Solution Structure of Human Secretory Component and Implications for Biological Function

Alexandra Bonner, Clémence Perrier, Blaise Corthésy, and Stephen J. Perkins

From the Department of Biochemistry and Molecular Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, United Kingdom and the Service d’Immunologie et d’Allergie, Centre Hospitalier Universitaire Vaudois, BH19-650, Rue du Bugnon, CH-1011 Lausanne, Switzerland

Secretory component (SC) in association with polymeric IgA (pIgA) forms secretory IgA, the major antibody active at mucosal surfaces. SC also exists in the free form, with innate-like neutralizing properties against pathogens. Free SC consists of five glycosylated variable (V)-type Ig domains (D1–D5), whose structure was determined by x-ray and neutron scattering, ultracentrifugation, and modeling. With a radius of gyration of 3.53–3.63 nm, a length of 12.5 nm, and a sedimentation coefficient of 4.0 S, SC possesses an unexpected compact structure. Constrained scattering modeling based on up to 13,000 trial models shows that SC adopts a J-shaped structure in which D4 and D5 are folded back against D2 and D3. The seven glycosylation sites are located on one side of SC, leaving known IgA-binding motifs free to interact with pIgA. This work represents the first analysis of the three-dimensional structure of full-length free SC and paves the way to a better understanding of the association between SC and its potential ligands, i.e. pIgA and pathogenic-associated motifs.

The mucosal epithelia lining the gastrointestinal, respiratory, and urogenital tracts represent the largest surface (400 m² for the human body) in contact with the external environment (1). Consequently, they are permanently exposed to foreign antigens and must therefore be efficiently protected to prevent possible damage by pathogenic agents (2). In response to this threat, a major immune defense system is mediated by secretory immunoglobulin A (SIgA) within the lumen (3), where SIgA exerts its function of antigen neutralization. IgA is produced locally as polymeric IgA (pIgA), which consists of two or four IgA molecules linked covalently by a J chain (4). Subsequent transport of pIgA (and to a lesser extent pentameric IgM) across the epithelium is ensured by the polymeric Ig receptor (pIgR), which is expressed on the basolateral surface of epithelial cells. Following cleavage at luminal surfaces, SIgA is released as a complex of pIgA and the cleaved extracellular portion of the pIgR which is termed secretory component (SC). Because secretory IgM can be present in external secretions with an accessory role as a neutralizing antibody, this suggests that SC has a generic antibody-binding ability, being able to bind to both pIgA and pentameric IgM (5). This selectivity may be required to allow only larger, polymerized antibodies protected by heavily glycosylated SCs to enter the harsh environment of external secretions.

SC consists of five immunoglobulin-like domains (D1–D5) and up to seven glycan chains. SC bound to pIgA protects the antibody against proteolytic digestion (6) and governs anchoring of SIgA at mucosal surfaces (7). Free SC is also found in secretions and is now recognized as an active antibacterial participant to protect mucosal surfaces against Helicobacter pylori (8), enteropathogenic Escherichia coli and Clostridium difficile toxin A (9), and Streptococcus pneumoniae choline binding protein A (10, 11). Recombinant human SC produced from transfected Chinese hamster ovary (CHO) cells behaves identically, and its neutralizing properties rely on the presence of oligosaccharides on SC (12).

Despite the importance of SC and its glycosylation, only partial information is available on its structure. SC comprises the first 585 residues in pIgR (Fig. 1A) (13). Homology modeling of rabbit SC domains 1 and 2 (D1 and D2) indicates that both exhibit the typical features of Ig variable (V)-type superfamily members with seven β-strands A–G and an extra two designated C′ and C″ (14, 15). A similar topology was determined in the modeling of domains 2 and 3 (D2 and D3) of mouse SC (16). These were confirmed by the crystal structure of D1 of human SC, which revealed a V-type fold (17). So far, the three-dimensional domain arrangement of full-length SC has not been considered, even though this is essential to appreciate molecular mechanisms that are involved in mucosal immunity. Its high glycosylation and the presence of long interdomain linkers suggest that SC may be too flexible to be crystallized intact. In a situation such as this, x-ray and neutron scattering and analytical ultracentrifugation in combination with constrained modeling techniques will reveal the domain arrangement of SC (18). Here, we determine a solution structure for recombinant human SC produced in CHO cells. We show that SC forms a...
Solution Structure of Secretory Component

compact domain arrangement, and this structure rationalizes many of the biological roles of both free and bound SC for the first time.

MATERIALS AND METHODS

Recombinant Secretory Component and Its Fragments—CHO cells producing human SC (7) were cultured in suspension in CHOS-SFMII medium (Invitrogen) using CELLine 350 bioreactors (Integra-Biosciences). Supernatants (8 ml) were collected twice a week, and immediately chromatographed on 1.6-cm x 1-meter long Superdex-200 column in PBS. Fractions containing human SC as confirmed by immunoblotting were pooled and filtered through 0.22-μm cartridges. Further purification of fragments D1–D3 and D4–D5 was achieved by repeating chromatography onto the Superdex-200 column. To identify the SC cleavage site, N-terminal sequencing of purified D4–D5 from a polyvinylidene difluoride blot was performed by Edman degradation at Alta Biosciences (University of Birmingham, UK).

For the x-ray scattering and ultracentrifugation experiments, samples were recovered in PBS. For neutron scattering, the buffer was Dulbecco’s PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na2HPO4, 1.5 mM KH2PO4, pH 7.5), made up in 100% H2O. The neutron samples were dialyzed at 6 °C into this 100% H2O buffer for 36 h with four buffer changes. The two buffer densities in H2O solvent were predicted to be 1.00485 g/ml and 1.00543 g/ml, respectively, from these compositions using SEDNTERP. The SC, D1–D3, and D4–D5 concentrations were determined using absorption coefficients of 12.3, 7.1, and 18.7 (1%, 280 nm, and 1-cm path length) that were calculated from its composition by assuming the presence of seven biantennary complex-type oligosaccharides Man3GlcNAc4Gal2NeuNAc2 (see Fig. 1) (19). The partial specific volume  for SC, D1–D3, and D4–D5 was calculated to be 0.717, 0.717, and 0.718 ml/g, respectively. The DCCT + g(s) time-derivative analyses and SEDFIT Lamm equation fits based on the non-interacting species model were performed as previously described (23).

