Eur Biophys J (2010) 39:469–480
DOI 10.1007/s00249-009-0432-2

ORIGINAL PAPER

Solution structure of the major (Spy0128) and minor (Spy0125 and Spy0130) pili subunits from *Streptococcus pyogenes*

Alexandra S. Solovyova · Jonathan A. Pointon · Paul R. Race · Wendy D. Smith · Michael A. Kehoe · Mark J. Banfield

Received: 14 November 2008 / Revised: 17 February 2009 / Accepted: 23 February 2009 / Published online: 17 March 2009

Abstract Adhesion of the serotype M1 *Streptococcus pyogenes* strain SF370 to human tonsil explants and cultured keratinocytes requires extended polymeric surface structures called pili. In this important human pathogen, pili are assembled from three protein subunits: Spy0125, Spy0128 and Spy0130 through the action of sortase enzymes. For this study, the structural properties of these pili proteins have been investigated in solution. Spy0125 and Spy0128 display characteristics of globular, folded proteins. Circular dichroism suggests a largely β-sheet composition for Spy0128 and Spy0125; Spy0130 appears to contain little secondary structure. Each of the proteins adopts a monodisperse, monomeric state in solution as assessed by analytical ultracentrifugation. Further, small-angle X-ray scattering curves for Spy0125, Spy0128 and Spy0130 suggest each protein adopts an elongated shape, likely comprised of two domains, with similar maximal dimensions. Based on the scattering data, dummy atom models of each of the pili subunits have been reconstructed ab initio. This study provides the first insights into the structure of *Streptococcus pyogenes* minor pili subunits, and possible implications for protein function are discussed.

Keywords Pili subunits · Circular dichroism · Analytical ultracentrifugation · Small-angle X-ray scattering · Dummy atom model · Structural disorder

A. S. Solovyova, J. A. Pointon and P. R. Race have contributed equally to this work.

AUC&HYDRO 2008—Contributions from 17th International Symposium on Analytical Ultracentrifugation and Hydrodynamics, Newcastle, UK, 11–12 September 2008.

A. S. Solovyova (✉) · J. A. Pointon · P. R. Race · W. D. Smith · M. A. Kehoe
Faculty of Medical Sciences, Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK
e-mail: alexandra.solovyova@newcastle.ac.uk

Present Address:
M. J. Banfield
Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

Introduction

Streptococcus pyogenes (also called Group A *Streptococcus*, or GAS) is an important Gram-positive bacterial pathogen that causes a wide variety of diseases in humans. These range in severity from mild self-limiting infections of the pharynx or skin, to more serious and potentially life-threatening infections such as severe pharyngitis, tonsillitis, streptococcal toxic-shock syndrome and necrotising fasciitis. A first critical step in the pathogenesis of many bacterial infections is specific adhesion to a particular host surface. A number of GAS cell-surface molecules have been implicated as specific adhesins by in vitro studies employing cultured mammalian cell lines, such as Hep-2 and A549 cells [reviewed in (Courtney et al. 2002)]. However, the extent to which these long-established cell-lines represent natural human host epithelial surfaces has been questioned (Abbot et al. 2007). Recent studies, using intact human tonsil explants and primary human keratinocytes, revealed that bacterial cell-surface structures called pili (also known as fimbriae) are required for efficient adhesion of the serotype M1 GAS strain SF370 (henceforth "GAS") to these clinically relevant host tissues (Abbot et al. 2007). Similar results were obtained from assessing the binding of GAS pilus-negative mutants to Detroit-562 pharyngeal cells (Edwards et al. 2008).

Pili are narrow (2–10 nm) polymeric protein structures that can extend for distances of a micron or more from many bacterial surfaces (Wu and Fives-Taylor 2001). Pili produced by Gram-negative bacteria have been studied for many decades, leading to detailed accounts of their structure and biogenesis [for reviews, see (Proft and Baker 2008; Wu and Fives-Taylor 2001)]. Studies of Gram-positive bacterial pili have lagged well behind their Gram-negative counterparts, partly because they have only recently been detected on the surfaces of an increasing number of important Gram-positive pathogens, including *Corynebacterium diphtheriae*, *Streptococcus agalactiae* (the Group B *Streptococcus*), *Streptococcus pneumoniae* and GAS (Proft and Baker 2008; Scott and Zähner 2006). Over the past 5 years intensive study has revealed that both the structures and the mechanisms of pili biogenesis in Gram-positive pathogens are markedly different from those found in Gram-negative bacteria. For example, pili in Gram-negative bacteria comprise subunits (often called pilins) held together through strong non-covalent interactions; in Gram-positive bacteria pilus subunits are linked covalently through isopeptide bonds to produce, in effect, a single large polymer extending from the bacterial cell-surface. These covalent linkages are formed by specialised membrane-associated transpeptidases called sortases, which recognise a specific cell-surface "sorting" motif, close to the C-terminal end of the subunit proteins and link this to free amino groups of either specific lysine side chains in other proteins, or cell-wall peptidoglycan [reviewed in (Marraffini et al. 2006)].

All Gram-positive bacterial pili studied to date are composed of a single major subunit plus either one, or more commonly two minor subunits (Mandlik et al. 2008b; Proft and Baker 2008; Scott and Zähner 2006). The major subunit forms the repeating unit of the pilus shaft and is therefore found in the highest abundance. Immuno-gold labelling has visualised the minor subunits either at the pilus tip, the pilus base and/or at occasional intervals along the pilus shaft (Dramsi et al. 2006; Mandlik et al. 2008b). The exact stoichiometry of the subunits has not been determined and is likely to vary between pili from different species. Functions for the minor subunits range from being essential for specific adhesion, to acting as a "linker" between the base of the polymerised pili and the bacterial cell wall (Mandlik et al. 2008a). In GAS, the major subunit is the product of *spy0128* (Spy0128, a 37-kDa protein); the two minor subunits are the products of *spy0125* (Spy0125, an 86-kDa protein) and *spy0130* (Spy0130, a 24-kDa protein). These three proteins only share very limited sequence identity (21% 125 vs. 128, 15% 128 vs. 130 and 12% 125 vs. 130). There is some evidence to suggest that the C-terminal domains of Spy0128 and Spy0125 may be related (Kang et al. 2007); beyond this no sequence or structural similarities are readily predictable.

Whilst the main players in the assembly of pili in GAS have been identified, the detailed mechanisms by which these adhesive macromolecular structures are formed on the bacterial cell-surface remains poorly understood. Recently, the crystal structure of the major GAS pilus subunit (Spy0128) has been determined (Kang et al. 2007). To date, no structural information on the minor pilus subunits is available (Spy0125 and Spy0130). In the absence of high resolution structural information, solution-based techniques [such as small-angle X-ray scattering (SAXS)] can provide valuable insights into protein conformation. One of the most widely used approaches for the analysis of SAXS scattering curves is ab initio shape restoration [for review see (Koch et al. 2003)]. This can provide a low resolution molecular envelope describing the overall shape of the macromolecule in solution and can compliment atomic analysis from either X-ray crystallography or nuclear magnetic resonance. Additionally, SAXS analysis is especially relevant for characterisation of proteins containing flexible and/or structurally disordered regions (Bernadó et al. 2007; Petoukhov et al. 2002).

In this study the low resolution solution structure of the three proteins that comprise pili in GAS (Spy0125, Spy0128 and Spy0130) have been determined in the form of dummy atom models (DAMs) restored from SAXS data. Combined with circular dichroism (CD) spectroscopy, analytical ultracentrifugation (AUC), using the sedimentation velocity method and bioinformatic analysis, these data reveal that Spy0125, Spy0128 and Spy0130 are elongated monomers in solution, and each is likely comprised of two domains. These results are important for informing on-going efforts targeting high resolution analysis of pili subunits and contributes to understanding of the structural organisation of Gram-positive pili at the molecular level.

