Conserved structure and domain organisation amongst bacterial Slc26 transporters

Emma L. R. Compton[*], Kimberly Page[*], Heather E. Findlay[†], Michael Haertlein[‡], Martine Moulin[‡], Ulrich Zachariae[*], David G. Norman[*], Frank Gabel[§] and Arnaud Javelle[*‖1]

[*]College of Life Sciences, University of Dundee, Dundee, DD1 5EH, UK.
[†]School of Biochemistry, University of Bristol, Bristol, BS8 1TD, UK
[‡]The Deuteration Laboratory, Institut Laue Langevin, CIBB, 6 rue Jules Horowitz, 38042 Grenoble cedex 9, France.
[§]Institut de Biologie Structurale CEA, CNRS, UJF, 41 rue Jules Horowitz, 38027 Grenoble, France.
[‖]Strathclyde Institute of Pharmacy and Biomedical Sciences, 161 Cathedral Street, Glasgow, G4 0RE, UK.

Address correspondence to: [1]Arnaud Javelle, Strathclyde Institute of Pharmacy and Biomedical Sciences, 161 Cathedral Street, Glasgow, G4 0RE
Phone: +44 (0)1415 482819
Fax: +44 (0)1415 522562
Email: ajavelle@dundee.ac.uk

Short (page heading) title: Structure of bacterial SLC26 transporters

Key Words: Membrane Proteins, Protein Purification, Protein Structure, SLC26/SulP, Small Angle Neutron Scattering, Transporter

Abbreviations used: DauA, Dicarboxylic uptake system A; D2O, deuterated water; DTT, dithiothreitol; IMAC, immobilized metal-ion-affinity chromatography; LB, Luria–Bertani; SANS, small-angle neutron scattering; STAS, Sulfate transporter and Anti Sigma; PELDOR: Pulsed ELectron–electron DOuble Resonance; SEC, size-exclusion chromatography; SPA, sequential peptide affinity; TM, Transmembrane; TMAO, trimethylamine N-oxide; TMH, Transmembrane Helices

ABSTRACT

The Slc26A proteins are a ubiquitous superfamily of anion transporters conserved from bacteria to man, amongst which four have been identified as human disease genes. Our functional knowledge of

this protein family has increased but limited structural information is available. These proteins contain a transmembrane (TM) domain and a C-terminal cytoplasmic Sulfate Transporter and Anti-Sigma factor (STAS) domain. In a fundamental step toward understanding the structure/function relationships within the family we have used Small Angle Neutron Scattering (SANS) on two distantly related bacterial homologues to show that there is a common structural architecture amongst Slc26A transporters. Pulsed ELectron–electron DOuble Resonance (PELDOR) spectroscopy confirms the proteins form a stable dimer. Using chimeric/truncated proteins we have determined the domain organisation, the STAS domains project away from the TM core and are essential for protein stability. We use the SANS generated envelopes to assess a homology model of the TM core.

INTRODUCTION

The Slc26A proteins are a ubiquitous family of secondary anion transporters conserved from bacteria to man [1] which belong to the amino acid-polyamine-organocation (APC) superfamily[2]. Slc26A proteins exhibit a wide variety of functions, transporting anions ranging from halides to carboxylic acids [3].

Human SLC26A proteins have been shown to play critical roles in cell physiology. They are implicated in genetic diseases such as diastrophic dysplasia, congenital chloride diarrhoea, Pendred syndrome and nonsyndromic deafness [4, 5], while the expression of SLC26A2 has been related to enhanced proliferation of colon cancer cells [6]. Additional pathological phenotypes have been reported in mice deficient in SLC26 gene expression, including male infertility, distal renal tubular acidosis, gastric achlorhydria, and nephrolithiasis [3]. In plants and fungi, Slc26A proteins are primarily sulphate uptake transporters, with mutations leading to auxotrophic phenotypes in fungi [7, 8].

Although these proteins are present ubiquitously amongst bacteria, their physiological functions are almost completely unknown. The only comprehensive physiological analyses concern the *Synechococcus* Slc26A protein, BicA, reported to act as a bicarbonate transporter [9] and our recent study demonstrating that the *Escherichia coli* DauA protein is a succinate transporter involved in C_4-dicarboxylic acid metabolism [10].

Members of the Slc26A family are diverse in terms of substrate specificity but are expected to share a common structure: these proteins comprise an integral membrane domain followed by a C-terminal cytoplasmic Sulphate Transporter and Anti-Sigma factor antagonist (STAS) domain. Topology predictions range from 10-14 transmembrane helices (TMH), although a topology of 12 transmembrane helices has been proposed across the family based upon a model of BicA [11, 12].

High resolution structural data is only available for the soluble STAS domain; structures of two bacterial (DauA and *Mycobacterium tuberculosis* Rc1739c) and the rat motor protein prestin STAS domains (Slc26A5) [13-15] have been published, the backbone traces of which all superimpose very closely. Recently a homology model of the TM domain of rat Prestin has been published based upon the uracil transporter, UraA [16]. The presence of a conserved stoichiometry within the family is the subject of debate as several members, across multiple species, have been reported to form dimers and/or tetramers [17-19]. Using Small Angle Neutron Scattering combined with contrast variation, we have recently proposed that a *Yersinia enterocolitica* Slc26A (YeSlc26A2) homologue forms a dimer [20].

In order to gain insight into the structural architecture and domain organisation of proteins from the Slc26A family, we have used Small Angle Neutron Scattering (SANS) to further investigate YeSlc26A2 domain organisation and compare this with the functionally characterised *E. coli* DauA, allowing us to define and propose a conserved structural architecture for Slc26A proteins. Pulsed ELectron–electron DOuble Resonance (PELDOR) spectroscopy confirms our initial hypothesis that

the purified proteins form stable dimers. SANS analysis on a truncated TM domain alone and an extended C-terminally GFP-fused variant locate the domains within the structure to reveal a multidomain organisation, consisting of a central transmembrane core which extends outwards over the STAS domain to further form two distal domains. Homology modelling allows us to evaluate the recently proposed atomic level structure for Slc26A proteins. Thus this work enables us to define a domain organisation conserved among the bacterial homologues and propose this continues across this important protein family.

EXPERIMENTAL

Expression and purification of proteins.

