empty volume (supplemental Fig. S6 and supplemental Table S6). To model the FL RrgB, we merged the D1 and D2–D4 structures into a single molecule by overlapping residues 188–191, shared by the C terminus of D1 and the N terminus of D2–D4. The relative orientation of D1 with respect to the others domains was then varied to best fit into the cryo-EM map (Fig. 6*B*). The final model, which represents only one of the possible orientations of D1 with respect to the rest of the protein, does not present steric clashes between D1 and the D2–D4 domains. The two linear epitopes, previously identified to be located within the D1 and D4 domains by peptide hybridization with specific protective polyclonal antibodies (Fig. 4), when mapped onto the FL model of RrgB, confirmed their superficial localization (Fig. 6*B*).

Lysine 183 of D1 Is Required for Intermolecular Isopeptide Bond Formation and Pilus Polymerization—Pili of Gram-positive bacteria are polymerized by means of intermolecular isopeptide bonds occurring between the Thr of the C-terminal LP*X*TG motif of a RrgB molecule and a Lys located at the N terminus of the following molecule (18, 29, 30). To identify the Lys implicated in the intermolecular isopeptide bond formation between two consecutive RrgB monomers, a TIGR4 RrgB deletion mutant (no longer able to assemble a pilus on its surface) was created. Subsequently, RrgB expression and pilus polymerization were restored in TIGR4ΔRrgB by transforming the mutant with plasmids expressing either wild-type RrgB or RrgB mutated in single Lys residues (Lys → Ala substitutions). Sequence analysis revealed that the D1 contains the canonical ^{181}YPKN^{184} pilin motif, with Lys-183 nicely superimposing onto the functional SpaA Lys-190, as shown in Fig. 6*A*. Noteworthy, D1 also presents the sequence fragment ^{160}GSKAVP^{165}, similar to the motif containing the Lys residue functional for the pilus assembly in *S. pyogenes* Spy0128 (31). Lys-183 and Lys-162, along with two additional Lys residues (Lys-138 and Lys-309), located on D1 and D3 domains, respectively, were selected and mutated.

WB analysis performed on whole cell bacterial lysates using rabbit polyclonal antisera raised against RrgA, RrgB, or RrgC revealed that all of the pilus proteins were expressed in all of the complemented mutants (Fig. 5). However, the typical pilus-associated high molecular weight ladder was revealed only in the case of TIGR4ΔRrgB complemented with RrgB wild type or RrgB mutated in Lys-162, Lys-138, or Lys-309. In contrast, in the mutant complemented with K183A, RrgB was detectable only as a monomer (Fig. 7), clearly indicating that Lys-183 is the residue implicated in intermolecular isopeptide bond formation.

Moreover, in the presence of the RrgB K183A substitution only hetero-oligomers composed of RrgA-RrgC (about 160 kDa) or RrgB-RrgC (about 110 kDa) could be detected by WB analysis (see *arrows* in Fig. 7, *A* and *C*). The absence of RrgA-RrgB heterodimers suggests that RrgA and RrgB are linked exclusively through intermolecular isopeptide bonds involving the D1 residue Lys-183 of RrgB.

DISCUSSION

Ever since their initial discovery, pili of Gram-positive bacteria have received considerable attention because they are associated with a number of different virulence mechanisms (15, 16, 33) and elicit protection in animal models (32, 49–52). In particular, the *S. pneumoniae* pilus was found to be implicated in the initial attachment to epithelial cells (15, 16, 33), and the pilus components, being protective in mouse models of infection, are regarded as potential vaccine candidates (32, 33). The major pilin RrgB is organized into four Ig-like domains (D1–D4), as shown by combined structural approaches (Ref. 24 and this work). We have tested the four RrgB domains in animal model experiments and have shown that the protective efficacy exerted by the combination of the four RrgB domains D1+D2+D3+D4 is comparable with that afforded by the FL

