
















# Bacterial protein domains with a novel Ig-like fold target human CEACAM receptors

Nina M van Sorge<sup>1,†</sup> , Daniel A Bonsor<sup>2</sup> , Liwen Deng<sup>3</sup> , Erik Lindahl<sup>4</sup> , Verena Schmitt<sup>5</sup> , Mykola Lyndin<sup>5,6</sup> , Alexej Schmidt<sup>7</sup> , Olof R Nilsson<sup>8</sup>, Jaime Brizuela<sup>9</sup>, Elena Boero<sup>1</sup>, Eric J Sundberg<sup>2,10</sup> , Jos A G van Strijp<sup>1</sup> , Kelly S Doran<sup>3,†</sup> , Bernhard B Singer<sup>5,†</sup> , Gunnar Lindahl<sup>8,11,\*</sup>  & Alex J McCarthy<sup>1,9,\*\*</sup> 

## Abstract

*Streptococcus agalactiae*, also known as group B *Streptococcus* (GBS), is the major cause of neonatal sepsis in humans. A critical step to infection is adhesion of bacteria to epithelial surfaces. GBS adhesins have been identified to bind extracellular matrix components and cellular receptors. However, several putative adhesins have no host binding partner characterised. We report here that surface-expressed  $\beta$  protein of GBS binds to human CEACAM1 and CEACAM5 receptors. A crystal structure of the complex showed that an IgSF domain in  $\beta$  represents a novel Ig-fold subtype called IgI3, in which unique features allow binding to CEACAM1. Bioinformatic assessment revealed that this newly identified IgI3 fold is not exclusively present in GBS but is predicted to be present in adhesins from other clinically important human pathogens. In agreement with this prediction, we found that CEACAM1 binds to an IgI3 domain found in an adhesin from a different streptococcal species. Overall, our results indicate that the IgI3 fold could provide a broadly applied mechanism for bacteria to target CEACAMs.

**Keywords** Adhesin; immunoglobulin superfamily; Igl; receptor; *Streptococcus agalactiae*

**Subject Categories** Cell Adhesion, Polarity & Cytoskeleton; Microbiology, Virology & Host Pathogen Interaction; Structural Biology

**DOI** 10.15252/embj.2020106103 | Received 30 June 2020 | Revised 8 December 2020 | Accepted 10 December 2020

**The EMBO Journal (2021) e106103**

## Introduction

The Gram-positive bacterium *Streptococcus agalactiae*, also known as group B *Streptococcus* (GBS), is a major cause of pneumonia, septicaemia and meningitis in human neonates (Seale *et al*, 2017; Hall *et al*, 2017). Despite public health interventions, GBS is estimated to cause 410,000 infant infections and 150,000 stillbirths per year (Seale *et al*, 2017). From a total of 10 different serotypes, GBS belonging to serotypes Ia, Ib, II, III and V are most commonly associated with disease cases worldwide (Edmond *et al*, 2012). Cellular adhesion to epithelial surfaces is the first critical step preceding infection with GBS and many other pathogens (Patras & Nizet, 2018). Bacterial colonisation is a multifactorial process that requires expression of adhesins that target extracellular matrix (ECM) constituents and/or host cell receptors (Pietrocola *et al*, 2018; Shabayek & Spellerberg, 2018). Accordingly, GBS express a diverse array of surface adhesins and the exact repertoire varies from strain to strain, reflective of high plasticity of the GBS genome (Tettelin *et al*, 2005; Chen, 2019; Gori *et al*, 2020). Consequently, individual GBS strains differ in the host receptors that can be targeted during colonisation. The adhesins likely function cooperatively to facilitate persistent contact between the bacterial cell and the host, promoting cellular adhesion and invasion.

GBS interacts with several ECM constituents and the molecular mechanisms underpinning these interactions are well defined. The adhesins currently known to interact with ECM components are the fibrinogen-binding adhesins Srr1, FsbA, FsbB and FsbC (Schubert

1 Department of Medical Microbiology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

2 Institute of Human Virology, University of Maryland School of Medicine, University of Maryland, Baltimore, MD, USA

3 Department of Immunology & Microbiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

4 Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm, Sweden

5 Institute of Anatomy, Medical Faculty, University Duisburg-Essen, Essen, Germany

6 Department of Pathology, Sumy State University, Sumy, Ukraine

7 Department of Medical Biosciences, Umeå University, Pathology, Umeå, Sweden

8 Department of Laboratory Medicine, Division of Medical Microbiology, Lund University, Lund, Sweden

9 Department of Infectious Disease, MRC Centre for Molecular Bacteriology & Infection, Imperial College London, London, UK

10 Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, USA

11 Department of Chemistry, Division of Applied Microbiology, Lund University, Lund, Sweden

\*Corresponding author. Tel: +46 735 342255; E-mail: gunnar.lindahl@tmb.lth.se

\*\*Corresponding author. Tel: +44 20 7594 3868; E-mail: a.mccarthy@imperial.ac.uk

†These authors contributed equally to this work

‡Present address: Department of Medical Microbiology, Infection Prevention and Netherlands Reference Laboratory for Bacterial Meningitis, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands

et al, 2002; Gutekunst et al, 2004; Seo et al, 2012; Buscetta et al, 2014), the keratin 4-binding adhesin Srr1 (Samen et al, 2007), the fibronectin-binding adhesins BsaB and SfbA (Jiang & Wessels, 2014; Mu et al, 2014), and the collagen-binding adhesin PilA (Banerjee et al, 2011). These interactions promote epithelial colonisation (Schubert et al, 2004; Samen et al, 2007; Sheen et al, 2011; Wang et al, 2014) as well as cellular invasion and/or invasive disease (Tenenbaum et al, 2005; van Sorge et al, 2009; Banerjee et al, 2011; Seo et al, 2012; Mu et al, 2014; Deng et al, 2019). Generally, our knowledge of the mechanisms that GBS utilises to directly adhere to host cells is limited (Bolduc & Madoff, 2007; Patras & Nizet, 2018; Pietrocola et al, 2018);  $\alpha$  adhesin binds  $\alpha 1\beta 1$ -integrin to promote epithelial cell internalisation (Bolduc & Madoff, 2007), BspC adhesin binds vimentin (Deng et al, 2019), and the BspA adhesin interacts with gp340, a mucin-like glycoprotein associated with the surface of mucosal tissues (Rego et al, 2016). However, the host receptor targets of several putative adhesins remain uncharacterised, including Rib, Sip, LrrG, HvgA and BibA proteins (Stålhammar-Carlemalm et al, 1993; Brodeur et al, 2000; Santi et al, 2007; Tazi et al, 2010). Consequently, it is expected that several GBS adhesin–host factor interactions remain uncharacterised. Their identification is important for development of a vaccine or anti-bacterial strategies that interfere with GBS mucosal colonisation (Michel et al, 1992; Larsson et al, 1996; Heath, 2016; Pietrocola et al, 2018).

We report here that a subset of GBS strains interact with carcinoembryonic antigen-related cell adhesion molecules (CEACAM) receptors. Human CEACAMs are a family of 12 cellular receptors belonging to the immunoglobulin (Ig) superfamily (IgSF) that are commonly expressed on epithelial cells, endothelial cells and leucocytes (Gray-Owen & Blumberg, 2006). CEACAMs regulate immune responses through formation of homophilic and heterophilic interactions (Kuespert et al, 2006; Gray-Owen & Blumberg, 2006; Bonsor et al, 2015b). These properties allow CEACAMs to regulate multiple physiological and pathophysiological processes including cell-to-cell communication, epithelial differentiation, apoptosis and regulation of pro-inflammatory reactions (Gray-Owen & Blumberg, 2006; Khairnar et al, 2015, 2018; Helfrich & Singer, 2019). Each CEACAM is composed of an N-terminal domain with V-set (IgV) fold, which promotes CEACAM–CEACAM interactions, and varying numbers of C2-set (IgC2) folds (Zhou et al, 1993; Watt et al, 2001; Bonsor et al, 2015a, 2015b). Interestingly, human CEACAMs have been identified as docking receptors for Gram-negative bacteria (Chen & Gotschlich, 1996; Virji et al, 1996; Hill et al, 2001; Connors et al, 2008; Tchoupa et al, 2014; Tchoupa et al, 2015; Javaheri et al, 2016; Königer et al, 2016; Brewer et al, 2019). The characterised bacterial adhesins include Opa of *Neisseria* spp., HopQ of *Helicobacter pylori*, UspA1 of *Moraxella catarrhalis*, P1 of *Haemophilus influenzae*, Dr adhesins of *Escherichia coli* and CbpF of *Fusobacterium* spp. Interaction of these adhesins with human CEACAMs promotes cellular adhesion, cellular invasion, translocation of virulence factors and tissue penetration (Billker et al, 2002; Korotkova et al, 2008; Tchoupa et al, 2014; Tchoupa et al, 2015; Königer et al, 2016; Javaheri et al, 2016; Islam et al, 2018). Of note, these CEACAM-binding bacterial adhesins are structurally distinct, implying that they arose through convergent evolution. Since CEACAMs have not previously been reported to interact with Gram-positive bacteria, it was therefore of considerable interest to characterise the mechanism by which GBS binds to CEACAM receptors.

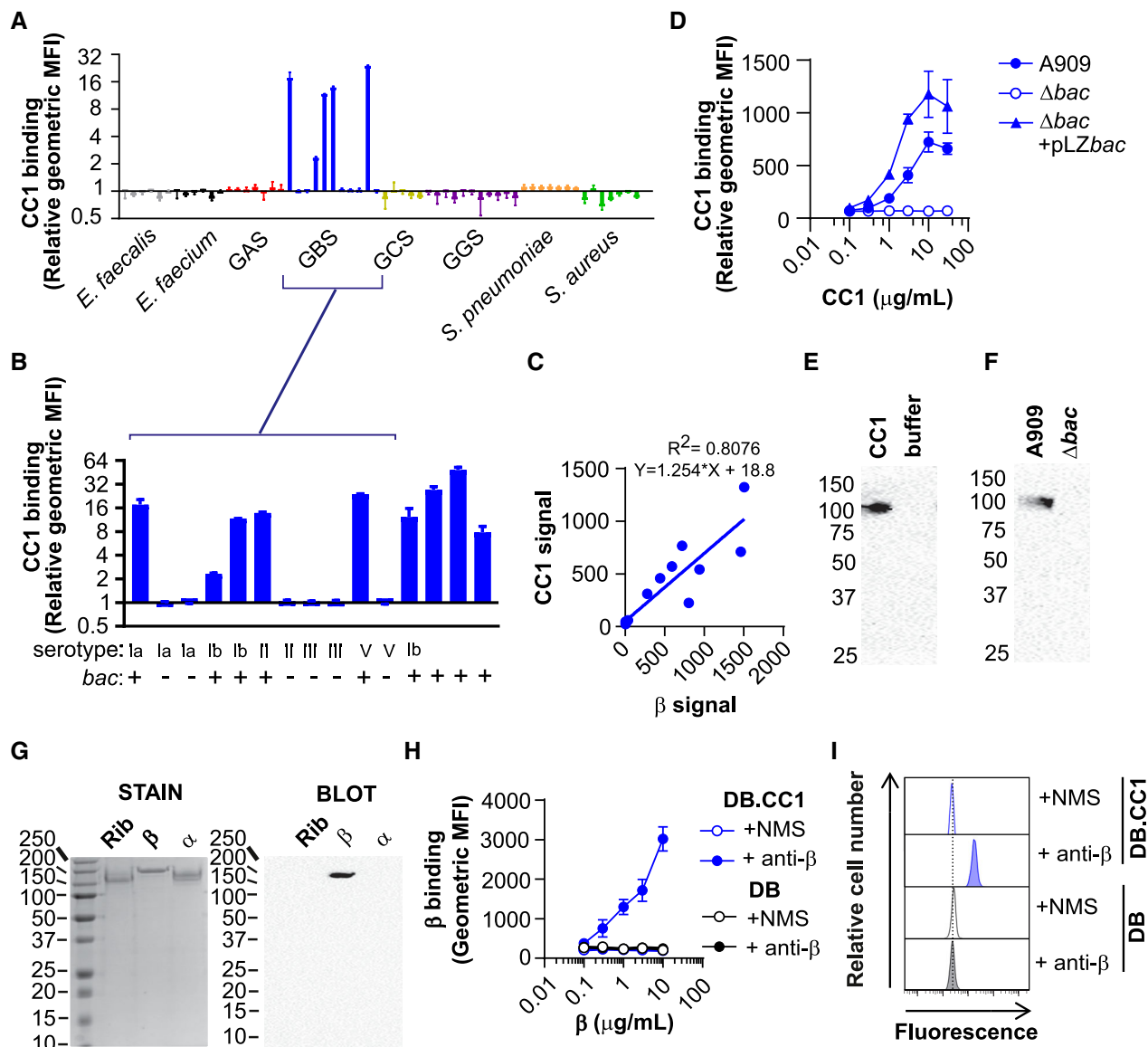
We used biochemical and cellular assays to reveal that the GBS surface protein  $\beta$  binds specifically to human CEACAM1 and CEACAM5 and that an IgSF domain in  $\beta$  promotes the binding. Through structural methods, we demonstrate that the IgSF domain in  $\beta$  protein contains unique structural features and represents a previously unrecognised variant of the IgI fold, which we termed the I3-set fold (IgI3). Homologs of  $\beta$ -IgI3 were identified in adhesins from several human bacteria including pathogens. We confirmed that CEACAM1 recognised one of these IgI3 variants and two streptococcal species. Together, these findings suggest that the interaction between bacterial IgI3 domains and CEACAMs may be of general importance for host barrier colonisation.

## Results

### The group B *Streptococcus* surface $\beta$ protein binds CEACAM1

GBS can successfully adhere to human epithelial cell lines, including vaginal, cervical and airway epithelial cell lines (Patras & Nizet, 2018). Yet, knowledge of the specific host factors that can be targeted by GBS for adhesion is limited. As CEACAM receptors are mucosal docking targets for microorganisms (Tchoupa et al, 2014), we hypothesised that GBS interacts with CEACAM receptors. Given that individual GBS strains are likely to vary in their molecular adhesion mechanisms due to genomic plasticity (Tettelin et al, 2005; Chen, 2019; Gori et al, 2020), we tested interaction of rCEACAM1 with a panel of genetically diverse genome-sequenced GBS strains. We used CEACAM1 in the screen as it binds all known bacterial ligands of CEACAMs. Using rCEACAM1 (Fig EV1A), we observed binding to GBS strains including reference strains A909 and H36B (Figs 1A and EV1A). To further test the specificity of the interaction, we also screened the ability of rCEACAM1 to interact with a variety of other Gram-positive species including *Streptococcus* spp., *Enterococcus* spp. and *Staphylococcus* spp. Notably, rCEACAM1 only bound to the indicated GBS strains (Fig 1A). These data reveal for the first time that GBS can bind to human CEACAM1 receptor.