Constrained Modeling of SC—SC corresponds to residues 19–603 of the polymeric immunoglobulin receptor (SwissProt code P01833). The start and end of D2–D5 is defined as residues 1–104 in Fig. 1. The homology modeling of D2–D5 employed MODELLER v7 (24) and was based on a topological sequence alignment with the D1 crystal structure identified using BLASTP (PDB code 1xed) (17). The models were structurally validated using PROCHECK. Seven biantennary oligosaccharide chains Man3GlcNAc4Gal2NeuNAc2 were incorporated in extended conformations in the D1–D5 models (25, 23).

The first SC residue Lys was added to the D1 crystal structure using MODELLER v7. To model the other interdomain linkers, residues 109–116, 219–221, 324–335, 440–446, and 543–585 (Fig. 1B) were created in extended β-strand conformations using INSIGHT II 98.0 with BIOPOLYMER, DISCOVER, HOMOLOGY, and DELPHI modules (Accelrys, San Diego, CA), and 5000 randomized conformations for each were generated using DISCOVER3 (22). Linker length constraints of 95–100% were applied to generate sufficient conformational variability when needed. Full SC models were created by superimpositions of the domains and the linkers.

Each SC, D1–D3, and D4–D5 model was converted into Debye spheres to calculate the x-ray and neutron scattering curves (22, 23). A cube side length of 0.542 nm in combination with a detector distance of 3.0 m, yielded a Q range from 0.07 to 2.1 nm⁻¹ (where Q = 4π sin θ/λ; 2θ = scattering angle; λ = wavelength). Samples were measured at concentrations between 0.62 and to 2.0 mg/ml (SC), 0.39 to 1.6 mg/ml (D1–D3), and 0.40 to 0.73 mg/ml (D4–D5). Data were acquired using five or ten time frames of 0.5, 1, 2, or 4 s to establish the absence of radiation damage. Neutron scattering data for SC were collected in one session in a Q range between 0.02 and 1.93 nm⁻¹ on Instrument LOQ at the ISIS facility at the Rutherford Appleton laboratory, Didcot, UK (21). The proton beam current was 172–173 μA, and acquisition times were 1–2 h for concentrations between 0.62 and 2.0 mg/ml, measured at 15 °C. Other details, including calibrations and data reduction, the Guinier analyses to determine the radius of gyration Rg and the radius of gyration of the cross-sectional structure Rsc, and the calculation of the distance distribution function P(r) using GNOM for which Q ranged between 0.08 and 2.06 nm⁻¹, are described previously (22, 23).

Analytical ultracentrifugation was performed at 20 °C on a Beckman XL-I instrument, equipped with an Antiscan rotor. Sedimentation equilibrium data sets for SC between 0.20 and 2.50 mg/ml were acquired over 45 h with column heights of 2 mm at rotor speeds of 8,000, 11,000, 14,000, 18,000, and 23,000 rpm, until equilibrium had been reached at each speed as shown by the perfect overlay of runs measured at 5-h intervals. Sedimentation velocity data for SC, D1–D3, and D4–D5 were acquired over 16 h at rotor speeds of 20,000, 25,000, 30,000, and 35,000 rpm (and 42,000 rpm for D1–D3 and D4–D5) with column heights of 12 mm. SC was studied at five concentrations from 0.30 to 2.50 mg/ml. D1–D3 and D4–D5 were each studied at two concentrations of 0.39–1.6 mg/ml and 0.40–0.73 mg/ml respectively. DCDDT + g(s) time-derivative analyses and SEDFIT Lamm equation fits based on the non-interacting species model were performed as previously described (23).

X-ray and Neutron Scattering and Ultracentrifugation Data—X-ray scattering data were obtained in two sessions on the Beamline ID02 at the European Synchrotron Radiation Facility (Grenoble, France), operating at 6.03 GeV (20). To reduce the incident flux, the experiments were performed in 16-bunch and 4-bunch mode with storage ring currents ranging between 92 and 53 mA (SC) and 4-bunch mode with currents ranging between 41 and 23 mA (D1–D3 and D4–D5). The sample-to-background ratio varied from 400 to 600 for SC and D1–D3, and 400 to 800 for D4–D5. Fractions containing human SC as confirmed by immunoblotting were collected twice a week, and immediately chromatographed on 1.6-cm x 1-meter long Superdex-200 column in PBS. Fractions containing human SC as confirmed by immunoblotting were collected twice a week, and immediately chromatographed on 1.6-cm x 1-meter long Superdex-200 column in PBS.
with a cutoff of 4 atoms consistently produced sphere models within 95% of the total unhydrated volume of 99.7, 56.9, and 43.8 nm³, respectively, calculated from their compositions. Hence the optimal totals of unhydrated spheres were 626, 268, and 353 respectively, which were then hydrated. The number of spheres N in the unhydrated and hydrated models after grid transformation was used to assess steric overlap between the D1, D2, D3, D4, and D5 domains.

The models were assessed by calculation of the Rₛ and Rₓₛ values in the same Q ranges used for the experimental Guinier fits. Models that passed these Rₛ and Rₓₛ filters were ranked using the R-factor goodness-of-fit agreement. A flat background correction of 2.7% of I(0) was applied to the final neutron scattering curve fits in Fig. 8 to allow for a uniform incoherent scattering of residual protons in the sample. Other details, including those of calibration studies used to validate this approach, are given elsewhere (22, 23).