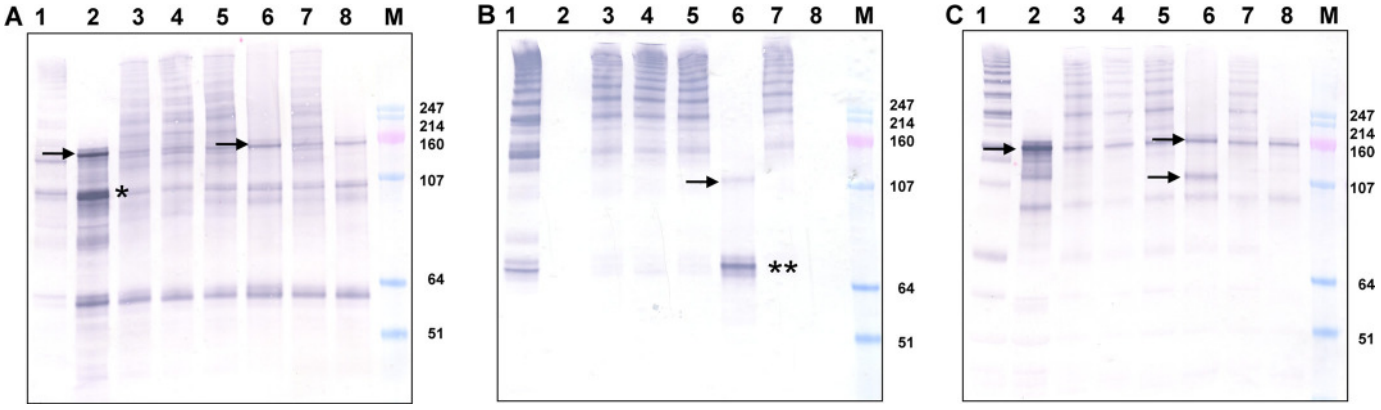


FIGURE 7. **Lys-183 of RrgB is implicated in intermolecular isopeptide bond formation.** WB analysis was performed using polyclonal rabbit antisera against RrgA (*A*), RrgB (*B*), and RrgC (*C*). In all *panels lanes* are loaded as follows: *1*, T4 WT; *2*, T4ΔRrgB; *3*, T4ΔRrgB▼RrgBWT; *4*, T4ΔRrgB▼RrgB(K138A); *5*, T4ΔRrgB▼RrgB(K162A); *6*, T4ΔRrgB▼RrgB(K183A); *7*, T4ΔRrgB▼RrgB(K309A); *8*, T4ΔRrgB▼pMU1328; *9*, molecular mass marker. *, RrgA monomer (estimated molecular mass of native monomeric RrgA is 92 kDa). **, RrgB monomer (estimated molecular mass of native monomeric RrgB is 65 kDa). *Arrows*, the band migrating at an apparent molecular mass of 160 kDa indicated by an *arrow* in *lane 2* and *6* of *A*, and *C* is compatible with an RrgA+RrgC complex; the band at 110 kDa present in *lane 6* of *B* and *C* could correspond to an RrgC+RrgB complex.

RrgB. Among the single domains, D1 is the most protective, and D4 retained an important part of the protective efficacy of the FL protein. The lower protection achieved by D2 and D3 compared with the FL protein, as well as with the D1 and D4, is probably because the antibodies elicited by the former two domains recognize the FL protein less efficiently, in both its native and recombinant forms.

This may be the result of smaller exposed surface areas experienced by D2 and D3 in the FL RrgB compared with the D2 and D3 isolated domains and with D1 and D4. It is possible, in fact, that the antibodies generated against the isolated domains are recognizing areas of D2 and D3 that are buried in the FL protein. Alternatively, D2 and D3 could assume a slightly different conformation when expressed as single domains, thus generating nonfunctional antibodies. On the other hand, the two linear epitopes (Fig. 6*B*) identified within D1 and D4 by PepScan analysis performed with protective polyclonal antibodies raised against the two domains (conformational epitopes are not detectable with this method) are well exposed on the surface of the RrgB molecule and could contribute to the protective activity exerted by D1 and D4. Further experiments are needed to understand to what extent these linear epitopes contribute to the protective activity exerted by the two domains. Taken together, these results suggest that RrgB contains multiple protective epitopes, thus confirming the potential of this vaccine candidate. Furthermore, although the existence of possible conformational epitopes involving residues from different domains cannot be excluded, their contribution to the overall protective efficacy might not be essential.