The *E. coli dauA* gene was cloned into pArno [20] between XhoI and BamHI sites using two oligonucleotides 5'-atgcctcgagaatgcctttccgcgctctgatcgac-3' and 5'-atgcggatcctaaatccgccatcgccgcgcgacg-3'. The Ye-TM domain was cloned into pWaldo-e [21] between XhoI and BamHI sites using two oligonucleotides 5'-atgcctcgagaatgtggcaggttttaaaatca-3' (XhoI site underlined) and atgcggatccgatacggcgagcaaacaggat-3' (BamHI site underlined). To express the two chimeric YeSlc26A2-GFP proteins Ye-3-GFP and Ye-11-GFP, the *Y. enterocolitica Yeslc26a2* gene (YE0973) was cloned into pWaldo-e between *Xho*I and *Bam*HI sites using two oligonucleotides: 5′-atgcctcgagaatgtggcaggttttaaaatca-3′ (*Xho*I site underlined) and 5′-atgcggatccctcaaatttccgctgagacg-3′ (*Bam*HI site underlined). Subsequently, the Stratagene Quickchange Lightening Multi Site-Directed Mutagenesis kit was used to add a Prescission Protease site between GFP and the histidine tag, then to modify the linker between Yeslc26A2 and GFP in pWaldo-e to produce Ye-3-GFP (linker GSE) and Ye-11-GFP (linker GSENLYFHSQF). The multiple cysteine variants of YeSlc26A2 were generated by using the Stratagene Quikchange II Site-Directed Mutagenesis Kit using Yeslc26A2-pArno as a template.

YeSlc26A2 and all derivatives were expressed and purified as previously described [20]. The histidine tag was removed from Ye-3-GFP and YE-11-GFP by Prescission Protease. GFP was removed from Ye-TM-GFP by TEV proteolysis as previously described for removal of the histidine tags. DauA was overexpressed in *E. coli* strain BL21(DE3) in TB media and purified using the same protocol as for YeSlc26A purification [20]. The samples purity and quality were estimated on SDS-PAGE gel and by gel filtration on superdex 200 (10/60) as previously described [20]. All SANS samples were prepared at the previously determined contrast match point of 11% D_2O.

Expression of random-fractional deuterated ("match-out labelled") YeSlc26A2 for neutron scattering was carried out in the ILL Deuteration Laboratory, a technical platform within the Partnership for Structural Biology (PSB), Grenoble, France. Recombinant BL21(DE3) cells were grown in a high-cell density culture at 30˚C in 85% (v/v) D_2O containing minimal medium using unlabelled glycerol as carbon source. Before induction with 0.2 mM IPTG at an OD_{600} of 15.2, the temperature was adjusted to 20˚C and the minimal medium supplemented with 10% of 85% (v/v) deuterated Silantes OD2 medium (Silantes GmbH, Munich, Germany). Expression tests in flask cultures had shown that YeSlc26A2 is not expressed in deuterated minimal medium without supplementation. Deuterated cells (70 g wet weight from 1.74 L of fermenter culture) were harvested after 20 hours expression at an OD_{600} of 15.6. The final deuteration level gave a contrast match point that corresponded to approximately 100% D_2O. For SANS experiments 11% D_2O was included in the gel filtration buffer and the purified proteins were concentrated with a Centricon 50-K (Millipore) to 5 mg/ml and dialysed against the same buffer.

SANS experiments, data reduction

All samples were measured on the large dynamic range diffractometer D22 at the Institute Laue-Langevin (Grenoble, France) in Hellma® quartz cuvettes 100QS with 1mm optical path-length. Sample temperature was kept at 6˚C during the exposure times. The dYeSlc26A2 and Ye-TM samples were recorded at a 4m/4m detector/collimator distance, all other samples at two configurations;

2m/2m and 5.6m/5.6m, using a neutron wavelength of λ = 6Å. At each configuration, the H_2O/D_2O-buffers, the empty beam, an empty quartz cuvette as well as a boron sample (electronic background) were measured. Exposure times varied between 20 minutes (empty cell, boron) and 1 hour for the protein samples and buffers. Transmissions were measured for 1 minute. The raw data were reduced (detector efficiency, electronic background, angular averaging) using a standard ILL software package [22]. Finally, the corrected scattered intensities $I(Q)$ ($Q = (4\pi/\lambda)\sin\theta$, where 2θ is the scattering angle) from the different Q-ranges were merged (in the case of the 2m/2m and 5.6m/5.6m setups) and the respective buffer signals subtracted using the program PRIMUS from the ATSAS suite.

SANS data analysis

Radii of gyration (Table S1). The radii of gyration, R_g of all samples were extracted by the Guinier approximation using the program PRIMUS:

$$\ln[I(Q)] = \ln[I(0)] - 1/3\, R_g^2 Q^2$$

The validity of the Guinier approximation, $R_g Q \leq 1.3$, was verified and fulfilled in each case.

Low-resolution models determined by ab initio *analysis using DAMMIN*. We used the program DAMMIN [23] to generate low-resolution envelopes of all samples. The DAMMIN input files were generated using the program GNOM imposing the restraints $p(r=0)=0$ and $p(r=D_{max})=0$ for the pair distance distribution function. As a quality control, the radii of gyration determined from the pair distribution analyses were compared to those determined by the Guinier analysis of the SANS data. P(R) curves are shown in supplementary materials for all samples.

Generation of averaged, 'most probable' models. For each protein at least 10 *ab initio* DAMMIN models were generated with no symmetry constraints applied (P1 symmetry), all were checked for both consistency in shape and the presence of symmetry. All proteins showed evidence of rotational symmetry and therefore at least 10 *ab initio* DAMMIN models were generated with the relevant symmetry constraints applied. Averaged models were generated by first using the programs DAMSEL and DAMSUP to remove outliers, select the most typical and align the DAMMIN models, then aligned models were averaged by DAMAVER. Subsequently DAMFILT was used to produce the final, most probable model. In all cases enantiomers were allowed where relevant and at least 10 *ab initio* models were used by DAMAVER. Final models were aligned using SUPCOMB with enantiomers allowed and symmetry applied where appropriate. SUPCOMB superimposes and aligns two models based upon the minimal Normalised Spatial Discrepancy (NSD), where NSD is 0 for identical structures and above 1 for molecules considered systematically different. Except where stated all P1 models showed P2 symmetry and hence the process was repeated with P2 symmetry imposed.