Materials and methods

Gene cloning

The DNA sequences encoding *spy0125*, *spy0128* and *spy0130* were amplified from M1 *Streptococcus pyogenes* strain SF370 genomic DNA using the primers 5′-TTACA TACATATGAAGACTGTTTTTGGTTTAGTAG-3′ (Forward) and 5′-AAGTTCTCGAGTTAAACTCCTGTAGG AACAAC-3′ (Reverse) for *spy0125*, 5′-TTACATACA TATGGAGACTGTTGTAAACGGAGC-3′ (Forward) and 5′-AAGTTCTCGAGTTATACTCCTGTTGGCACTTC-3′ (Reverse) for *spy0128* and 5′-TTACATACATATGGA GAACCTCACTGCAAGCATT-3′ (Forward) and 5′-AA

GTTCTCGAGTTAACCAGTTGATGGCAAAATACC-3′ (Reverse) for *spy0130*. These primers introduced appropriate restriction sites for subsequent cloning (*Nde*1 and *Xho*1 sites, underlined). The primers were designed to remove the predicted signal sequences from the N-terminus of each protein, and the sequences to the C-terminal side of the sorting signal. PCR products were digested with the appropriate enzymes and cloned directly into pET28a (cut with the same enzymes, Novagen). The resulting constructs were verified by DNA sequencing and transformed into *Escherichia coli* BL21 (DE3) cells for protein expression.

Protein expression and purification

Cultures of BL21 (DE3) cells harbouring the *spy0125*:pET28a, *spy0128*:pET28a or *spy0130*:pET28a plasmids were grown in 1 l of Luria–Bertani (LB) media supplemented with kanamycin (to a final concentration of 50 μg/ml) at 37°C with shaking. When the cell density reached an $A_{600} \sim 0.4$–0.6, protein expression was induced by the addition of isopropyl 1-thio-β-D-galactopyranoside to a final concentration of 1 mM. Cells were grown for an additional 3–4 h before harvesting by centrifugation. Cell pellets were resuspended in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 150 mM NaCl, 10 mM imidazole, pH 7.5, supplemented with 5 mM 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride and lysed by sonication. The cell lysate was clarified by centrifugation and the supernatant applied to a pre-equlibrated Ni^{2+}-IMAC column (GE Healthcare). Proteins were eluted with an imidazole gradient. Fractions containing protein (as identified by SDS-PAGE) were pooled and concentrated. The proteins were then injected onto a Hi-Load 16/60 Superdex 75 column (GE Healthcare) for Spy0128 and Spy0130, or a Hi-Load 16/60 Superdex 200 (GE Healthcare) for Spy0125, pre-equilibrated in 20 mM trishydroxymethylaminomethane–HCl (Tris), 150 mM NaCl, pH 7.5. Fractions containing purified protein (as identified by SDS-PAGE) were pooled and concentrated to 10 mg/ml by ultrafiltration. Protein yields were ~20 mg for Spy0125, ~100 mg for Spy0128 and ~15 mg for Spy0130 l^{-1} bacterial cell culture. Initial characterisation revealed all recombinant proteins to be monodisperse [dynamic light scattering (DLS)], single species (SDS-PAGE, DLS, size-exclusion chromatography) and of > 95% purity.

CD spectroscopy

Circular dichroism spectra of Spy0125, Spy0128 and Spy0130 were collected on a JASCO-810 spectropolarimeter (Jasco, Tokyo, Japan, fitted with a Peltier temperature controller) from 190 to 250 nm at 25°C, using a 0.2 mm pathlength cuvette. Proteins were buffered in 20 mM sodium phosphate, pH 7.5 for Spy0125 and Spy0128; 20 mM Tris, 150 mM NaCl, pH 7.5 was used for Spy0130. The concentration of each stock protein solution was determined by absorbance at 280 nm using the extinction coefficients 78,745 M^{-1} cm^{-1} for Spy0125, 26,360 for Spy0128 M^{-1} cm^{-1} and 11,920 M^{-1} cm^{-1} for Spy0130. Based on these measurements, proteins were diluted to 0.2 mg ml^{-1} immediately prior to data collection. Final spectra are averages of five repeat scans, with appropriate protein-free buffer spectra subtracted. The data were plotted without smoothing using Excel. All data were truncated to 194 nm on the wavelength axis corresponding to the value where the high-tension voltage rose above 600 V.

Analytical ultracentrifugation

Sedimentation velocity (SV) experiments were carried out in a Beckman Coulter (Palo Alto, CA, USA) ProteomeLab XL-I analytical ultracentrifuge using both absorbance at 280 nm and interference optics. All AUC runs were carried out at a rotation speed of 48,000 rpm and an experimental temperature 4°C using an 8-hole AnTi50 rotor and double-sector aluminium-epon centerpieces. The sample volume was 400 μl for the SV experiments, and the sample concentrations ranged between 0.15 and 2 mg ml^{-1} (usually up to seven sample dilutions were used). The partial specific volumes ($\bar{v}$) for the proteins were calculated from the protein amino acid sequence, using the program SEDNTERP (Laue et al. 1992) and extrapolated to the experimental temperature following the method described in (Durchschlag 1986). The density and viscosity of the buffer (20 mM Tris, 150 mM NaCl, pH 7.5) at the experimental temperature was also calculated using SEDNTERP.

The distributions of sedimenting material were modelled as a distribution of Lamm equation solutions (Schuck 1998) where the measured boundary a(r,t) was modelled as an integral over differential concentration distributions c(s):

$$a(r,t) = \int c(s)\chi(s,D,r,t)ds + \varphi \quad (1)$$

where φ is the noise component, r is the distance from the centre of rotation and t is time. The expression $\chi(s, D, r, t)$ denotes the solution of the Lamm equation (Lamm 1929) for a single species by finite element methods (Demeler and Saber 1998). Equation 1 is solved numerically by discretisation into a grid of 200 sedimentation coefficients for both absorbance and interference data and the best-fit concentrations for each plausible species are calculated according to a linear least squares fit [as implemented in

SEDFIT (http://www.analyticalultracentrifugation.com)]. The sedimentation velocity profiles were fitted using a maximum entropy regularisation parameter of $P = 0.95$. The data were fitted for the position of the meniscus, value of the friction coefficient (f/f_0) and both the time- and radial-independent noise. The weight average sedimentation coefficient was calculated by integrating the differential sedimentation coefficient distribution (Schuck 2003):

$$s_w = \frac{\int c(s)ds}{\int cds} \tag{2}$$

Sedimentation coefficients were extrapolated to 0 concentration and converted to standard conditions, i.e. those that would be measured at 20°C in water. The diffusion coefficient (D) corresponding to each sedimentation coefficient value was estimated from a weight-average frictional ratio $(f/f_0)_w$:

$$D(s) = \frac{\sqrt{2}}{18\pi} kTs^{-1/2}\left[\eta(f/f_0)_w\right]^{-3/2}\left[(1-\bar{v}\rho)/\bar{v})\right]^{-1/2} \tag{3}$$

Conversion to a molecular mass distribution from s and D expressed above can be made using the Svedberg equation:

$$M = \frac{sRT}{D(1-\bar{v}\rho)} \tag{4}$$

The integration of the mass distribution $c(M)$ was calculated in a similar manner to Eq. 2 to determine the weight average molecular mass of solute.

The two-dimensional "size-and-shape" distribution model [c(s,f/f_0)] was applied to determine the value of molecular mass from sedimentation velocity data (Brown and Schuck 2006). In this model a differential distribution of sedimentation coefficients and frictional ratios (f_r, i.e. f/f_0) is defined as:

$$a(r,t) = \iint c(s,f_r)\chi(s,D(s,f/f_0),r,t)dsdf_r \tag{5}$$

where all symbols are the same as in Eq. 1. In this model, the sedimentation coefficient and friction ratio (i.e. mass) for sedimenting species are fitted as independent variables. The numerical solutions of Eq. 4 were obtained in a discrete grid of 50 sedimentation coefficients and 10 values of f/f_0.

