

The modular architecture of meningococcal factor H-binding protein

Peter T. Beernink and Dan M. Granoff

Correspondence
Dan M. Granoff
dgranoff@chori.org

Center for Immunobiology and Vaccine Development, Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA

Meningococcal factor H binding protein (fHbp) is a promising vaccine antigen that binds the human complement downregulatory molecule factor H (fH), and this binding enhances the survival of the organism in serum. Based on sequence variability of the entire protein, fHbp has been divided into three variant groups or two subfamilies. Here, we present evidence based on phylogenetic analysis of 70 unique fHbp amino acid sequences that the molecular architecture is modular. From sequences of natural chimeras we identified blocks of two to five invariant residues that flanked five modular variable segments. Although overall, 46% of the fHbp amino acids were invariant, based on a crystal structure, the invariant blocks that flanked the modular variable segments clustered on the membrane surface containing the amino-terminal lipid anchor, while the remaining invariant residues were located throughout the protein. Each of the five modular variable segments could be classified into one of two types, designated α or β , based on homology with segments encoded by variant 1 or 3 fHbp genes, respectively. Forty of the fHbps (57%) comprised only α ($n=33$) or β ($n=7$) type segments. The remaining 30 proteins (43%) were chimeras and could be classified into one of four modular groups. These included all 15 proteins assigned to the previously described variant 2 in subfamily A. The modular segments of one chimeric modular group had 96% amino acid identity with those of fHbp orthologs in *Neisseria gonorrhoeae*. Collectively, the data suggest that recombination between *Neisseria meningitidis* and *N. gonorrhoeae* progenitors generated a family of modular, antigenically diverse meningococcal fHbps.

Received 10 April 2009
Revised 7 June 2009
Accepted 29 June 2009

INTRODUCTION

Meningococcal factor H-binding protein (fHbp) [previously referred to as GNA1870 (Masignani *et al.*, 2003) or LP2086 (Fletcher *et al.*, 2004)] is a promising vaccine antigen for prevention of disease caused by *Neisseria meningitidis*. Two recombinant protein vaccines containing fHbp are in clinical development (Jansen *et al.*, 2008; Rappuoli, 2008). Both vaccines were recently reported to elicit serum bactericidal responses in children and adults against genetically diverse meningococcal strains (Jansen *et al.*, 2008; Marshall *et al.*, 2008; Richmond *et al.*, 2008; Snape *et al.*, 2008; Plested *et al.*, 2009). An important function of fHbp is to bind human complement factor H

(fH) (Granoff *et al.*, 2009; Madico *et al.*, 2006), which downregulates complement activation. Binding of fH to the bacterial surface is an important mechanism by which the pathogen survives in non-immune human serum or blood and evades innate host defences (Madico *et al.*, 2006; Schneider *et al.*, 2006; Welsch *et al.*, 2008). Serum antibodies directed at fHbp both activate classical complement-mediated bacteriolysis (Welsch *et al.*, 2004) and block binding of fH (Madico *et al.*, 2006; Welsch *et al.*, 2008). Blocking binding of fH to the bacterial surface would be expected to increase susceptibility of the organism to complement-mediated killing. The main limitations of fHbp as a vaccine antigen are low density on the surface of some strains (Welsch *et al.*, 2008) and variability (Beernink *et al.*, 2007; Fletcher *et al.*, 2004; Masignani *et al.*, 2003).

Abbreviations: fH, factor H; fHbp, factor H binding protein; MLST, multilocus sequence type; peptide ID no., peptide identification number from <http://neisseria.org>.

Four supplementary tables (with supplementary references), listing source strains and characteristics of unique fHbp variants sorted by modular group, parameters used for phylogenetic analyses, distribution of fHbp variants in each clonal complex, and unique sequences in fHbp variable segments and numbers from each variant group, and a supplementary figure, showing the mean number of peptides and percentage amino acid identity for each of the five modular variable segments, are available with the online version of this paper.

In their original description of fHbp, Masignani and co-workers classified fHbp into three variant groups based on amino acid sequence diversity and lack of cross-reactive serum bactericidal antibody responses of immunized mice (Masignani *et al.*, 2003). Using similar analyses, Fletcher and co-workers assigned fHbp variants into two subfamilies, designated A and B (Fletcher *et al.*, 2004). Subfamily B

corresponded to variant group 1 of Masignani and co-workers, and subfamily A included strains with fHbp in the variant 2 or 3 groups. Thus far, there has not been a consensus as to whether variant groups 2 and 3 are antigenically or phylogenetically distinct.

In an effort to develop fHbp antigens capable of eliciting broad bactericidal antibody responses against strains with fHbp from different variant groups, our laboratory has conducted epitope mapping studies of anti-fHbp bactericidal mAbs (Beernink *et al.*, 2008, 2009b). Based on the locations of the epitopes, we prepared recombinant chimeric fHbp antigens containing domains from different variant groups (Beernink & Granoff, 2008). As part of these studies we noted that certain portions of the fHbp molecule shared amino acid sequence identity and cross-reactivity between variant 1 and 2 proteins (Beernink *et al.*, 2009b), while other portions of the protein shared identity between variant 2 and 3 proteins (Beernink *et al.*, 2008). Furthermore, we noted examples of naturally occurring chimeric fHbps that all had in common certain blocks of invariant amino acids, which appeared to flank segments containing variable amino acids. The results of these preliminary analyses suggested the hypothesis that the fHbp architecture is 'modular', with distinct variable segments derived from different progenitor sequences. This architecture had not been described in the published fHbp phylogenetic studies (Fletcher *et al.*, 2004; Jacobsson *et al.*, 2009; Masignani *et al.*, 2003), which were largely based on analyses of the variability of the overall sequences of the mature protein. To investigate the modular fHbp hypothesis, in the present study we performed phylogenetic analyses of different segments of 70 distinct fHbps. Our goals were to increase our understanding of the evolution of fHbp variation and to provide a basis for subclassification of fHbps into different modular variant groups.

METHODS

fHbp sequencing. We amplified the *fHbp* gene from genomic DNA prepared with the DNeasy Tissue kit (Qiagen) by PCR using primers A1 and B2 and the thermal cycling parameters described by Masignani *et al.* (2003). The PCR products were purified using a QiaQuick PCR purification kit (Qiagen) and eluted in 30 µl sterile deionized H₂O. The *fHbp* DNA sequences were determined by a commercial sequencing facility using primers A1 and 22 described by Masignani *et al.* (2003).

Source of data. We analysed protein sequences encoded by 76 fHbp genes determined in our laboratory from *N. meningitidis* case isolates from the USA (Beernink *et al.*, 2007), Europe (Beernink & Granoff, 2008; Beernink *et al.*, 2008) and Africa (Beernink *et al.*, 2009a). The dataset included 48 sequences determined as part of our previous studies and 28 new sequences performed for the present study. We obtained 95 additional *fHbp* gene sequences from GenBank (<http://www.ncbi.nlm.nih.gov>) by performing translated BLAST (tblastn) searches with fHbp amino acid sequences from strains MC58 (variant 1/subfamily B) and M1239 (variant 3/subfamily A). Among the 171 nucleotide sequences from our collection and those in GenBank, we identified *fHbp* genes that encoded 64 unique protein sequences. These 64, plus six additional unique fHbp amino acid sequences

obtained from the Neisseria.org fHbp peptide database (<http://neisseria.org>), were used for our analyses of 70 unique fHbp peptides. The respective GenBank accession numbers and/or peptide identification numbers and the characteristics of the source strains are listed in Supplementary Table S1.