Labelling of single cysteine mutants with Methoxypolyethylene glycol maleimide (MAL-PEG)

For the labelling experiment, the cysteine variants were purified by IMAC only. MAL-PEG (Sigma) was used at 300 μM to label 6 μM of purified protein. After incubation for 1 hr at room temperature, the reaction was quenched with 33 mM DTT and analysed by SDS-PAGE gel. MAL-PEG labelling was detected by *colloidal coomassie* G-250 staining or western blot using primary antibody anti-His (Qiagen) and secondary antibody goat anti-mouse conjugated to horseradish peroxidise; (HRP) (Biorad).

Continuous Wave Electron Paramagnetic Resonance Spectroscopy

Spectra were taken using a Bruker EMX spectrometer with an ER 4122SHQE resonator. MW power was 10mW, frequency was 9.876 GHtz and Q value was 7900. The Field centre was 351.910 mT and the Sweep Range was 10 mT (346.9 mT to 356.9 mT). Spectra were gathered using 2048 data points.

Data was gathered at room temperature. Spin label concentration was estimated by comparison to a standard curve calculated using serial dilutions of TEMPO.

Circular Dichroism (CD) Spectroscopy

Data were collected on an Aviv 410 CD Spectrometer using 0.05mm cell, 1nm steps and 0.5s averaging time at 25^{o}C. Data were analysed with CDtools software and Dichroweb [24] using CDSSTR and reference data set 7.

Pulsed Electron-Electron Double Resonance Spectroscopy

The protein YeSlc26A2-D416C (4mg/ml) was mixed with a 10-fold molar excess of (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)methanethiosulfonate (MTSSL) and incubated for 12 hr at 4 C. MTSSL excess was subsequently removed by successive dialysis and buffer exchange into 200 mM Tris, 300 mM NaCl, 0.2 % Foscholine-12 in D_2O (Sigma). Experiments were performed at 50 % D-glycerol using a Bruker ELEXSYS E580 pulsed spectrometer operating at X-band with a standard flexline probe hosting a dielectric ring resonator (MD4) and a Bruker 400U second microwave source unit. All measurements were made at 50 K with an overcoupled resonator giving a quality factor Q of approximately 100. The video bandwidth was set to 20 MHz. The four pulse, dead-time free PELDOR sequence was used, with the pump pulse frequency positioned at the centre of the nitroxide spectrum. The frequency of the observer pulses was set 80 MHz higher, relative to the pump position, which coincides with the low-field local maximum of the spectrum. The observer sequence used a 32 ns p-pulse; the pump p-pulse was typically 16 ns. The experiment repetition time was 4 ms, and the number of shots at each time point was 50. The number of time points and scans used were 233 and 109 respectively. Data were analysed using the DeerAnalysis 2011 software package. The dipolar coupling evolution data were corrected for background decay using a homogeneous three-dimensional spin distribution model. The distance distribution was determined by fitting of a single Gaussian distribution model, utilising a 1.5nm linewidth target.

Homology modelling and molecular dynamics

Sequences were submitted to the HHPred and MODELLER servers at the Max-Planck Institute for Developmental Biology to select homologous proteins and generate homology models, standard search options were implemented. The STAS domain model was generated using an optimal template based upon STAS domains from *E.coli* DauA, *Rhodobacter sphaeroides*, *Vibrio Cholerae, Wolinella Succinogenes* Slc26 proteins and *Geobacillus stearothermophilus* SPOIIAA.

All-atom molecular dynamics simulations were performed with GROMACS4.6[25], using the Amber99sb-ildn force field for the protein[26], the Berger model for POPC lipids [27] and the SPC/E model for water[28]. The structural models were equilibrated for 20 ns with position restraints on protein heavy atoms. Three independent 100 ns simulations were then performed in the NpT-ensemble (integration time-step: 2 fs) at 0.15 M NaCl, 1 bar and a temperature of 300 K held constant by Berendsen [29]and velocity-rescale temperature coupling[30], respectively.

RESULTS

A conserved structural architecture amongst bacterial Slc26A transporters

We have studied two distantly related bacterial homologues to investigate the existence of a common architecture within the Slc26A family. Analysis of prokaryotic proteins in order to gain insights into their eukaryotic counterparts is an approach used successfully in the study of many transporter families[31, 32]. We chose *E.coli* DauA, 53kDa, and *Y. enterocolitica* Slc26A2 (YeSlc26A2), 59kDa. Phylogenetic analysis demonstrates they are from two distinct clades (Figure 1) with an identity of 28%, yet along with other Slc26A proteins, they are predicted to share the common architecture. These proteins are of particular interest as YeSlc26A2 is the only bacterial homologue for which there is full-length structural information available (low resolution SANS envelope) [20]and which can be

purified in high enough yields for deuteration, while DauA is the only bacterial homologue for which there is both functional and partial structural information available (high resolution structure of the STAS domain)[10, 13].

We determined low resolution structural envelopes for DauA and YeSlc26A2 by SANS with contrast variation. SANS is ideally suited to this purpose as contributions from the detergent to the signal can be reduced to a minimum by the presence of D_2O at the correct contrast match point, i.e. H_2O/D_2O ratio in the solvent. Further, deuteration of a protein allows for a greater reduction in the relative contribution of detergent to the signal of the protein, thus increasing the signal to noise ratio. Proteins can be deuterated for SANS by growth in deuterated minimal media using high cell density cultures. YeSlc26A2 is expressed readily (0.1mg protein/g cells) in M9 minimal media; DauA expression, however, in minimal media is extremely low (data not shown) and is therefore not suitable for deuteration. Thus, we performed SANS experiments on deuterated YeSlc26A2 (dYeSlc26A2) and non-deuterated (hydrogenated) DauA.

We collected scattering curves from both proteins, as expected dYeSlc26A2 has a dramatically improved scattering signal when compared to the non-deuterated DauA (Figure 2A), but both sets of data are suitable for *ab initio* modelling. The proteins have very similar radii of gyration, 47.7±3.6Å (DauA) and 46.7±0.4Å (dYeSlc26A2) (Table S1), suggestive of an overall similar shape.