Small-angle X-ray scattering

X-ray scattering experiments on solutions of Spy0125 and Spy0128 were carried out at the Synchrotron Radiation Source, beamline 2.1 (Daresbury, UK) with beam currents of between 80 and 170 mA, an electron energy of 2 GeV and a wavelength of 1.54 Å. Two camera lengths were used in the experiments: 1 m to cover a momentum transfer range of $0.03 < s < 0.6$ Å^{-1} and 4 m for $0.015 < s < 0.2$ Å^{-1}, where $s = (4\pi\sin\theta)/\lambda$ and 2θ is the scattering angle. The detector positioned at 4 m from the sample was calibrated using a sample of wet rat tail collagen; while silver behenate was used for calibration of the detector positioned at the distance of 1 m. Sample concentrations used were 10 and 5 mg ml^{-1} for the short and long camera lengths respectively. The measurements were carried out at 4°C. Experimental data were collected and averaged as 5 × 60 s frames and 10 × 60 s frames for sample-detector distances of 1 and 4 m, respectively. The data were normalised for buffer scattering and detector response using XOTOKO (Boulin et al. 1986). To check for radiation damage and aggregation during SAXS experiments data in the first and last frame were compared.

SAXS data for Spy0130 were collected on the X33 camera at the EMBL Hamburg Outstation on the storage ring DORIS III of DESY. Scattering curves were recorded at a wavelength of 1.5 Å at a sample-detector distance 2.7 m covering the momentum transfer range $0.05 < s < 0.5$ Å^{-1}. Sample concentrations used were 10 and 5 mg ml^{-1}. Data were normalised to the buffer scattering and scaled for concentration using PRIMUS (Konarev et al. 2003). As for Spy0125 and Spy0128, data in the first and last frame were compared to check for radiation damage.

The radius of gyration was determined using the Guinier approximation:

$$\ln \mathrm{I}(s) = \ln \mathrm{I}(0) - \frac{1}{3}R_g^2 s^2 \tag{6}$$

as implemented in the program PRIMUS. The linear least-squares fit of the experimental data was calculated in the range of $R_g s \leq 1$.

The distance distribution function [P(r)] for each protein was calculated with GNOM (Semenyuk and Svergun 1991). Low-resolution molecular shapes were restored as DAMs using DAMMIN (Svergun 1999). The resolution of the dummy models was determined as $2\pi/s_{max}$ where s_{max} is the reciprocal spacing of the highest resolution data point used in the restoration process. DAMs were generated 20 times from single scattering data sets and corresponding $P(r)$ functions used to establish a degree of similarity for the models which were then averaged to find the most common shape. All 20 models were superimposed and averaged using the DAMAVER package (http://www.embl-hamburg.de/ExternalInfo/Research/Sax/damaver.html). Low- and high-resolution structures were superimposed using the program SUBCOMB20 (Kozin and Svergun 2001).

Hydrodynamic calculations

The value of the sedimentation coefficient for an anhydrous monomer was estimated based on the amino acid sequence of the proteins (i.e. calculated mass of the monomer) assuming a spherical shape with (Lebowitz et al. 2002):

$$s = 0.012\frac{M^{2/3}(1-\bar{v}\rho)}{\bar{v}^{-1/3}} \quad (7)$$

where M is the mass of monomer, ρ is the solvent density and $\bar{v}$ is the partial specific volume of the protein. The protein hydration number (δ) of 0.4 g_w/g_p was assumed, which gives a decrease in the value of anhydrous sedimentation coefficient by:

$$s_{hydr} = s_{anh}\left(\frac{\bar{v}_2 + \delta\bar{v}_1}{\bar{v}_2}\right)^{1/3} \quad (8)$$

where $\bar{v}_1$ and $\bar{v}_2$are the partial specific volumes of solvent and protein, respectively.

HYDROPRO (Garcia de la Torre et al. 2000) was used to calculate sedimentation coefficients for DAMs generated from SAXS data, following the approach described by Ackerman et al. (2003).

Structural disorder prediction

Protein sequences were submitted to the PONDR web engine (www.pondr.com) using the neural network predictor VL-XT (Romero et al. 2001). Access to PONDR was provided by Molecular Kinetics (Indianapolis, IN, USA).

Results and discussion

Secondary structure of pili subunits

The far ultra-violet (UV) CD spectra of Spy0125, Spy0128 and Spy0130 are shown in Fig. 1. The spectrum of Spy0125 displays a non-classical β_{II} spectrum with a minimum at 208 nm and a near zero value 194 nm (Seerama and Woody 2003). The shape of the curve is reminiscent of porcine elastase, but with a shift in the position of the minima. Porcine elastase is an example of a $\beta_{II\text{-type}}$ protein, which has sufficient amounts of polyproline II (PPII) structure that it obscures the weaker β CD signal. The spectrum of Spy0128 displays a classical β_I-type sheet protein, with a single minimum at 215 nm and a maximum at 202 nm. This is consistent with the published crystal structure (Kang et al. 2007), which reveals that Spy0128 comprises two discrete all-β domains.

The CD data for Spy0130 suggests the protein contains little secondary structure. Despite this, the spectrum shows no evidence of a negative shoulder near 222 nm, or an observable minimum at 203 nm, both characteristic features of intrinsic disorder (Receveur-Brechot et al. 2006). Although unusual, it is established that globular, folded proteins can display CD spectra that lack definitive secondary structural features [e.g. SH3 domains (Viguera et al. 1994), chymotrypsin (Volini and Tobias 1969) and anthrax protective antigen (Chalton et al. 2006)]. This is especially prevalent for all-β-type proteins, for which CD spectra can vary significantly (unlike α-containing proteins where the signal is quite distinct). Spy0130 is predicted bioinformatically to be a predominantly β-protein, so in this case CD may not be a good indicator of the folded/unfolded state. DLS and size exclusion chromatography indicate a folded state for this protein (data not shown), albeit an extended one. Further, the f/f_0 ratio of 1.38, the experimentally determined hydrodynamic radius of 29 Å (as determined by integration of the two-dimensional "size-and-shape" distribution (Brown and Schuck 2006)) and the observation that this calculated value of the R_S for this molecular mass lies towards the "folded" end of the range of expected values [21 Å (fully folded) to 39 Å (for fully disordered) (Uversky 2002)], suggests that whilst the protein is elongated, it retains some form of compactness or globularity. Finally, and for illustration, the C-terminal domain of the *Enteropathogenic Escherichia coli* translocated intimin receptor, a protein of mass comparable with that of Spy0130 (19.7 kDa) but characterised as intrinsically disordered, sediments at 1.2–1.4 S (Race et al. 2007) while Spy0130 sediments at 1.76 S, again indicating the globularity of the protein.