Next, we aimed to identify the GBS adhesin responsible for binding rCEACAM1. We utilised the GBS genome comparison dataset of Tettelin et al. (2005), which includes rCEACAM1 binding (A909 and H36B) and non-binding (18RS21 and NCTC10/84) strains, to identify genes encoding cell wall-anchored proteins that were associated with rCEACAM1-binding phenotype. We identified that genomic island of diversity region 3.1 was present in rCEACAM1-binding strains (A909 and H36B) and absent in non-binding strains (515, COH1, NEM316 and NCTC10/84). This region encodes a cell wall-anchored protein known as  $\beta$  protein, or simply  $\beta$ , that is encoded by *bac*. To test the hypothesis that  $\beta$  protein was responsible for CEACAM1 interaction, we screened rCEACAM1 binding to a broader collection of GBS isolates (Fig 1B) and observed that CEACAM1 binding correlated with carriage of *bac* (Hedén et al, 1991; Lindahl et al, 2005). The intensity of rCEACAM1 binding differed between interacting GBS strains, which may reflect differences in  $\beta$  expression. Indeed,  $\beta$  protein expression levels, detected using anti- $\beta$  serum (Fig EV1B), correlated with rCEACAM1-binding capacity (Fig 1C). Finally, we tested whether CEACAM1 binding was dependent on  $\beta$  protein expression. Deletion of *bac* in the GBS A909



**Figure 1. Human CEACAM1 binds to group B *Streptococcus* (GBS) via the surface-expressed  $\beta$  protein.**

- A, B Binding of recombinant (r)CEACAM1 (CC1)-HIS (10  $\mu$ g/ml) to (A) a panel of Gram-positive bacteria, namely *Enterococcus faecalis*, *Enterococcus faecium*, group A *Streptococcus* (GAS), group B *Streptococcus* (GBS), group C *Streptococcus* (GCS), group G *Streptococcus* (GGS), *Streptococcus pneumoniae*, *Staphylococcus aureus*, (B) an expanded collection of GBS isolates, with serotype and carriage of *bac* gene shown. rCC1 binding to live bacteria was quantified by a secondary anti-His-FITC monoclonal antibody (mAb). Mean and standard deviation (SD) values are reported for  $n = 3$  independent experiments. In (B), GBS strains from left to right are A909, BS39, 515, BS22, H36B, BS29, 18RS21, BM110, COH1, SBL3066, NCTC10/84, SB35, SB10, SB20, BS26.
- C Correlation between rCC1-binding capacity and  $\beta$  protein expression in GBS strains, respectively.  $\beta$  protein expression was quantified using rabbit anti- $\beta$  protein serum and a secondary PE-conjugated goat anti-rabbit-IgG. rCC1 binding to live bacteria was quantified by a secondary anti-His-FITC mAb. Mean binding values from  $n = 3$  independent experiments.
- D Concentration-dependent binding of rCC1-HIS to wild type (WT),  $\Delta bac$  and  $\Delta bac + pLZ.bac$  complemented A909 strains. Mean and SD values are reported for  $n = 3$  independent flow cytometry experiments.
- E Pull-down of human rCC1-HIS (30  $\mu$ g/ml) by GBS strain A909 strain, detected using Western blot analysis.
- F Human rCC1-HIS (30  $\mu$ g/ml) binding to GBS A909 and A909 $\Delta bac$  strains analysed by pull-down experiments followed by Western blot analysis.
- G Western blot analysis of rCC1-HIS (30  $\mu$ g/ml) binding to purified  $\beta$  protein, but not to control GBS surface proteins  $\alpha$  and Rib. Panel on the left show staining of proteins after separation by SDS-PAGE.
- H, I Binding of purified  $\beta$  protein to dynabeads (DB) coated with rCC1 (DB.CC1) or control. Binding of  $\beta$  protein was detected using mouse anti- $\beta$  serum or normal mouse serum (NMS), and PE-conjugated goat anti-mouse-IgG. (H) shows mean and SD values compiled from  $n = 3$  independent replicates, and (I) shows representative flow cytometry plots using 10  $\mu$ g/ml rCC1.

Data information: Fluorescence of bacteria (A, B, C and D) and DB (H and I) was measured by flow cytometry. Western Blot in (E, F and G) employed mouse anti-CC1 mAb (clone C5-1X/8) and HRP-conjugated goat anti-mouse-IgG pAb.

reference strain abolished rCEACAM1 binding, whilst complementation with a vector carrying the *bac* gene re-established rCEACAM1 binding (Figs 1D and EV1C). Mutation of genes encoding other GBS surface adhesins or capsule did not influence interaction with rCEACAM1 (Fig EV1C). Additionally, we were able to pull down rCEACAM1 with GBS A909 (Fig 1E), but not the  $\Delta bac$  strain (Fig 1F). Direct interaction of rCEACAM1 with purified  $\beta$  protein, but not with purified GBS surface proteins Rib or  $\alpha$ , was demonstrated by Western blot analysis (Fig 1G). Finally,  $\beta$  protein bound to CEACAM1-coated dynabeads (DB), but not control beads, in a concentration-dependent manner (Fig 1H and I). Together, these data unequivocally show that GBS specifically binds to human CEACAM1 through expression of  $\beta$  protein.

### CEACAM1 binds to an IgSF domain in $\beta$ protein

$\beta$  protein is a multifunctional protein and interacts with several proteins of the human immune system, including factor H, Siglec-5, Siglec-7, Siglec-14 and IgA, through distinct domains (Fig 2A) (Lindahl *et al.*, 1990; Hedén *et al.*, 1991; Areschoug *et al.*, 2002b; Carlin *et al.*, 2009; Nordström *et al.*, 2011; Ali *et al.*, 2014; Fong *et al.*, 2018). To pinpoint the  $\beta$  protein domain interacting with CEACAM1, we expressed and purified recombinant  $\beta$  protein domains (B6N, IgABR, B6C, IgSF and  $\beta$ 75KN) from *E. coli* (Fig 2B) and tested their interaction with rCEACAM1. These domains include a part of  $\beta$  called IgSF because the domain was predicted to adopt an immunoglobulin superfamily (IgSF) fold (Bateman *et al.*, 1996; Lindahl *et al.*, 2005). IgSF folds are composed of around 100 amino acids comprising two  $\beta$  sheets that pack face-to-face. rCEACAM1 bound in a concentration-dependent manner to DB coated with the biotinylated IgSF domain, but not to DB coated with other  $\beta$  protein domains or biotinylated-HSA (Fig 2C). As a control, we included rSiglec-5, which interacted only with DB coated with the B6N domain as described previously (Fig 2C; Nordström *et al.*, 2011). Furthermore, DB coated with rCEACAM1 interacted with the IgSF domain, but not HSA, coupled to streptavidin (Fig 2D). Thus, the IgSF domain in  $\beta$  specifically binds to human CEACAM1. We assessed whether the IgSF domain shared structural homology with other resolved IgSF structures. Structural predictions of the  $\beta$ -IgSF domain suggested resemblance to a V-set Ig (IgV) domain identified in SrpA, a cell surface-anchored protein

in *Streptococcus sanguinis* (Fig 2E; Bensing *et al.*, 2016). However, rCEACAM1 did not bind to the surface of a SrpA-expressing *S. sanguinis* strain (Appendix Fig S1). We denoted the IgSF fold in  $\beta$  as  $\beta$ -IgSF.

The CEACAM1-4L isoform possesses four Ig-like domains (Appendix Fig S2A; Gray-Owen & Blumberg, 2006), including the N-terminal IgV-like domain that forms the homo- and heterophilic interactions implicated in the CEACAM-mediated functions (Bonsor *et al.*, 2015b). All bacterial ligands characterised to date for CEACAM1 target the N-terminal IgV-like domain. To identify the CEACAM1 domain targeted by  $\beta$ -IgSF, we tested whether domain-specific CEACAM1 monoclonal antibodies (mAb) could inhibit the interaction of  $\beta$ -IgSF with CEACAM1-coated DB. Only a mAb that blocks the N-terminal domain of CEACAM1 abolished the interaction with  $\beta$ -IgSF (Fig 2F; Appendix Fig S2B). Similarly, the *H. pylori* protein HopQ, which binds to the N-terminal domain (Bonsor *et al.*, 2018), blocked the  $\beta$ -IgSF-CEACAM1 interaction (Appendix Fig S2C). To test for direct interaction with the N-terminal domain, we coated DB with rCEACAM1-N or rCEACAM1-A1B1A2 (Appendix Fig S2D) and observed concentration-dependent binding of  $\beta$ -IgSF only to DB coated with rCEACAM1-N (Fig 2G). In the reciprocal experiment, rCEACAM1-N, but not rCEACAM1-A1B1A2, bound to DB coated with  $\beta$ -IgSF (Appendix Fig S2E).

### The $\beta$ -IgSF domain binds with high affinity to human CEACAM1 and CEACAM5

Bacterial adhesins of CEACAM1 often also bind to the N-terminal domain of CEACAM3, CEACAM5 and CEACAM6, reflecting the high sequence (~90% identity) and structural homology between CEACAM family members (Popp *et al.*, 1999; Gray-Owen & Blumberg, 2006). However, no bacterial adhesins to date bind CEACAM8. To ascertain the CEACAM-binding profile for  $\beta$ -IgSF, we measured the affinity of  $\beta$ -IgSF for unglycosylated N-terminal domains of these five CEACAMs, purified from *E. coli*, by isothermal calorimetry (ITC) (Fig 3A; Appendix Table S3).  $\beta$ -IgSF bound with high affinity to rCEACAM1-N ( $K_D = 96 \pm 2$  nM) and rCEACAM5-N ( $K_D = 152 \pm 27$  nM), but did not bind to rCEACAM3-N, rCEACAM6-N or rCEACAM8-N. The affinities of the ( $\beta$ -IgSF)-(CEACAM1-N) and ( $\beta$ -IgSF)-(CEACAM5-N) interactions were comparable to those reported for other bacterial ligands (Korotkova

**Figure 2. The IgSF domain in the  $\beta$  protein of GBS binds to the N-terminal domain of CEACAM1.**

- Schematic representation of the  $\beta$  protein (Hedén *et al.*, 1991). The first residue of each domain is indicated. XPZ is a proline-rich repeat region (Areschoug *et al.*, 2002a). Known human ligand binding partners are indicated above each respective domain. IgA binds to a region denoted IgABR. Factor H binds the  $\beta$ 75K region, but does not bind via XPZ.
- Staining of recombinant  $\beta$  protein domains (B6N, IgABR, B6C,  $\beta$ -IgSF and  $\beta$ 75KN) after separation by SDS-PAGE. Recombinant proteins were expressed in *E. coli*.  $\beta$  protein purified from GBS is used as a control. Dotted boxes signal proteins of expected size.
- Binding of rCC1-His or rSiglec5-Fc to dynabeads (DB) coated with  $\beta$  protein domains. Mean and SD values are reported for  $n = 2$  replicates.
- Binding of varying concentrations of  $\beta$ -IgSF or human serum albumin (HSA) coupled to streptavidin-PE (SPE), to DB coated with CC1 (DB.CC1) or control protein (DB). Mean and SD values are reported for  $n = 6$  replicates.
- Structure of the relevant (Siglec-like sialic acid-binding) domain of the highest-scoring template 5KIQ (SrpA binding region), and predicted three-dimensional model of the IgSF domain from the  $\beta$  protein, constructed from the Hhpred alignment using MODELLER (Webb & Sali, 2016).
- Inhibition of  $\beta$ -IgSF binding to DB.CC1 by anti-CC1-N mAb (clone CC1/3/5-Sab), but not by anti-CC1-A1B1 mAb (clone B3-17).  $\beta$ -IgSF-biotin was coupled to SPE. Mean and SD values are reported for  $n = 6$  replicates.
- Concentration-dependent binding of  $\beta$ -IgSF to DB coated with 100  $\mu$ g/ml CC1 (DB.CC1), CC1-N (DB.CC1-N) or CC1-A1B1A2 (DB.CC1-A1B1A2). Assays utilised proteins coupled to SPE. Mean and SD values are reported for  $n = 6$  replicates.

Data information: Fluorescence of DB in (C, D and F) was measured by flow cytometry.

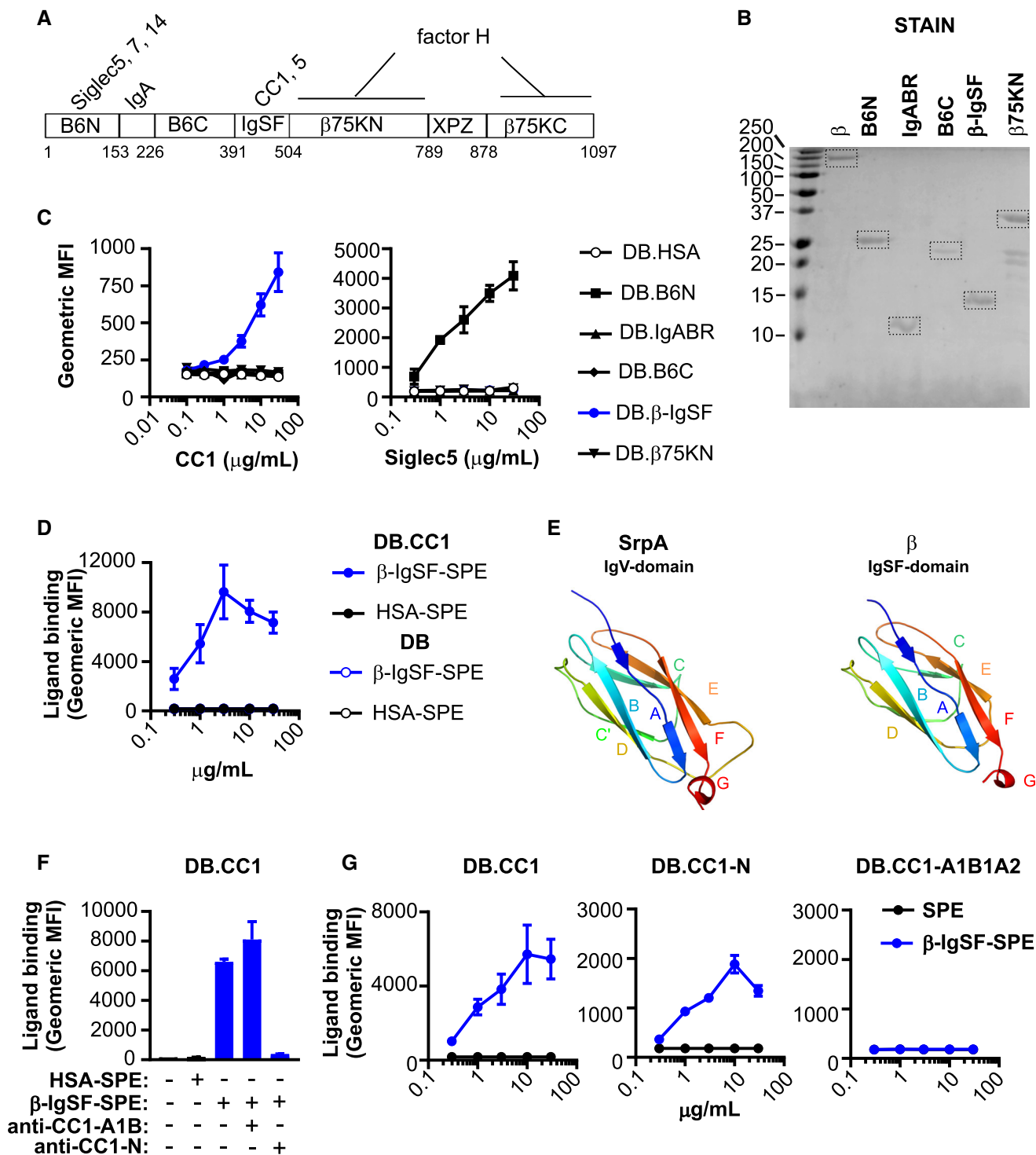
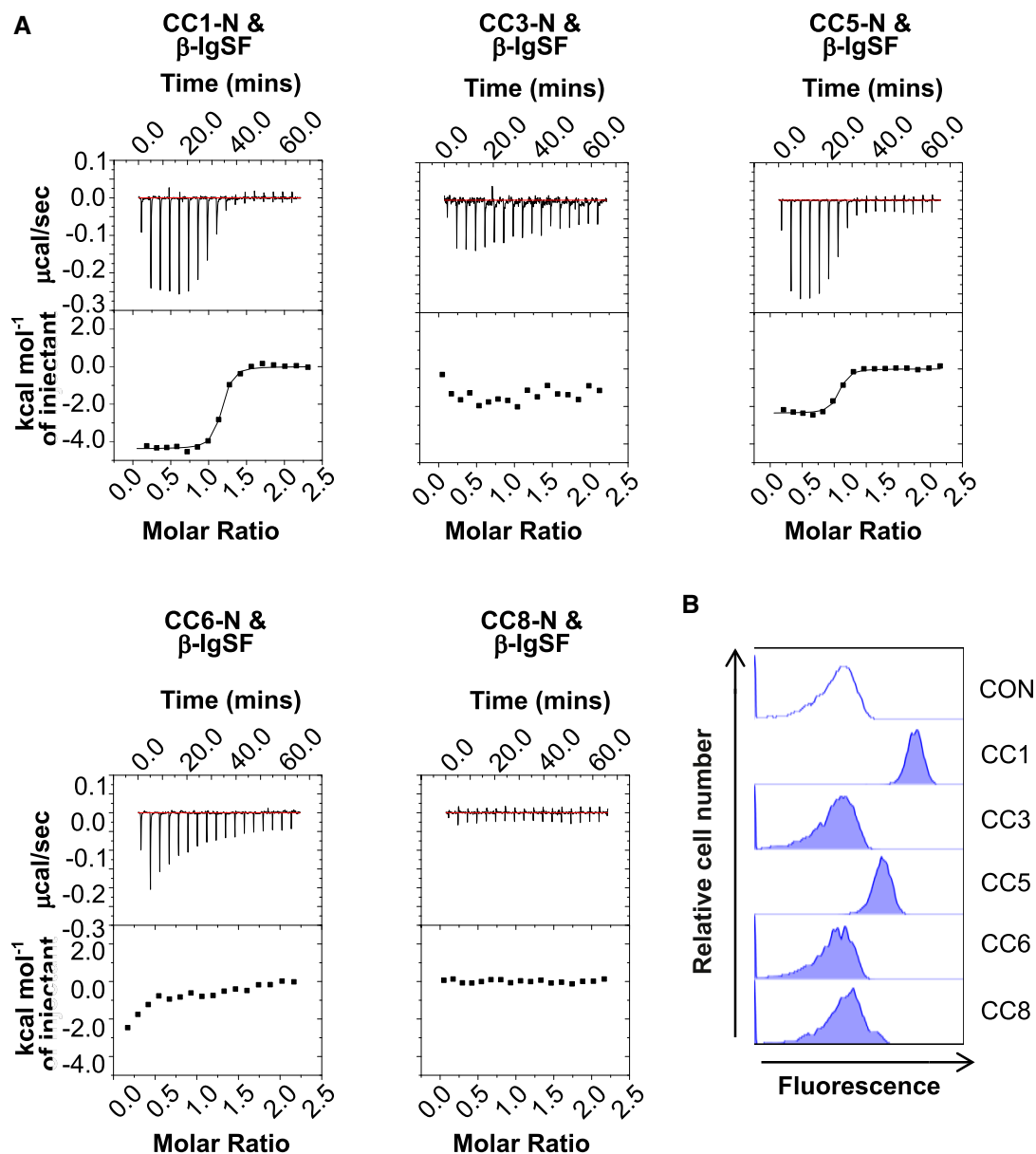