The *ab initio* modelling program DAMMIN [23] was used in conjunction with the DAMAVER suite of programs [33] to generate low resolution envelope models of the proteins as described in the Experimental Procedures section. Initial *ab initio* models were generated with no symmetry constraints (P1 symmetry); individual DauA models are presented in Figure 2B. DAMAVER was then used to generate averaged, filtered models. Both the *ab initio* and final P1 (Figure 2C) DauA models appeared to have P2 symmetry, thus the process was repeated with P2 symmetry applied to produce a final averaged, filtered model (Figure 2D). Figure 2E shows the fit of predicted scattering curves of the most typical P1 and P2 DAMMIN models (as selected by DAMSEL) to the SANS data, the curves are indistinguishable with χ^2 values of 0.679, thus fitting the SANS data well. Both final P1 and P2 DauA models are extremely similar in appearance and have a bulky central domain with two distal lobes projecting away. An alignment of the models using the program SUPCOMB (Figure 2F) was used to compare the similarity of the two models, the NSD value of 0.483 confirms that DauA has genuine P2 symmetry.

The same process was used to generate a final model of dYeSlc26A2 with P2 symmetry imposed (Figure 3A); the final, most probable P1 model is shown in Supplementary Figure S2B. Fits of the predicted scattering curves from the DAMSEL-selected, most typical individual P1 and P2 models to the scattering data are shown in Figure 3B, as for DauA they fit the SANS scattering data well and are indistinguishable. The χ^2 values, 2.57 (P1) and 2.56 (P2), are higher than for the non-deuterated proteins as the improved signal to noise ratio means that a small discrepancy between the experimental data and the predicted DAMMIN scattering curve is heavily weighted and penalised within the χ^2 value. The P1 and P2 models superimpose well (Supplementary Figure S2C) with an NSD of 0.505.

The overall shape, architecture and dimensions of DauA and dYeSlc26A2 are very similar and the two can be aligned with an NSD of 0.549 (Figure 3C). DauA is a slightly larger molecule than YeSlc26A2, as reflected by the dimensions of the particles 130x80x65Å for DauA and 120x75x65Å for dYeSlc26A2, the central domain of DauA appears bulkier with a more prominent central extension than that of dYeSlc26A2. Overall the dimensions and architecture of the two proteins demonstrate that there is a conserved structural architecture and domain organisation within the bacterial Slc26A family.

Oligomeric state

The presence of two-fold rotational symmetry in both the DauA and YeSlc26A2 models suggests the proteins may exist as dimers, however SANS data alone are not ideal to determine the oligomeric

state of a protein. To investigate this directly, we used Pulsed Electron-Electron Double Resonance (PELDOR) spectroscopy coupled with Site-Directed Spin Labelling. By creating a YeSlc26A2 mutant containing only one cysteine residue the protein can be specifically labelled with the nitroxide spin-label necessary for PELDOR analysis. The number of coupled spins in the molecule, and hence its oligomeric state, can be determined by examination of the modulation depth of the[34]. The distance between the spin labels can also be measured by extraction of distance information from the spin-spin dipole interactions observed after pulsed microwave radiation. These properties enable us to validate our models, both the presence of dimerisation and the accuracy of the envelopes.

By inspection of the YeSlc26A2 STAS domain model we published previously [20], position 416 was chosen as the site for nitroxide spin labelling as it is predicted to be both accessible to the solvent, thus maximising the likelihood of labelling, and within an α-helix, reducing the possibility of locating the site in a highly mobile loop region (Figure 4A). The mutation D416C was made in a construct where the three native cysteines (C15, C51 and C272) were replaced by serine to form Ye-D416C. Ye-D416C was purified to homogeneity in Fos-choline12 and labelled with the nitroxide spin label, MTSSL. The free spin label was removed by dialysis and successive rounds of buffer exchange.

Continuous Wave Electron Paramagnetic Resonance (CW EPR) was used to confirm that the mutant can be fully labelled and that the bound label behaves in a way that is consistent with it being attached to the protein at a solvent exposed position (Supplementary Figure S3A). CD spectral analysis demonstrates that the labelled mutant folds correctly with a predicted secondary structure content of 51% α-helix, 11% β-sheet, 16% turn and 21% disordered, almost identical to that of YeSlc26A2 (53% α-helix, 10% β-sheet, 15% turn and 20% disordered) (Supplementary Figure S3B).

PELDOR measurements were made at x-band (Figure 4B). The resultant data, after background correction, showed a modulation depth of about 0.6 (Figure 4C), typically seen for a dimeric spin pair interaction. The lack of a defined oscillation on the data negated the use of Tikhonov regularization for distance distribution determination, however, a Gaussian distribution was fitted to this data. This cannot give a defined distance but suggests a broad distance distribution centred at about 30-35Å (Figure 4D).

In conclusion, the D416C could be fully labelled and the mutation had no effect on the folding of the protein. The results obtained by PELDOR spectroscopy analysis firstly demonstrate that YeSlc26A2 forms a dimer and secondly indicates that the distance between the two STAS domains in the dimer is approximately 30-35Å, in good agreement with that of 35-40 Å in the SANS model.

Slc26A2 domain organisation

SLC26 proteins possess two domains: a TM domain predicted to comprise 10-14 TMHs and a soluble, cytoplasmic STAS domain. To investigate the positioning of these domains within the conserved structural architecture we considered how the TM and STAS domains could be arranged. The dimerisation interface could be formed by either the TM domains or the STAS domains. To investigate this arrangement, we generated two classes of YeSlc26A2 variants: one with green fluorescent protein (GFP) fused after the STAS domain, forming a C-terminal extension, and the other with the STAS domain removed, forming a C-terminal truncation. We anticipated two possible outcomes, if the dimerisation occurs via the TM domain the extra bulk of the GFP molecule would be seen as an extension to the distal regions in the C-terminally extended protein, while these regions, or parts of them, would be missing in the C-terminally truncated protein, leaving only the symmetrical, central domain. However, if dimerisation occurs via the STAS domains we expected that the central region would increase in size when GFP is present whereas the TM domain alone would form much smaller monomers with no rotational symmetry.

GFP increases the size of the distal lobes

To form the C-terminally extended YeSlc26A2, two different chimeric GFP-tagged YeSlc26A2 proteins were tested with 3 and 11 amino acid linkers (GSE and GSENLYFHSQF) between the

STAS domain and GFP to form Ye-3-GFP ($MW_{monomer}$ 80.2kDa) and Ye-11-GFP ($MW_{monomer}$ 81.2kDa) respectively. The two differing lengths of linkers were chosen to negate anticipated negative effects of either a highly mobile GFP molecule causing a loss of scattering signal (in the case of Ye-11-GFP) or a short linker (Ye-3-GFP) affecting the overall folding of the protein.