Thirty-eight (54.3%) of the 70 peptides were classified in the variant 1 group of Masignani *et al.* (2003), 15 (21.4%) in the variant 2 group, and 17 (24.3%) in the variant 3 group. Of the 70 source strains, one was capsular group A, 57 were group B, seven were group C, two were group W-135, two were group X and one was group Y. Multilocus sequence type (MLST) information was available for 59 of the strains, of which 15 were from the ST-269 clonal complex, 12 were from the ST-11 complex, 10 were from the ST-41/44 complex, five each were from the ST-162 and ST-213 complexes, and three each were from the ST-8 and ST-32 complexes. Six strains were from other clonal complexes and had sequence types ST-4, ST-35, ST-751, ST-4821, ST-5403 and ST-6874. The 11 strains without MLST information were not available for testing.

Phylogenetic analysis. The analysis of complete or partial protein sequences was performed on the platform at <http://www.phylogeny.fr> (Dereeper *et al.*, 2008) and comprised the following steps. Sequences were aligned with MUSCLE (v3.7) (Edgar, 2004) configured for highest accuracy, and the alignments were checked for accuracy using Gblocks (Castresana, 2000). The respective alignments, which contained up to three sites of insertions or deletions, were inspected and verified by adjacent invariant sequences. The phylogenetic trees were reconstructed using the maximum-likelihood method implemented in the PhyML program (v3.0 aLRT) (Guindon & Gascuel, 2003). Statistical tests for branch support were performed using the bootstrapping method (100 replicates). Phylograms were displayed with MEGA 4.0 (Tamura *et al.*, 2007). The percentage sequence identities within and between variable segment types of sequences were determined using CLUSTAL W (Larkin *et al.*, 2007). Additional information on the parameters used is provided in Supplementary Table S2.

RESULTS

Phylogenetic analysis of mature proteins

We aligned the 70 unique fHbp amino acid sequences to generate a phylogram based on variability of the entire mature protein, which was arbitrarily rooted on peptide 1 (Fig. 1). The analysis showed the two major branches previously designated subfamilies A and B (Fletcher *et al.*, 2004). Subfamily A contained fHbp sequences in antigenic variant groups 2 and 3, and subfamily B corresponded to fHbps in the antigenic variant group 1 (Masignani *et al.*, 2003). For some clonal complexes, there were examples of strains with fHbp in each of the variant groups (for example, for ST-11, peptide ID nos 2, 3, 6, 9, 10, 11 and 78 in the variant 1 group, peptides 17, 22, 23 and 27 in the variant 2 group, and peptide 59 in the variant 3 group). The strains from the ST-32 clonal complex had fHbp in two of the variant groups (peptides 1 and 89 in the variant 1 group and peptide 76 in the variant 3 group). The ST-8 clonal complex had fHbp only in variant group 2 (peptides 16, 50, and 77). The distribution of fHbp variants in all of the observed clonal complexes is given in Supplementary Table S3.

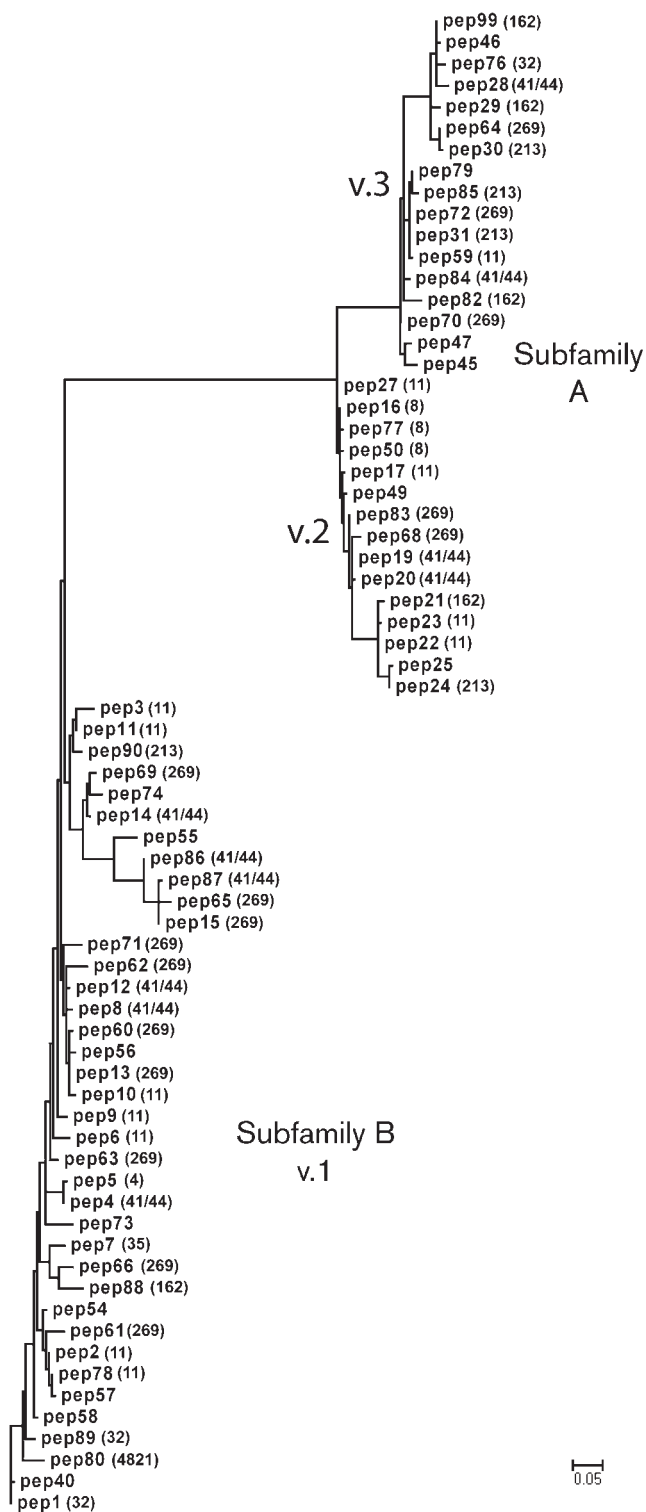


Fig. 1. Phylogram of fHbp based on 70 unique amino acid sequences. For each sequence, the peptide identification number assigned in the fHbp peptide database at <http://Neisseria.org> is shown and, if known, the MLST clonal complex is shown in parentheses. The lower left branch shows variant group 1 as defined by Massignani *et al.* (2003) (subfamily B of Fletcher *et al.*, 2004); subfamily A contained two branches, variant groups 2 and 3. The phylogram was constructed by multiple sequence alignment as described in Methods and Supplementary Table S2. Bar, five amino acid changes per 100 residues.

of two to five invariant residues that flanked five modular variable segments. With the exception of the amino-terminal CSSG, each of the invariant blocks was also present in two *N. gonorrhoeae* orthologues (GenBank accession nos AE004969 and CP001050) (Fig. 2a). We mapped the variable and invariant residues onto a molecular model based on the published coordinates from the fHbp crystal structure (Fig. 2b) (Schneider *et al.*, 2009). The subset of invariant amino acids flanking each of the modular variable segments formed a distinct cluster on the surface of the protein predicted to be anchored to the cell wall (visible on the models in the centre and on the right). This location suggests that there are structural constraints involving these invariant blocks, perhaps a requirement for a partner protein, for anchoring and/or orienting fHbp on the bacterial cell membrane.