Figure 2.

*et al*, 2008; Bonsor *et al*, 2018). As the binding of β-IgSF is weaker to the isolated N-terminal domain compared to glycosylated CEACAM1 (Fig 2G), other CEACAM1 domains may stabilise the interaction, as reported for the HopQ-CEACAM1 interaction (Bonsor *et al*, 2018). To confirm that the CEACAM-binding profile of the protein is valid when expressed on the surface of GBS bacteria, we tested binding of the full-length glycosylated rCEACAMs to GBS by

flow cytometry. The β-expressing GBS A909 strain only recognised CEACAM1 and CEACAM5 (Fig 3B). Specificity for this subset was generally confirmed in a wider panel of β-expressing GBS strains (Appendix Fig S3), though variation existed in CEACAM5 binding. Although several β-expressing GBS strains did not bind CEACAM5, the β-IgSF domain displayed 100% sequence conservation amongst 57 GBS proteins identified through BLAST search (Appendix Fig



**Figure 3.**  $\beta$ -IgSF binds with high affinity to human CEACAM1 and CEACAM5.

A Representative isothermal titration calorimetry (ITC) binding curves of CEACAM-N domains and  $\beta$ -IgSF duplicates. Experiments were performed using  $\beta$ -IgSF and N domains of (r)CEACAM1 (CC1), CEACAM3 (CC3), CEACAM5 (CC5), CEACAM6 (CC6) and CEACAM8 (CC8).

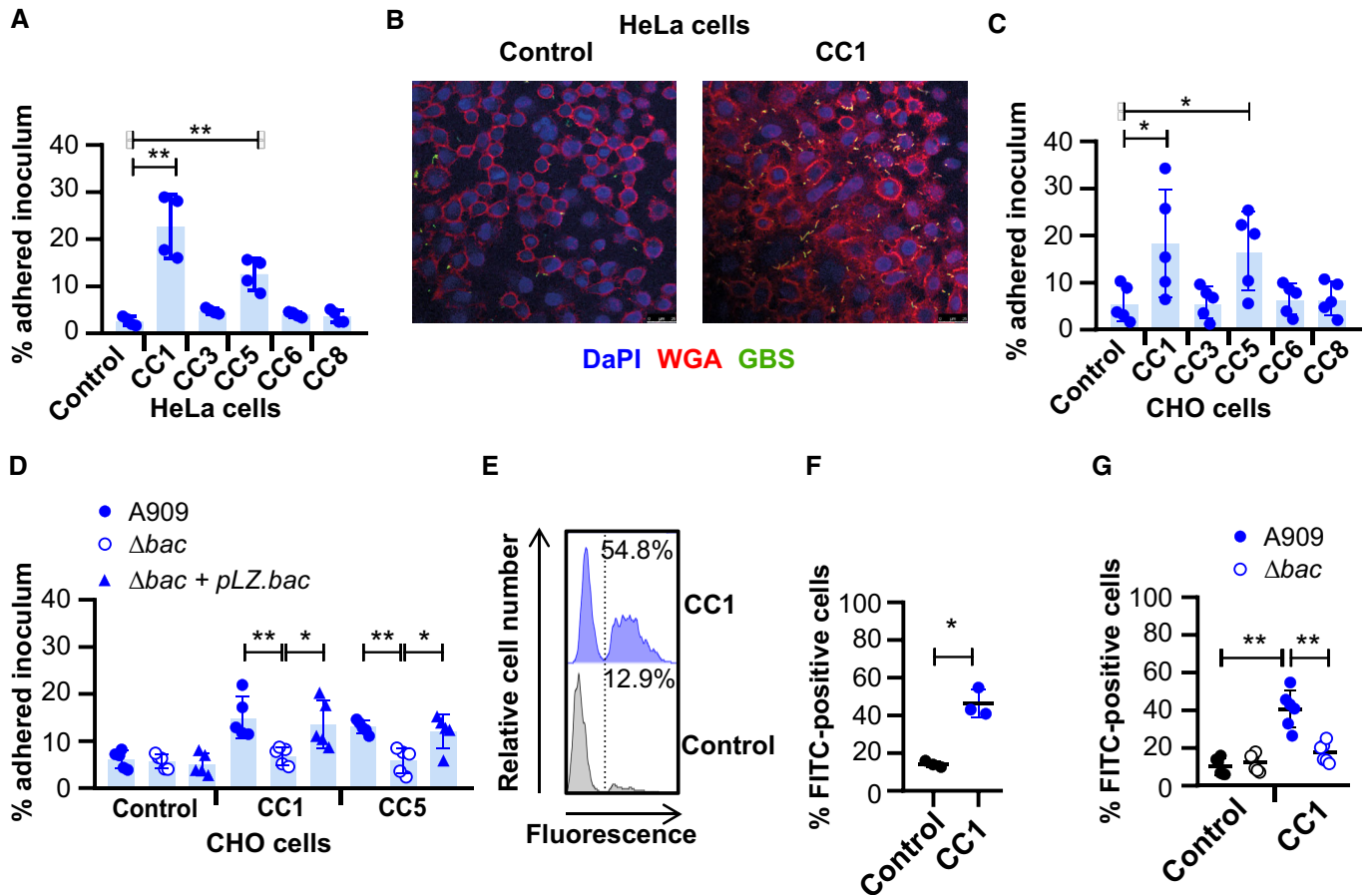
B Representative flow cytometry plots analysing the interaction of rCC1, CC3, CC5, CC6 or CC8 (10  $\mu\text{g/ml}$ ) with GBS A909. Fluorescence of bacteria was measured by flow cytometry. Representative data from  $n = 3$  independent experiments.

S4), indicating that intrastain variation in CEACAM5 binding is most likely due to differential  $\beta$  expression.

#### Expression of human CEACAMs on epithelial cells enhances GBS adhesion

As CEACAM engagement leads to enhanced cellular adhesion of Gram-negative microorganisms, we hypothesised that CEACAMs could represent a novel cellular adhesion mechanism for GBS. Therefore, we tested the binding of strain A909 to well-

characterised CEACAM-expressing HeLa cells (Fig 4A; Bos *et al*, 1998). Consistent with rCEACAM binding, a higher percentage of the A909 inoculum was recovered from the CEACAM1- and CEACAM5-expressing HeLa cells in comparison to all other HeLa cell lines (Fig 4A). Increased binding of GBS to human CEACAM1-expressing HeLa cells was confirmed by confocal microscopy (Fig 4B). To ensure that the CEACAM-observed binding was not influenced by the background of the cell line, we assessed GBS adhesion to CEACAM-expressing CHO cells (Hollandsworth *et al*, 2020). Also, a higher percentage of the GBS inoculum was



**Figure 4. GBS exploit human CEACAM for enhanced adhesion to epithelial cells.**

- A Adherence of GBS A909 to human CEACAM1 (CC1)-, CEACAM3 (CC3)-, CEACAM5 (CC5)-, CEACAM6 (CC6)- or CEACAM8 (CC8)-expressing, or an empty vector control, HeLa transfectants at MOI of 10 for 30 min. Mean and SD for  $n = 4$  independent replicates are shown.
- B Confocal imaging of FITC-labelled GBS A909 strain adhesion to human CC1-expressing or empty vector control HeLa transfectants. Cell membranes were stained with AF-647-conjugated wheat germ agglutinin (WGA) (red) and nuclei with DAPI (blue). Scale bar: 26  $\mu\text{m}$ .
- C Adherence of A909 to human CC1-, CC3-, CC5-, CC6- or CC8-expressing, or an empty vector control, CHO transfectants at MOI of 10 for 30 min. Mean and SD from  $n = 5$  independent replicates is displayed.
- D Adhesion of WT,  $\Delta bac$  and  $\Delta bac + pLZ.bac$  complemented A909 strains to human CC1- or CC5-expressing CHO transfectants, or empty vector control cells. Mean and SD from  $n = 5$  independent replicates is displayed.
- E, F Adhesion of FITC-labelled A909 to human CC1-expressing, or control, CHO transfectants at an MOI of 10. Cell lines were detached and then incubated with FITC-labelled A909 for 30 min at 4°C. The percentage of fluorescent cells was calculated for each population. Representative flow cytometry plots are shown (E), and the integrated results displaying the mean and SD from 3 independent replicates are reported in (F).
- G Adhesion of FITC-labelled A909 strains to human CC1-expressing or control CHO transfectants at MOI of 10. Cell lines were detached and then incubated with FITC-labelled A909 strains for 30 min at 4°C. Mean and SD for each population for  $n = 6$  independent replicates is displayed.

Data information: Fluorescence of cells in (E, F and G) was measured by flow cytometry analysis, where the percentage of fluorescent cells was calculated. The number of bacteria adhered to cell monolayers in (A, C and D) were quantified by growth on Todd-Hewitt agar and enumeration of colony forming units (CFU). Data in (A, C, D, F and G) were analysed by repeated one-way ANOVA with Sidak's multiple comparisons. \* $P < 0.05$  and \*\* $P < 0.01$ .

recovered from the CEACAM1- and CEACAM5-expressing CHO cells in comparison to all other CHO cell lines (Fig 4C). For unknown reasons, there was considerable variation in GBS adhesion to the CEACAM1- and CEACAM5-expressing CHO cells. This variation could not be attributed to unstable CEACAM expression, as our cell lines expressed CEACAM at consistent levels (Fig EV2A and B). It is possible that the variation in GBS adhesion could be due to differences in CEACAM-expression density by CHO cells or due to differences in  $\beta$  expression by GBS strains in replicated experiments.

To confirm that adhesion of the GBS strain A909 to the CEACAM1-expressing CHO cell line was dependent on  $\beta$  protein, we also assessed binding of the *bac* deletion mutant and the complemented strain in this same system. GBS adhesion was abolished by mutation of the *bac* gene, and the phenotype could be recapitulated by the complemented mutant (Fig 4D). This result was confirmed by confocal microscopy (Fig EV2C). In agreement with the results obtained with purified proteins (Fig 2F), the binding of  $\beta$ -expressing GBS to CEACAM1-expressing CHO cells was inhibited by blocking the CEACAM1-N domain with a specific mAb (Fig EV2D).

Moreover, pre-incubation of A909 with rCEACAM1-N, but not rCEACAM8-N, impaired adhesion to CEACAM1-expressing CHO transfectants (Fig EV2E). This indicates that the CEACAM1-N and  $\beta$ -IgSF domains were both responsible for the cellular adhesion phenotype.

To rule out the possibility that differences in adhesion of A909 wild-type and  $\Delta bac$  strains to CHO cells reflected interstrain variation in growth rates during adhesion at 37°C, we developed an alternative assay in which we assessed adhesion of FITC-labelled GBS strains to detached cell lines during incubation at 4°C. A higher percentage of CEACAM1-expressing CHO cells were FITC-positive upon incubation with A909 in comparison to control cells (Fig 4E and F). As expected, adhesion of FITC-labelled A909 to the CEACAM1-expressing CHO cell was abolished by mutation of *bac* (Figs 4G and Fig EV2F). Collectively, the data therefore indicate that binding to human CEACAM receptors, via  $\beta$ , is a novel cellular adhesion mechanism for GBS.

### Crystallography reveals that the IgSF domain in $\beta$ adopts a novel Ig fold, the IgI3 fold

To gain insights into  $\beta$ -IgSF binding mechanisms, we aimed to solve its structure. We solved the structure in complex with CEACAM1-N at a resolution of 3.25 Å (Fig 5A). The asymmetric unit contains two molecules of the ( $\beta$ -IgSF)-(CEACAM1-N) complex, which are similar to each other when superimposed (r.m.s.d 0.20 Å). All analysis of the ( $\beta$ -IgSF)-(CEACAM1-N) complex has been performed with chains B and C of the co-crystal complex. The  $\beta$ -IgSF domain has the characteristic features of an Ig fold, principally a pair of  $\beta$  sheets built of anti-parallel  $\beta$  strands that surround a hydrophobic core (Fig 5B). IgSF domains can be classified into (variable) V-set, (constant) C-set or (intermediate) I-set, with differentiation based on the number and placement of  $\beta$ -strands between the conserved cysteine residue disulphide bridge (Fig 5C; Wang & Springer, 1998; Wang, 2013). The V-set Ig (IgV) domains contain ten  $\beta$  strands with four strands found on one sheet (ABED) and six strands on the other (A'GFCC'C'). C-set Ig (IgC) domains lack the A' and C' strands and are further grouped into C1 or C2 based on presence or absence of the D strand, respectively. I-set Ig (IgI) domains lack the C' strand and are classified into I1 or I2 based on presence or absence of a D strand, respectively. The  $\beta$ -IgSF domain has two  $\beta$ -sheets labelled ABED and A'GFC, with sheets connected by the BC, EF, CD and AA' loops (Fig 5B). Therefore,  $\beta$ -IgSF has an I-set fold topology that most closely resembles an I1-set domain (Fig 5C). However,  $\beta$ -IgSF lacks cysteines and disulphide bridges that are characteristic for I-set folds. Furthermore, the  $\beta$ -IgSF domain

possesses a truncated C strand that is directly followed by 1.5-turn  $\alpha$ -helix (Fig 5C). These features were not observed in the structurally characterised IgI1 domains in the DALI or PDBeFOLD databases that  $\beta$ -IgSF most closely resembled, including macrophage colony stimulating factor 1 (MCS-F) and intracellular adhesion molecules 3 (ICAM-3) (Figs 5D and Fig EV3A and B). Therefore, the topology of  $\beta$ -IgSF domain represents a previously unrecognised IgI fold subtype, denoted here as I3-set Ig (IgI3) domain. Accordingly, the domain in  $\beta$  will now be referred to as  $\beta$ -IgI3. The unique features of this domain are (a) the absence of cysteine residues, (b) absence of C' and C'' strands and (c) a truncated C strand that is directly followed by a 1.5-turn  $\alpha$ -helix. Of note, the unique  $\beta$ -IgI3 region stretching from the C to the D strand possesses protruding hydrophobic residues, such as F42, located between the C strand and the  $\alpha$ -helix, L46 located in the  $\alpha$ -helix, and V53 located in the CD loop (Fig 5E).