Both proteins were purified to homogeneity by gel filtration. Ye-3-GFP and Ye-11-GFP dimers would be expected to be larger than YeSlc26A2 by ~52 and 54 kDa respectively. In practice, however, it is not possible to measure exact differences for membrane proteins by gel filtration due to the unknown quantity of detergent bound to the protein. Elution volumes can be used qualitatively to give an indication of relative masses. The elution volume for Ye-3-TM from a Superdex 200 gel filtration column indicates a molecular weight of around 200 kDa and Ye-11-TM around 215 kDa, both appearing larger than the full length dimeric YeSlc26A2 (~175 kDa).

SANS data were collected for both chimeric proteins as described previously. The scattering curves (Figure 5A), and thus radii of gyration (51.9±2.2 Å Ye-3-GFP and 53.1±2.4 Å Ye-11-GFP – Table S1), are almost indistinguishable demonstrating that the two molecules are identical and the presence of GFP does not impact upon the folding of Ye-3-GFP.

To identify the location of the GFP, Ye-3-GFP *ab initio* models were constructed and used to generate a most probable final structure. As the scattering data for both Ye-3-GFP and Ye-11-GFP are very similar, we present here only the modelling analysis for Ye-3-GFP. Examples of the predicted scattering curves from individual, most typical DAMMIN models are shown in Figure 5B, the fits of the P1 and P2 models to the SANS data are excellent with χ^2 values of 0.344 and 0.343 respectively. P2 symmetry was observed in the initial models (Supplementary Figure S4B) and was therefore applied to produce the final (P2) model (Figure 5C). The P1 and P2 models align well (Supplementary Figure S4C) with an NSD of 0.609.

Ye-3-GFP molecule is indeed larger than YeSlc26A2, with dimensions of 165x95x100Å (Figure 5C-D), however when the shape of the molecules are compared it can be seen that Ye-3-GFP has extended distal domains with the central region relatively unchanged as would be expected if GFP, and thus the STAS domain, were located in the distal arms. In agreement with this observation, the radii of gyration are larger for both Ye-3-TM and Ye-11-TM than for YeSlc26A2 alone (46.7±0.4Å – Table S1), also suggestive of a more elongated molecule.

We do not see a single position for the GFP molecule as might have been initially expected, however GFP and YeSlc26A2 are not expected to specifically interact with each other. It is likely that the GFP molecule is relatively mobile around the linker attaching it to the static STAS domain giving rise to multiple positions which are all represented and hence averaged in the scattering data. This would give rise to a general increase in the observed volume of the domain rather than seeing an extra single, globular domain, exactly as we observe. Importantly nevertheless, only an increase in the volume of the distal domains and not the central dimerization domain is observed, which qualitatively supports a model where the dimerisation interface is formed by the TM domain and the STAS domains are located in the distal regions.

The TM domain forms the dimerisation domain

In parallel to investigating C-terminal GFP fusion we studied the effect of removing the C-terminal STAS domain to leave only the TM domain (Ye-TM). The TM domain could not be expressed alone, however it was readily expressed with a C-terminally fused GFP placed after a TEV protease cleavable linker. The GFP proteolytically removed after initial IMAC purification and the TM domain alone further purified to homogeneity. Its apparent molecular mass can be estimated from the size-exclusion chromatogram as ~125kDa, smaller than for the full-length protein (~175kDa) as expected, but much larger than that predicted for the monomer (42kDa) suggesting that the TM domain, like the full length protein, may purify as a stable oligomer.

SANS data were collected for Ye-TM (Figure 6A) resulting in a radius of gyration of 38.9±3.8 Å (Table S1), as expected smaller than that for the full length protein and suggestive of a more compact particle. Unexpectedly *ab initio* models generated with no symmetry constraints (P1) appeared to have P222 rotational symmetry (Figure 6B), rather than the P2 symmetry seen in the full length and extended proteins. The presence of P222 symmetry however does not fit either of the anticipated outcomes: if Ye-TM were forming the dimerisation domain, as predicted by the Ye-3-GFP model, it would be expected to have P2 symmetry like YeSlc26A2 and Ye-3-GFP or if the STAS domain forms the dimerisation domain then Ye-TM would be expected to be monomeric with no symmetry.

Firstly we performed CD spectroscopy to confirm that removal of the STAS domain had not compromised the folding of Ye-TM. The spectral analysis clearly shows that the Ye-TM is fully folded and stable in detergent micelles with a composition of 53% α-helix, 11% β-sheet, 18% turn and 19% disordered. To investigate the possibility that Ye-TM genuinely has P222 symmetry, *ab initio* models were generated with both the predicted P2 symmetry as well as the observed P222 symmetry. The predicted scattering curves are indistinguishable from each other and the P1 model (Figures 6C and S5B), fitting the SANS data well with χ^2 values of 0.475 (P2) and 0.474 (P222). The P2 and P222 models are also almost identical to each other (Figure 6D), forming bow tie-like structures and can be aligned to the P1 model with excellent NSD values of 0.480 (P2) and 0.481 (P222). It is not unexpected that the P2/P222 models are very similar as P2 symmetry is also present in the P222 model, however from Figure 6E it can be seen that the P222 model is almost identical to the P1 model suggesting that Ye-TM could have genuine P222 symmetry.

To investigate the possibility of a biological basis for the unexpected presence of P222 symmetry, the Ye-TM envelope was superimposed with that of dYeSlc26A2 (Figure 6F). It can be seen that the TM envelope is larger even than the full length protein supporting the possibility that Ye-TM forms an oligomer of higher order than full-length YeSlc26A2. Examination of the solvent excluded volumes, derived from a POROD analysis of the SANS data (240000+-30000Å^3) and the DAMMIN models (227000Å^3), being 4.5 and 4.3 times the theoretical value for a monomer (53000 Å^3), plus the presence of P222 symmetry which is more usually found in tetramers than dimers, suggests the possibility that Ye-TM forms a tetramer.