Modular variable segments

Based on the positions of the blocks of invariant amino acids, the overall architecture could be divided into an amino-terminal repetitive element and five modular variable segments, which we designated V_A – V_E (Fig. 2a). Segments V_B and V_D contained 15 and 19 amino acids, respectively, while segments V_A , V_C and V_E contained 69, 62 and 71 amino acids, respectively. Within each of the modular variable segments, there were both invariant and variable amino acids. The percentages of variable amino acids in segments V_A and V_D were 49 and 47%, respectively, while the percentages of variable amino acids in segments V_B , V_C and V_E ranged from 60 to 68%.

The amino-terminal repetitive element

For all 70 sequences, the mature fHbp began with a cysteine residue that is lipidated by signal peptidase II, which was followed by three invariant amino acid residues, SSG. This invariant sequence was followed by a repetitive element consisting of one to six glycine and/or serine residues and then by two invariant glycine residues (Fig. 2a). The variable portion of the amino-terminal element consisted of a single glycine residue for 34 of the 70 proteins (nearly all in the previously described variant 1 group, Table 1), or GG residues for 12 of the proteins (all in the variant 2 group). The other common repetitive element sequence was GGGSGG (eight in the variant 1 group and seven in the variant 3 group).

Evidence of a modular architecture: invariant amino acids

Overall, 46% of the amino acids in the 70 fHbps were invariant. However, based on examples of naturally occurring chimeric fHbps (see below), we identified blocks

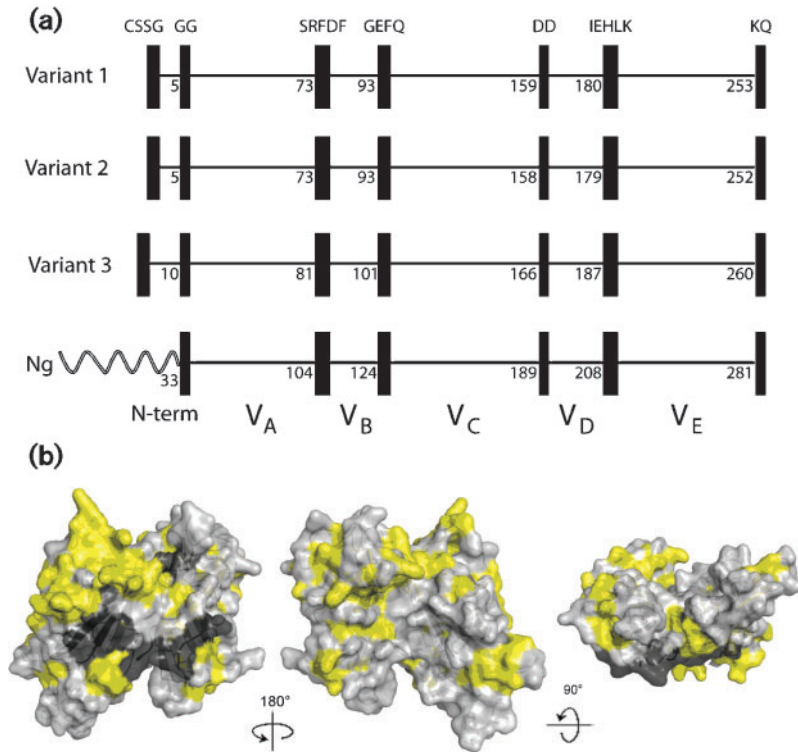


Fig. 2. (a) Schematic representation of fHbp, showing positions of blocks of invariant residues (shown as black vertical rectangles). The top three panels show representative architectures of three *N. meningitidis* fHbp variants in groups 1, 2 and 3 (peptide ID nos 1, 16 and 28, respectively). The amino acid positions of the last residue in each variable segment are shown. With the exception of a longer, unrelated, amino-terminal element, two *N. gonorrhoeae* orthologues (Ng, GenBank accession nos AE004969 and CP001050) had the identical six invariant blocks of residues that flanked segments V_A to V_E. (b) Space-filling structural models of fHbp based on the coordinates of the protein in a complex with a fragment of human fH (Schneider *et al.*, 2009). The light-grey residues represent the variable amino acids located within the modular variable segments. The invariant blocks of residues separating each of the variable segments are shown in dark grey. The yellow residues represent the invariant amino acids outside of these blocks. The model on the left is the surface predicted to be anchored to the cell wall. The model in the centre has been rotated 180° on the y axis from the corresponding model on the left, while the model on the right has been rotated 90° on the x axis as compared with the model in the centre. The figure was constructed with PyMol (<http://www.pymol.org>).

Downstream modular variable segments

We analysed separately the phylogeny of each of the five modular variable segments of the 70 distinctive meningococcal fHbp peptide variants.

Table 1. Sequence and distribution of amino-terminal repetitive elements

Variable sequence*	Number of fHbps in variant group		
	Variant 1	Variant 2	Variant 3
G	29	3	2
GG	0	12	0
SGG	0	0	1
GGGS	0	0	2
SGSGG	0	0	1
GGGSGG	8	0	7
GGGSGS	1	0	4

*The mature fHbp begins with an invariant tetrapeptide, CSSG, which is followed by a repetitive variable sequence consisting of one to six glycine and/or serine residues, which is then followed by two invariant glycine residues.

In these analyses, the numbering of the amino acid residues was based on the mature fHbp peptide 1 encoded by a gene from MC58 in the previously assigned variant 1 group (Masignani *et al.*, 2003). The lengths of the respective variant 2 and 3 proteins differed from that of the variant 1 protein by -1 and +7, respectively (Fig. 2).

Segment V_A began at amino acid residue 8 immediately after the invariant GG sequence and extended to position 73, while segment V_B began at position 79 immediately after an invariant SRFDF sequence and extended to position 93 (Fig. 2a). Based on the phylogram of the V_A segments (Fig. 3, upper panel), 33 of the 38 fHbps assigned to Masignani variant 1 group (shown in black) clustered together with those of the 15 fHbps in the variant 2 group (red), whereas all 17 fHbps in the variant 3 group (aqua) were located in a separate cluster. In the phylogram of the V_B segments (Fig. 3, lower panel), there was similar clustering of the variant 1 and 2 proteins, which were separate from those of the variant 3 proteins.