### Identification of interacting residues in the ( $\beta$ -IgI3)-(CEACAM1-N) complex

CEACAMs form their homophilic and heterophilic interactions through the A'GFCC'C' face of the V-set domain (Figs 5A and 6A). Bacterial ligands also bind to the A'GFCC'C' face (Virji *et al.*, 1999; Villullas *et al.*, 2007; Connors *et al.*, 2008; Korotkova *et al.*, 2008; Bonsor *et al.*, 2018; Moonens *et al.*, 2018; Brewer *et al.*, 2019). The *H. pylori* ligand HopQ uses a coupled folding and binding mechanism (Bonsor *et al.*, 2018), and simulated docking indicated that the *E. coli* ligand AfaE binds to the CEACAM dimerisation interface through the BE strands and DE loop of an incomplete Ig fold (Anderson *et al.*, 2004; Korotkova *et al.*, 2008).

The structure determination by X-ray crystallography (Fig 6B; Appendix Table S4) of the complex of  $\beta$ -IgI3 and CEACAM1-N at 3.25 Å resolution showed that  $\beta$ -IgI3 also binds to the A'GFCC'C' face of CEACAM1 through residues located in the C to D strand region including the  $\alpha$ -helix (Figs 6C and EV3C). To identify the residues forming contact sites between  $\beta$ -IgI3 and CEACAM1-N within the co-crystal, we used the NCONT sub-program of CCP4 (Appendix Table S5). Figure 6B shows the interacting residues on the surfaces of the two molecules. Specifically, L46 in the  $\alpha$ -helix of  $\beta$ -IgI3 interacts by van der Waals forces with F29 and L95 (distances 3.34 Å and 3.59 Å) located in the C and G strands of CEACAM1, respectively (Fig EV3D; Appendix Table S5). In addition, the protruding  $\beta$ -IgI3 residue V53 is within close contact of I91 (distance 3.36 Å, F strand), Q44 (distance 2.50 Å, C to C' loop) and S32 (distance 3.41 Å, C strand) of CEACAM1, D55 is within close contact of Y34 (distance 3.50 Å, C strand) and G41 (distance

**Figure 5. The Ig-like domain of  $\beta$  protein represents a new IgI fold subtype.**

- Dimer co-crystal structure of  $\beta$ -IgSF (blue) and CEACAM1-N (CC1-N; red).
- Molecular structure of  $\beta$ -IgSF, in which each  $\beta$ -strand is coloured differently. The 1.5-turn  $\alpha$ -helix shown in grey is located between the C strand and the D strand.
- Illustration of the secondary structure organisation of the V-set, C1-set, C2-set, I1-set, I2-set and I3-set domains. Each secondary structure is composed of varying combinations of the  $\beta$ -strands A, A', B, C, C', C'', D, E, F, G. Disulphide bridges between the B strand and F strand are represented as a black dashed lines. The C-set domains lack the A' and C' strands and are subtyped into C1-set and C2-set based on presence and absence of the D strand, respectively. The I-set domains lack C' strand and are subtyped into I1-set and I2-set based on presence and absence of the D strand, respectively. The IgSF fold from  $\beta$  protein has the I3-set domain, characterised by absence of disulphide bridge, presence of the D strand, and presence of a truncated C strand followed by a 1.5-turn  $\alpha$ -helix.
- Comparison of  $\beta$ -IgI3 (blue) with the most similar folds; the IgI1 domain of MCS-F (turquoise) and the IgI1 domain of ICAM-3 (magenta).
- The molecular surface mesh of  $\beta$ -IgI3 residues showing protruding residues located between the C and D strands in blue.



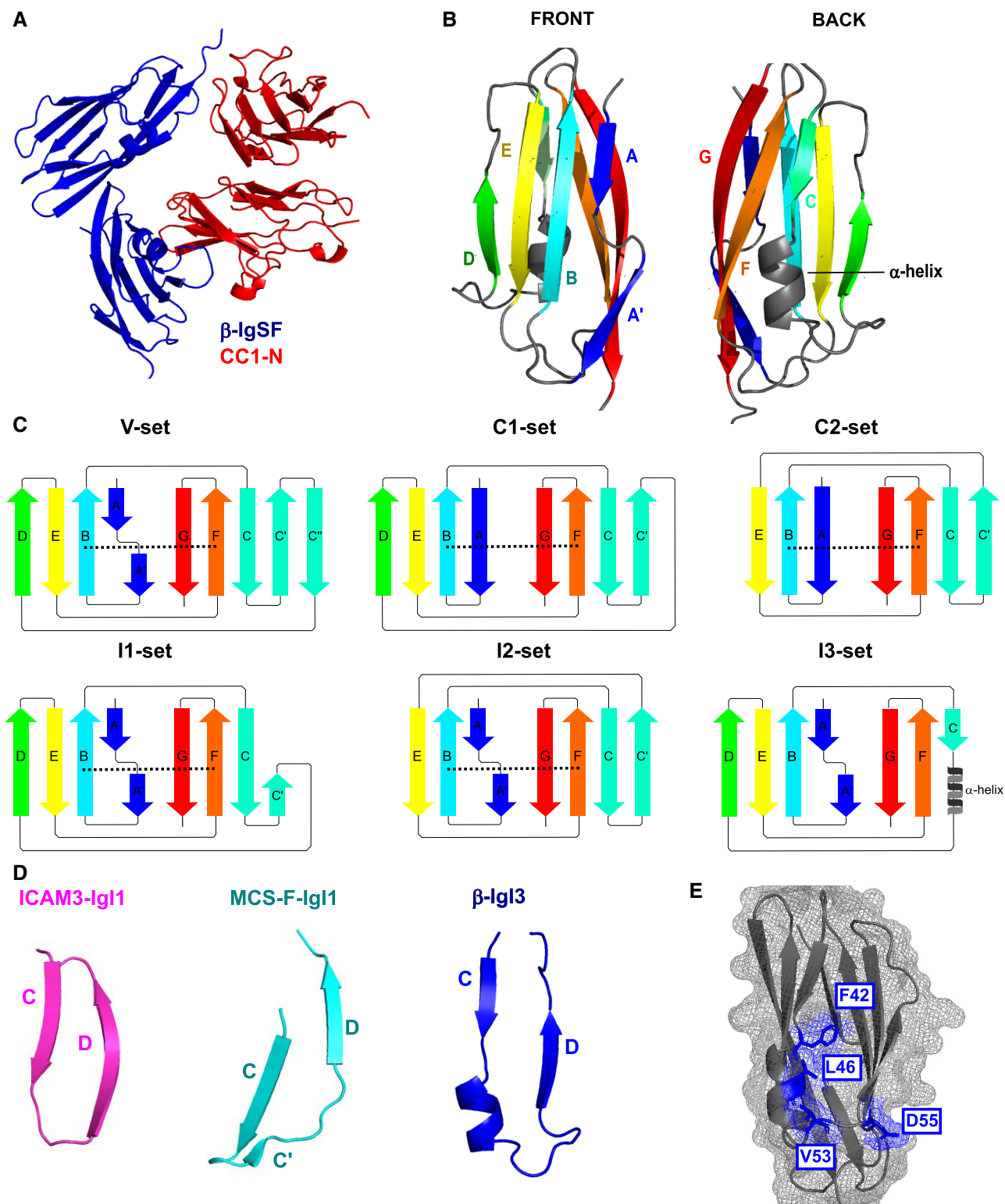
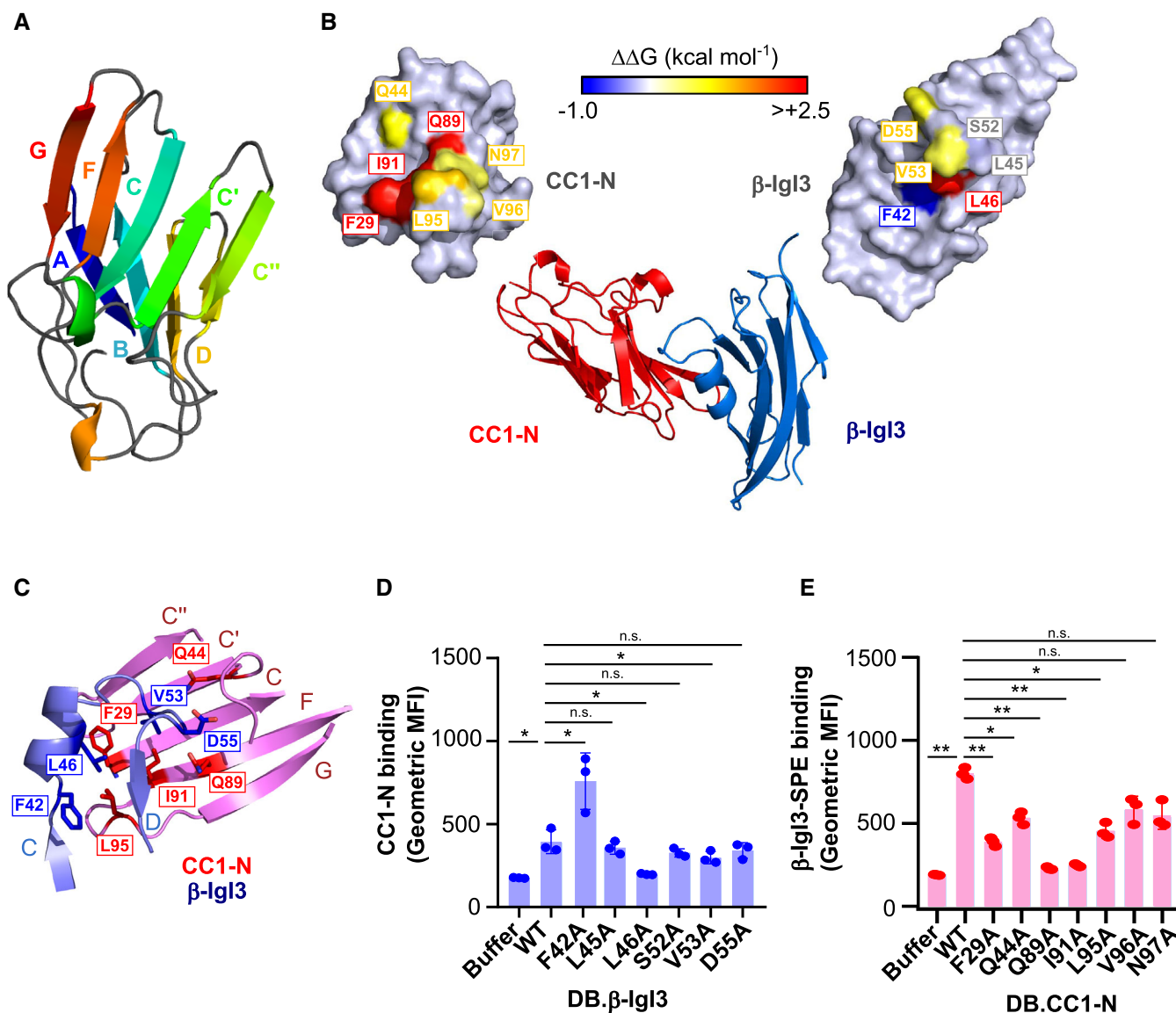


Figure 5.

2.90 Å, C to C' loop) of CEACAM1, and F42 is in close contact of L95 (distance 2.87 Å, G strand) of CEACAM1 (Appendix Table S5). Further contacts are made by β-Ig13 through residues L40, D41, S43, T47, N50, P51, S52, S54, I57, T59 and Y61 and CEACAM1 residues G30, Y31, Y34, D40, A49, T56, I91, S93, D94 and V96 of

CEACAM1, respectively (Appendix Table S5). No conformational changes were observed in CEACAM1 upon binding of β-Ig13 (Fig EV3E).

Based on the co-crystal data, we hypothesised that β-Ig13 residue L46 was critical for contacting CEACAM1-N via F29 and L95



(Figs 6C and EV4B). Additionally,  $\beta$ -Ig3 residue V53 appeared critical for contacting CEACAM1 residue I91, and  $\beta$ -Ig3 residue F42 was critical for contacting CEACAM1-N residue L95. We generated alanine mutations in  $\beta$ -Ig3 at these positions, and others, based on their contact with CEACAM1. ITC binding studies of these  $\beta$ -Ig3 mutants to “wild-type” unglycosylated CEACAM1-N revealed that the mutant  $\beta$ -Ig3<sup>L46A</sup> failed to bind to rCEACAM1-N. Two additional mutants,  $\beta$ -Ig3<sup>V53A</sup> and  $\beta$ -Ig3<sup>D55A</sup>, had reduced affinity to bind rCEACAM1-N ( $K_D = 562 \pm 44$ ,  $690 \pm 52$  nM, respectively; Fig EV4A; Appendix Table S6). In contrast,  $\beta$ -Ig3<sup>F42A</sup> bound with higher affinity ( $K_D = 16 \pm 15$  nM). We also tested the ability of unglycosylated rCEACAM1-N to bind DB coated with the  $\beta$ -Ig3 variants (Fig 6D). In this analysis, rCEACAM1-N displayed significantly reduced binding to DB coated with  $\beta$ -Ig3<sup>L46A</sup> and  $\beta$ -Ig3<sup>V53A</sup> and

significantly enhanced binding to DB coated with  $\beta$ -IgI3<sup>F42A</sup>. Together, these data indicate that the L46, V53 and D55 residues of  $\beta$ -IgI3 are critical for the binding to CEACAM1. These residues were observed to be conserved in 57  $\beta$  protein sequences (Appendix Fig S4).