### RESULTS

#### Characterization of SC and Its Fragments

In SC, sequence similarities between D1 and the other four SC domains (D2–D5) show that all belong to the immunoglobulin (Ig) variable (V)-type classification (Fig. 1A). For clarification, the domains and interdomain linkers are depicted in Fig. 1B. As glycosylation and proper disulfide bonding are essential features of SC, we chose to produce the protein in CHO cells, which yields SC that is indistinguishable when compared with colostrum SC (12, 28). Purified recombinant human SC resulted in a major band at an apparent molecular mass of 75 kDa when analyzed by silver staining and immunodetection (Fig. 1C, lanes 1, 3, 5, and 7). This molecular mass agrees well with the sequence-derived value of 79.6 kDa. Upon storage for several months at 4 °C, two additional bands at 45 and 30 kDa appear (Fig. 1C, lanes 2, 4, 6, and 8). In Western blots, detection with polyclonal antiserum specific for D1 and D5 established that the bands at 45 and 30 kDa represent cleavage of full-length SC into D1–D3 and D4–D5 fragments. This was confirmed by N-terminal sequencing of fragment D4–D5 that identified the sequence SPTTVKGAGVAG corresponding to the beginning of domain 4 and indicating cleavage between Arg³³⁶ and Ser³³⁷. The PeptideCutter web tool (available from the ExPASy server) suggested that the enzymatic cleavage may be caused by one of trypsin, thrombin, or other related proteases. This unexpected observation provided us with three different polypeptides that were used to generate the results presented below.

#### X-ray and Neutron Scattering

X-ray scattering identified the domain arrangements in SC, D1–D3, and D4–D5 at concentrations between 0.40 and 2.7 mg/ml (see “Materials and Methods”). Because no radiation damage was detectable, the ten time frames of each acquisition were averaged for analyses. SDS-PAGE analyses before and after runs showed that, although SC and D4–D5 generally remained intact, further cleavages occurred in D1–D3 to give single domain fragments as well as the three-domain fragment. All the SC and D4–D5 results below correspond to samples shown to be intact by SDS-PAGE before and after the experiment.

At low Q values, Guinier analyses of ln I(Q) against Q² resulted in linear plots for SC and its two fragments, yielding the radius of gyration (Rₛ) (Fig. 2, A, E, and F). At larger Q
values, analyses of \( \ln(I(Q)) \) against \( Q^2 \) yielded the \( R_G \) of the cross-section (\( R_{XS} \)). For SC, a mean x-ray \( R_G \) of 3.53 (± 0.03) nm and mean \( R_{XS} \) of 1.76 (± 0.08) nm were recorded from seven runs (Table 1). Another SC sample gave a similar mean x-ray \( R_G \) of 3.64 (± 0.17) nm and \( R_{XS} \) of 1.63 (± 0.12) nm from six runs. A partially cleaved sample with ~33% SC, 33% D1–D3, and 33% D4–D5 showed a small increase in the mean \( R_G \) to 3.77 (± 0.17) nm and a small decrease in the mean \( R_{XS} \) to 1.60 (± 0.61) nm. From five runs, the partially cleaved sample of D1–D3 showed an apparent \( R_G \) of 3.44 (± 0.27) nm and an apparent \( R_{XS} \) of 1.07 (± 0.07) nm (supplemental Fig. S1). From five runs, D4–D5 was found to have a mean \( R_G \) of 3.16 (± 0.04) nm and a mean \( R_{XS} \) of 1.11 (± 0.08) nm (supplemental Fig. S1). The Guinier analyses were confirmed by the \( I(0)/c \) values, which are proportional to molecular masses, where \( c \) is the concentration (18). The observed \( I(0)/c \) values were in the ratio of 1:0.30:0.53, respectively, to Lupolen. The sequence-derived molecular mass ratios were 1:0.57:0.43 in that order. \( I(0)/c \) for D1–D3 was lower than expected, which is attributed to its cleavage, whereas that for D4–D5 was in fair agreement with the sequence.

Interestingly SC appeared more compact than either D1–D3 or D4–D5. Shape information was determined from the anisotropy ratio, \( R_G/R_O \) (where \( R_G \) is the \( R_G \) value of a sphere of equal volume to that of the hydrated glycoprotein). The ratios were 1.44–1.49, 1.69, and 1.70 for SC, D1–D3, and D4–D5 in that order (Table 1). Because \( R_G/R_O \) for globular proteins is close to 1.28 (29), all three proteins are elongated. Assuming that the proteins possess cylindrical shapes, the lengths of SC, D1–D3, and D4–D5 were calculated from its \( R_G \) and \( R_{XS} \) values to be 10.6–11.3, 11.3, and 10.2 nm, respectively. From the \( I(0) \) and \( [I(Q)_Q]_{Q=0} \) values, the lengths were calculated to be 11.9, 9.1, and 10.8 nm, respectively (Table 1). Despite the cleavage in the D1–D3 sample, the similarity of these three lengths is explained by postulating that SC is folded back upon itself. If the end-to-end length of the SC domains is taken to be 3.7 nm from the D1 crystal structure (17), SC, D1–D3, and D4–D5 are predicted to be 18.5, 11.1, and 7.4 nm long if they are fully extended but without linkers. The disparity between the observed SC length of 10.6–11.9 nm and the 18.5 nm prediction shows that SC must be folded back. That D4–D5 (Table 1) is observed to be longer than the predicted value may result from the 42 C-terminus residues in SC (Fig. 1A). The observed lengths for D1–D3 are in fair agreement with the predicted length of 11.1 nm.

Neutron scattering data for SC confirm the x-ray \( R_G \) values (Table 1). These were obtained as a control of radiation damage by x-rays, because these effects are absent in neutron scattering. The neutron data are complementary in that the hydration shell surrounding SC is not detectable (26). The high negative protein-solvent scattering contrast acts as a control of any scattering inhomogeneity effects caused by the high proportion of carbohydrate.

**TABLE 1**

Experimental x-ray and neutron scattering and ultracentrifugation analyses for SC, D1–D3, and D4–D5

| Protein | \( R_G \) | \( R_{XS} \) | \( R_G/R_O \) | \( I(0)/c \) | \( R_G \) | \( M \) | \( S \) |
|---------|----------|----------|----------------|----------------|----------|-------|-------|
| SC (x-ray) | 3.53 ± 0.03 | 1.76 ± 0.08 | 1.44 | 10.6 | 11.9 | 3.66 ± 0.19 | 3.8 ± 0.2 | 13.0 | 4.2 | 4.0 (int) | 1.55 |
| SC (x-ray) | 3.64 ± 0.17 | 1.65 ± 0.12 | 1.49 | 11.3 | NAa | 3.71 ± 0.12 | NA | NA | 3.9 (abs) | 3.1 | NA | 1.42 |
| SC (neutron) | 3.63 ± 0.28 | 1.30 ± 0.10 | NA | 11.7 | 13.1 | 3.73 ± 0.17 | 3.2 | 12.0 | 2.4 | 2.4 (int) | 2.3 (abs) | 1.55 |
| D1–D3 (x-ray) | 3.44 ± 0.27b | 1.07 ± 0.07b | 1.69 | 11.3 | 9.1 | 3.52 ± 0.27 | 2.9 | 12.0 | 3.1 | NA | 1.42 |
| D4–D5 (x-ray) | 3.16 ± 0.04 | 1.11 ± 0.08 | 1.70 | 10.2 | 10.8 | 3.15 ± 0.18 | 2.7 | 10.0 | 2.4 | 2.4 (int) | 2.3 (abs) | 1.48 |