In contrast to the V_A and V_B segments, in the phylograms of the respective V_C (residues 98–159) and V_E (residues 186–253) segments, all of the fHbps assigned to

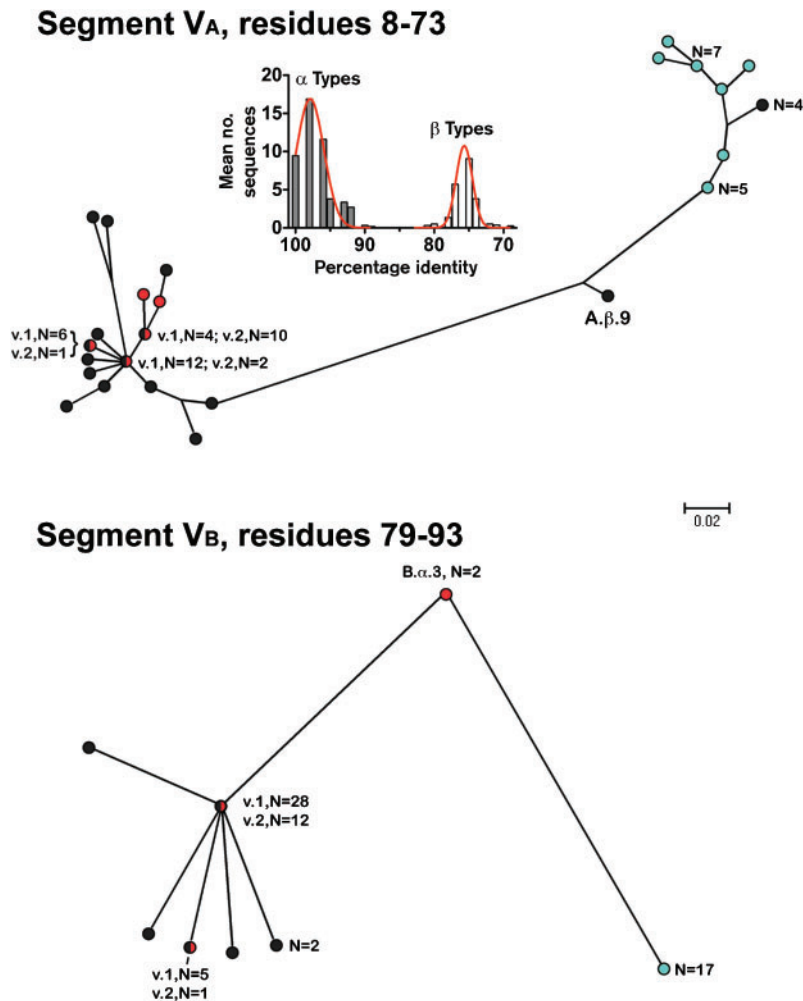


Fig. 3. Phylograms of unique fHbp amino acid sequences in modular variable segments V_A (residues 8–73) and V_B (residues 79–93). Segments derived from fHbp in the variant group 1 are shown as black circles, the variant 2 group as red circles, and the variant 3 group as light-blue circles. Where multiple proteins possessed an identical sequence in a segment, the number of proteins is given. A. β .9 and B. α .3 refer to proteins with exceptional junctional points (see text). Bar, two amino acid changes per 100 residues. The histogram (Segment V_A) shows the mean number of protein sequences (y axis) and percentage amino acid identity (x axis) generated by comparing each of the 48 α A segments with the corresponding α - and β -type A segments of all 70 proteins. The corresponding histograms comparing each of the α segments with the corresponding α and β types, and each of the β segments with the corresponding α and β types, for all five modular variable segments, V_A to V_E , are shown in Supplementary Fig. S1.

Masignani's variant 2 and 3 groups clustered together and were separated from those of the variant 1 group 1 (Fig. 4). The phylogram of the V_D segments (residues 162–180) had a different pattern from those of the other segments: all of the variant 1 proteins clustered with a subset of 10 of the 15 variant 2 proteins and 10 of the 17 variant 3 proteins (Fig. 5). The remaining five variant 2 proteins and seven variant 3 proteins were located together in a separate cluster.

Each of the five modular variable segments segregated into one of two types (Table 2). One of the types had signature amino acid residues and sequence similarity to proteins in the antigenic variant 1 group. The other type had signature amino acid residues and sequence similarity to proteins in the antigenic variant 3 group. The latter were also similar to those of the gonococcal orthologue (see Fig. 2 and Discussion). For purposes of classification, we designated the first group of segments as α types and the second group as β types.

For each segment, we calculated the percentages of the amino acid identity of each of the α -type segments with each of the corresponding α or β -type segments of the 70

fHbps. We performed a second calculation of the percentages of the amino acid identity of each of the β -type segments with the corresponding α and β types. We generated histograms showing the respective mean frequencies of peptide variants with different percentages of amino acid identity. The inset panel on Fig. 3 shows a representative histogram comparing each of the 48 α V_A segments with the corresponding α and β -type V_A segments of the 70 proteins. The corresponding histograms comparing each of the α segments, and each of the β segments, with the corresponding α and β types for all five modular variable segments, V_A – V_E , are shown in Supplementary Fig. S1. For each of the modular variable segments, there was clear separation between the percentage amino acid identities of the respective α and β types.

For each modular variable segment, distinct sequence variants were assigned a unique identifier beginning with a letter, A to E, to represent the modular variable segment, followed by an α or β to indicate the presence of residues with the respective types described above, followed by a number for each distinct sequence (listed in Supplementary Table S4). For example, the V_A segment contained 16