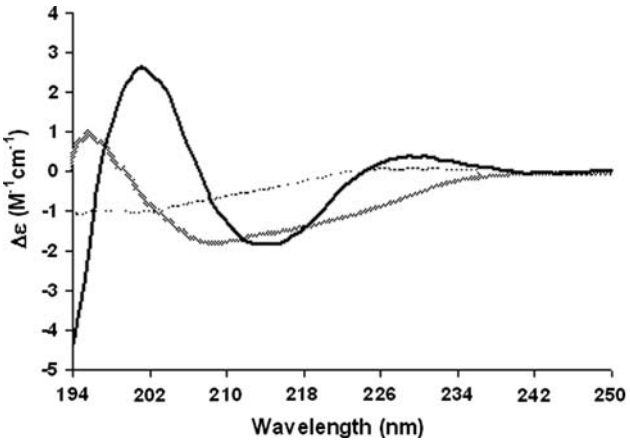


Fig. 1 Far UV CD spectra of Spy0125 (*hashed line*), Spy0128 (*solid line*) and Spy0130 (*dotted line*). All experiments were performed as described in "Materials and methods", with a description of the observed features provided in the text

Stoichiometry of isolated pili components in solution

The heterogeneity and oligomeric state of Spy0125, Spy0128 and Spy0130 in solution were investigated by AUC using the continuous $c(s)$ distribution model (Fig. 2, panel B). An example showing the fit of the data is presented in Fig. 2 (panel A) for Spy0125, Spy0128 and Spy0130; the root mean square deviation of the fit is given in Table 1. As seen in Fig. 2 (panel A) the residuals of the data fit are randomly distributed and comparable with the noise of interference data. Resulting from general c(s) size-distribution model one dominant species was evident as a single peak at 4.30 S, 2.63 S and 1.76 S for Spy0125, Spy0128 and Spy0130, respectively. The integration of these peaks gave a weight average sedimentation coefficient (extrapolated to zero concentration) for each protein as shown in Table 1. The experimentally determined sedimentation coefficients for Spy0125 and Spy0130 are significantly lower than would be calculated from a typically hydrated ($\delta = 0.4\ g_w/g_p$) sphere of equivalent molecular mass (5.65 and 2.32 S, respectively), implying substantial elongation of Spy0125 and Spy0130 in solution. The value of sedimentation coefficient of Spy0128 based on its high-resolution structure [calculated using HYDROPRO (Garcia de la Torre et al. 2000)] is 2.54 S (see Table 1), consistent with that determined experimentally (2.63 S).

Non-linearly fitted values of the frictional ratio comprise two factors: the geometrical shape and hydration expansion. In order to reveal the shape asymmetry of Spy0125,

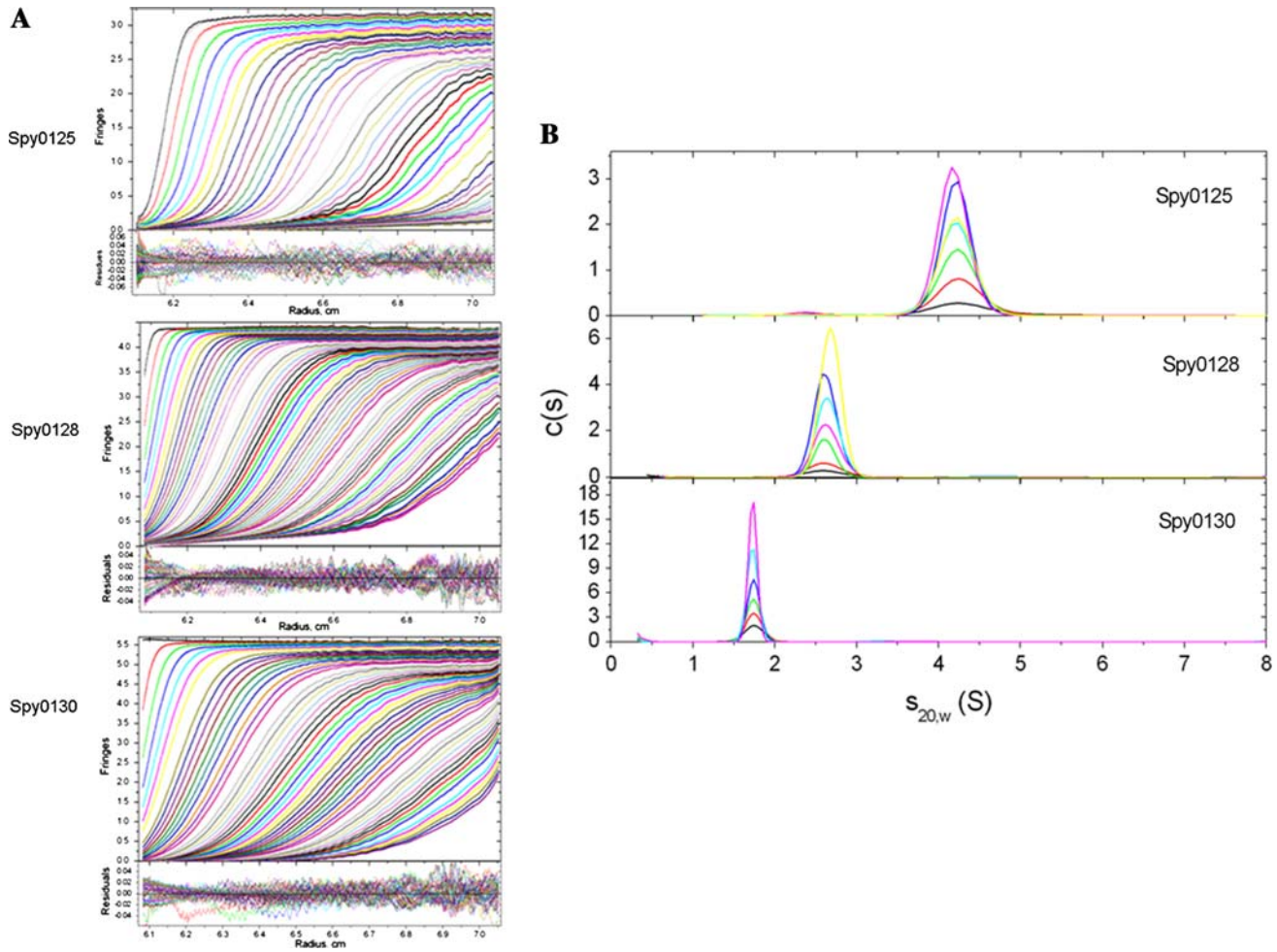


Fig. 2 Sedimentation velocity study of pili subunits. *Panel* **a** a representative SV interference data fit for Spy0125 (c = 0.76 mg ml^{-1}), Spy0128 (c = 1.34 mg ml^{-1}), Spy0130 (c = 2 mg ml^{-1}). *Panel* **b** Size distribution c(s). *Top panel* Spy0125, the concentrations examined were: 0.15 mg ml^{-1} (*black line*), 0.31 mg ml^{-1} (*red line*), 0.51 mg ml^{-1} (*green line*), 0.76 mg ml^{-1} (*blue line*), 0.65 mg ml^{-1} (*cyan line*), 0.87 mg ml^{-1} (*magenta line*), 0.68 mg ml^{-1} (*yellow line*); *Middle panel* Spy0128, the concentrations examined were: 0.15 mg ml^{-1} (*black line*), 0.26 mg ml^{-1} (*red line*), 0.43 mg ml^{-1} (*green line*), 1.34 mg ml^{-1} (*blue line*), 1.01 mg ml^{-1} (*cyan line*), 0.72 mg ml^{-1} (*magenta line*), 1.65 mg ml^{-1} (*yellow line*); *Bottom panel* Spy0130, the concentrations examined were: 0.32 mg ml^{-1} (*black line*), 0.61 mg ml^{-1} (*red line*), 0.91 mg ml^{-1} (*green line*), 1.23 mg ml^{-1} (*blue line*), 1.57 mg ml^{-1} (*cyan line*), 2.04 mg ml^{-1} (*magenta line*)

Table 1 Sedimentation velocity results

Protein	Calculated parameters		SV experimental data[d]				
	M (kDa)[a]	s (S)	RMSD[e]	f/f_0^{shape}	$s_{20,w}$ (S)	Mass (kDa)	R_S, Å[f]
Spy0125	78.43	5.65[b]	0.0161	1.41	4.30 ± 0.035	81.81 ± 1.6	46.96 ± 0.74
Spy0128	34.19	2.54[c]	0.0109	1.23	2.63 ± 0.019	34.48 ± 1.35	32.34 ± 0.50
Spy0130	20.23	2.32[b]	0.0121	1.45	1.76 ± 0.003	20.47 ± 0.35	29.01 ± 0.47

[a] Calculated from amino acid sequence

[b] Calculated on the basis of hydrated sphere approximation

[c] Calculated by HYDROPRO on the basis of atomic coordinates

[d] Integrated interference data

[e] RMSD (root mean square deviation) is given for the highest sample concentration

[f] Determined using two-dimensional "size-and-shape" distribution

Spy0128 and Spy0130, the hydration term [for details see (Lebowitz et al. 2002)] was excluded and the derived values of f/f_0^{shape}, greatly exceeded 1 ($f/f_0 = 1$ for a sphere), indicating elongated particles (Table 1). The highest asymmetry (with regard to a monomeric mass) was observed for Spy0130.