Ye-TM forms a dimer of dimers

Close inspection of a superimposition of Ye-TM and dYeSlc26A2 reveals that the upper half of the molecules are very similar, it is the lower halves that differ (Figure 6F). Given the suspected P222 symmetry within Ye-TM, where the lower half of the molecule would be a rotation of the upper half, this led us to speculate that Ye-TM consists of a 'dimer of dimers'. We hypothesised that the STAS domain is located in the lower portion of the distal arms and hence its removal leaves the upper portion of the native dimer, the same structure as the upper portion of dYeSlc26A2. To investigate whether this was possible we made a model of the upper portion of the dYeSlc26A2 P2 model by removing beads and fitting this and the model of the STAS domain within the dYeSlc26A2 P2 envelope (Figure 7A). If the region which interacted with the STAS domain retains the potential to form protein-protein interactions, oligomerisation with another Ye-TM dimer, is possible, hence forming a tetramer with P222 symmetry (Figure 7B). This hypothesis predicts that the three YeSlc26A2 native cysteine residues, C15, C51 and C272, all localised within the TM domain and predicted to be on or towards the cytoplasmic face of the protein, would be accessible to the solvent in the full length protein, but not in Ye-TM. We tested this prediction by attempting to label the native cysteines with the large thiol specific probe, methoxypolyethylene glycol maleimide (PEG-MAL).

To ensure that any observed labelling was specific to these cysteine residues a cysteine-less mutant was constructed as well as variants each containing only one of the three native cysteines. The cysteine-less protein and the three single cysteine variants were incubated with PEG-MAL and the effects of the label investigated by SDS-PAGE. When PEG-MAL reacts with a cysteine it causes a band shift upon SDS-PAGE analysis. All three cysteine mutants could be labelled by PEG-MAL, whereas the cysteine-less protein remained unmodified (Supplementary Figure S6), demonstrating

that each single cysteine is accessible to PEG-MAL in the full length protein and confirming that any labelling seen in the full length or Ye-TM proteins is specific to the native cysteine residues.

After incubation of the full length YeSlc26A2 with PEG-MAL three bands appear of increasing molecular weight corresponding to successive labelling of 1, 2 and 3 cysteines (Figure 7C). When Ye-TM was incubated with PEG-MAL there was no change in the band pattern, indicating that the cysteines are no longer accessible to the labelling reagent as we predicted if Ye-TM forms a dimer of dimers via the cytoplasmic face of the protein. Thus we conclude that the TM domains form the YeSlc26A2 dimerisation interface with the STAS domain residing in the lower portion of the distal arms.

Homology Modelling

To complement our experimental study of the domain organisation within Slc26A proteins, we sought to build homology models for YeSlc26A2 and DauA. The HHPred server [35] for homology prediction was used in conjunction with the MODELLER homology modelling program [36] to generate atomic models. Sequences for full length YeSlc26A2, TM (1-429) and STAS (430-481) domains were submitted separately, as were sequences for the full length and TM domain (1-435) of DauA. UraA, a uracil transporter of the NCS2 family, was given a 100% probability of being homologous to the TM domain of YeSlc26A2 and DauA when either the full length protein or the TM domains were submitted. Hence TM domain models were built using UraA as a template. As expected the STAS domain was found to be homologous to STAS domains of other Slc26 proteins as well as anti-sigma factor antagonists and a model was built using the optimal top five matches as selected by HHPred.

To allow us to confidently compare the YeSlc26A2 homology model with our SANS envelopes we tested the structural model by subjecting it to 100-ns all-atom molecular dynamics simulations in a fully hydrated POPC membrane bilayer. The model did not show large deviations from the initial structure, with a root-mean-square deviation (RMSD) that remained below 1.7 Å in all three simulations, indicating that the model does not suffer from major structural weaknesses (See Figure 8 and Supplementary Figure 7).

DISCUSSION

Conserved structural architecture within the Slc26A family

The Slc26A protein family belongs to the APC superfamily and is present from bacteria to man. Its members are anion transporters, variously able to function as symporters, antiporters and channels, moving a wide variety of substrates from halides to carboxylic acids across cell membranes [3]. While a variety of mechanisms may be responsible for this range of functions, there is expected to be a conserved fold and, therefore, overall structural architecture across the family [12]. Slc26A proteins are predicted to comprise two domains: a TM domain and a cytoplasmic STAS domain. However, to date there has been no thorough study to investigate the structural architecture and its conservation within the family. We have used SANS to investigate both the conservation of, and the domain arrangement within, the structural architecture of two distantly related bacterial Slc26A proteins, DauA and YeSlc26A2.

As hypothesised SANS derived envelopes for DauA and YeSlc26A2 appear very similar, consisting of a bulky central domain with two distal lobes (Figure 3C). They align with an excellent NSD of 0.549, with only minor differences in the central region, thus confirming that there is indeed a conserved structural architecture throughout the bacterial members of the Slc26A family. Based upon the proposal from Shelden *et al,* 2010 [11] we infer that this architecture is conserved across the whole Slc26A family, as has been demonstrated in other transporter families where structural information has been determined from bacterial homologues[31, 32].

We sought to clearly define the oligomeric state of the molecule and the domain arrangements within this envelope, allowing us to define a model for the Slc26A proteins. Firstly, as suggested by the P2

symmetry present in the SANS models and as we previously proposed [20], PELDOR analysis confirmed that YeSlc26A2 is a dimer. Given the structural similarity between YeSlc26A2 and DauA it can be assumed that DauA is also dimeric. To date in all cases where the oligomeric state of an Slc26A protein has been determined they have been shown to be dimeric [18], except for the motor protein Prestin for which dimeric and tetrameric states have been reported [17-19, 37, 38].

Secondly, by the use of truncated and extended mutants we investigated the positioning of the TM and STAS domains within the SANS envelope and therefore which domain is responsible for forming the dimerisation interface. The addition of GFP at the C-terminus of YeSlc26A2 created an extended molecule demonstrating that the STAS domain resides within the distal lobes. This is in agreement with studies that have expressed the isolated STAS domain of YeSlc26A2 [20] and other Slc26A proteins as monomers with no evidence of dimerization [39]. The location of the TM domain in the central dimerisation domain was confirmed by inspection of the SANS envelope of the TM domain alone. This envelope suggests that the TM domain forms not only the central region but also extends into the upper portion of the distal lobes.

Interestingly, the TM domain could only be expressed as a GFP fusion protein, appearing to form a 'dimer of dimers' once the GFP was removed. This suggests that a soluble C-terminal domain is necessary for correct expression and stability of the protein, however it is not necessary for this to be the STAS domain itself. Perhaps providing an insight into SLC26-related diseases where trafficking impairment due to mutation in, and thus misfolding of, the STAS domain has been identified as one of the causes.