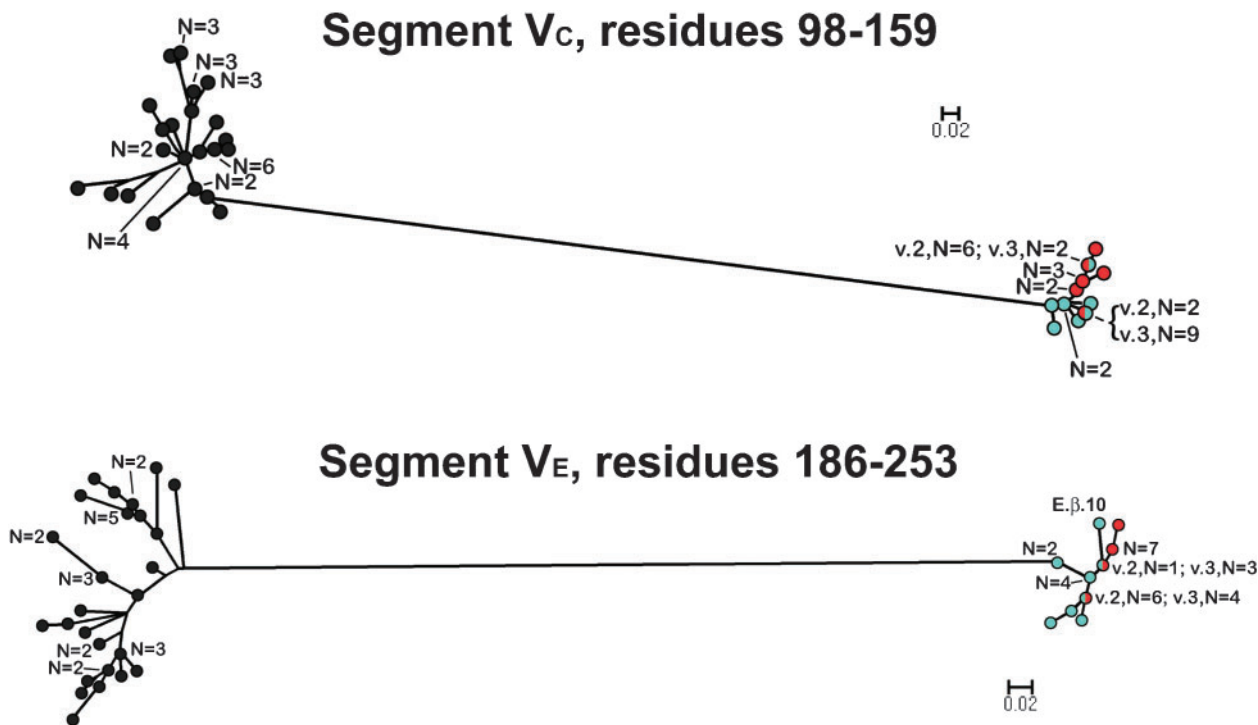


Fig. 4. Phylograms of unique fHbp amino acid sequences in variable segments V_C (residues 98–159) and V_E (residues 186–253). The colours of the circles for segments in each variant group correspond to those described in the legend of Fig. 3. E.β.10 refers to peptide ID no. 82 with an exceptional junctional point (see text). Bars, two amino acid changes per 100 residues.

distinct α variants and nine β variants. The most common α variant ($n=14$) was designated A.α.1. The most common β variant of segment V_A was designated A.β.1. Using this identifier system, each distinct fHbp can be described by a combination of five specific peptides (for example, peptide 1, A.α.2, B.α.1, C.α.5, D.α.5, E.α.8). These designations are provided for each of the 70 proteins in Supplementary Table S1.

Exceptional chimeric fHbps

Four of the 70 fHbp sequences had junctional points between two of the segments that utilized alternative invariant sequences to those described in Fig. 2. The V_A segment of peptide ID no. 55 switched from a β -type

sequence to an α -type sequence at an invariant AQGAE sequence starting at residue 50 rather than at SRFDF, which started at residue 74. This exceptional V_A segment was designated A.β.9 (β because of its higher sequence identity to other β -type A segments than to the α type A segments). The V_B segment of two other fHbp peptides, ID nos 24 and 25, switched from an α -type to a β -type sequence at an invariant IEV sequence beginning at residue 82 instead of GEFQ at position 94. This exceptional V_B segment, designated B.α.3, was categorized as an α type because of its higher sequence identity to other α -type V_B segments than to the β -type V_B segments. Finally, the V_E segment of peptide ID no. 82 had an exceptional crossover point from an α -type sequence to a β -type at residue A196 instead of at IEHLK starting at position 181. This segment was designated E.β.10

Segment V_D, residues 162-180

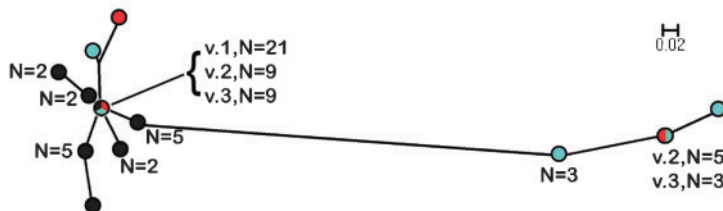


Fig. 5. Phylogram of unique fHbp amino acid sequences in variable segment V_D (residues 162–180). The colours of the circles for peptides in each variant group correspond to those described in the legend of Fig. 3. Bar, two amino acid changes per 100 residues.

Table 2. Amino acid identity within and between sequence types by segment

Variable segment	Residue range*	α Sequence types			β Sequence types			Identity between α and β types (%)
		Signature residues†	Number‡	Identity range (%)	Signature residues†	Number‡	Identity range (%)	
V _A	8–73	QSV	48	89–100	DSI and KDN	21	87–100	69–83
V _B	79–93	IRQ	51	86–100	VQK	17	100	60–66
V _C	98–159	QDS	38	83–100	NNP	32	93–100	37–45
V _D	162–180	AG or AS	58	89–100	PN	12	84–100	52–63
V _E	186–253	SLGI	38	83–100	HLAL	31	92–100	48–57

*Amino acid numbering based on the mature fHbp from strain MC58. (peptide ID no. 1)

†See Supplementary Table S4 for complete sequences of each variable segment.

‡Number of sequences in each sequence type for each variable segment ($n=70$ proteins). Data from four segments with exceptional crossover points were excluded from calculations of percentage identity [segments A. β .9, B. α .3 ($n=2$) and E. β .10; see Figs 4, 5 and 6 and text].

fHbp modular groups

Based on the phylogenetic analysis of the five modular variable segments described above, we could categorize each of the 70 different fHbp variants into one of six distinct fHbp modular groups (Fig. 6). Of the 70 fHbps, 40 (57%) comprised only α ($n=33$) or β ($n=7$) type segments, which were designated fHbp modular groups I and II, respectively (Fig. 6). The remaining 30 proteins (43%), which included the four with exceptional junctional points described above, could be classified into one of four chimeric groups derived from recombination of different α or β segments (designated fHbp modular groups III, IV, V or VI).

Of the 38 fHbps in the Masignani variant 1 group, 33 were in modular group I and five were chimeras in modular group IV. Of the 17 proteins in the variant 3 group, seven were in modular group II and 10 were chimeras in modular group V. Two *N. gonorrhoeae* orthologues (GenBank accession nos AE004969 and CP001050) were also in modular group V. The five modular segments of the orthologues had 96% amino acid sequence identity with

the respective segments of meningococcal fHbp peptide ID 79. All 15 variant 2 proteins were chimeras in modular groups III or VI.

Structural features of the modular variable fHbp segments

As described above for the invariant residues, we mapped the respective variable segments of fHbp onto a space-filling molecular model based on the published coordinates from the crystal structure of the fHbp–fH complex (Fig. 7) (Schneider *et al.*, 2009). Each colour on the model represents a specific modular variable segment and the white segments indicate the blocks of invariant residues that flank the variable segments.