Converting the $c(s)$ distribution to c(M) (as described in Sect. "Materials and methods", Eq. 2–4), followed by

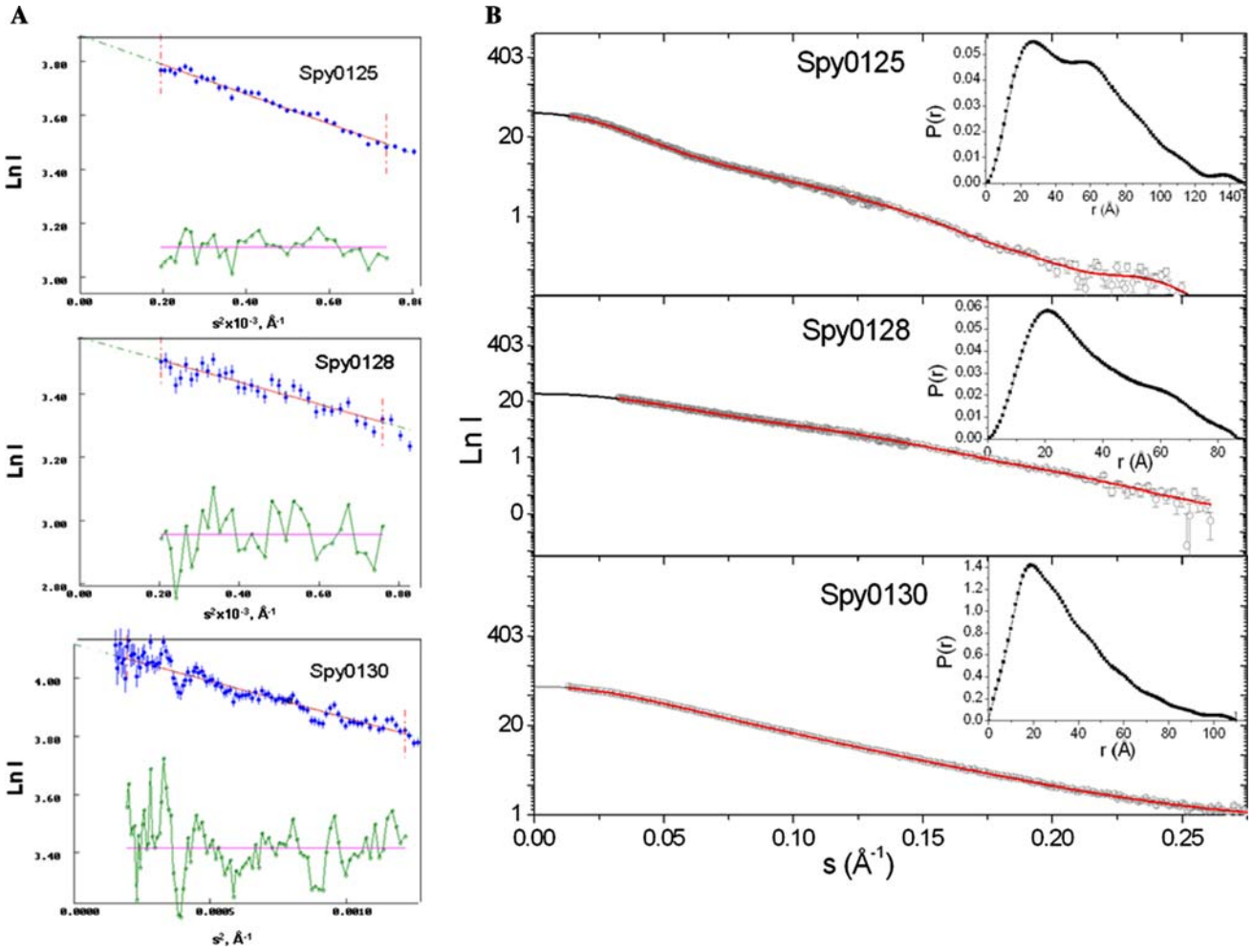


Fig. 3 Small-angle X-ray scattering results. *Panel* **a** Guinier approximation for Spy0125, Spy0128 and Spy0130. *Panel* **b** Scattering curves (*grey dots*), GNOM fit (*black line*) and DAMMIN fit (*red line*) fit for pilus proteins. *Inserts*—the distance distribution function ($P(r)$) derived from each scattering curve

Table 2 Summary of SAXS integral parametes and hydrodynamic calculations of DAM's

Protein	Calculated $V_{anhydrous}$ (nm^3)	AUC $s_{20,w}$ (S)	SAXS						HYDROPRO calculations based on DAM's coordinates					
			Guinier approximation R_g (Å)	*GNOM*		*DAMMIN*								
				R_g (Å)	D_{max} (Å)	R_{DAM} (Å)	V_{hydr} (nm^3)	δ (g_w/g_{prot})	AER (Å)	*s* (S)	V_{hydr} (nm^3)	δ (g_w/ g_{prot})	R_g (Å)	R_S (Å)
Spy0125	95.19	4.30	40.5 ± 0.37	42.99 ± 0.17	148.00	3.75	152.5	0.440	4.0	4.3	122.2	0.207	40.51	42.93
Spy0128	41.46	2.63	32.6 ± 0.94	28.09 ± 0.09	87.00	2.25	56.6	0.267	3.8	2.7	66.48	0.441	28.44	30.14
Spy0130	24.45	1.76	27.7 ± 2.61	28.30 ± 0.01	107.00	2.50	33.57	0.272	3.3	1.8	39.64	0.452	27.02	20.00

peak integration, results in experimentally determined molecular masses of 81.8, 34.5 and 20.5 kDa for Spy0125, Spy0128 and Spy0130. The calculated molecular mass of expressed constructs are 78.4 kDa (Spy0125), 34.2 kDa (Spy0128) and 20.2 kDa (Spy0130), respectively, confirming that each of the proteins are monomeric in solution.

X-ray scattering analysis of Spy0125, Spy0128 and Spy0130

The SAXS scattering curves for Spy0125, Spy0128 and Spy0130 are shown in Fig. 3. Also shown is the analysis of the Guinier region (panel A) and the entire scattering curve (panel B). Each scattering curve is almost featureless, with only that of Spy0125 revealing a slight "feature" within the 0.10–0.13 $Å^{-1}$ range. Such "featureless" contours of scattering curves are indicative of elongated molecular shape (Volkov and Svergun 2003). Further, the presence of an extended shoulder in the $P(r)$ function at distances greater than the radius of gyration (R_g) (Fig. 3, inserts) indicates elongated shapes [indirect Fourier transformation of the reciprocal space scattering data generates the real space function $P(r)$ (Svergun 1992)]. Also, the $P(r)$ for Spy0125 and Spy0128 has a pronounced two-maxima form, which implies a two-domain structure for these proteins. The $P(r)$ function for Spy0130 also displays this feature but to a significantly lesser extent. The radius of gyration (R_g) values obtained from both a linear approximation and indirect Fourier transformation [implemented into program GNOM (Semenyuk and Svergun 1991)] with a particle maximal dimension (D_{max}) for Spy0125, Spy0128 and Spy0130 are listed in Table 2. Interestingly, despite an ~ 15 kDa difference in molecular mass, the R_g values of Spy0128 and Spy0130 are very similar and D_{max} for Spy0130 actually exceeds that for Spy0128.