Homology Modelling

An experimentally verified homology model of rat Prestin (SLC26A5) has recently been published based upon the high resolution structure of UraA from the NCS2 family [16]. Both Slc26A and NCS2 families belong to the APC superfamily[40]. We generated a homology model of YeSlc26A2 based on the UraA structure using a similar method to compare with our experimentally determined structural architecture (Figure 8 and Supplementary Figure S7). All-atom molecular dynamics simulations in a solvated bilayer were used to determine the structural integrity of this model. The final model has a secondary structure content of 55% α-helix and 11% β-sheet in excellent agreement with our experimentally determined composition of 53% α-helix and 11% β-sheet for the TM domain. Furthermore, the trypsin, and hence solvent, accessible sites identified previously [20] cluster on the predicted cyto- and periplasmic edges of the molecule (Figure 8B).

In common with the prestin model and UraA structure (refs) the YeSlc26A2 homology model comprises a 'core' domain containing two unusual, short β-strands which are thought to play a role in the transport pathway and a 'gate' domain that is proposed to undergo conformational change during the UraA transport cycle. As in the Prestin model, helices 1, 2 and 9 which contain the Slc26A signature sequences are situated in the core domain clustered together around the important β-strand motif. Residues E252 and N260 which are highly conserved and implicated in human disease are also positioned in close proximity to this β-strand motif.

Comparison of SANS Envelopes and Homology Models

We placed the YeSlc26A2 TM domain model within the deuterated YeSlc26A2 SANS envelope alongside a homology model of the YeSlc26A2 STAS domain, positioning the STAS domain within the lower portion of the distal arms (Figure 8). Whilst the TM domain model appears to fit the central region of the SANS model relatively well, it does not fill the upper portion of the distal lobe as predicted by the TM domain SANS envelopes and consequently there is a steric clash with the STAS domain. It is possible that the Slc26A proteins, or the bacterial members at least, deviate from the UraA structure in this distal region. However, it is tempting to speculate that the SANS structures presented here represent the 'extended' conformation whereas the homology models represent the 'contracted' conformation predicted to occur within these proteins.

The reasons for the differences between the YeSlc26A2 and DauA envelopes, where the DauA upper central region is more prominent and the lower portion bulkier than in YeSlc26A2, are clear when the homology models of the TM domains are aligned (Supplementary Figure S7). The intramembranous regions are essentially identical, however DauA has larger periplasmic loops, a longer cytoplasmic N-terminus as well as extra residues within the STAS domain. Thus, when the proposed TM domain envelopes are positioned within a theoretical membrane bilayer the extra bulk of DauA is solely in the extramembranous regions.

Together the oligomeric state and domain organisation allow us to propose a model for the conserved, dimeric architecture of Slc26A proteins where the STAS domains reside in the lower portion of the distal lobes and the TM domain occupies the central region forming the dimerisation domain. It would be interesting in the future to investigate this structure using a complete PELDOR analysis.

ACKNOWLEDGEMENTS

We would like to thank the ILL Block Allocation Group (BAG) system for SANS beamtime at D22 and Dr. Anne Martel for help with the setup of the instrument. ***Author contributions:*** ELRC/AJ designed the research, ELRC/KP/HF/UL/DGM/FG/AJ performed the experiments, ELRC/HF/DGM/FG/AJ analysed the data, ELRC/AJ wrote the paper, MH/MM produced the deuterated cells.

FUNDING

This work was supported by a grant from Medical Research Council (G1000054). AJ was supported by a Scottish government/Royal Society of Edinburgh personal research fellowship.

REFERENCES

FIGURES LEGENDS

Figure 1. Phylogenetic tree of bacterial DauA homologues

Modified from Karinou *et al*, 2013**.** Neighbour-joining tree of 766 DauA-related sequences retrieved from the NCBI database BLASTP. Cluster A is a separate group without any functional annotation. Bootstrap values are reported only for main clusters. The scale marker represents 0.1 substitutions per residue.

Figure 2. *Ab initio* DauA envelope.

(A) Experimentally determined SANS scattering data for DauA and dYeSlc26A2. Q is the scattering vector and I represent the intensity. (B) Individual DauA models produced by DAMMIN with no symmetry constraints (P1). (C) Averaged DauA model with no symmetry constraints (P1). (D) Averaged DauA model generated with P2 symmetry. (E) Experimental DauA SANS data and predicted scattering curves for typical DAMMIN DauA models generated with no symmetry (P1) and P2 symmetry constraints over the Q-range used for DAMMIN modelling. (F) Superimposition of averaged DauA P1 (beads) and P2 (outline) models.

Figure 3. *Ab initio* dYeSlc26A2 envelope.

(A) Averaged dYeSlc26A2 model generated with P2 symmetry. (B) Experimental dYeSlc26A2 SANS data and predicted scattering curves for typical DAMMIN dYeSlc26A2 models generated with no symmetry (P1) and P2 symmetry over the Q-range used for DAMMIN modelling. Q is the scattering vector and I represents the intensity. (C) Superimposition of averaged P2 dYeSlc26A2 (grey beads) and DauA (cyan outline) models.

Figure 4. PELDOR analysis of Ye-D416C

(A) Model of dYeSlc26A2 STAS domain, the position of single cysteine residue, D416C, selected for PELDOR measurements is highlighted. (B) Uncorrected PELDOR data showing baseline fit (black line). (C) Background corrected PELDOR data. (D) Distance distribution, single Gaussian fit. See also Supplementary Figure S3.

Figure 5. ***Ab initio* Ye-3-GFP envelope.**

(A) Experimentally determined SANS scattering data for Y-3-GFP and Ye-11-GFP. Q is the scattering vector and I represents the intensity. (B) Experimental Ye-3-GFP SANS data and predicted scattering curves for typical DAMMIN Ye-3-GFP models generated with no symmetry (P1) and P2 symmetry (P2) constraints over the Q-range used for DAMMIN modelling. Q is the scattering vector and I represents the intensity. (C) Averaged Ye-3-GFP model generated with P2 symmetry. (D) Superimposition of averaged P2 dYeSlc26A2 (grey beads) and Ye-3-GFP (green outline) models. See also Supplementary Figure S4.

Figure 6. *Ab initio* Ye-TM envelope.