The amino acid residues reported to be in contact with fH are depicted in black. These residues formed clusters on variable segments V_A (blue), V_C (green) and V_E (purple), and are visible in the models shown in the middle and right of Fig. 7. Since fH is known to bind to fHbp on live bacteria (Madico *et al.*, 2006; Schneider *et al.*, 2006), this

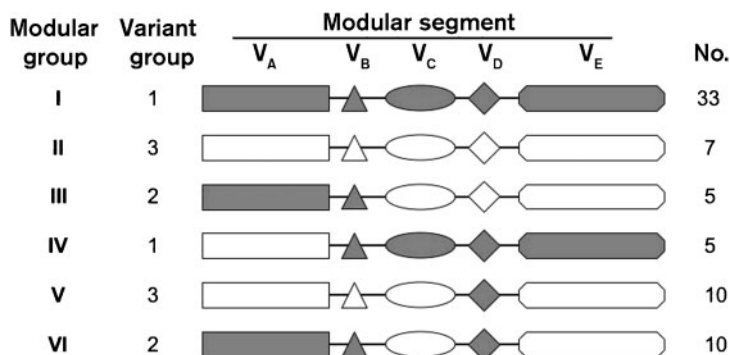


Fig. 6. Schematic representation of six fHbp modular groups deduced from phylogenetic analysis. Forty of the 70 proteins contained only α -type segments or β -type segments, and were designated modular groups I and II, respectively. The remaining 30 proteins were chimeras with different combinations of α (shown in grey) and β segments (shown in white), and could be assigned to one of four modular groups (III–VI). The relationship between the modular group and Masignani variant group designation, and the number of unique sequences observed within each fHbp modular group, are shown.

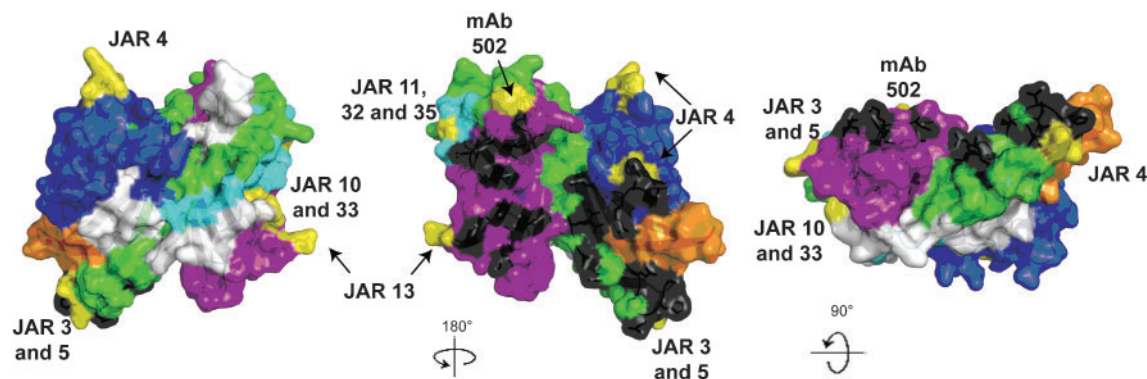


Fig. 7. Space-filling structural models of fHbp based on the coordinates of the protein in a complex with a fragment of human fH (Schneider *et al.*, 2009). The five variable segments V_A to V_E are depicted in different colours (V_A , blue; V_B , orange; V_C , green; V_D , light blue; V_E , violet) and the invariant blocks of residues separating each of the variable segments are shown in white. The model in the middle was generated by rotation of 180° around the y axis relative to the respective model on the far left. The model on the right was generated by a 90° rotation around the x axis relative to the model in the middle. The fH contact residues are depicted in black, and the residues affecting the epitopes of anti-fHbp mAbs are shown in yellow. The figure was constructed with PyMol (<http://www.pymol.org>).

binding site must be surface-exposed. Previous studies have mapped the epitopes of 10 bactericidal anti-fHbp mAbs (Beernink *et al.*, 2008, 2009b; Giuliani *et al.*, 2005; Scarselli *et al.*, 2009). The amino acids that affect expression of each of these epitopes are highlighted in yellow. With the exception of segment V_B , all of the modular variable segments contained epitopes recognized by bactericidal mAbs. Note that the amino acids that affect epitope expression tended to be located on the periphery of the surface-exposed portion of the molecule (best visualized on the models at the middle and far right of the figure). In contrast, the residues in contact with fH were located in clusters in the central portions. The epitopes of certain mAbs, such as JAR 3, 5 and 13, which have been previously reported to inhibit binding of fH to fHbp (Beernink *et al.*, 2008, 2009b), involved amino acids located in proximity to some of the fH contact residues. However, there was no example of overlap between the two sets of contact residues. This finding suggests that even low-affinity binding of mouse fH to fHbp affects the portion of fHbp recognized by the immune system.

DISCUSSION

Based on previous analyses of the sequences of the mature protein, fHbp was classified into two subfamilies (Fletcher *et al.*, 2004) or three variant groups (Masignani *et al.*, 2003) (Fig. 1). However, neither of these classification systems, nor those used in subsequent studies (Bambini *et al.*, 2009; Jacobsson *et al.*, 2009), is entirely satisfactory, since our analyses show that the overall architecture of fHbp is mosaic. The similar modular architectures of meningococcal and gonococcal fHbps (Fig. 2a) suggest that during evolution the respective fHbp genes recombined to create

antigenically diverse meningococcal fHbps. The mosaic architecture of fHbp is not entirely surprising, since *Neisseria* are known to transfer genes horizontally (Bentley *et al.*, 2007; Hotopp *et al.*, 2006), and other neisserial proteins have been reported to have mosaic structures (Arreaza *et al.*, 2003; Bilek *et al.*, 2009; Callaghan *et al.*, 2008; Ochiai *et al.*, 2008; Rokbi *et al.*, 1995; Szczesny & Lupas, 2008).

For each of the fHbp modular variable segments there was clear separation of the sequences into one of two types, designated α or β , based on the presence of 'signature' amino acids and a high percentage of amino acid identity with variant 1 or variant 3 proteins (Table 2, Supplementary Fig. S1). For a protein comprising five modular variable segments, each of which can be of one of two types, there are $2^5=32$ theoretical independent modular combinations. That we identified only six modular groups (Fig. 6) suggests that there are functional or structural constraints on the molecule that select for certain combinations of the variable modules.

Note that within a modular variable segment, only half to two-thirds of the amino acids were variable, and that within an α or β type, the respective sequences of a segment were conserved. If conservative amino acid substitutions are considered, the modular variable segments could have been defined at different boundary residues than those shown in Fig. 2. However, the boundary residues were selected based on the invariant blocks observed in natural fHbp chimeras. Our hypothesis is that these invariant blocks represent junctional points for recombination that result in fit fHbp mutants. As additional data become available to correlate fHbp variants with strain susceptibility to anti-fHbp bactericidal activity, the ability to bind complement fH, or changes in fHbp amino acid sequence

over time, it may be possible to redefine the modular variable segments based on these functional or evolutionary considerations.