Pili subunit solution structures modelled ab initio

Low-resolution structures based on DAMs were restored ab initio from the experimental SAXS data with a resolution of 25 Å for all proteins examined using DAMMIN (Svergun 1999). After averaging 20 ab initio constructed models the approximate dimensions of the protein envelopes were: 136 Å × 51 Å × 55 Å; 66 Å × 37 Å ×

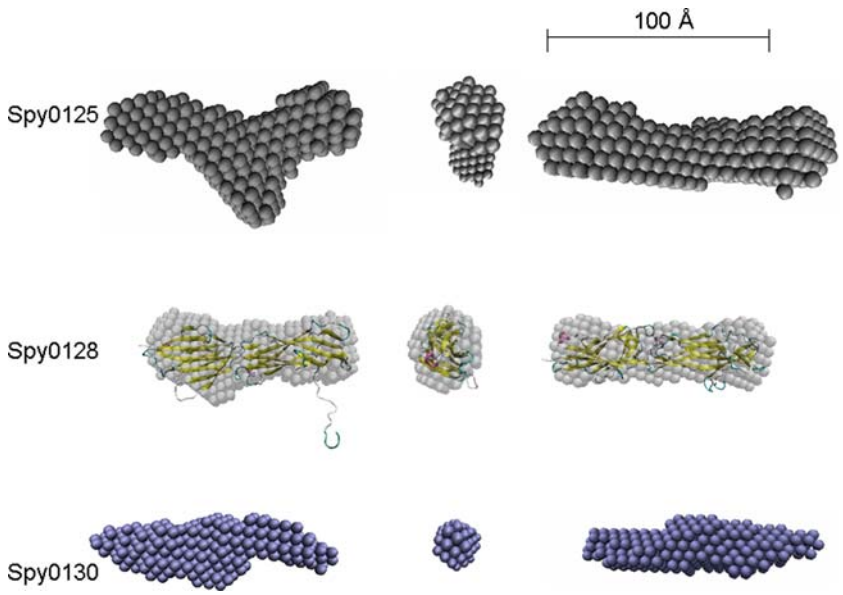


Fig. 4 Averaged dummy atom models of Spy0125 (*top panel*), Spy0128 (*middle panel*, superimposed with the crystal structure (PDB code: 3B2 M)) and Spy0130 (*bottom panel*)

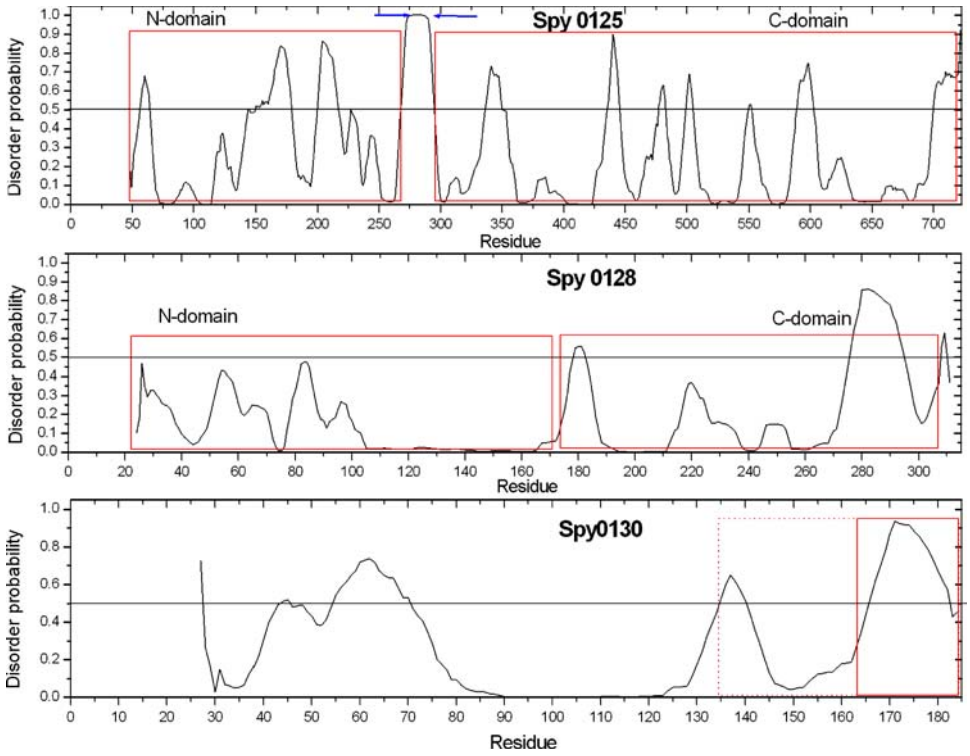


Fig. 5 PONDR (http://www.pondr.com/) prediction of structural disorder based on the sequence of each pilus subunit. Sequences are numbered as originally translated with residue 1 representing the N-terminal Met, but here the N-terminal signal peptide (which is removed during protein secretion), and the C-terminal sequences downstream from the LP × TG motif that are removed during pilus assembly are omitted

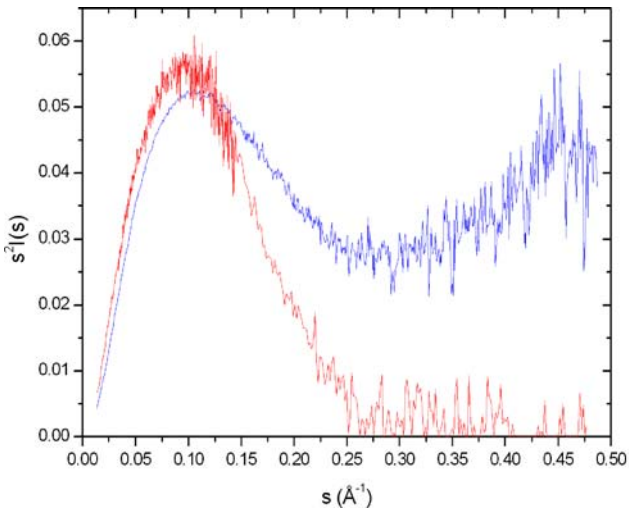


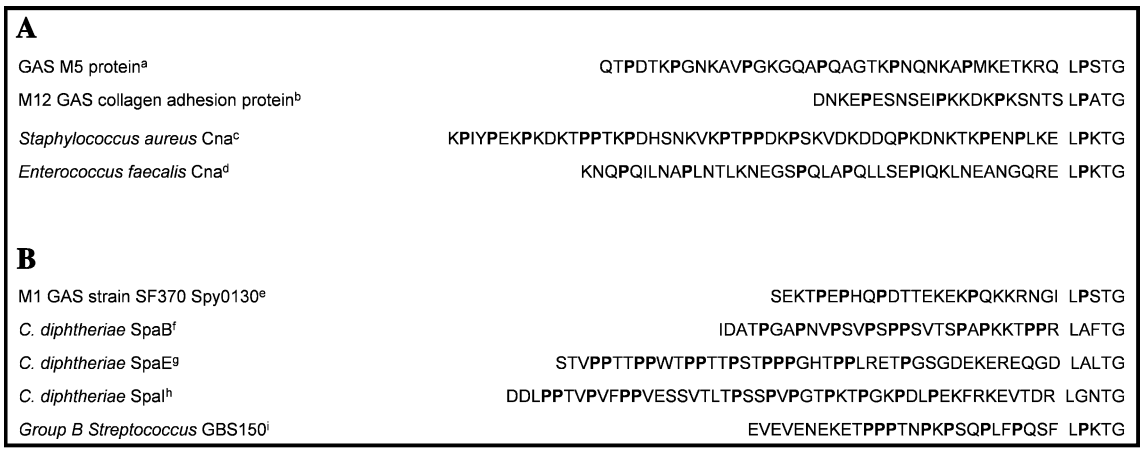
Fig. 6 Kratky plot for Spy0128 (*red line*, as an example of a folded protein) and Spy0130 (*blue line*)