For CEACAM1-N, the crystal structure of the complex indicated that  $\beta$ -IgI3 interacts with the dimer interface, contacting several residues which are important for CEACAM1 homodimerisation and for binding to other bacterial adhesins. We hypothesised that residues F29, I91 and L95 of CEACAM1 are critical for contacting  $\beta$ -IgI3 (Figs 6C and EV3C and D) and generated alanine mutants at these and several other positions in the homodimerisation interface. In ITC analysis with unglycosylated CEACAM1-N proteins, both CEACAM1-N<sup>F29A</sup> and CEACAM1-N<sup>I91A</sup> failed to bind  $\beta$ -IgI3 (Fig EV4B; Appendix Table S6). Additionally, CEACAM1-N<sup>Q89A</sup> also lacked ability to bind  $\beta$ -IgI3. Two mutants, CEACAM1-N<sup>Q44A</sup> and CEACAM1-N<sup>L95A</sup>, resulted in a 10-fold decrease in binding affinities ( $K_D = 996 \pm 116$  and  $1350 \pm 460$  nM, respectively) whilst two further mutants, CEACAM1-N<sup>V96A</sup> and CEACAM1-N<sup>N97A</sup>, displayed only a modest decrease (4-fold) in binding affinities ( $K_D = 370 \pm 4$  and  $490 \pm 120$  nM, respectively). In addition, we tested the ability of  $\beta$ -IgI3 coupled to streptavidin to interact with DB coated with the unglycosylated wild-type or mutant rCEACAM1-N (Fig 6E). Binding was significantly reduced for CEACAM1-N<sup>F29A</sup>, CEACAM1-N<sup>Q89A</sup> and CEACAM1-N<sup>I91A</sup>, as well as for CEACAM1-N<sup>Q44A</sup> and CEACAM1-N<sup>L95A</sup>, confirming that residues F29, Q89 and I91 are major targets of  $\beta$ -IgI3 binding. I91 has also been identified as a critical CEACAM1 residue for interaction with *M. catarrhalis* (Connors et al, 2008), *Neisseria* spp. (Virji et al, 1999; Villullas et al, 2007), *H. influenzae* (Hill et al, 2001) and *Fusobacterium* spp (Wang, 2013). Though Q89 in CEACAM1 is critical for interaction with  $\beta$ -IgI3 and other bacterial ligands, Q89 was not in close contact with any  $\beta$ -IgI3 residues. It is possible that mutation of the CEACAM1 residue Q89 to an alanine forms a cavity on the surface that I91 and surrounding residues of CEACAM1 attempt to fill through alternative rotamer conformations that subsequently prevents interaction with  $\beta$ -IgI3. In summary, contact of residue L46 in  $\beta$ -IgI3 with CEACAM1 residue F29, and residue V53 in  $\beta$ -IgI3 with CEACAM1 residue I91, provide the critical interactions. Additional stability is gained through  $\beta$ -IgI3 residues F42 and D55.

### Comparison of $\beta$ -IgI3- and HopQ-bound CEACAM1 structures

Comparison of the ( $\beta$ -IgI3)-(CEACAM1-N) complex and the (HopQ)-(CEACAM1-N) complex, revealed that the same set of CEACAM1 residues are contacted by both  $\beta$ -IgI3 and HopQ (Fig EV3F). However, HopQ is structurally completely unrelated to  $\beta$ -IgI3 (Bonsor et al, 2018). HopQ uses an intrinsically disordered loop, that folds into a  $\beta$  hairpin and a small helix upon binding to CEACAM1. The  $\beta$  hairpin extends the CEACAM1 A'GFCC' face whilst the small helix straddles the face (Fig EV3F). In contrast,  $\beta$ -IgI3 interacts through the small  $\alpha$  helix and loop, with a single residue, L46, critical for  $\beta$ -IgI3 to fit into a pocket on CEACAM1-N formed by F29 and I91 (Fig 6C). Mutation of any of these residues in  $\beta$ -IgI3 or CEACAM1-N causes a hole in the protein-protein interface or collapse of the pocket, respectively, abolishing binding as observed with our ITC experiments.

### $\beta$ -IgI3 homologs are broadly distributed in human Gram-positive bacteria

To gain a broader perspective on the possible role of the unique  $\beta$ -IgI3 structure in bacteria-host interactions, we investigated the distribution of this domain in the bacterial kingdom through BLAST analysis. A sequence similar or related to  $\beta$ -IgI3 was identified in 296 proteins predicted to be expressed by human Gram-positive bacteria, including several human pathogens, such as *S. pyogenes*, *S. dysgalactiae* and *S. pneumoniae*. These sequences formed several different clades upon maximum likelihood analysis (Fig 7A). The  $\beta$ -IgI3 domain was present in a large clade, clade I, which also included sequences from *S. oralis*. A second major clade, clade II, contained sequences from GBS and from other human streptococcal pathogens including *S. pyogenes* (also known as group A *Streptococcus*), *S. dysgalactiae* and *S. mitis*. Additionally, sequences closely related to clade II were detected in *S. mitis*, *S. milleri* and *S. intermedius*, forming clade III. Finally, many homologs of  $\beta$ -IgI3 were present in proteins from a diverse panel of Gram-positive human pathogens including *S. pneumoniae*, *S. pseudopneumoniae* and *Gemella haemolysans*. An additional homolog was identified in *Gardnerella vaginalis*, a Gram-variable bacterium. The sequences of  $\beta$ -IgI3 homologs were all located in proteins predicted to be surface-localised and anchored to the bacterial cell wall.

We predicted the tertiary structure of 11 representative homolog sequences based on the  $\beta$ -IgI3 crystal structure. These structures maintained the overall IgI3 fold, including an  $\alpha$ -helix and loop located between the truncated C strand and the D strand (Appendix Fig S5), except clade XV sequence from *G. vaginalis* that lacked the C strand. These data suggest that the IgI3 structure is broadly distributed in bacterial cell wall-anchored adhesins. We focused further analysis on clade II that maintains key IgI3 structures whilst sharing 40% amino acid identity with  $\beta$ -IgI3 (Fig 7B; Appendix Fig S6A-C). This clade includes the R28 protein, which is found in *S. pyogenes* and in GBS (where the same protein is also called Alp3). R28 is a cell wall-anchored adhesin that is unrelated to  $\beta$  protein except for the IgI3 domain (Stålhammar-Carlemalm et al, 1999; Lindahl et al, 2005). To understand the CEACAM1 binding capacity of the putative IgI3 domain in R28, we inspected whether the critical protruding residues located between C and D strands in  $\beta$ -IgI3 (Fig 7C) were conserved in the IgI3 domain of R28. Though the CEACAM1-binding residues of  $\beta$ -IgI3 (F42, L46, V53 and D55) are not conserved in R28-IgI3 (Fig 7C), two of the R28-IgI3 residues found at corresponding positions were of hydrophobic nature as in  $\beta$ -IgI3. Interestingly, additional residue variation was evident amongst other IgI3 domains (Fig EV5). Since this result suggested that IgI3 homologs are likely to differ in their properties, it became critical to analyse whether the R28-IgI3 homolog binds CEACAM1.

### Human CEACAMs bind Gram-positive bacteria expressing $\beta$ -IgI3 homologs

To determine whether the IgI3 domain in R28 interacts with CEACAM1, we purified rR28-IgI3 protein domain, that was expressed in *E. coli*, and tested interaction with rCEACAM1. DB coated with biotinylated R28-IgI3 and  $\beta$ -IgI3, but not HSA, interacted with rCEACAM1 (Fig 7D). In the reverse assay, DB coated with

rCEACAM1 interacted with R28-IgI3 and  $\beta$ -IgI3, but not HSA, coupled to streptavidin (Fig 7E). We also confirmed that rCEACAM1 bound to an R28-expressing *S. pyogenes* strain (Fig 7F). In addition,

rCEACAM1 bound to an Alp3-expressing strain of GBS, which was expected given that R28 and Alp3 proteins are identical in sequence. To ascertain whether R28-IgI3 and CEACAM1 interactions were

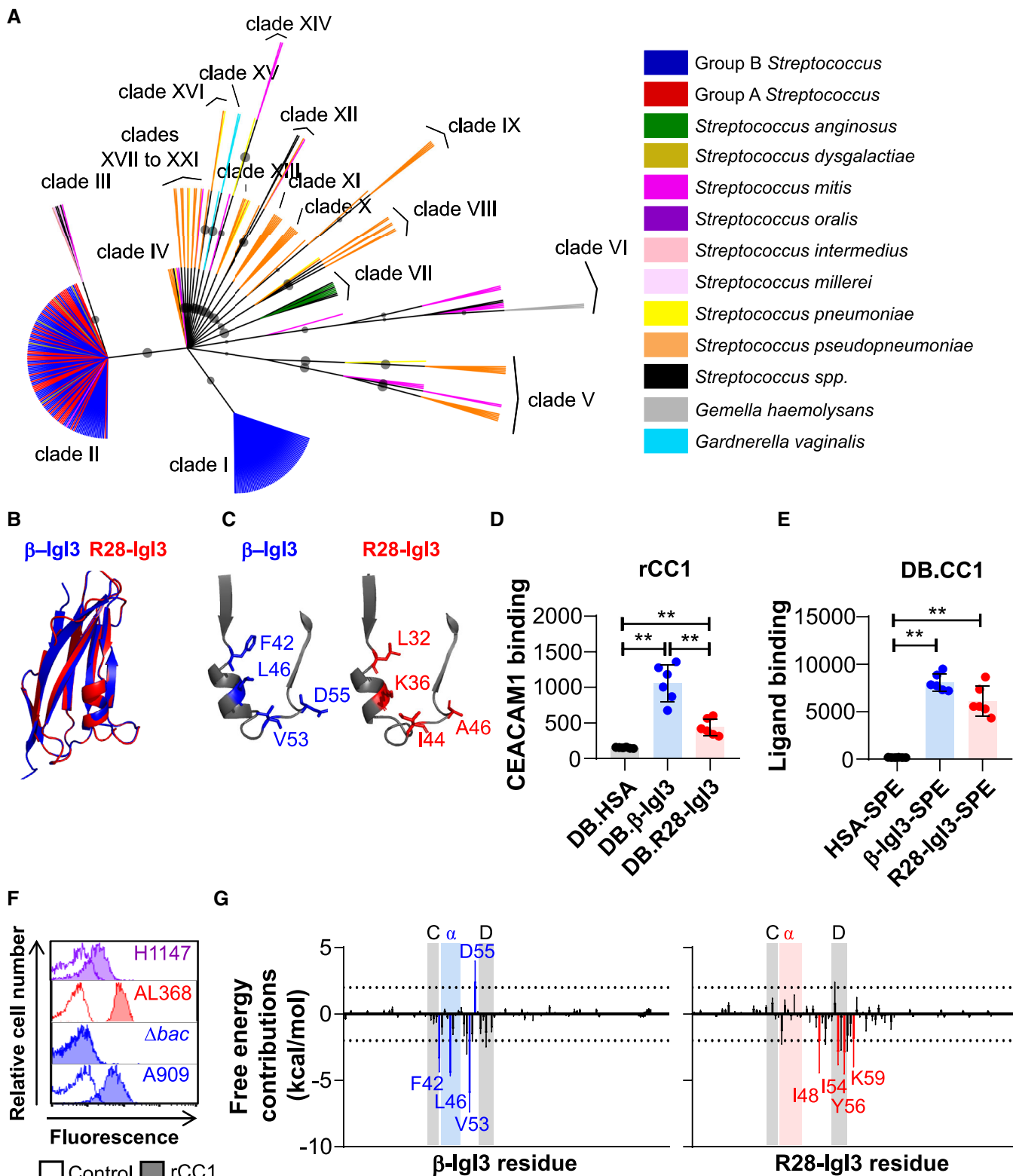


Figure 7.

**Figure 7. The unique IgI3 structure is present in Gram-positive bacterial species and evidence for binding CEACAM1.**

- A Amino acid based maximum likelihood tree of  $\beta$ -IgI3 homologs. Clades of related IgI3 domains are shown, with clade I representative of  $\beta$ -IgI3 domain sequences. Branches are coloured by bacterial species from which the sequence originates, except those with no species designated that are shown in black. Maximum likelihood bootstrap values > 60 are shown as a circle, where larger circles indicate a higher bootstrap support value.
- B Superposition of the IgI3 domain from  $\beta$  protein (blue) onto predicted IgI3 domain structure of clade II sequence of *S. pyogenes* R28 protein (red).
- C Close-up of the truncated C strand,  $\alpha$ -helix and D strand of  $\beta$ -IgI3 (blue) and R28-IgI3 (red), indicating the corresponding R28-IgI3 residues present at sites in  $\beta$ -IgI3 that bind CEACAM1 (CC1).
- D Binding of recombinant (r)CC1-HIS to dynabeads (DB) coated with  $\beta$ -IgI3, R28-IgI3 or control. Mean and SD is shown for  $n = 6$  independent replicates.
- E Binding of  $\beta$ -IgI3, R28-IgI3 or human serum albumin (HSA) coupled to PE-conjugated streptavidin (SPE), to DB coated with CC1 (DB.CC1). Mean and SD is shown for  $n = 6$  independent replicates.
- F Binding of recombinant rCC1-HIS (10  $\mu$ g/ml) to R28-expressing GBS strain H1147 and R28-expressing *S. pyogenes* strain AL368. Data representative of  $n = 3$  replicates.
- G The per residue free energy contribution spectrums of  $\beta$ -IgI3/CC1-N interface and R28-IgI3/CC1-N interface. The residues with binding free energy contributions lower than  $-2.0$  kcal/mol are identified as key (favourable) residues and coloured blue (for  $\beta$ -IgI3) or red (for R28-IgI3). Free energy contributions at  $-2.0$  kcal/mol and  $2.0$  kcal/mol are indicated by dashed lines. The position of the truncated C strand,  $\alpha$ -helix and D strand is shaded for each protein. The mean and standard deviation is representative of estimates from 50 docked models.

Data information: Statistical significance was calculated using a one-way ANOVA with Sidak's multiple comparisons for (D and E), where \*\*,  $P < 0.01$ . Fluorescence of DB (D and E) and bacteria (F) was measured by flow cytometry.

mediated through the N-terminal domain of CEACAM1, we used ITC to measure the affinity of R28-IgI3 for unglycosylated N-terminal domains of CEACAM1 (Appendix Fig S7). In this analysis, R28-IgI3 bound with intermediate affinity ( $K_D = 1050 \pm 18$  nM,  $\Delta H = -4.8 \pm 0.3$  kcal/mol) to CEACAM1-N. Taken together with the structural predictions described above, these data indicate that IgI3 folds present in adhesins from a wider range of Gram-positive bacterial pathogens can interact with the N-terminal domain of human CEACAM1, but that individual IgI3 domains may interact with human CEACAM1 differently.

To dissect whether  $\beta$ -IgI3 and R28-IgI3 interact differently with CEACAM1, we simulated the docking of IgI3 domains onto human CEACAM1 and examined the free energy calculations to ascertain the key IgI3 residues in the complex formation (Appendix Fig S6C). Residues with binding free energy contributions lower than  $-2.0$  kcal/mol and greater than  $2$  kcal/mol are identified as key residues and unfavourable residues, respectively. Simulation of  $\beta$ -IgI3 and CEACAM1 docking correctly predicted that residues F42, L46 and V53 were key determinants of CEACAM1 binding, whilst D55 was an unfavourable determinant (Fig 7G; Appendix Fig S6D). The binding free energy of the (R28-IgI3)-(CEACAM1-N) complexes was strong ( $-31.02 \pm 6.70$  kcal/mol,  $n = 50$  simulations). In contrast to  $\beta$ -IgI3, residues located in the  $\alpha$ -helix of R28-IgI3 were not identified as key determinants of CEACAM1 binding (Fig 7G; Appendix Fig S6E). Instead, critical R28-IgI3 residues (K48, I54, I56 and K59) were located in the CD loop and the D strand only. Mapping of key residues on the surface structure therefore suggested that  $\beta$ -IgI3 and R28-IgI3 target CEACAM1 through different faces on the IgI3 fold (Appendix Fig S6D and E). Thus, IgI3 domains in different bacterial proteins probably interact with human CEACAM1 through mechanisms that are not identical but use the same structural fold.

## Discussion

We show here that a novel Ig-like fold in different streptococcal adhesins promotes binding to human CEACAM receptors. Molecular analysis of this domain in the  $\beta$  protein from GBS demonstrated that the Ig-like domain binds to the N-terminal domain of human CEACAM1 and CEACAM5. We solved the crystal structure of the Ig-like domain in  $\beta$  and revealed that it represents a novel Ig-fold

structure related to the I-set that is characterised by an absence of cysteine residues and the presence of a truncated C strand that is directly followed by a unique 1.5-turn  $\alpha$ -helix. As this Ig-like fold of the I-set has unique features, it was designated IgI3. Whilst cysteine residues and disulphide bridges are found in Ig folds of prokaryotic and eukaryotic proteins (Halaby & Mornon, 1998; Bodelón *et al*, 2013), the absence of cysteine residues has not been documented in IgI folds to date. In addition, the partial replacement of the C strand with a 1.5-turn  $\alpha$ -helix was not identified in any other Ig fold in a search of PDB using HHpred (Zimmermann *et al*, 2018). I-set folds are rarely found within bacterial proteins as observed through a search of the SMART database (Letunic & Bork, 2018). This work therefore presents the first representative structure of a novel Ig-fold subtype, IgI3, and highlights that unique subtypes of Ig folds remain to be discovered in biology.