In previous studies, serum antibodies to recombinant fHbp in the variant 1 group were bactericidal primarily only against strains with variant 1 proteins (Beernink *et al.*, 2007; Fletcher *et al.*, 2004; Masignani *et al.*, 2003), while antibodies to fHbp in the variant 2 or 3 groups had activity primarily against strains with homologous variant 2 or 3 proteins but had no activity against strains with fHbp in the variant 1 group. These observations suggest that epitopes in the V_C and V_E modular segments of fHbp, for which variant 1 proteins were phylogenetically separated from those of variant 2 and 3 proteins (Fig. 4), are more important for eliciting bactericidal antibodies than those in the V_A, V_B or V_D modular segments, where variant 1 proteins clustered together with variant 2 and/or 3 proteins (Figs 3 and 5). However, all of the variable segments with the exception of segment V_B contained residues previously identified as affecting epitopes recognized by murine bactericidal anti-fHbp mAbs (Beernink *et al.*, 2008, 2009b; Giuliani *et al.*, 2005). Furthermore, some of the mAbs inhibit binding of fH to the bacterial surface (Madico *et al.*, 2006), and some combinations of mAbs that individually are not bactericidal elicit cooperative bactericidal activity (Beernink *et al.*, 2008, 2009b; Welsch *et al.*, 2008). The latter include mAb JAR 4, which is specific for an epitope on variable segment V_A (Beernink *et al.*, 2009b). The implications of the different modular groups for the susceptibility of strains to the bactericidal activity of anti-fHbp antibodies elicited by different fHbp vaccines will require additional studies.

Approximately 40 % of disease-causing isolates have fHbp in the previously described subfamily A (Fletcher *et al.*, 2004), which includes both variant 2 and variant 3 as described by Masignani *et al.* (2003). However, <3 % of strains are in the fHbp variant 3 group, which means that the vast majority of subfamily A strains are in the variant 2 group, and these are all natural chimeras of variants 1 and 3 (the variant 2 group includes modular groups III and VI; Fig. 6). The modular architecture of fHbp thus leads to difficulties in classification based on overall sequence. For epidemiological and evolutionary purposes, classification of fHbp based on the six observed modular groups appears to be a more useful system. For complete annotation of an individual protein, its modular group can be supplemented by the peptide ID (from the Neisseria.org website) and the segment ID numbers of each of the five modular variable segments (Supplementary Table S4). For vaccine development, additional data are needed correlating immunogenicity of fHbp from different modular groups and strain susceptibility to bactericidal activity.

The 70 fHbp sequences analysed in the present study reflect the known amino acid diversity of fHbp reported to date. Since minor fHbp variants that are not widely transmitted in human populations were likely over-represented, our data do not reflect the prevalence of fHbp variants among disease-

causing strains. Also, sequences from carriage isolates were not represented. In the future, population-based surveillance studies will be required to define the prevalence of fHbp from different modular groups among carriage and disease isolates, as well as changes over time, and the relationship between the fHbp modular group and other strain characteristics such as capsular group, clonal complex and PorA variable regions. Also, analyses at the DNA level are needed, particularly to determine which codons encoding fHbp may be under diversifying selective pressure.

In conclusion, our analyses showed that the overall architecture of fHbp is mosaic, consisting of five modular variable segments. Collectively, the data suggest that recombination occurred between *N. meningitidis* and *N. gonorrhoeae* progenitor sequences to generate an antigenically diverse family of meningococcal fHbps. Our data thus provide insights into the evolution of fHbp variants, and may provide a rational basis for classification of fHbp variants that is more useful than previously described schemes based on overall amino acid identity of the entire mature protein.

ACKNOWLEDGEMENTS

We thank Maurizio Comanducci, Novartis Vaccines, Siena, Italy; Lee Harrison, University of Pittsburgh, PA; Jay Lucidarme, Health Protection Agency, UK; and Leonard Mayer, US Centers for Disease Control and Prevention, Atlanta, GA, for providing demographic and MLST information on some of the meningococcal strains described in this study. Drs Rolando Feyt, University of Calgary, and Duccio Medini, Novartis Vaccines, Siena, Italy, provided helpful comments on the manuscript. The work was supported by Public Health Service R01 grants AI 70955 (P. T. B.) and R01 AI 46464 (D.M.G.) from the National Institute of Allergy and Infectious Diseases, NIH. The work was performed in a facility funded by Research Facilities Improvement Program grant C06 RR 16226 from the National Center for Research Resources, NIH.

REFERENCES

- Arreaza, L., Alcalá, B., Salcedo, C., de la Fuente, L. & Vazquez, J. A. (2003). Dynamics of the *penA* gene in serogroup C meningococcal strains. *J Infect Dis* **187**, 1010–1014.
- Bambini, S., Muzzi, A., Olcen, P., Rappuoli, R., Pizza, M. & Comanducci, M. (2009). Distribution and genetic variability of three vaccine components in a panel of strains representative of the diversity of serogroup B meningococcus. *Vaccine* **27**, 2794–2803.
- Beernink, P. T. & Granoff, D. M. (2008). Bactericidal antibody responses induced by meningococcal recombinant chimeric factor H-binding protein vaccines. *Infect Immun* **76**, 2568–2575.
- Beernink, P. T., Welsch, J. A., Harrison, L. H., Leipus, A., Kaplan, S. L. & Granoff, D. M. (2007). Prevalence of factor H-binding protein variants and NadA among meningococcal group B isolates from the United States: implications for the development of a multicomponent group B vaccine. *J Infect Dis* **195**, 1472–1479.
- Beernink, P. T., Welsch, J. A., Bar-Lev, M., Koeberling, O., Comanducci, M. & Granoff, D. M. (2008). Fine antigenic specificity and cooperative bactericidal activity of monoclonal antibodies directed at the meningococcal vaccine candidate, factor H-binding protein. *Infect Immun* **76**, 4232–4240.