24 Å and 105 Å × 25 Å × 14 Å for Spy0125, Spy0128 and Spy0130, respectively, see Fig. 4. The averaged DAM for Spy0128 is presented with the crystal structure of Spy0128 superimposed using the program SUBCOMB (middle panel of Fig. 4, Kozin and Svergun 2001). To confirm the molecular shapes of the averaged DAMs are consistent with the hydrodynamic data obtained from the SV experiments, the hydrodynamic parameters of these models were calculated with HYDROPRO (Garcia de la Torre et al. 2000). To approximate the amount of bound water for each DAM, the anhydrous protein volume ($V_A = M\bar{v}/N_A$, where M is the protein mass and N_A is Avogadro's number) was subtracted from the Porod volume (the volume of the hydrated protein) of the model. The effective hydration of Spy0125 in HYDROPRO was slightly underestimated compared with the typical hydration of proteins [0.21 $g_{water}/g_{protein}$ against 0.3 $g_{water}/g_{protein}$ (Lebowitz et al. 2002)]; the hydration of Spy0128 and Spy0130 were slightly over-estimated (0.44 $g_{water}/g_{protein}$ and 0.45 $g_{water}/g_{protein}$, respectively). However, these are still within the acceptable range. The sedimentation coefficients calculated with HYDROPRO for all proteins were in good agreement with the experimental SV data (see Table 2).

Two domain organisation of Spy0128 and Spy0125

The crystal structure of Spy0128 comprises distinct N-terminal (residues 18–171) and C-terminal (residues 173–307) domains (Kang et al. 2007). Features suggestive of a two domain structure are observed in the SAXS distance distribution function of this protein (as mentioned above) and the averaged DAM constructed ab initio. Prediction of disordered regions in Spy0128 using PONDR (Romero et al. 2001) indicates two fully folded domains, connected by a short disordered region (residues 178–183).

The significant branching observed in the Spy0125 DAM could be interpreted as indicating a multi-domain structure for this protein (Fig. 4, top panel). Seven days



GenBank accession numbers are as follows: [a]SpyM51682, [b]MGAS2096_Spy0119, [c]M81736, [d]EF1099, [e]SPy_0130, [f]DIP2011, [g]DIP0237, [h]DIP2223, [i]SAG1404

Fig. 7 Examples of pro-rich sequences upstream of the LP × TG-like sorting motifs of various Gram-positive wall-anchored proteins (*panel* **a**) and selected proteins found in Gram-positive pilus gene clusters (*panel* **b**)

after purification, Spy0125 is observed to spontaneously break down into distinct N- and C-terminal domains; the composition of these regions has been determined by N-terminal sequencing and mass spectrometry (J.A. Pointon and M.A. Kehoe, unpublished results). Prediction of disordered regions for Spy0125 using PONDR reveals a region of high disorder probability for residues 268–295, which may suggest a two domain configuration for this protein connected by a flexible linker, akin to Spy0128. The N to C-terminal domain boundary as identified through spontaneous breakdown also locates to this region [residues 280–290, marked by the blue arrows on the disorder plot (Fig. 5, top panel)]. Structural studies of Spy0125 are of significant interest as recent studies in our laboratory (M.A. Kehoe, unpublished) reveal that Spy0125 has a role in cell adhesion, and is located to the tip of pilus. Interestingly, limited sequence homology has been observed between the C-terminal domains of Spy0125 and Spy0128 (Kang et al. 2007). This may indicate that Spy0125 is attached to the Spy0128 shaft in a manner analogous to Spy0128 polymerisation. It should be noted however, that the DAM does not allow unambiguous localisation of the Spy0128 C-terminal domain within the Spy0125 SAXS envelope. A full molecular description of Spy0125 will require crystal structure determination of the intact protein or its domains.

Structurally disordered regions in Spy130

Despite a lack of apparent secondary structure, as observed by CD, the sedimentation behaviour of Spy0130 is suggestive of a compact particle. Encouraged by this, analysis of the SAXS data has enabled the generation of a DAM (i.e. rigid body approach) for this protein. The hydrodynamic parameters of this model (calculated with HYDROPRO) are consistent with the experimentally determined sedimentation coefficient (see Table 2). However, it is clear that this protein does not adopt a fully folded structure. The Kratky plot [a measure of protein compactness (Feigin and Svergun 1987; Vachette et al. 2002), see Fig. 6) derived from the Spy0130 SAXS scattering curve demonstrates the presence of disordered regions. Whilst the Kratky plot for Spy0128 (red line in Fig. 6) presents an inverted parabola (bell shape), indicating a well-folded structure, the Kratky plot for Spy0130 forms only half of this parabola and the tail adopts a form indicative of a "random walk" polymer (Perez et al. 2001). This shape of Kratky plot is characteristic of the presence of both folded and unfolded regions (Longhi et al. 2003). Further, PONDR (Romero et al. 2001) predicts regions of disorder towards the N-terminus and at the C-terminus of this protein (Fig. 5, bottom panel). In fact, the repeating disordered-ordered-disordered region at the C-terminus (residues 134–183, boxed by the red dotted line in Fig. 5) may form a region with a low compactness level. It is important to note that due to the apparent structural flexibility of Spy0130 it is unlikely that this protein will prove tractable for structure determination by X-ray crystallography and SAXS may provide an opportunity for visualising the molecular shape of this protein.

The presence of a repeating disordered-ordered-disordered region at the C-terminus of Spy0130 may be relevant for this protein's function as the cell-wall anchoring pilus subunit (W.D. Smith and M.A. Kehoe, unpublished data). In this context the C-terminus will be embedded in the peptidoglycan layer of the cell-wall, and it has previously been noted that flexibility at this junction may be an advantage in accommodating the intrinsic dynamics of the bacterial wall (Kehoe 1994). Analysis of the protein

sequence in this region of Spy0130 reveals it is rich in the disorder-promoting amino acid proline (see Fig. 7). Further searches of proteins encoded in pilus gene clusters in several Gram-positive bacteria reveal similar pro-rich sequences upstream of the LPXTG-like sorting motif, as do other wall-anchored proteins (Fig. 7). This may therefore represent a conserved signature sequence that could facilitate wall-anchoring.

Conclusions

In this study the hydrodynamic properties and solution structures of pilus subunits from the M1 *Streptococcus pyogenes* strain SF370 have been determined. These results represent the first structural information on any GAS pilus minor subunit (Spy0125 and Spy0130 in this case). This analysis reveals each of the subunits adopts an extended conformation in solution. The low resolution structure of Spy0125 is useful in guiding further structural studies of this protein and suggests the overall conformation of potentially adhesive surfaces. Solution and bioinformatic analysis of Spy0130 reveals a globular protein that displays some regions of disorder, which may be functionally relevant.

Acknowledgments This work was supported, in part, by MRC project grant G0400849 (to MAK) and a Royal Society University Research Fellowship to Mark J. Banfield. Jonathan A. Pointon is supported by a studentship from the Medical Research Council (MRC), UK. Alexandra S. Solovyova and Mark J. Banfield are grateful to Newcastle University for funding. The authors thank the Daresbury-SRS (UK) and the EMBL-Hamburg Outstation (Germany) for the provision of beamtime and the expert assistance of beamline staff and software developers during data acquisition and data treatment [specifically Günter Grossmann (SRS), Dmitri Svergun and Peter Konarev (EMBL-DESY)].