The co-crystal structure of the complex demonstrated that  $\beta$ -IgI3 targets the dimerisation interface of CEACAM1 through a unique mechanism involving residues in the C strand to D strand loop of  $\beta$ -IgI3 including the  $\alpha$ -helix. In contrast, Opa proteins of *Neisseria* spp. are membrane proteins that bind to the CEACAM dimerisation interface through two hypervariable loops (Billker *et al*, 2000), UspA1 of *M. catarrhalis* forms a trimeric coiled-coil that binds to CEACAM dimerisation interface through a small bend (Connors *et al*, 2008), HopQ of *H. pylori* uses a disordered loop to bind the CEACAM dimerisation interface (Bonsor *et al*, 2018), and AfaE of *E. coli* uses an incomplete Ig-like domain to target the CEACAM dimerisation interface (Anderson *et al*, 2004; Korotkova *et al*, 2008).

The finding that  $\beta$ -IgI3 homologs are broadly distributed in Gram-positive bacteria raises many fascinating questions about the evolutionary history of IgI3-folds and their role in bacterial pathogenesis. What is the origin of the IgI3-fold? How conserved is the IgI3-fold structure? How did the IgI3-fold become so widely distributed? Do all bacterial adhesins containing IgI3 homologs bind CEACAMs and how is binding conferred? Whilst IgI3 homologs were found in putative adhesins from several Gram-positive bacteria, there was low conservation of the critical  $\beta$ -IgI3 residues in the predicted structures of homologs, suggesting that each sequence may possess different properties. Our study demonstrated that R28-IgI3 and  $\beta$ -IgI3 both bind the N-terminal domain of CEACAM1, though stimulated docking suggests that these IgI3 domains bind through alternative mechanisms. As R28-IgI3-like sequences were found in proteins from multiple human pathogens, CEACAM-

binding may represent an unrecognised adhesion strategy for a range of Gram-positive bacteria. Of note, the gene encoding  $\beta$  is encoded on mobile genetic elements that can move between bacteria by horizontal gene transfer (Tettelin *et al*, 2005). This could provide opportunities for rapid acquisition of CEACAM binding in other bacterial species leading to the emergence of strains with enhanced colonisation properties or virulence. Thus, the exploration of IgI3-fold and CEACAM receptor interactions is further warranted.

Adhesion to epithelial cells at mucosal surfaces is a prerequisite for streptococcal colonisation and disease development. For GBS bacteria, the data presented here show that the IgI3 domain in the  $\beta$  protein promotes binding to CEACAM1 and CEACAM5 present on epithelial cell surfaces. This finding is consistent with reports of CEACAM-interacting Gram-negative bacteria (Virji *et al*, 1996; Bos *et al*, 1998; Tchoupa *et al*, 2014, 2015; Gutbier *et al*, 2015; Königer *et al*, 2016; Javaheri *et al*, 2016). This is of general importance as it expands the spectrum of cellular adhesion mechanisms that may be utilised by GBS. Whilst the currently characterised interactions include the ability of the  $\alpha$  adhesin to bind  $\alpha 1\beta 1$ -integrin, BspC to bind vimentin and the binding of BspA to gp340 (Bolduc & Madoff, 2007; Rego *et al*, 2016; Deng *et al*, 2019), our demonstration that the  $\beta$  protein binds to CEACAM receptors identifies a novel type of interaction that potentially has important implications for pathogenesis. It is unclear whether the different GBS adhesins operate independently or in synergy with one another. Conceivably, they operate in synergy with adhesins that bind ECM components, such as Srr1, Srr2, FsbA, FsbB, FsbC, PilA or SfbA. As GBS can invade and survive within epithelial cells (Patras *et al*, 2015), we speculate that CEACAM engagement may also enhance virulence by promoting cellular invasion, as reported for CEACAM-binding proteins of Gram-negative pathogens (Billker *et al*, 2002; Korotkova *et al*, 2008; Tchoupa *et al*, 2015; Islam *et al*, 2018; Behrens *et al*, 2020). In addition, the possibility that binding to human CEACAMs mediated by  $\beta$  could modify epithelial cell responses and promote translocation of virulence factors, as described for *H. pylori* (Königer *et al*, 2016; Javaheri *et al*, 2016; Behrens *et al*, 2020), requires further investigation.

The capacity of individual pathogens to cause invasive systemic infections is dependent on their capacity to evade and/or subvert the host immune response.  $\beta$  protein interferes with antibody opsonisation by binding IgA and disrupting interaction with Fc $\alpha$ R (Pleass *et al*, 2001) and interferes with complement opsonisation by binding factor H, which acts as a cofactor in the degradation of C3b (Areschoug *et al*, 2002b). Moreover, interactions with the inhibitory receptors Siglec-5 and Siglec-7, which are expressed on leucocytes, could enhance the immunosuppressive effects of  $\beta$  through induction of tyrosine phosphorylation (Carlin *et al*, 2009; Fong *et al*, 2018). As CEACAM1 is an inhibitory receptor, the binding of  $\beta$  might also induce tyrosine phosphorylation and CEACAM1-dependent immunosuppression, as been reported for Opa of *Neisseria* spp., UspA1 of *M. catarrhalis* and HopQ of *H. pylori* (Boulton & Gray-Owen, 2002; Slevogt *et al*, 2008; Javaheri *et al*, 2016; Gur *et al*, 2019). Thus, dual- or multi-engagement of  $\beta$  with different human ligands may have potent immunosuppressive effects. However, dissecting how  $\beta$  affects immune escape by interacting with different ligands will be challenging, particularly because the CEACAM and Siglec receptors have different cellular expression profiles. Moreover, the human specificity of the known ligand interactions presents further challenges for characterising the function of  $\beta$

protein *in vivo*. Nonetheless, the data collectively suggest that  $\beta$  protein may act as a multitool protein that assists epithelial cell adhesion, cellular invasion and immunosuppression. Whilst not all strains of GBS express  $\beta$ , this protein is commonly expressed by strains of serotypes Ia, Ib, II and V (Lindahl *et al*, 2005) and it has been reported that expression of  $\beta$  at a high level is associated with increased virulence (Nagano *et al*, 2006). It therefore seems likely that  $\beta$  plays an important role in a large proportion of the serious infections caused by GBS, through its ability to bind CEACAMs and other human ligands.

In summary, our data demonstrate that an adhesin containing a domain with unique Ig-like fold, the IgI3 fold, promotes binding of GBS to human CEACAM1 and CEACAM5 receptors. Moreover, homologous domains were identified in a variety of predicted surface proteins from Gram-positive bacterial pathogens, suggesting that many Gram-positive bacterial pathogens employ IgI3 domains to engage human CEACAMs. Characterisation of these domains may lead to a better understanding of pathogenicity mechanisms and could pave the way for the development of novel therapeutic approaches to prevent infections.

## Materials and Methods

### Bacterial growth conditions

Organisms used in this study are listed in Appendix Table S1. Bacteria were cultured in Todd-Hewitt Broth (TH; *S. agalactiae*), TH Broth + 1% yeast (*S. pyogenes*), Tryptic Soy Broth (TSB; *S. aureus*) or Lysogeny Broth (LB; *E. coli*) and supplemented with antibiotics as listed in Appendix Table S1.

### Expression and purification of glycosylated and unglycosylated rCEACAMs

Glycosylated CEACAMs were expressed and purified from Expi293F cells (Life Technologies). In brief, gBlocks containing open reading frames (ORFs) encoding the extracellular domain of CEACAMs, and a C-terminal LPETGGS-6xHis tag, were cloned into the pcDNA3.4 expression vector (Invitrogen) (Appendix Table S2). Recombinant CEACAM-His proteins were expressed in Expi293F cells (Life Technologies) cultured in Expi293 Expression Medium (Life Technologies) and purified by affinity chromatography (ÄKTA Pure, GE Healthcare Life Sciences) using a Nickel column (GE Healthcare Life Sciences) as previously described (Zhao *et al*, 2020). Eluate was dialysed against 300 mM NaCl 50 mM Tris pH 7.8 at 4°C. Unglycosylated CEACAMs were expressed and purified from *E. coli* cultures. In brief, gBlocks containing ORFs encoding the N-terminal or A1B1A2 domains of CEACAM1, and a C-terminal LPETGGS-6xHis tag, were cloned into the pRSET-C vector (Appendix Table S2). Vectors were transformed into the Rosetta Gami (RG) *E. coli* strain. RG strains were cultured in LB supplemented with 100  $\mu$ g/ml ampicillin and 1 mM D-glucose at 37°C overnight and subcultured 1:20 into LB supplemented with 1 mM D-glucose at 37°C until OD<sub>600</sub> = 0.4. Protein expression was induced by culturing for 4 h at 37°C following the addition of 1 mM IPTG. After lysis of bacteria, proteins were purified using a Nickel column (GE Healthcare Life Sciences) and affinity chromatography (ÄKTA Pure, GE Healthcare

Life Sciences). Eluates were dialysed against 300 mM NaCl 50 mM Tris pH 7.8 at 4°C. Large-scale expression of tag-less CEACAM1-N domain and point mutants for use in isothermal titration calorimetry (ITC) and crystallisation were expressed using pET21d vectors in *E. coli* and purified as previously described (Bonsor *et al*, 2015a).

### Expression and purification of bacterial proteins

$\beta$ ,  $\alpha$  and Rib proteins were purified from GBS cultures as previously described (Lindahl *et al*, 1990; Stålhammar-Carlemalm *et al*, 1993). Domains of the  $\beta$  proteins (B6N, IgABR, B6C, IgSF and  $\beta$ 75KN) and R28 (R28-IgI3) were cloned and expressed in *E. coli* (Appendix Table S2). The IgSF domain structure was found to be in a region overlapping the previously published B6C (amino acids 226–434) and  $\beta$ 75KN domains (amino acids 434–788). (Areschoug *et al*, 2002b; Lindahl *et al*, 2005) Therefore, we have updated domain designations as follows: B6C represents amino acids 226–390, IgSF represents amino acids 391–503 and  $\beta$ 75KN represents amino acids 504–788. Briefly, gBlocks containing ORFs and a C-terminal LPETGGGS-6xHis tag were cloned into the pRSET-C vector using BamHI and NdeI. Vectors were transformed into the RG strain. RG strains were cultured, induced, harvested and lysed as described above for unglycosylated CEACAMs. Bacterial supernatants were filtered, supplemented with 10 mM imidazole and passed over a Nickel column (GE Healthcare Life Sciences). Proteins were purified by affinity chromatography (ÄKTA Pure, GE Healthcare Life Sciences), eluted with a gradient of imidazole (10–500 mM) and were dialysed against 300 mM NaCl 50 mM Tris pH 7.8 at 4°C using 6–8 kDa dialysis tubing.

### Binding of glycosylated and non-glycosylated recombinant CEACAM to bacteria

$6 \times 10^6$  of mid-logarithmic phase bacteria were incubated with rCEACAM (1, 3, 5, 6 or 8), rCEACAM1-N or rCEACAM1-A1B1A2 at 4°C for 1 h with shaking. After washing in PBS + 0.1% bovine serum albumin (BSA), rCEACAM on the bacterial surface was detected by incubation with FITC-conjugated anti-HIS monoclonal antibody (mAb) at 4°C for 1 h with shaking. Bacterial fluorescence was measured by flow cytometric analysis following washing and fixation in 1% paraformaldehyde (PFA). To detect rCEACAM binding to GBS using pull-down Western blot analysis, bacterial pellets were resuspended in 1 $\times$  sample buffer and heated to 95°C for 10 min. Lysates were separated by SDS-PAGE in a 12.5% polyacrylamide gel at 270V, blotted onto nitrocellulose membranes, blocked with 5% milk solution and probed with anti-CEACAM1 (clone C51X/8) mAb (BB Singer, Essen, Germany). Membranes were probed using rabbit anti-mouse-IgG-HRP and developed using ECL substrate.

### Measurement of $\beta$ protein expression by GBS strains

Rabbit antiserum was raised against  $\beta$  protein as previously described (Lindahl *et al*, 1990).  $6 \times 10^6$  of mid-logarithmic phase bacteria were incubated with heat-inactivated 0.1% rabbit anti- $\beta$  serum or normal rabbit serum at 4°C for 1 h with shaking, washed and incubated in the presence of PE-conjugated goat anti-rabbit-IgG at 4°C for 1 h with shaking. Fluorescence of bacteria was measured by flow cytometric analysis after washing and fixation in 1% PFA.

### Detection of CEACAM1 and $\beta$ protein interaction

#### Western blot analysis

Purified GBS proteins, or  $\beta$  protein domains (B6N, IgABR, B6C, IgSF or  $\beta$ 75KN), were separated by non-reducing SDS-PAGE, transferred to nitrocellulose membranes, blocked with 5% milk solution overnight at 4°C, probed with 10  $\mu$ g/ml rCEACAM1-His (4°C) and horse radish peroxidase (HRP)-conjugated mouse anti-His-IgG and developed using Pierce ECL Western Blot Substrate Reagent (Thermo Fisher Scientific).

#### Binding of bacterial proteins to CEACAM1-coated dynabeads

rCEACAM-His was attached to nickel dynabeads (DB; Dynabead His-tag pull-down and isolation, Invitrogen) following standard protocols. Presence of CEACAM1 was confirmed by incubating DB in the presence of 5  $\mu$ g/ml anti-CEACAM or isotype control mAb, and detection with PE-conjugated goat anti-mouse-IgG mAb. To detect binding of purified GBS proteins, 40  $\mu$ g of DB was incubated in the presence of 10  $\mu$ g/ml of  $\beta$  protein and probed with 0.1% mouse antiserum or normal mouse serum, and secondary goat anti-mouse-IgG-PE mAb. Biotinylated proteins (including  $\beta$ -IgSF-biotin or HSA-biotin amongst others) were incubated with PE-conjugated Streptavidin on ice for 10 min. This generates biotin-streptavidin complexes on the remaining 3 binding pockets of streptavidin. To detect binding of proteins coupled with streptavidin ( $\beta$ -IgSF-biotin or HSA-biotin formed with Streptavidin-PE), 20  $\mu$ g of DB was incubated with varying concentrations of protein-streptavidin. All dilutions and DB washes employed PBS + 0.1% BSA, all incubations were performed for 1 h at 4°C, and all fixations used 1% PFA. Fluorescence of DB was measured by flow cytometry. For inhibition experiments, CEACAM-coated DB were pre-incubated with 5  $\mu$ g/ml anti-CEACAM1-N (clone CC1/3/5-Sab, LeukoCom, Essen, Germany), anti-CEACAM1-A1B1 mAb (clone B3-17, BB Singer, Essen, Germany) or 10  $\mu$ g/mL rHopQ for 1 h at 4°C.

#### Binding of CEACAM to bacterial protein coated dynabeads

Proteins with C-terminal biotin tags were attached to streptavidin-coated dynabeads (Dynabead M-280 Streptavidin, Invitrogen) following standard procedures. To detect binding of rCEACAM-His or rCEACAM-N-His,  $3 \times 10^5$  DB was incubated in the presence of 9  $\mu$ l of proteins (concentrations range 0.1–30  $\mu$ g/ml) and probed with FITC-conjugated mouse anti-His-IgG. In the case of rSiglec-5-Fc, binding was detected using anti-IgG-AF488. All dilutions and DB washes employed PBS + 0.1% BSA, all incubations were performed for 1 h at 4°C, and all fixations used 1% PFA. Fluorescence of DB was measured by flow cytometry.