- Beernink, P. T., Caugant, D. A., Welsch, J. A., Koeberling, O. & Granoff, D. M. (2009a). Meningococcal factor H-binding protein variants expressed by epidemic capsular group A, W-135, and X strains from Africa. *J Infect Dis* **199**, 1360–1368.
- Beernink, P. T., Lopasso, C., Angiolillo, A., Felici, F. & Granoff, D. (2009b). 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.
- Bentley, S. D., Vernikos, G. S., Snyder, L. A., Churcher, C., Arrowsmith, C., Chillingworth, T., Cronin, A., Davis, P. H., Holroyd, N. E. & other authors (2007). Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genet* **3**, e23.
- Bilek, N., Ison, C. A. & Spratt, B. G. (2009). Relative contributions of recombination and mutation to the diversification of the *opa* gene repertoire of *Neisseria gonorrhoeae*. *J Bacteriol* **191**, 1878–1890.
- Callaghan, M. J., Buckee, C. O., Jolley, K. A., Kriz, P., Maiden, M. C. & Gupta, S. (2008). The effect of immune selection on the structure of the meningococcal Opa protein repertoire. *PLoS Pathog* **4**, e1000020.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540–552.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V. & other authors (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, W465–W469.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- Fletcher, L. D., Bernfield, L., Barniak, V., Farley, J. E., Howell, A., Knauf, M., Ooi, P., Smith, R. P., Weise, P. & other authors (2004). Vaccine potential of the *Neisseria meningitidis* 2086 lipoprotein. *Infect Immun* **72**, 2088–2100.
- Giuliani, M. M., Santini, L., Brunelli, B., Biolchi, A., Aricò, B., Di Marcello, F., Cartocci, E., Comanducci, M., Maignani, V. & other authors (2005). The region comprising amino acids 100 to 255 of *Neisseria meningitidis* lipoprotein GNA 1870 elicits bactericidal antibodies. *Infect Immun* **73**, 1151–1160.
- Granoff, D. M., Welsch, J. A. & Ram, S. (2009). Binding of complement factor H (fH) to *Neisseria meningitidis* is specific for human fH and inhibits complement activation by rat and rabbit sera. *Infect Immun* **77**, 764–769.
- Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.
- Hotopp, J. C., Grifantini, R., Kumar, N., Tzeng, Y. L., Fouts, D., Frigimelica, E., Draghi, M., Giuliani, M. M., Rappuoli, R. & other authors (2006). Comparative genomics of *Neisseria meningitidis*: core genome, islands of horizontal transfer and pathogen-specific genes. *Microbiology* **152**, 3733–3749.
- Jacobsson, S., Hedberg, S. T., Molling, P., Unemo, M., Comanducci, M., Rappuoli, R. & Olcen, P. (2009). Prevalence and sequence variations of the genes encoding the five antigens included in the novel 5CVMB vaccine covering group B meningococcal disease. *Vaccine* **27**, 1579–1584.
- Jansen, K. U., McNeil, L. K., Dragalin, V., Anderson, A. S., Hoiseth, S. K., Arora, A., Emimi, E. E., Zlotnick, G. W. & Jones, T. (2008). Bivalent recombinant LP2086 vaccine to provide broad protection against *Neisseria meningitidis* B disease: immunological correlates of protection and how to assess coverage against invasive MnB strains. Abstract number 064 in *16th International Pathogenic Neisseria Conference*, Rotterdam, The Netherlands, September 7–12 2008. Edited by L. van Alphen, P. van der Ley & G. van den Dobbelen (http://neisseria.org/ipnc/2008/Abstracts_oral_presentations_IPNC_2008.pdf).
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL_X version 2.0. *Bioinformatics* **23**, 2947–2948.
- Madico, G., Welsch, J. A., Lewis, L. A., McNaughton, A., Perlman, D. H., Costello, C. E., Ngampasutadol, J., Vogel, U., Granoff, D. M. & Ram, S. (2006). The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance. *J Immunol* **177**, 501–510.
- Marshall, H., Nissen, M. D., Richmond, P., Lambert, S. B., Robertson, D. M., Jones, T., Lockhart, S., Gruber, W. & Arora, A. (2008). A randomized, placebo-controlled, double-blind, phase 1 trial of ascending doses of meningococcal group B rLP2086 vaccine in healthy adults. Abstract number P213 in *16th International Pathogenic Neisseria Conference*, Rotterdam, The Netherlands, September 7–12 2008. Edited by L. van Alphen, P. van der Ley & G. van den Dobbelen. (http://neisseria.org/ipnc/2008/Abstracts_oral_presentations_IPNC_2008.pdf).
- Maignani, V., Comanducci, M., Giuliani, M. M., Bambini, S., Adu-Bobie, J., Arico, B., Brunelli, B., Pieri, A., Santini, L. & other authors (2003). Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA1870. *J Exp Med* **197**, 789–799.
- Ochiai, S., Ishiko, H., Yasuda, M. & Deguchi, T. (2008). Rapid detection of the mosaic structure of the *Neisseria gonorrhoeae penA* gene, which is associated with decreased susceptibilities to oral cephalosporins. *J Clin Microbiol* **46**, 1804–1810.
- Plested, J. S., Welsch, J. A. & Granoff, D. M. (2009). Ex vivo model of meningococcal bacteremia using human blood for measuring vaccine-induced serum passive protective activity. *Clin Vaccine Immunol* **16**, 785–791.
- Rappuoli, R. (2008). The application of reverse vaccinology, Novartis MenB vaccine developed by design. Abstract number 065 in *16th International Pathogenic Neisseria Conference*, Rotterdam, The Netherlands, September 7–12 2008. Edited by L. van Alphen, P. van der Ley & G. van den Dobbelen. (http://neisseria.org/ipnc/2008/Abstracts_oral_presentations_IPNC_2008.pdf).
- Richmond, P., Marshall, H., Nissen, M. D., Lambert, S., Jones, T., Gruber, W. & Arora, A. (2008). A randomized, observer-blinded, active control, phase 1 trial of meningococcal serogroup B rLP2086 vaccine in healthy children and adolescents aged 8 to 14 years. Abstract number P2132 in *16th International Pathogenic Neisseria Conference*, Rotterdam, The Netherlands, September 7–12 2008. Edited by L. van Alphen, P. van der Ley & G. van den Dobbelen. (http://neisseria.org/ipnc/2008/Abstracts_oral_presentations_IPNC_2008.pdf).
- Rokbi, B., Maitre-Wilmotte, G., Mazarin, V., Fourrichon, L., Lissolo, L. & Quentin-Millet, M. J. (1995). Variable sequences in a mosaic-like domain of meningococcal *tbp2* encode immunoreactive epitopes. *FEMS Microbiol Lett* **132**, 277–283.
- Scarselli, M., Cantini, F., Santini, L., Veggi, D., Dragonetti, S., Donati, C., Savino, S., Giuliani, M. M., Comanducci, M. & other authors (2009). Epitope mapping of a bactericidal monoclonal antibody against the factor H binding protein of *Neisseria meningitidis*. *J Mol Biol* **386**, 97–108.
- Schneider, M. C., Exley, R. M., Chan, H., Feavers, I., Kang, Y. H., Sim, R. B. & Tang, C. M. (2006). Functional significance of factor H binding to *Neisseria meningitidis*. *J Immunol* **176**, 7566–7575.
- Schneider, M. C., Prosser, B. E., Caesar, J. J., Kugelberg, E., Li, S., Zhang, Q., Quoraishi, S., Lovett, J. E., Deane, J. E. & other authors (2009). *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature* **458**, 890–893.
- Snape, M. D., Dawson, T., Morant, A., John, B., Ohene-Kena, R., Borrow, R., Oster, P. & Pollard, A. J. (2008). Immunogenicity and reactogenicity of a novel serogroup B *Neisseria meningitidis* vaccine

administered from 6 months of age. Abstract number 069 in *16th International Pathogenic Neisseria Conference*, Rotterdam, The Netherlands, September 7–12 2008. Edited by L. van Alphen, P. van der Ley & G. van den Dobbelsteen. (http://neisseria.org/ipnc/2008/Abstracts_oral_presentations_IPNC_2008.pdf).

Szczesny, P. & Lupas, A. (2008). Domain annotation of trimeric autotransporter adhesins – daTAA. *Bioinformatics* **24**, 1251–1256.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

Welsch, J. A., Rossi, R., Comanducci, M. & Granoff, D. M. (2004). Protective activity of monoclonal antibodies to genome-derived neisserial antigen 1870, a *Neisseria meningitidis* candidate vaccine. *J Immunol* **172**, 5606–5615.

Welsch, J. A., Ram, S., Koeberling, O. & Granoff, D. M. (2008). Complement-dependent synergistic bactericidal activity of antibodies against factor H-binding protein, a sparsely distributed meningococcal vaccine antigen. *J Infect Dis* **197**, 1053–1061.

Edited by: P. van der Ley