a: NA, data not available.
b: Apparent (see text).
in SC. The mean of three \( R_G \) and \( R_{XS} \) values for SC were 3.63 (± 0.28) nm and 1.30 (± 0.10) nm, respectively (Fig. 2). The decrease in the neutron \( R_{XS} \) compared with the x-ray \( R_{XS} \) value is attributed to the absence of hydration effects on the cross-section of SC (25). The length of SC from the neutron \( R_G \) and \( R_{XS} \) was 11.7 nm and from \( I(0) \) and \( |I(Q)\|_{0\rightarrow\infty} \) was 13.1 nm, which is consistent with the x-ray determined lengths. The neutron data gives molecular masses from the expression \( I(0)/c \times 9 \times 10^5 \) (22). The observed \( I(0)/c \) of 0.087 (± 0.009) resulted in 78 (± 8) kDa, which is close to the sequence-derived value of 79.6 kDa.

The transformation of \( I(Q) \) into the distance distribution function, \( P(r) \) confirmed that SC has a folded back conformation in solution. The \( P(r) \) curves for SC (x-rays and neutrons) and for D1–D3 and D4–D5 (x-rays) were reproducible with single maxima (Fig. 3). The \( R_G \) from these were 3.66 (± 0.19) nm and 3.71 (± 0.12) nm (two SC x-ray samples), 3.73 (± 0.17) nm (SC neutrons), 3.52 (± 0.27) nm (D1–D3), and 3.15 (± 0.18) nm (D4–D5), all of which were consistent with the Guinier analyses (Table 1). The maximum, \( M \), corresponds to the most frequently occurring interatomic distance. These were 3.8 (± 0.2) nm and 3.2 nm, respectively, for the x-ray and neutron data for SC and were slightly smaller at 2.9 nm and 2.7 nm for D1–D3 and D4–D5, respectively. The length, \( L \), is determined when \( P(r) \) becomes zero at large \( r \) values and is not dependent on shape assumptions. For SC, \( L \) was determined to be 13 (± 1) nm (x-rays) and 12 (± 1) nm (neutrons). For D1–D3 and D4–D5, \( L \) was determined to be 12 (± 1) nm and 10 (± 1) nm, respectively. Although the D1–D3 analysis may be affected by cleaved products, this analysis will reveal the maximum length of the intact protein as any cleavage products will be shorter in length. These lengths were consistent with the Guinier-derived analyses (Table 1).

**Analytical Ultracentrifugation**—The oligomerization of SC at concentrations between 0.2 and 2.5 mg/ml was determined by sedimentation equilibrium to show that this is monomeric. Assuming a single species was present, good curve fits with small random residuals were obtained (Fig. 4). SDS-PAGE analyses before and after runs confirmed that any SC cleavage was not detectable. A small decrease in the molecular mass of SC was observed with increasing rotor speed, which is attributable to sample polydispersity (30). This is most likely to result from heterogeneous occupation at the seven \( N \)-linked glycosylation sites, and/or any cleavage products (Fig. 1). Thus the presence of seven bi-antennary oligosaccharides would give a molecular mass of 79,600 Da, whereas seven tetra-antennary ones would give 88,800 Da. In addition, the molecular masses slightly increased on dilution (lower panels, Fig. 4). This was attributed to decreased interactions between SC molecules at lower concentrations, because any oligomerization would result in a decrease in molecular masses on dilution. Regressions resulted in molecular masses of 79,500 (± 2,900) Da (interference) and 81,600 (± 5,500) Da (absorbance) at zero concentration. These agree well with the sequence-derived molecular mass. The equilibrium data therefore showed that all seven putative glycosylation sites are filled.

The sedimentation coefficient, \( s_{20,w}^0 \), monitors the macromolecular elongation, independently from \( R_G \). Sedimentation velocity was used to examine SC, D1–D3, and D4–D5, as well as SC that was cleaved into an equimolar mixture of the three forms, by incubating this at room temperature. The advantage of SEDFIT \( c(s) \) distribution plots is that these identify the \( s_{20,w}^0 \) values of any species present in each sample (Table 1), hence these values are independent of degradation effects. All the boundary fits showed good agreements (Fig. 5, A–D).
plots showed that SC sedimented as a single species at 4.2 S, whereas major species were seen at 3.1 S for D1–D3 and 2.4 S for D4–D5 (Fig. 5, E–G). This interpretation was confirmed by the three major peaks seen in the cleaved SC sample (Fig. 5H). The c(s) plot for D4–D5 showed a minor peak at 0.7 S (Fig. 5G). In contrast, the D1–D3 c(s) analysis showed peaks at 0.7 S and 2.2 S to indicate more extensive cleavage, in agreement with the x-ray scattering and SDS-PAGE analyses above. The c(M) analyses represent the c(s) peaks as molecular mass values. The three major peaks in Fig. 5H correspond to 24,500 (± 5,600), 44,700 (± 6,200), and 79,000 (± 7,200) Da. These agree well with the sequence-derived values of 34,100, 45,500, and 79,600 Da.

The SEDFIT analyses for SC and D4–D5 were confirmed by the good fits obtained in time-derivative g(s*) analyses using DCDT+ (Fig. 6 and supplemental Fig. S2). The \(g_{20,w}'\) values were determined to be 4.0 S and 3.9 S for SC (interference and absorbance optics, respectively) and 2.4 S and 2.3 S for D4–D5 (interference and absorbance optics, respectively (Table 1)). D1–D3 could not be analyzed by DCDT+.