To obtain more information about the structural role of D1 and to try to correlate it with the protection data discussed above, we solved the solution structure of this domain by using NMR spectroscopy. D1 shows an Ig-like fold, does not contain any intramolecular isopeptide bond, and has many flexible regions. The observed D1 flexibility, indeed, could play a fundamental role in the specific antigen-antibody recognition process (53), thus accounting for D1 enhanced protection capability with respect to the more rigid D2–D4 domains, each one containing an intramolecular isopeptide bond. In fact, the protein structural plasticity could be related to the ability of D1 to undergo local conformational changes and to adapt its structure to optimize the interactions with the antibodies and increase the affinity and the specificity of the antigen-antibody recognition process. The dynamics of D1 could therefore strongly contribute to the interface adaptation for molecular recognition such that the antibody can select an optimal conformer from a wide distribution of possible D1 conformations. The rigid structure of the D2–D4 region prevents such an effective conformational selection for these domains. The above described phenomenon has been observed for other protein-protein or protein-DNA interaction processes (53, 54).

To shed light on the molecular mechanism driving pilus polymerization in *S. pneumoniae*, we investigated which of the lysine residues of D1 was engaged in the intermolecular isopeptide bond formation. Site-directed mutagenesis followed by complementation identified Lys-183 as crucial for the pilus assembly. This result is consistent with the observation that the spatial position of RrgB Lys-183 can be superimposed onto Lys-190 of the *C. diphtheriae* pilus backbone subunit SpaA, known to be involved in the intermolecular isopeptide linkage (30). Interestingly, as shown in Fig. 6*A*, both lysines are not fully available to form an external bond. In particular, the average relative solvent accessibility of the Lys-183 over the D1 family of conformers is 27.2% ± 4.8, as the Lys side chain projects into a cleft between the main body of the protein and the segment 40–78 containing the mobile loop 56–69 (27). This suggests that pilus backbone proteins, to be polymerized, might undergo conformational changes, probably involving not only the Lys residue (Lys-183) but also the flexible regions spatially close to it (49–69, 152–167, and 183–193), to allow the formation of the covalent intermolecular isopeptide bond. NMR mobility data indicate that the C terminus of the D1, where Lys-183 is located, and other loops are highly mobile and that such dynamics could be relevant for the intermolecular isopeptide bond formation (supplemental Fig. S2). Consistently, the absence of stabilizing intramolecular isopeptide bonds renders D1, unlike the other domains, less rigid and prone to conformational rearrangements. Furthermore, the occurrence of a struc-

tural rearrangement of D1 within the native pilus structure is also in line with the partial fitting of its NMR structure onto the molecular shape of the native pilus determined by cryo-EM.

Mutagenesis and complementation data were used to analyze the covalent links established among the three pilus proteins either in the absence of RrgB or in the presence of the RrgB K183A mutant, in an attempt to provide further insights into the possible organization of the native pilus. In the presence of the nonpolymerizing RrgB K183A mutant, the lack of RrgA-RrgB heterodimers provides evidence that RrgA and RrgB are unidirectionally linked only through the Lys-183 of RrgB and the C-terminal Thr of RrgA. This arrangement is in accordance with the model proposed by Hilleringmann *et al.*, positioning RrgA at the pilus terminus, thus ruling out the alternative possibility of RrgA being incorporated along the pilus shaft (22). Conversely, an RrgA-RrgC multimer is detectable either when RrgB is not expressed, as already reported by Falker *et al.* (20) and Le Mieux *et al.* (55), or in the presence of nonpolymerizing RrgB K183A. In this case, the mutated RrgB is competing with RrgA for the linkage to RrgC, as demonstrated by the presence of RrgB-RrgC hetero-oligomers (Fig. 7, *B* and *C*). Concomitant detection of RrgA-RrgC hetero-oligomers even under these conditions further strengthens the idea that in wild-type bacteria, although not detectable by electron microscopy analysis of purified pili, a fraction of RrgA and RrgC might be directly linked to each other.

In conclusion, this study provides additional information elucidating pilus proteins features and also paves the way to the rational design of new RrgB-based molecules to implement a protein-based vaccine against pneumococcal disease. Moreover, the newly acquired structural and dynamic information on the RrgB molecule provided by this study suggests that the conformational flexibility of D1 is pivotal for the protein-antibody recognition process. These findings together with the new functional information could be used to better understand pilus functions and its role in pathogenesis.

Acknowledgments—We thank Giacomo Matteucci and Tommaso Pasquali, who managed Animal Resources; Marco Tortoli, Stefania Torricelli, Luigi Manganelli, and Elena Amantini, who took care of the animal treatments; Silvia Maccari, Esmeralda Bizzarri, and Alessia Corrado, who lent technical assistance for the in vivo experiments; and Morena Lo Sapio for technical assistance in FACS and ELISA experiments. Intercell AG kindly provided us with pMU1328. Luisa Lozzi and Luisa Bracci (University of Siena) synthesized the peptides onto the glass fiber membranes used for Pepscan analysis. Finally, we thank Mariagrazia Pizza for a critical review of the manuscript.