References

Abbot EL, Smith WD, Siou GP, Chiriboga C, Smith RJ, Wilson JA, Hirst BH, Kehoe MA (2007) Pili mediate specific adhesion of *Streptococcus pyogesnes* to human tonsil and skin. Cell Microbiol 9:1822–1833. doi:10.1111/j.1462-5822.2007.00918.x

Ackerman CJ, Harnett MM, Harnett W, Kelly SM, Svergun DI, Byron O (2003) 19 A solution structure of the filarial nematode immunomodulatory protein, ES-62. Biophys J 84:489–500. doi:10.1016/S0006-3495(03)74868-1

Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664. doi:10.1021/ja069124n

Boulin C, Kempf R, Koch MHJ, Mc Laughlin SM (1986) Data appraisal, evaluation and display for synchrotron radiation experiments: hardware and software. Nucl Instrum Methods A249:399–407

Brown PH, Schuck P (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. Biophys J 90:4651–4661. doi:10.1529/biophysj.106.081372

Chalton DA, Musson JA, Flick-Smith H, Walker N, McGregor A, Lamb HK, Williamson ED, Miller J, Robinson JH, Lakey JH (2006) Immunogenicity of a Yersinia pestis vaccine antigen monomerized by circular permutation. Infect Immun 74:6624–6631. doi:10.1128/IAI.00437-06

Courtney HS, Hasty DL, Dale JB (2002) Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci. Ann Med 34:77–87. doi:10.1080/07853890252953464

Demeler B, Saber H (1998) Determination of molecular parameters by fitting sedimentation data to finite element solution of the Lamm equation. Biophys J 74:444–454. doi:10.1016/S0006-3495(98)77802-6

Dramsi S, Caliot E, Bonne I, Guadagnini S, Prévost MC, Kojadinovic M, Lalioui L, Poyart C, Trieu-Cuot P (2006) Assembly and role of pili in group B streptococci. Mol Microbiol 60:1401–1413. doi:10.1111/j.1365-2958.2006.05190.x

Durchschlag H (1986) Specific volumes of biological macromolecules and some other molecules of biological interest. In: Hinz H-J (ed) Thermodynamic data for biochemistry and biotechnology. Springer, Berlin, pp 45–128

Edwards AM, Manetti AG, Falugi F, Zingaretti C, Capo S, Buccato S, Bensi G, Telford JL, Margarit I, Grandi G (2008) Scavenger receptor gp340 aggregates group A streptococci by binding pili. Mol Microbiol 68:1378–1394. doi:10.1111/j.1365-2958.2008.06220.x

Feigin LA, Svergun DI (1987) Structure analysis by small-angle X-ray and neutron scattering. Plenum Press, New York

de Garcia la Torre J, Huertas ML, Carrasco B (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys J 78:719–730. doi:10.1016/S0006-3495(00)76630-6

Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. Science 318:1625–1628. doi:10.1126/science.1145806

Kehoe MA (1994) Cell-wall-associated proteins in Gram-positive bacteria. In: Ghuysen JM, Hakenbeck R (eds) Bacterial cell wall. Elsevier, Amsterdam, pp 217–261

Koch MH, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys 36:147–227. doi:10.1017/S0033583503003871

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS—a Windows-PC based system for small-angle scattering data analysis. J Appl Cryst 36:1277–1282. doi:10.1107/S0021889803012779

Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. J Appl Cryst 34:33–41. doi:10.1107/S0021889800014126

Lamm O (1929) Die Differentialgleichung der Ultrazentrifugierung. Ark Mater Astron Fys 21B:1–4

Laue TM, Shah BD, Ridgeway TM, Pelletier S (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: Analytical ultracentrifugation in biochemistry and polymer science. Redwood Press Ltd, Melksham, pp 90–125

Lebowitz J, Lewis MS, Schuck P (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci 11:2067–2079. doi:10.1110/ps.0207702

Longhi S, Receveur-Bréchot V, Karlin D, Johansson K, Darbon H, Bhella D, Yeo R, Finet S, Canard B (2003) The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem 278:18638–18648. doi:10.1074/jbc.M300518200

Mandlik A, Das A, Ton-That H (2008a) The molecular switch that activates the cell wall anchoring step of pilus assembly in Gram-positive bacteria. Proc Natl Acad Sci USA 105:14147–14152. doi:10.1073/pnas.0806350105

Mandlik A, Swierczynski A, Das A, Ton-That H (2008b) Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends Microbiol 16:33–40. doi:10.1016/j.tim.2007.10.010

Marraffini LA, Dedent AC, Schneewind O (2006) Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria. Microbiol Mol Biol Rev 70:192–221. doi:10.1128/MMBR.70.1.192-221.2006

Perez J, Vachette P, Russo D, Desmadril M, Durand D (2001) Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. J Mol Biol 308:721–743. doi:10.1006/jmbi.2001.4611

Petoukhov MV, Eady NA, Brown KA, Svergun DI (2002) Addition of missing loops and domains to protein models by x-ray solution scattering. Biophys J 83:3113–3125. doi:10.1016/S0006-3495(02)75315-0

Proft T, Baker EN (2008) Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. Cell Mol Life Sci. doi:10.1007/s00018-008-8477-4:-8

Race PR, Solovyova AS, Banfield MJ (2007) Conformation of the EPEC Tir protein in solution: investigating the impact of serine phosphorylation at positions 434/463. Biophys J 93:586–596. doi:10.1529/biophysj.106.101766

Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B, Longhi S (2006) Assessing protein disorder and induced folding. Proteins 62:24–45. doi:10.1002/prot.20750

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. Proteins 42:38–48. doi:10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3

Schuck P (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. Biophys J 75:1503–1512. doi:10.1016/S0006-3495(98)74069-X

Schuck P (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal Biochem 320:104–124. doi:10.1016/S0003-2697(03)00289-6

Scott JR, Zähner D (2006) Pili with strong attachments: Gram-positive bacteria do it differently. Mol Microbiol 62:2320–2330. doi:10.1111/j.1365-2958.2006.05279.x

Semenyuk AV, Svergun DI (1991) GNOM—a program package for small-angle scattering data processing. J Appl Crystallog 24:537–540

Sreerama N, Woody RW (2003) Structural composition of betaI- and betaII-proteins. Protein Sci 12:384–388. doi:10.1110/ps.0235003

Svergun DI (1992) Determination of regularisation parameter in direct-transform methods using perceptual criteria. J Appl Cryst 25:495–503. doi:10.1107/S0021889892001663

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 76:2879–2886. doi:10.1016/S0006-3495(99)77443-6

Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11:739–756. doi:10.1110/ps.4210102

Vachette P, Dainese E, Vasyliev VB, Di Muro P, Beltramini M, Svergun DI, De Filippis V, Salvato B (2002) A key structural role for active site type 3 copper ions in human ceruloplasmin. J Biol Chem 277:40823–40831. doi:10.1074/jbc.M207188200

Viguera AR, Martínez JC, Filimonov VV, Mateo PL, Serrano L (1994) Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. Biochemistry 33:2142–2150. doi:10.1021/bi00174a022

Volini M, Tobias P (1969) Circular dichroism studies of chymotrypsin and its derivatives. Correlation of changes in dichroic bands with deacylation. J Biol Chem 244:5105–5109

Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. J Appl Cryst 36:860–864. doi:10.1107/S0021889803000268

Wu H, Fives-Taylor PM (2001) Molecular strategies for fimbrial expression and assembly. Crit Rev Oral Biol Med 12:101–115. doi:10.1177/10454411010120020101