The \(g_{20,w}'\) values lead to frictional ratios \(f/f_o\) that confirm the folded back structure of SC, where \(f_o\) is the frictional coefficient of the sphere of volume equal to that of the hydrated protein. Many globular proteins show \(f/f_o\) of 1.2 to 1.3. For SC, D1–D3, and D4–D5, the \(f/f_o\) ratios were determined to be 1.55, 1.42, and 1.48. These show that all three glycoproteins possess similar degrees of elongation. They are consistent with the \(R_g/R_o\) ratios of 1.44–1.49, 1.69, and 1.70 in that order.

Constrained Modeling—The constrained modeling of the solution data utilized the D1 crystal structure, which has a \(\beta\)-sheet sandwich formed from the DEBA and A’GFCC’C’ \(\beta\)-strands (17). These \(\beta\)-strands are defined in Fig. 1A. D1 showed sequence identities from 24% to 36% with D2–D5, where over 30% is considered significant (31). Cys\(^{22}\)-Cys\(^{92}\) forms the internal disulfide bridge between \(\beta\)-strands B and F and was conserved in D1–D5. Sequence gaps and insertions were minimal, except that \(\beta\)-strand C’ is deleted in D5. The alignment using BLASTP thus defined the D1–D5 domains and their six flanking peptides. Homology models for the D2–D5 domains were constructed and validated, and oligosaccharides
were attached ("Materials and Methods"). The four linkers were defined to run from the end of H9252-strand G to H9252-strand A between domain pairs (25). Linkers (boxed in Fig. 1) were created from conformational libraries, from which random selections were used to generate full SC models in all orientations. Secondary structure analyses of the 42-residue C-terminal peptide predicted little secondary structure (32). BLASTP sequence searches of over 39,000 structures in the Protein Data Bank did not reveal any significant structural similarity for this C-terminal peptide (33). Its amino acid composition showed a high proportion of 12 negatively charged Asp/Glu residues and nine positively charged Arg/Lys residues, together with 14 small Gly/Ala/Ser residues, all of which promote random structures (27). Accordingly the C-terminal peptide was modeled as another unconstrained structure.

In the D4–D5 modeling, for which the x-ray scattering curve was free of cleavage, the only conformational variables were the D4–D5 linker and the C-terminal peptide. Unconstrained linkers resulted in too many models that were overlapped between D4 and D5. Accordingly, the D4–D5 linker was constrained to be between 95 and 100% of its maximum length to permit full domain reorientations. The C-terminal peptide was set to be at minimum lengths of 5, 7, 9.8, and 11.8 nm to enforce the generation of a wide range of conformations. The resulting 6000 models spanned a wide range of conformational separations of the N terminus and C terminus from 0.6 to 20.6 nm (Table 2). Filtering to retain a minimum of 346 hydrated spheres (i.e. those with no significantly overlap) and a R_G between 3.00 and 3.20 nm gave 1011 satisfactory models (Table 2). Sorting based on the R-factor goodness-of-fit parameter identified the best nine models. The best D4–D5 model has an R_G of 3.15 nm and an R_XS of 1.14 nm, giving the closest agreement with the experimental values (Table 2), and gave an excellent curve fit. Its sedimentation coefficient of 2.04 S is close to the experimental values of 2.3–2.4 S. All nine best-fit models showed that the D4 and D5 domains were V-shaped with an extended C-terminal peptide. The constrained modeling parameters and curve fits are presented in panels D–F of supplemental Figs. S3 and S4.

A poor-fit D4–D5 model is exemplified by that with the largest R_G of 3.71 nm, an R_XS of 1.45 nm, and a sedimentation coefficient of 1.94 S. This has a linear D4–D5 arrangement and showed a large discrepancy with the experimental I(Q) curve (supplemental Fig. S4F). Another model was created from the best-fit model in which the oligosaccharides were folded back against the protein surface of D4–D5 (green in supplemental Fig. S4, D–F; see "CHO in" in Table 2). Although the R_G and sedimentation coefficient from this model were similar, the R_XS was significantly reduced from 1.14 to 0.86 nm. Good D1–D3 curve fits were obtained by postulating that the observed x-ray scattering corresponded to a mixture of D1–D3 and single-domain fragments (cf.: the SDS-PAGE and c(s) anal-

FIGURE 5. Sedimentation velocity fits for SC, D1–D3, and D4 –D5 by SEDFIT. The boundary fits are shown to the left and the c(s) plots are shown to the right. A and E, SC at 1.24 mg/ml and 30,000 rpm, showing every 5th boundary of the first 100; B and F, D1–D3 at 1.6 mg/ml and 42,000 rpm, showing every 16th boundary of 500; C and G, D4–D5 at 0.40 mg/ml and 42,000 rpm, showing every 5th boundary of the first 55; and D and H, cleaved SC at 0.31 mg/ml and 42,000 rpm, showing every 5th boundary of the first 110.

FIGURE 6. Sedimentation velocity fits for SC by DCDT+. In the g(s*) plots, the s_{20,w} values are arrowed. SC at 2.5 mg/ml (A) and 1.2 mg/ml (B), using 35,000 rpm. The residuals are shown above each g(s*) analysis. For the sedimentation velocity fit for D4–D5 by DCDT+, see supplemental Fig. S2.
The D1–D3 modeling was achieved in five searches. The use of three unconstrained linkers yielded few good fits. Hence the linkers were constrained to be 95–100% extended, and this gave models that spanned a sufficiently wide range of \( R_G \) but with \( R_{S_S} \) that were too high. The final search was based on the randomization of only the D1–D2 linker. The D2–D3 linker proved too short to permit movement without steric overlap, and the C-terminal D3–D4 linker was held fixed in an extended conformation. Even with this, whereas 65 models showed compatible \( R_G \) values, none showed a low enough \( R_{S_S} \) of 1.07 nm. All ten best-fit models exhibited linear domain arrangements. Alterations of the oligosaccharide conformations were also unable to account for the reduced \( R_{S_S} \) value. The final good D1–D3 curve fits were obtained only by computing composite scattering curves from the best-fit D1–D3 model with the lowest R-factor of 13.2% and a single domain D1 model. The sum of 74% D1–D3 and 26% D1 gave the best curve fit with an R-factor of 6.0%. The linear D1–D3 model gave a sedimentation coefficient of 3.02 S that agrees well with the observed value of 3.1 S. Other poor-fit D1–D3 models confirmed this interpretation. For example, combining a U-shaped D1–D3 model with the D1 model did not reduce the R-factor or \( R_{S_S} \), even though it gave a good \( R_G \). The D1–D3 constrained modeling parameters and curve fits are shown in panels A–C in the supplemental Figs. S3 and S4.