REFERENCES

1. Fletcher, M. A., and Fritzell, B. (2007) *Vaccine* **25,** 2507–2512
2. Kim, K. S. (2010) *Lancet Infect. Dis.* **10,** 32–42
3. O'Brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., McCall, N., Lee, E., Mulholland, K., Levine, O. S., and Cherian, T. (2009) *Lancet* **374,** 893–902
4. Pelton, S. I., and Leibovitz, E. (2009) *Pediatr. Infect. Dis. J.* **28,** S133–137
5. Ryan, M. W., and Antonelli, P. J. (2000) *Laryngoscope* **110,** 961–964
6. van der Poll, T., and Opal, S. M. (2009) *Lancet* **374,** 1543–1556
7. Melegaro, A., Gay, N. J., and Medley, G. F. (2004) *Epidemiol. Infect.* **132,** 433–441
8. Melegaro, A., Choi, Y. H., George, R., Edmunds, W. J., Miller, E., and Gay, N. J. (2010) *BMC Infect. Dis.* **10,** 90
9. Millar, E. V., O'Brien, K. L., Zell, E. R., Bronsdon, M. A., Reid, R., and Santosham, M. (2009) *Pediatr. Infect. Dis. J.* **28,** 711–716
10. Barocchi, M. A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A. R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D. W., Normark, S., Rappuoli, R., and Henriques-Normark, B. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 2857–2862
11. Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prévost, M. C., Kojadinovic, M., Lalioui, L., Poyart, C., and Trieu-Cuot, P. (2006) *Mol. Microbiol.* **60,** 1401–1413
12. Lauer, P., Rinaudo, C. D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A. R., Mora, M., Rappuoli, R., Grandi, G., and Telford, J. L. (2005) *Science* **309,** 105
13. LeMieux, J., Hava, D. L., Basset, A., and Camilli, A. (2006) *Infect. Immun.* **74,** 2453–2456
14. Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A. G., Maggi, T., Taddei, A. R., Grandi, G., and Telford, J. L. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 15641–15646
15. Hilleringmann, M., Giusti, F., Baudner, B. C., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M. A., and Ferlenghi, I. (2008) *PLoS Pathog.* **4,** e1000026
16. Nelson, A. L., Ries, J., Bagnoli, F., Dahlberg, S., Fälker, S., Rounioja, S., Tschöp, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M., Holden, D. W., Rappuoli, R., Normark, S., Barocchi, M. A., and Henriques-Normark, B. (2007) *Mol. Microbiol.* **66,** 329–340
17. Telford, J. L., Barocchi, M. A., Margarit, I., Rappuoli, R., and Grandi, G. (2006) *Nat. Rev. Microbiol.* **4,** 509–519
18. Ton-That, H., and Schneewind, O. (2004) *Trends Microbiol.* **12,** 228–234
19. El Mortaji, L., Terrasse, R., Dessen, A., Vernet, T., and Di Guilmi, A. M. (2010) *J. Biol. Chem.* **285,** 12405–12415
20. Fälker, S., Nelson, A. L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S., and Henriques-Normark, B. (2008) *Mol. Microbiol.* **70,** 595–607
21. Manzano, C., Contreras-Martel, C., El Mortaji, L., Izoré, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A. M., and Dessen, A. (2008) *Structure* **16,** 1838–1848
22. Hilleringmann, M., Ringler, P., Müller, S. A., De Angelis, G., Rappuoli, R., Ferlenghi, I., and Engel, A. (2009) *EMBO J.* **28,** 3921–3930
23. Izoré, T., Contreras-Martel, C., El Mortaji, L., Manzano, C., Terrasse, R., Vernet, T., Di Guilmi, A. M., and Dessen, A. (2010) *Structure* **18,** 106–115
24. Spraggon, G., Koesema, E., Scarselli, M., Malito, E., Biagini, M., Norais, N., Emolo, C., Barocchi, M. A., Giusti, F., Hilleringmann, M., Rappuoli, R., Lesley, S., Covacci, A., Masignani, V., and Ferlenghi, I. (2010) *PLoS One* **5,** e10919
25. Budzik, J. M., Poor, C. B., Faull, K. F., Whitelegge, J. P., He, C., and Schneewind, O. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 19992–19997
26. Kang, H. J., and Baker, E. N. (2009) *J. Biol. Chem.* **284,** 20729–20737
27. Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H., and Baker, E. N. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 16967–16971
28. Alegre-Cebollada, J., Badilla, C. L., and Fernández, J. M. (2010) *J. Biol. Chem.* **285,** 11235–11242
29. Guttilla, I. K., Gaspar, A. H., Swierczynski, A., Swaminathan, A., Dwivedi, P., Das, A., and Ton-That, H. (2009) *J. Bacteriol.* **191,** 5603–5612
30. Ton-That, H., and Schneewind, O. (2003) *Mol. Microbiol.* **50,** 1429–1438
31. Kang, H. J., Coulibaly, F., Clow, F., Proft, T., and Baker, E. N. (2007) *Science* **318,** 1625–1628
32. Gianfaldoni, C., Censini, S., Hilleringmann, M., Moschioni, M., Facciotti, C., Pansegrau, W., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M. A., and Ruggiero, P. (2007) *Infect. Immun.* **75,** 1059–1062
33. Moschioni, M., Emolo, C., Biagini, M., Maccari, S., Pansegrau, W., Donati, C., Hilleringmann, M., Ferlenghi, I., Ruggiero, P., Sinisi, A., Pizza, M., Norais, N., Barocchi, M. A., and Masignani, V. (2010) *Infect. Immun.* **78,** 5033–5042
34. Achen, M. G., Davidson, B. E., and Hillier, A. J. (1986) *Gene* **45,** 45–49
35. Alloing, G., Martin, B., Granadel, C., and Claverys, J. P. (1998) *Mol. Microbiol.* **29,** 75–83

36. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13,** 289–302
37. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) *J. Mol. Biol.* **319,** 209–227
38. Pearlman, A. D., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatman, T. E., III, DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) *Comp. Phys. Commun.* **91,** 1–41
39. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph.* **14,** 29–32
40. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33,** 5984–6003
41. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc.* **115,** 12593–12594
42. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104,** 4546–4559
43. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104,** 4559–4570
44. Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) *J. Biomol. NMR* **16,** 23–28
45. Goddard, T. D., Huang, C. C., and Ferrin, T. E. (2007) *J. Struct. Biol.* **157,** 281–287
46. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., and Ferrin, T. E. (2006) *BMC Bioinformatics* **7,** 339
47. Goddard, T. D., Huang, C. C., and Ferrin, T. E. (2005) *Structure* **13,** 473–482
48. Holm, L., and Sander, C. (1996) *Methods Enzymol.* **266,** 653–662
49. Abbot, E. L., Smith, W. D., Siou, G. P., Chiriboga, C., Smith, R. J., Wilson, J. A., Hirst, B. H., and Kehoe, M. A. (2007) *Cell. Microbiol.* **9,** 1822–1833
50. Pezzicoli, A., Santi, I., Lauer, P., Rosini, R., Rinaudo, D., Grandi, G., Telford, J. L., and Soriani, M. (2008) *J. Infect. Dis.* **198,** 890–898
51. Rosini, R., Rinaudo, C. D., Soriani, M., Lauer, P., Mora, M., Maione, D., Taddei, A., Santi, I., Ghezzo, C., Brettoni, C., Buccato, S., Margarit, I., Grandi, G., and Telford, J. L. (2006) *Mol. Microbiol.* **61,** 126–141
52. Margarit, I., Rinaudo, C. D., Galeotti, C. L., Maione, D., Ghezzo, C., Buttazzoni, E., Rosini, R., Runci, Y., Mora, M., Buccato, S., Pagani, M., Tresoldi, E., Berardi, A., Creti, R., Baker, C. J., Telford, J. L., and Grandi, G. (2009) *J. Infect. Dis.* **199,** 108–115
53. Mittag, T., Kay, L. E., and Forman-Kay, J. D. (2010) *J. Mol. Recognit.* **23,** 105–116
54. Dyson, H. J., and Wright, P. E. (2005) *Nat. Rev. Mol. Cell Biol.* **6,** 197–208
55. LeMieux, J., Woody, S., and Camilli, A. (2008) *J. Bacteriol.* **190,** 6002–6013