### Isothermal titration calorimetry

Duplicate ITC measurements were performed on an iTC200 instrument (GE Healthcare). Typically, 500  $\mu$ M of unglycosylated CEACAM-N domains were loaded into the syringe of the calorimeter and 50  $\mu$ M  $\beta$ -IgC<sub>2</sub> protein was loaded into the syringe. All measurements were performed at 25°C, with a stirring speed of 750 rpm, in 30 mM Tris-HCl, 150 mM NaCl, pH 7.5. Data were analysed using the Origin 7.0 software.

## Adhesion of GBS to epithelial cells

HeLa and Chinese Hamster Ovary (CHO) cell lines expressing CEACAMs were cultured as previously described (Daniel *et al*, 1993; Bos *et al*, 1998; Hollandsworth *et al*, 2020). Cells were seeded into 24-well tissue culture plates. Tightly confluent monolayers were infected with mid-logarithmic phase (absorbance at OD<sub>600</sub> = 0.4) GBS strains at a multiplicity of infection (MOI) of 10. Assays were commenced by centrifugation for 2 min at 700 rpm and incubated at 37°C with 5% CO<sub>2</sub> for 30 min. Cells were washed five times with DMEM, detached with 0.25% trypsin and lysed (PBS + 0.025% Triton-X). Adherent bacteria were enumerated using serial dilution plating and growth on Todd-Hewitt agar plates. In specific assays, cells were pre-incubated with 5 µg/ml mAb (anti-CEACAM1-N or anti-CEACAM1-A1B1A2 mAb) for 60 min, or bacteria were pre-incubated with rCEACAM1-N for 60 min before co-culture. For confocal microscopy, experiments were performed with FITC-labelled GBS at a MOI of 10 for 30 min. After washing five times with PBS, monolayers were fixed with 4% PFA overnight prior to staining of nuclei with DaPI (Sigma Aldrich) and cell membranes with AF647-conjugated wheat germ agglutinin (Thermo Fisher Scientific). To perform assays with detached CHO cell lines, adherent monolayers were detached by incubation at 37°C for 5 min with Accutase Cell Detachment Solution (BioLegend). After washing, cells were resuspended to 3 × 10<sup>6</sup> cells/ml in DMEM containing 10% FCS. To measure CEACAM expression, cells were incubated with mouse mAb clone CC1/3/5-Sab (detects N-terminal domain of human CC1, CC3 and CC5) for 30 min, washed and incubated with PE-conjugated rabbit anti-mouse-IgG. To measure GBS binding, cells were incubated with FITC-labelled GBS at a MOI of 10 for 30 min at 4°C. After washing with PBS, cells were fixed in 4% PFA and fluorescence was measured by flow cytometry.

An overview of all cell lines used in this study is provided in Appendix Table S7. An overview of all antibodies used in this study is provided in Appendix Table S8.

## Crystallography of unglycosylated CEACAM1-N and β-IgSF

Initially, the β-IgSF and CEACAM1-N complex was prepared by mixing the proteins with a slight excess of CEACAM1-N (1:1.1 ratio). Protein was concentrated using an amicon Ultra-4 10,000 NMWL before purification by size exclusion chromatography using a Superdex 200 column (GE Healthcare) equilibrated with 20 mM Tris, 150 mM NaCl, pH 7.5. The complex was concentrated to 9 mg/ml and screened against the commercial Wizards I/II/III/IV (Rigaku) and JCSG<sup>+</sup> (QIAGEN) screens using a Crystal Gryphon Protein Crystallography System (Art Robbins Instruments). Crystals were found in several conditions but diffraction was never found beyond 8.0 Å in the best condition of 1.8 M Ammonium Sulphate and 0.1 M Sodium Citrate pH 5.5. β-IgSF contains both a C-terminal Sortase and 6xHis tag. This was removed by first incubating β-IgSF by itself with Carboxypeptidase A (Sigma). 2 ml of 6 mg/ml of β-IgSF was incubated with 1:100 w/w of Carboxypeptidase A for 24 h at 4°C before the addition of another 1:100 w/w of Carboxypeptidase A and incubation for 24 h at room temperature. The reaction was terminated by the addition of EDTA to a final concentration of 5 mM. The protein was concentrated using an Amicon Ultra-4

10,000 NMWL before purification by size exclusion chromatography using a Superdex 200 column (GE Healthcare) equilibrated with 20 mM Tris, 150 mM NaCl, pH 7.5. CEACAM1-N was added in a slight excess (1:1.05 ratio) and purified by size exclusion as described above. The complex was concentrated and screened again at a concentration of 7.5 mg/ml. Crystals again formed in the condition of 1.8 M Ammonium Sulphate and 0.1 M Sodium Citrate pH 5.5; however, their morphology was different. Crystals were cryo-protected in 30 % v/v glycerol. A dataset was collected at beamline 12-2 at the Stanford Synchrotron Radiation Lightsource. Data were processed, merged, scaled and converted to structure factors using Aimless. Matthew's coefficient suggests five complexes in the asymmetric unit. Molecular replacement using MOLREP and CEACAM1-N (PDB entry 2gk2) as a search model in MOLREP only found two CEACAM1 molecules. Two copies of the complex would suggest a high solvent content (78 %). Refinement using REFMAC5 and visualisation in Coot showed that CEACAM1-N was well-placed and residual electron density was present for the β-Ig3 protein. Density modification was performed using Parrot (solvent flattening, histogram matching and NCS averaging) to improve the maps. The β-IgSF backbone was built manually. A PDBFold search of the backbone showed that it was an IgI-like fold, with the closest fold by R.M.S.D being the IgC2 domain of Perlecan (PDB entry 1gl4). The Ig fold of β-IgSF possesses a previous unrecognised modification (Fig 5B and C). We denote this as the I3-set domain and the domain within β protein as β-IgI3. This was used to successfully position sidechains, lock the registry and provide extra restraints during refinement in REFMAC5 with ProSMART. The (β-IgI3)-(CEACAM1-N) complex has been deposited to the PDB with the entry code, 6V3P. Atom contacts in the (β-IgI3)-(CEACAM1-N) complex interface were identified using NCONT (CCP4) with a cut-off of 4.0 Å (Winn *et al*, 2011).

## Bioinformatics analysis

BLAST analysis of the β-Ig3 amino acid sequence was performed to identify homologs in bacterial proteins and subsequently aligned using ClustalW. Phylogenetic analysis was performed using maximum likelihood approach and 1,000 bootstrap replications in MEGA (Kumar *et al*, 2016), and the resulting tree displayed with interactive tree of life version 4 (Letunic & Bork, 2019). The structure of IgI3 homologs was predicted using the β-Ig3 structure as input in SWISS-MODEL (Webb & Sali, 2016). Only models passing a QMEAN score of < -4.00 were further analysed. Docking of β-Ig3 structure or the R28-Ig3 predicted structure to CEACAM1-N was simulated 50 times using ZDOCK server (Pierce *et al*, 2014), in which CEACAM1-N residues F29, Q44, Q89, I91 and N95 were included as contact sites. The free energy contribution of each simulation was interpedently calculated using MM/GBSA analysis on the HAWKDOK server (Weng *et al*, 2019).

## Data availability

The datasets produced in this study are available in the following databases:

- Protein interaction (β-IgI3)-(CEACAM1-N) data: PDB 6V3P (<https://www.rcsb.org/structure/unreleased/6V3P>).



- The raw diffraction data have been deposited to the Integrated Resource for Reproducibility in Macromolecular Crystallography (<https://proteindiffraction.org/>).

All other data supporting the findings of this study are available within the paper and its supplementary information files and are available from the corresponding author on reasonable request.

**Expanded View** for this article is available online.

## Acknowledgments

The authors thank Carla J.C. de Haas, Piet Aerts, Kok P.M. van Kessel (UMC Utrecht, Utrecht, The Netherlands), Birgit Maranca-Hüwel, Bärbel Gobs-Hevelke (University of Duisburg-Essen, Germany) and Margaretha Stålhammar-Carlemalm (Lund University, Sweden) for excellent technical support and invaluable help. We thank Paul Sullam (University of California, USA) for providing *S. sanguinis* strains. We also wish to thank the support staff of Beamlines of 12-2 and ID23-D at the Stanford Synchrotron Radiation Light-source and the Advanced Photon Source for their aid in data collection, respectively. We thank Irene M. Mavridis (Institute of Child Health, Greece) and José R. Penadés (Imperial College London, UK) for critical reading of the manuscript. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 700862 (to A.J.M., J.A.G.), Deutsche Forschungsgemeinschaft DFG grant SI-1558/3-1 (to B.B.S.), The Swedish Research Council grant K2011-56X-09490-21-6 (to G.L.), The Foundation Olle Engkvist Byggmästare (to G.L.) and the National Institutes of Health R01 NS116716 (to K.S.D.).

## Author contributions

Conception and design of the work (NMS, LD, DAB, EJS, JAGS, KSD, BBS, GL, AJM). Acquisition, analysis or interpretation of data (NMS, LD, DAB, JB, VS, ML, AS, ORN, EB, EL, JAGS, KSD, BBS, GL, AJM). Drafting of manuscript (NMS, LD, DAB, JAGS, KSD, BBS, GL, AJM). All authors read and commented on the final version of the manuscript.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Ali SR, Fong JJ, Carlin AF, Busch TD, Linden R, Angata T, Areschoug T, Parast M, Varki N, Murray J *et al* (2014) Siglec-5 and Siglec-14 are polymorphic paired receptors that modulate neutrophil and amnion signaling responses to group B *Streptococcus*. *J Exp Med* 211: 1231–1242
- Anderson KL, Billington J, Pettigrew D, Cota E, Simpson P, Roversi P, Chen HA, Urvil P, du Merle L, Barlow PN (2004) An atomic resolution model for assembly, architecture, and function of the Dr adhesins. *Mol Cell* 15: 647–657
- Areschoug T, Linse S, Stålhammar-Carlemalm M, Hedén LO, Lindahl G (2002a) A proline-rich region with a highly periodic sequence in streptococcal  $\beta$  protein adopts the polyproline II structure and is exposed on the bacterial surface. *J Bacteriol* 184: 6376–6383
- Areschoug T, Stålhammar-Carlemalm M, Karlsson I, Lindahl G (2002b) Streptococcal  $\beta$  protein has separate binding sites for human factor H and IgA-Fc. *J Biol Chem* 277: 12642–12648
- Banerjee A, Kim BJ, Carmona EM, Cutting AS, Gurney MA, Carlos C, Feuer R, Prasadarao NV, Doran KS (2011) Bacterial pili exploit integrin machinery to promote immune activation and efficient blood-brain barrier penetration. *Nat Commun* 2: 462
- Bateman A, Eddy SR, Chothia C (1996) Members of the immunoglobulin superfamily in bacteria. *Protein Sci* 5: 1
- Behrens IK, Busch B, Ishikawa-Ankerhold H, Palamides P, Shively JE, Stanners C, Chan C, Leung N, Gray-Owen S, Haas R (2020) The HoPq-CEACAM interaction controls CagA translocation, phosphorylation, and phagocytosis of *Helicobacter pylori* in neutrophils. *mBio* 11: e03256-19
- Bensing BA, Loukachevitch LV, Mcculloch KM, Yu H, Vann KR, Wawrzak Z, Anderson S, Chen X, Sullam PM, Iverson TM (2016) Structural basis for sialoglycan binding by the *Streptococcus sanguinis* SrpA adhesin. *J Biol Chem* 291: 7230–7240
- Billker O, Popp A, Gray-Owen SD, Meyer TF (2000) The structural basis of CEACAM-receptor targeting by neisserial Opa proteins. *Trends Microbiol* 8: 258–260
- Billker O, Popp A, Brinkmann V, Wenig G, Schneider J, Caron E, Meyer T (2002) Distinct mechanisms of internalization of *Neisseria gonorrhoeae* by members of the CEACAM receptor family involving Rac1- and Cdc42-dependent and -independent pathways. *EMBO J* 21: 560–565
- Bodelón G, Palomino C, Fernández LÁ (2013) Immunoglobulin domains in *Escherichia coli* and other enterobacteria: From pathogenesis to applications in antibody technologies. *FEMS Microbiol Rev* 37: 204–250
- Bolduc GR, Madoff LC (2007) The group B streptococcal alpha C protein binds  $\alpha 1\beta 1$ -integrin through a novel KTD motif that promotes internalization of GBS within human epithelial cells. *Microbiology* 153: 4039–4049
- Bonsor DA, Beckett D, Sundberg EJ (2015a) Structure of the N-terminal dimerization domain of CEACAM7. *Acta Crystallogr F Struct Biol Commun* 71: 1169–1175
- Bonsor DA, Günther S, Beadenkopf R, Beckett D, Sundberg EJ (2015b) Diverse oligomeric states of CEACAM IgV domains. *Proc Natl Acad Sci USA* 112: 13561–13566
- Bonsor DA, Zhao Q, Schmidinger B, Weiss E, Wang J, Deredge D, Beadenkopf R, Dow B, Fischer W, Beckett D *et al* (2018) The *Helicobacter pylori* adhesin protein HopQ exploits the dimer interface of human CEACAMs to facilitate translocation of the oncoprotein CagA. *EMBO J* 37: e98664
- Bos MP, Kuroki M, Krop-Watorek A, Hogan D, Belland RJ (1998) CD66 receptor specificity exhibited by neisserial Opa variants is controlled by protein determinants in CD66 N-domains. *Proc Natl Acad Sci USA* 95: 9584–9589
- Boulton IC, Gray-Owen SD (2002) Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4+ T lymphocytes. *Nat Immunol* 3: 229–236
- Brewer ML, Dymock D, Brady RL, Singer BB, Virji M, Hill DJ (2019) *Fusobacterium* spp. target human CEACAM1 via the trimeric autotransporter adhesin CbpF. *J Oral Microbiol* 11: 1565043
- Brodeur BR, Boyer M, Charlebois I, Hamel JE, Couture F, Rioux MR, Martin D (2000) Identification of group B streptococcal Sip protein, which elicits cross-protective immunity. *Infect Immun* 68: 5610–5618
- Buscetta M, Papasergi S, Firon A, Pietrocola G, Biondo C, Mancuso G, Midiri A, Romeo L, Teti G, Speciale P *et al* (2014) FbsC, a novel fibrinogen-binding protein, promotes *Streptococcus agalactiae*-host cell interactions. *J Biol Chem* 289: 21003–21015
- Carlin AF, Chang YC, Areschoug T, Lindahl G, Hurtado-Ziola N, King CC, Varki A, Nizet V (2009) Group B *Streptococcus* suppression of phagocyte functions by protein-mediated engagement of human Siglec-5. *J Exp Med* 206: 1691–1699
- Chen SL (2019) Genomic insights into the distribution and evolution of group B *Streptococcus*. *Front Microbiol* 10: 1447
- Chen T, Gotschlich EC (1996) CGM1a antigen of neutrophils, a receptor of gonococcal opacity proteins. *Proc Natl Acad Sci USA* 93: 14851–14856