Intact SC (for which the x-ray scattering curve was shown to be free of cleavage products) was modeled using either five unconstrained linkers or the best-fit D1–D3 and D4–D5 models. The unconstrained search is summarized by the filled cyan circles in Fig. 7 (A–F). Convex distributions of \( R_{S_S} \) as a function of \( R_G \) were observed. Their maxima agreed well with the observed \( R_G \) of 3.53 nm (x-rays) and 3.63 nm (neutrons) in Fig. 7 (A and D), and their minima corresponded to the lowest R-factors in Fig. 7 (B and E). However, many models showed

![FIGURE 7. Constrained modeling of SC. The parameters for 10,000 models for SC are shown (A–F). The top panels compare the \( R_{S_S} \) and \( R_G \) values. The middle panels compare the R-factors with the \( R_G \). The bottom panels compare the N-terminal and C-terminal \( \alpha \)-carbon separation with the \( R_G \). The dashed lines correspond to the experimental \( R_G \) and \( R_{S_S} \). The 5,000 SC models represented by blue circles show the unrestrained linker conformations. The final SC analyses are represented by 5,000 unfilled circles. In each analysis, the best-fit model is shown in red, related best-fit models in yellow, models with compact oligosaccharides in green, and selected poor fit models in blue. For the results of the constrained modeling for D1–D3 and D4–D5 see supplemental Fig. S3.](image-url)
steric overlap; accordingly, this search was not pursued further. The second search varied only the D3–D4 linker between the best-fit D1–D3 and D4–D5 models. The unfilled circles in Fig. 7(A–F) showed that these followed the same trends as the cyan circles, but resulted in fewer sterically overlapping models with too low R_g values. Filtering showed that 268 SC models agreed with the experimental x-ray R_G of 3.53 nm, of which 70 agree with the experimental R_XS of 1.76 nm (Table 2). The removal of partially overlapping SC models gave the 10 best-fit SC models (yellow in Fig. 7, A–C). All 10 best-fit SC models showed a folded-back arrangement in which D1–D3 is in proximity to D4–D5 in an overall J-shaped conformation. This arrangement was also found in the best-fit models from the unconstrained search, although their R-factors were slightly higher. The sedimentation coefficient of 4.06 S is close to the experimental values of 3.9–4.0 S. The neutron modeling in Fig. 7(D–F) also resulted in the same J-shaped best-fit models and a good curve fit in Fig. 8D. A linear domain arrangement showed significantly worsened fits (blue in Figs. 7(A–F), 8C, and 8F). When the extended oligosaccharide orientations in the best-fit model analyses were conducted in near-physiological solution conditions using complementary methods based on x-ray and neutron scattering and analytical ultracentrifugation. When analyzed with constrained modeling, the three-dimensional structure of free SC was found to be a compact, folded-back structure. The ten best-fit models of Fig. 9A are very similar to each other. As all adopted a J-shaped structure, this emphasized the robustness of the procedure. A stereo view of the best-fit model is presented in Fig. 9B. The J-shaped structure is confirmed by the lengths of Fig. 3, the sedimentation coefficients (Figs. 5 and 6) and the constrained modeling (Fig. 8). The point at which SC is folded back to form a compact J-shaped structure was located to the junction between the D3 and D4 domains in both modeling searches of all five SC domains. In support of this model, the length and sedimentation coefficient data obtained with the D1–D3 and D4–D5 fragments both fit perfectly with the SC structure (Figs. 3 and 5) and reinforce the accuracy of the SC model.

This compact, folded-back structure of SC was unexpected in the light of the extended domain arrangements of molecular modeling and crystal structures of SC domains from rabbit, mouse, and human have already contributed insight into the topology of SC and pIgA complexes (14, 15, 17). However, the analyses of only protein fragments, the use of strictly predictive computer-based modeling, and the absence of information on carbohydrate residues all represent limitations when compared with the determination of the three-dimensional structure of full-length SC. Given the demonstrated importance of free SC in innate immunity and its role in endowing gDNA with optimal biological activity, our study of intact SC provides several novel insights on the structure-function relationship in the whole protein.

In view of the importance of glycosylation, and inter-domain flexibility to the SC structure, all analyses were conducted in near-physiological solution conditions using complementary methods based on x-ray and neutron scattering and analytical ultracentrifugation. When analyzed with constrained modeling, the three-dimensional structure of free SC was found to be a compact, folded-back structure. The ten best-fit models of Fig. 9A are very similar to each other. As all adopted a J-shaped structure, this emphasized the robustness of the procedure. A stereo view of the best-fit model is presented in Fig. 9B. The J-shaped structure is confirmed by the lengths of Fig. 3, the sedimentation coefficients (Figs. 5 and 6) and the constrained modeling (Fig. 8). The point at which SC is folded back to form a compact J-shaped structure was located to the junction between the D3 and D4 domains in both modeling searches of all five SC domains. In support of this model, the length and sedimentation coefficient data obtained with the D1–D3 and D4–D5 fragments both fit perfectly with the SC structure (Figs. 3 and 5) and reinforce the accuracy of the SC model.

This compact, folded-back structure of SC was unexpected in the light of the extended domain arrangements of...
Solution Structure of Secretory Component

A

D1
D2
D3
D4
D5

I
II
III
IV
V
VI
VII
VIII
IX
X

B

C

D

CDR I
CDR II
CDR III
Cys502

16978 JOURNAL OF BIOLOGICAL CHEMISTRY

VOLUME 282 • NUMBER 23 • JUNE 8, 2007
the structurally related seven Ig-like domains in carcinoembryonic antigen and four FnIII domains in anosmin-1 (25, 34). In fact, the linker between D3 and D4 comprises 10 amino acids and is unusually long (Fig. 1). This length provides the requisite flexibility needed to form the J-shaped conformation. It is noteworthy that the cleavage site between Arg^{336} and Ser^{337} is exposed on the surface in all ten best-fit structures (Fig. 9A), thus accounting for its susceptibility to protease(s) and justifying the appearance of the two D1–D3 and D4–D5 polypeptides. The Poisson-Boltzmann electrostatic maps of the carbohydrate-free surfaces in D1–D3 and D4–D5 (Fig. 9C) suggest that there are spatially well defined areas of negative and positive charges, respectively, on these that can attract each other to form the bent structure of SC. The short linker between D2–D3 may reduce flexibility between D2 and D3, hence facilitating the binding of D4 and D5 to these.