- Conners R, Hill DJ, Borodina E, Agnew C, Daniell SJ, Burton NM, Sessions RB, Clarke AR, Catto LE, Lammie D *et al* (2008) The *Moraxella* adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil. *EMBO J* 27: 1779–1789
- Daniel S, Nagel G, Son JP, Lobo FM, Hirn M, Jantschke P, Kurok M, von Kleff S, Grunert F (1993) Determination of the specificities of monoclonal antibodies recognizing members of the CEA family using a panel of transfectants. *Int J Cancer* 55: 303–310
- Deng L, Spencer BL, Holmes JA, Mu R, Rego S, Weston TA, Hu Y, Sanches GF, Yoon S, Park N *et al* (2019) The group B streptococcal surface antigen I/II protein, BspC, interacts with host vimentin to promote adherence to brain endothelium and inflammation during the pathogenesis of meningitis. *PLoS Pathog* 15: e1007848
- Edmond KM, Kortsalioudaki C, Scott S, Schrag SJ, Zaidi AK, Cousens S, Heath PT (2012) Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. *Lancet* 379: 547–556
- Fong JJ, Tsai CM, Saha S, Nizet V, Varki A, Buif JD (2018) Siglec-7 engagement by GBS  $\beta$ -protein suppresses pyroptotic cell death of natural killer cells. *Proc Natl Acad Sci USA* 115: 10410–10415
- Gori A, Harrison OB, Mlia E, Nishihara Y, Chan JM, Msefula J, Mallewa M, Dube Q, Swarthout TD, Nobbs AH *et al* (2020) Pan-GWAS of *Streptococcus agalactiae* highlights lineage-specific genes associated with virulence and niche adaptation. *mBio* 11: e00728–20
- Gray-Owen SD, Blumberg RS (2006) CEACAM1: Contact-dependent control of immunity. *Nat Rev Immunol* 6: 433–446
- Gur C, Maalouf N, Gerhard M, Singer BB, Emgård J, Temper V, Neuman T, Mandelboim O, Bachrach G (2019) The *Helicobacter pylori* HopQ outermembrane protein inhibits immune cell activities. *Oncol Immunology* 8: e1553487
- Gutbier B, Fischer K, Doehn J-M, von Lachner C, Herr C, Klaike E, Frischmann U, Singer BB, Riesbeck K, Zimmermann W *et al* (2015) *Moraxella catarrhalis* induces an immune response in the murine lung that is independent of human CEACAM5 expression and long-term smoke exposure. *Am J Physiol Lung Cell Mol Physiol* 309: 250–261
- Gutekunst H, Eikmanns BJ, Reinscheid DJ (2004) The novel fibrinogen-binding protein FbsB promotes *Streptococcus agalactiae* invasion into epithelial cells. *Infect Immun* 72: 3495–3504
- Halaby DM, Mornon JPE (1998) The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. *J Mol Evol* 46: 389–400
- Hall J, Adams NH, Bartlett L, Seale AC, Lamagni T, Bianchi-Jassir F, Lawn JE, Baker CJ, Cutland C, Heath PT *et al* (2017) Maternal disease with group B *Streptococcus* and serotype distribution worldwide: systematic review and meta-analyses. *Clin Infect Dis* 65: S112–S124
- Heath PT (2016) Status of vaccine research and development of vaccines for GBS. *Vaccine* 34: 2876–2879
- Hedén L-O, Frithz E, Lindahl G (1991) Molecular characterization of an IgA receptor from group B streptococci: sequence of the gene, identification of a proline-rich region with unique structure and isolation of N-terminal fragments with IgA-binding capacity. *Eur J Immunol* 21: 1481–1490
- Helfrich I, Singer BB (2019) Size matters: The functional role of the CEACAM1 isoform signature and its impact for NK cell-mediated killing in melanoma. *Cancers* 11: E356
- Hill DJ, Toleman MA, Evans DJ, Villullas S, van Alphen L, Virji M (2001) The variable P5 proteins of typeable and non-typeable *Haemophilus influenzae* target human CEACAM1. *Mol Microbiol* 39: 850–862
- Hollandsworth HM, Amifakhri S, Filemoni F, Schmitt V, Wennemuth G, Schmidt A, Hoffma RM, Singer BB, Bouvet M (2020) Anti-carcinoembryonic antigen-related cell adhesion molecule antibody for fluorescent visualization of primary colon cancer and metastases in patient-derived orthotopic xenograft mouse models. *Oncotarget* 11: 429–439
- Islam EA, Anipindi VC, Francis I, Shaik-Dasthagirisaheb Y, Xu S, Leung N, Sintsova A, Amin M, Kaushic C, Wetzler LM *et al* (2018) Specific binding to differentially expressed human carcinoembryonic antigen-related cell adhesion molecules determines the outcome of *Neisseria gonorrhoeae* infections along the female reproductive tract. *Infect Immun* 86: e00092–e118
- Javaheri A, Kruse T, Moonens K, Mejías-Luque R, Debraekeleer A, Asche CI, Tegtmeyer N, Kalali B, Bach NC, Sieber SA *et al* (2016) *Helicobacter pylori* adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. *Nature Microbiology* 2: 16189
- Jiang S, Wessels MR (2014) BsaB, a novel adherence factor of group B *Streptococcus*. *Infect Immun* 82: 1007–1016
- Khairnar V, Duhan V, Maney SK, Honke N, Shaabani N, Pandya AA, Seifert M, Pozdeev V, Xu HC, Sharma P *et al* (2015) CEACAM1 induces B-cell survival and is essential for protective antiviral antibody production. *Nat Commun* 6: 6217
- Khairnar V, Duhan V, Patil AM, Zhou F, Bhat H, Thoens C, Sharma P, Adomati T, Friedrich SK, Bezgovsek J *et al* (2018) CEACAM1 promotes CD8<sup>+</sup> T cell responses and improves control of a chronic viral infection. *Nat Commun* 9: 2561
- Königer V, Holsten L, Harrison U, Busch B, Loell E, Zhao Q, Bonsor DA, Roth A, Kengmo-Tchoupa A, Smith SI *et al* (2016) *Helicobacter pylori* exploits human CEACAMs via HopQ for adherence and translocation of CagA. *Nature Microbiology* 2: 16188
- Korotkova N, Yang Y, le Trong I, Cota E, Demeler B, Marchant J, Thomas WE, Stenkamp RE, Moseley SL, Matthews S (2008) Binding of Dr adhesins of *Escherichia coli* to carcinoembryonic antigen triggers receptor dissociation. *Mol Microbiol* 67: 420–434
- Kuespert K, Pils S, Hauck CR (2006) CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol* 18: 565–571
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874
- Larsson C, Stålhammar-Carlemalm M, Lindahl G (1996) Experimental vaccination against group B *Streptococcus*, an encapsulated bacterium, with highly purified preparations of cell surface proteins Rib and alpha. *Infect Immun* 64: 3518–3523
- Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* 46: D493–D496
- Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47: W256–W259
- Lindahl G, Akerström B, Vaerman JP, Stenberg L (1990) Characterization of an IgA receptor from group B streptococci: specificity for serum IgA. *Eur J Immunol* 20: 2241–2247
- Lindahl G, Stålhammar-Carlemalm M, Areschoug T (2005) Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. *Clin Microbiol Rev* 18: 102–127
- Michel JL, Madoff LC, Olson K, Kling DE, Kasper DL, Ausubel FM (1992) Large, identical, tandem repeating units in the C protein alpha antigen gene, bca, of group B streptococci. *Proc Natl Acad Sci USA* 89: 10060–10064
- Moonens K, Hamway Y, Neddermann M, Reschke M, Tegtmeyer N, Kruse T, Kammerer R, Mejías-Luque R, Singer BB, Backert S *et al* (2018) *Helicobacter pylori* adhesin HopQ disrupts trans dimerization in human CEACAMs. *EMBO J* 37: e98665
- Mu R, Kim BJ, Paco C, del Rosario YD, Courtney HS, Doran KS (2014) Identification of a group B streptococcal fibronectin binding protein, SfbA, that contributes to invasion of brain endothelium and development of meningitis. *Infect Immun* 82: 2276–2286

- Nagano N, Nagano Y, Nakano R, Okamoto R, Inoue M (2006) Genetic diversity of the C protein  $\beta$ -antigen gene and its upstream regions within clonally related groups of type Ia and Ib group B streptococci. *Microbiology* 152: 771–778
- Nordström T, Møvert E, Olin AI, Ali SR, Nizet V, Varki A, Areschoug T (2011) Human Siglec-5 inhibitory receptor and immunoglobulin A (IgA) have separate binding sites in streptococcal  $\beta$  protein. *J Biol Chem* 286: 33981–33991
- Patras KA, Nizet V (2018) Group B Streptococcal maternal colonization and neonatal disease: Molecular mechanisms and preventative approaches. *Frontiers in Pediatrics* 6: 27
- Patras KA, Rösler B, Thoman ML, Doran KS (2015) Characterization of host immunity during persistent vaginal colonization by Group B *Streptococcus*. *Mucosal Immunol* 8: 1339–1348
- Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z (2014) ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 30: 1771–1773
- Pietrocola G, Arciola CR, Rindi S, Montanaro L, Speziale P (2018) *Streptococcus agalactiae* non-pilus, cell wall-anchored proteins: Involvement in colonization and pathogenesis and potential as vaccine candidates. *Front Immunol* 9: 602
- Pleass RJ, Areschoug T, Lindahl G, Woof JM (2001) Streptococcal IgA-binding proteins bind in the C $\alpha$ 2-C $\alpha$ 3 interdomain region and inhibit binding of IgA to human CD89. *J Biol Chem* 276: 8197–8204
- Popp A, Dehio C, Grunert F, Meyer TF, Gray-Owen SD (1999) Molecular analysis of neisserial Opa protein interactions with the CEA family of receptors: identification of determinants contributing to the differential specificities of binding. *Cell Microbiol* 1: 169–181
- Rego S, Heal TJ, Pidwill GR, Till M, Robson A, Lamont RJ, Sessions RB, Jenkinson HF, Race PR, Nobbs AH (2016) Structural and functional analysis of cell wall-anchored polypeptide adhesin BspA in *Streptococcus agalactiae*. *J Biol Chem* 291: 15985–16000
- Samen U, Eikmanns BJ, Reinscheid DJ, Borges F (2007) The surface protein Srr-1 of *Streptococcus agalactiae* binds human keratin 4 and promotes adherence to epithelial HEp-2 cells. *Infect Immun* 75: 5405–5414
- Santi I, Scarselli M, Mariani M, Pezzicoli A, Masignani V, Taddei A, Grandi G, Telford JL, Soriani M (2007) BibA: A novel immunogenic bacterial adhesin contributing to group B *Streptococcus* survival in human blood. *Mol Microbiol* 63: 754–767
- Schubert A, Zakikhany K, Schreiner M, Frank R, Spellerberg B, Eikmanns BJ, Reinscheid DJ (2002) A fibrinogen receptor from group B *Streptococcus* interacts with fibrinogen by repetitive units with novel ligand binding sites. *Mol Microbiol* 46: 557–569
- Schubert A, Zakikhany K, Pietrocola G, Meinke A, Speziale P, Eikmanns BJ, Reinscheid DJ (2004) The fibrinogen receptor FbsA promotes adherence of *Streptococcus agalactiae* to human epithelial cells. *Infect Immun* 72: 6197–6205
- Seale AC, Bianchi-Jassir F, Russell NJ, Kohli-Lynch M, Tann CJ, Hall J, Madrid L, Blencowe H, Cousens S, Baker CJ et al (2017) Estimates of the burden of group B streptococcal disease worldwide for pregnant women, stillbirths, and children. *Clin Infect Dis* 65: S200–S219
- Seo HS, Mu R, Kim BJ, Doran KS, Sullam PM (2012) Binding of glycoprotein Srr1 of *Streptococcus agalactiae* to fibrinogen promotes attachment to brain endothelium and the development of meningitis. *PLoS Pathog* 8: e1002947
- Shabayek S, Spellerberg B (2018) Group B streptococcal colonization, molecular characteristics, and epidemiology. *Front Microbiol* 9: 437–
- Sheen TR, Jimenez A, Wang NY, Banerjee A, van Sorge NM, Doran KS (2011) Serine-rich repeat proteins and pili promote *Streptococcus agalactiae* colonization of the vaginal tract. *J Bacteriol* 193: 6834–6842
- Slevogt H, Zabel S, Opitz B, Hocke A, Eitel J, N'Guessan PD, Lucka L, Riesbeck K, Zimmermann W, Zweigner J et al (2008) CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. *Nat Immunol* 9: 1270–1278
- van Sorge NM, Quach D, Gurney MA, Sullam PM, Nizet V, Doran KS (2009) The group B streptococcal serine-rich repeat 1 glycoprotein mediates penetration of the blood-brain barrier. *J Infect Dis* 199: 1479–1487
- Stålhammar-Carlemalm M, Areschoug T, Larsson C, Lindahl G (1999) The R28 protein of *Streptococcus pyogenes* is related to several group B streptococcal surface proteins, confers protective immunity and promotes binding to human epithelial cells. *Mol Microbiol* 33: 208–219
- Stålhammar-Carlemalm M, Stenberg L, Lindahl G (1993) Protein Rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. *J Exp Med* 177: 1593–1603
- Tazi A, Disson O, Bellais S, Bouaboud A, Dmytruk N, Dramsi S, Mistou MY, Khun H, Mechler C, Tardieux I et al (2010) The surface protein HvgA mediates group B *Streptococcus* hypervirulence and meningeal tropism in neonates. *J Exp Med* 207: 2313–2322
- Tchoupa AK, Lichtenegger S, Reidl J, Hauck CR (2015) Outer membrane protein P1 is the CEACAM-binding adhesin of *Haemophilus influenzae*. *Mol Microbiol* 98: 440–455
- Tchoupa AK, Schuhmacher T, Hauck CR (2014) Signaling by epithelial members of the CEACAM family - Mucosal docking sites for pathogenic bacteria. *Cell Commun Signal* 12: 27
- Tenenbaum T, Bloier C, Adam R, Reinscheid DJ, Schroten H (2005) Adherence to and invasion of human brain microvascular endothelial cells are promoted by fibrinogen-binding protein FbsA of *Streptococcus agalactiae*. *Infect Immun* 73: 4404–4409
- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NI, Angiuoli SV, Crabtree J, Jones AI, Durkin AS et al (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”. *Proc Natl Acad Sci USA* 102: 13950–13955
- Villullas S, Hill DJ, Sessions RB, Rea J, Virji M (2007) Mutational analysis of human CEACAM1: The potential of receptor polymorphism in increasing host susceptibility to bacterial infection. *Cell Microbiol* 9: 329–346
- Virji M, Evans D, Hadfield A, Grunert F, Teixeira AM, Watt SM (1999) Critical determinants of host receptor targeting by *Neisseria meningitidis* and *Neisseria gonorrhoeae*: identification of Opa adhesin sites on the N-domain of CD66 molecules. *Mol Microbiol* 34: 538–551
- Virji M, Makepeace K, Ferguson DJP, Watt SM (1996) Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic neisseriae. *Mol Microbiol* 22: 941–950
- Wang JH (2013) The sequence signature of an Ig-fold. *Protein Cell* 4: 569–572
- Wang J-H, Springer TA (1998) Structural specializations of immunoglobulin superfamily members for adhesion to integrins and viruses. *Immunol Rev* 163: 197–215
- Wang NY, Patras KA, Seo HS, Cavaco CK, Rösler B, Neely MN, Sullam PM, Doran KS (2014) Group B streptococcal serine-rich repeat proteins promote interaction with fibrinogen and vaginal colonization. *J Infect Dis* 210: 982–991
- Watt SM, Teixeira AM, Zhou G-Q, Doyonnas R, Zhang Y, Grunert F, Blumberg RS, Kuroki M, Skubitz KM, Bates PA (2001) Homophilic adhesion of human CEACAM1 involves N-terminal domain interactions: structural analysis of the binding site. *Blood* 98: 1469–1479
- Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics* 54: Unit-5.6.

Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T (2019) HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res* 47: W322–W330

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A *et al* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235–242

Zhao Y, van Woudenberg E, Zhu J, Heck A, van Kessel K, de Haas C, Aerts P, van Strijp J, McCarthy A (2020) The Orphan immune receptor LILRB3 modulates Fc receptor-mediated functions of neutrophils. *J Immunol* 204: 954–966

Zhou H, Fuks A, Alcaraz G, Bolling TJ, Stanners CP (1993) Homophilic adhesion between Ig superfamily carcinoembryonic antigen molecules involves double reciprocal bonds. *J Cell Biol* 122: 951–960

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 430: 2237–2243



**License:** This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.