The modeling showed that the C-terminal polypeptide linker connecting SC with the transmembrane domain of the pLgR exhibits no particular folding (Figs. 1 and 9). Hence D1–D5 of SC can freely move with respect to this C-terminal linker, thus enhancing their likelihood to encounter and interact with their cargo represented by polymeric Ig present in the lamina propria. In a more physiological context, this length facilitates access to proteases in order that SC or SIgA can be released at mucosal surfaces. The unfolded nature of the linker also suggests that the structure of SC identified in this work will be similar to that when incorporated in its pLgR precursor, because steric interference with components on the cell membrane will be limited.

The seven Asn residues that constitute the attachment points for carbohydrate residues mostly appear on the same side of SC, implying that the glycans are clustered within a restricted part of the SC structure (Fig. 9B). The curve fits showed a preference for modeling the seven oligosaccharide chains in conformations extending away from the protein surface (middle panels of Fig. 8). Although their precise spatial locations remain unassigned by constrained scattering modeling, a preferential distribution of oligosaccharides on one face of SC is not ruled out. If this is so, this can contribute to tether free and/or IgA-bound SC efficiently to antigenic structures found at the surface of microorganisms (12).

In the context of the interaction between SC and pLgA, the location of the oligosaccharides on one face of SC leave uncovered, and thus accessible, the three CDR-like motifs present in D1, as well as Cys^{502} in D5 (Fig. 9D). These constitute two critical elements for the interaction with pLgA (35, 36). Indeed, the oligosaccharides of SC are not involved in the interaction with pLgA but are nonetheless important for its resistance to proteases and its correct anchoring within the mucus layer on the epithelium (7). Studies comparing deglycosylated and partly glycosylated D1 revealed nearly equal binding affinities for pLgA (17). Recovery of the Fd fragment after digestion of pLgA with proteases from intestinal fluids indicates that the Fc portion is particularly sensitive in the absence of bound SC (6). The determined length of SC fits very well with its likely association with pLgA without affecting the F(ab')_{2} and the hinge region. This is supported by the observation that pLgA and SlgA of the same specificity show the same affinity for their cognate antigen (37). Additionally, the labile link between D3 and D4 of SC is likely to be concealed within pLgA, thus providing an explanation for the mutual protection previously observed between the two partners (6). This translates into the generation of a complex molecule optimally designed to operate at mucosal surfaces.

Although not covalent, the interaction between the C-terminal part of the J chain in pLgA and SC is a critical step for transport of the pLgR complex to mucosal surfaces (4). Pentameric IgM and pLgA are both transported by pLgR, but it is known that D1 can discriminate among Ig subclasses in favor of IgA (35), and that D1–D2 and D1–D3 bind pLgA, but not mIgA nor IgG (38), explaining the predominance of SlgA in secretions. As far as pLgA is concerned, the integrity of D2 and D3 is necessary for a covalent interaction to take place (28), yet without establishing close contact with the pLgA backbone (39, 28). This argues in favor of the “zipper effect” model initiated by the interaction of D1 with one Ca chain of pLgA, progression to the J chain, and establishment of the disulfide bond between Cys^{311} of the Ca2 domain of IgA and Cys^{502} of D5 (36). In the J-shaped structure model, D1 remains appropriately exposed to initiate this “zipper” binding interaction.

Several unsolved issues remain, such as the identification of the residues in SC that are involved in the interactions with the J chain and with the residues 402–410 of the Ca3 domain in the Fc fragment of pLgA (40). Likewise, even though the CDR-like motifs of D1 have been extensively described, their ligands on the pLgA molecule remain unidentified. In this respect, based on electron micrographs of human pLgA displaying a partly bent tail-to-tail arrangement (41), one can envisage that the Fc domain spacing is suitable for the binding of at least one SC molecule. Whether the compact conformation of SC opens up upon binding to pLgA is another issue. Moreover, structural differences between human IgA1 and IgA2 exist (22, 23) that are consistent with recent findings showing distinct binding modes of human SC to pLgA1 and pLgA2 (42). Future experiments aimed at unraveling the contribution of SC in terms of the architecture of the whole SlgA molecule are now being planned. Knowledge of the three-dimensional structure of SC represents a mandatory first step toward the resolution of the complete structure of SlgA. If we assume that the model of SC is not that different from the extracellular portion of pLgR, it might also be helpful per se to map binding site(s) to

**FIGURE 9. Best-fit models for SC.** The domains D1–D5 are colored from blue (N terminus) to red (C terminus). A, the ten best-fit models / to X are shown with D1–D3 in the same orientation. Model I is the best-fit model from Fig. 8A.8, stereo view of the best-fit model of Fig. 8A. Oligosaccharides here and in D are shown in light blue, C, electrostatic views of the continuous protein surfaces seen in the D1–D3 and D4–D5 models. The numbering corresponds to the oligosaccharide chains (Fig. 18). Red, acidic; blue, basic. D, the best-fit model of Fig. 8A is rotated by 180° to show the CDR I, II, and III peptides in D1 (in pink) and Cys^{502} in D5 (in yellow).
Solution Structure of Secretory Component

permit epithelial drug targeting through specific interaction with plgR itself (43, 44, 45).

Acknowledgments—We gratefully thank R. E. Saunders for computational support, J. Gor for ultracentrifugation support, Dr. S. Finet for x-ray support, and Drs. S. M. King and R. K. Heenan for neutron support.

REFERENCES

1. Brandtzaeg, P. (1995) APMIS 103, 1–19
2. Lamm, M. E. (1997) Annu. Rev. Microbiol. 51, 311–340
3. Nagler-Anderson, C. (2001) Nat. Rev. Immunol. 1, 59–67
4. Johansen, F. E., Braathen, R., and Brandtzaeg, P. (2001) Annu. Assoc. Immunol. 167, 5185–5192
5. Mach, J. P. (1970) Nature 228, 1278–1282
6. Crottet, P., and Corthey, B. (1998) J. Immunol. 161, 5445–5453
7. Phalipon, A., Cardona, A., Kraehenbuhl, J. P., Edelman, L., Sansonetti, P. J., and Corthey, B. (2002) Immunity 17, 107–115
8. Boren, T., Falk, P., Roth, K. A., Larson, G., and Normark, S. (1993) Science 262, 1892–1895
9. Dallas, S. D., and Rolfe, R. D. (1998) J. Med. Microbiol. 47, 879–888
10. Lu, L., Lamm, M. E., Li, H., Corthey, B., and Zhang, I. R. (2003) J. Biol. Chem. 278, 48178–48187
11. Elm, C., Braathen, R., Bergmann, S., Frank, R., Vaerman, J. P., Kaetzel, C. S., Chhatwal, G. S., Johansen, F. E., and Hammerschmidt, S. (2004) J. Biol. Chem. 279, 6296–6304
12. Perrier, C., Sprenger, N., and Corthey, B. (2006) J. Biol. Chem. 281, 14280–14287
13. Hughes, G. J., Frutiger, S., Savoy, L. A., Reason, A. J., Morris, H. R., and Jaton, J. C. (1995) FEBS Lett. 410, 443–446
14. Coyne, R., Siebrecht, M., Peitsch, M. C., and Casanova, J. E. (1994) J. Biol. Chem. 269, 31620–31625
15. Corthey, B., Kaufmann, M., Phalipon, A., Peitsch, M., Neutra, M., and Kraehenbuhl, J. P. (1996) J. Biol. Chem. 271, 33670–33677
16. Crottet, P., Cottet, S., and Corthey, B. (1999) Biochem. J. 341, 299–306
17. Hamburger, A. E., West, A. P., and Bjorkman, P. J. (2004) Structure 12, 1925–1935
18. Perkins, S. I., Ashton, A. W., Boehm, M. K., and Chamberlain, D. (1998) Int. J. Mol. Macromol. 22, 1–16
19. Perkins, S. J. (1986) Eur. J. Biochem. 157, 169–180
20. Narayanan, T., Diat, O., and Bosecke, P. (2001) Nucl. Instrum. Methods Phys. Res. A 467–468, 1005–1009
21. Heenan, R. K., Penfold, J., and King, S. M. (1997) J. Appl. Crystallogr. 30, 1140–1147
22. Boehm, M. K., Wool, J. M., Kerr, M. A., and Perkins, S. J. (1999) J. Mol. Biol. 286, 1421–1447
23. Furtado, P. B., Whitty, P. W., Robertson, A., Eaton, J. T., Almogren, A., Kerr, M. A., Wool, J. M., and Perkins, S. I. (2004) J. Mol. Biol. 338, 921–941
24. Marti-Renom, M. A., Stuart, A., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000) Annu. Rev. Biophys. Biomol. Struct. 29, 291–325
25. Boehm, M. K., and Perkins, S. J. (2000) FEBS Lett. 475, 11–16
26. Perkins, S. J. (2001) Biophys. Chem. 93, 129–139
27. Sun, Z., Almogren, A., Furtado, P. B., Chowdhury, B., Kerr, M. A., and Perkins, S. J. (2005) J. Mol. Biol. 353, 155–173
28. Crottet, P., and Corthey, B. (1999) J. Biol. Chem. 274, 31456–31462
29. Perkins, S. J. (1988) Biochem. J. 254, 313–327
30. McRorie, D. K., and Voelker, P. J. (1993) Self-associating Systems in the Analytical Ultracentrifuge, pp. 9–12, Beckman Instruments, Palo Alto, CA
31. Sternberg, M. J. E. (1996) Protein Structure Prediction, pp. 111–180, IRL Press, Oxford
32. Rost, B., and Liu, J. (2003) Nucleic Acids Res. 31, 3300–3304
33. Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. (2001) Nucleic Acids Res. 29, 2994–3005
34. Hu, Y., Sun, Z., Eaton, J. T., Bouloux, P. M. G., and Perkins, S. J. (2005) J. Mol. Biol. 350, 553–570
35. Bakos, M. A., Widen, S. G., and Goldblum, R. M. (1994) Mol. Immunol. 31, 165–168
36. Fallgreen-Gebauer, E., Gebauer, W., Bastian, A., Kratzin, H. D., Eiffert, H., Zimmermann, B., Karas, M., and Hilschmann, N. (1993) Biol. Chem. Hoppe Seyler, 374, 1023–1028
37. Lüllau, E., Heyse, S., Vogel, H., Marison, L. von Stockkar, U., Kraehenbuhl, J. P., and Corthey, B. (1996) J. Biol. Chem. 271, 16300–16309
38. Bakos, M. A., Kurosky, A., Woodard, C. S., Denney, R. M., and Goldblum, R. M. (1991) J. Immunol. 146, 162–168
39. Bakos, M. A., Kurosky, A., and Goldblum, R. M. (1991) J. Immunol. 147, 3419–3426
40. Hexham, J. M., White, K. D., Carayannopoulos, L. N., Mandecki, W., Brietsette, R., Yang, Y. S., and Capra, J. D. (1999) J. Exp. Med. 189, 747–751
41. Nezlin, R. (1998) The Immunoglobulins: Structure and Function, pp. 4–62, Academic Press, New York
42. Almogren, A., Senior, B. W., and Kerr, M. A. (2006) Immunology 120, 273–280
43. Ferkol, T., Perales, J. C., Eckman, E., Kaetzel, C. S., Hanson, R. W., and Davis, P. B. (1995) J. Clin. Invest. 95, 493–502
44. White, K. D., and Capra, J. D. (2002) J. Exp. Med. 196, 551–555
45. Braathen, R., Sandvik, Å., Berntzen, G., Hammerschmidt, S., Fleckenstein, B., Sandle, I., Brandtzaeg, P., Johansen, F. E., and Lavrak, V. (2006) J. Biol. Chem. 281, 7075–7081