(A) Experimentally determined SANS scattering data for Ye-TM. Q is the scattering vector and I represents the intensity. (B) Averaged Ye-TM model with no symmetry constraints (P1). (C) Experimental Ye-TM SANS data and predicted scattering curves for typical DAMMIN Ye-TM models generated with no symmetry (P1) and P222 symmetry (P222) constraints over the Q-range used for DAMMIN modelling. Q is the scattering vector and I represents the intensity. (D) Averaged Ye-TM models generated with P2 and P222 symmetry. (E) Superimposition of averaged Ye-TM models with no symmetry constraints (beads) and P222 symmetry (outline). (F) Superimposition of averaged P2 dYeSlc26A2 (grey beads) and P222 Ye-TM models (outline). See also Supplementary Figure S5.

Figure 7. Oligomerisation of Ye-TM.

(A) Models of the YeSlc26A2 STAS domain and a theoretical TM domain fitted into the envelope of dYeSlc26A2. (B) Theoretical dimerisation of TM domain dimers fitted inside the envelope of P222 Ye-TM model. (C) SDS-PAGE analysis of full length YeSlc26A2 and Ye-TM labelled with PEG-MAL. Number of labelled cysteines indicated by arrows. See also Supplementary Figure S6.

Figure 8. YeSlc26A2 Homology Model.

Homology model of YeSlc26A2 TM domain after all-atom molecular dynamics simulation. The core domain (light pink), containing the β-strand motif (red), and gate domain (light blue) are highlighted. (A) Viewed from the periplasm. (B) Viewed from the plane of the membrane, trypsin cleavable sites (green) are shown in stick representation. (C) YeSlc26A2 TM and STAS (cyan) models placed inside the dYeSlc26A2 SANS envelope (grey outline). Slc26A signature motif regions are highlighted in blue.

1 Saier Jr, M. H., Eng, B. H., Fard, S., Garg, J., Haggerty, D. A., Hutchinson, W. J., Jack, D. L., Lai, E. C., Liu, H. J., Nussinew, D. P., Omar, A. M., Pao, S. S., Paulsen, I. T., Quan, J. A., Sliwinski, M., Tseng, T. T., Wachi, S. and Young, G. B. (1999) Phylogenetic characterisation of novel transport protein families revealed by genome analysis. Biochem.Biophys.Acta. **1422**, 1-56

2 Wong, F. H., Chen, J. S., Reddy, V., Day, J. L., Shlykov, M. A., Wakabayashi, S. T. and Saier, M. H., Jr. (2012) The amino acid-polyamine-organocation superfamily. J Mol Microbiol Biotechnol. **22**, 105-113

3 Alper, S. L. and Sharma, A. K. (2013) The SLC26 gene family of anion transporters and channels. Mol Aspects Med. **34**, 494-515

4 Dorwart, M. R., Shcheynikov, N., Yang, D. and Muallem, S. (2008) The solute carrier 26 family of proteins in epithelial ion transport. Physiology (Bethesda). **23**, 104-114

5 Mount, D. B. and Romero, M. F. (2004) The SLC26 gene family of multifunctional anion exchangers. Pflugers Arch. **447**, 710-721

6 Yusa, A., Miyazaki, K., Kimura, N., Izawa, M. and Kannagi, R. (2010) Epigenetic silencing of the sulfate transporter gene DTDST induces sialyl Lewisx expression and accelerates proliferation of colon cancer cells. Cancer research. **70**, 4064-4073

7 Marzluf, G. A. (1997) Molecular genetics of sulfur assimilation in filamentous fungi and yeast. Annu.Rev.Microbiol. **51**, 73-96

8 Hawkesford, M. J. and De Kok, L. J. (2006) Managing sulphur metabolism in plants. Plant Cell Environ. **29**, 382-395

9 Price, G. D., Woodger, F. J., Badger, M. R., Howitt, S. M. and Tucker, L. (2004) Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. Proc.Natl.Acad.Sci.U.S.A. **101**, 18228-18233

10 Karinou, E., Compton, E. L., Morel, M. and Javelle, A. (2013) The Escherichia coli SLC26 homologue YchM (DauA) is a C(4)-dicarboxylic acid transporter. Mol Microbiol. **87**, 623-640

11 Shelden, M. C., Howitt, S. M. and Price, G. D. (2010) Membrane topology of the cyanobacterial bicarbonate transporter, BicA, a member of the SulP (SLC26A) family. Mol Membr Biol. **27**, 12-23

12 Price, G. D. and Howitt, S. M. (2011) The cyanobacterial bicarbonate transporter BicA: its physiological role and the implications of structural similarities with human SLC26 transporters. Biochem Cell Biol. **89**, 178-188

13 Babu, M., Greenblatt, J. F., Emili, A., Strynadka, N. C., Reithmeier, R. A. and Moraes, T. F. (2010) Structure of a SLC26 anion transporter STAS domain in complex with acyl carrier protein: implications for E. coli YchM in fatty acid metabolism. Structure. **18**, 1450-1462

14 Pasqualetto, E., Aiello, R., Gesiot, L., Bonetto, G., Bellanda, M. and Battistutta, R. (2010) Structure of the cytosolic portion of the motor protein prestin and functional role of the STAS domain in SLC26/SulP anion transporters. J Mol Biol. **400**, 448-462

15 Sharma, A. K., Ye, L., Baer, C. E., Shanmugasudaram, K., Alber, T., Alper, S. L. and Rigby, A. C. (2010) Solution structure of the guanine nucleotide-binding STAS domain of SLC26-related SulP protein Rv1739c from *M. tuberculosis*. J Biol Chem

16 Gorbunov, D., Sturlese, M., Nies, F., Kluge, M., Bellanda, M., Battistutta, R. and Oliver, D. (2014) Molecular architecture and the structural basis for anion interaction in prestin and SLC26 transporters. Nat Commun. **5**, 3622

17 Mio, K., Kubo, Y., Ogura, T., Yamamoto, T., Arisaka, F. and Sato, C. (2008) The motor protein prestin is a bullet-shaped molecule with inner cavities. J Biol Chem. **283**, 1137-1145

18 Detro-Dassen, S., Schanzler, M., Lauks, H., Martin, I., zu Berstenhorst, S. M., Nothmann, D., Torres-Salazar, D., Hidalgo, P., Schmalzing, G. and Fahlke, C. (2008) Conserved dimeric subunit stoichiometry of SLC26 multifunctional anion exchangers. J Biol Chem. **283**, 4177-4188

19 Wang, X., Yang, S., Jia, S. and He, D. Z. (2010) Prestin forms oligomer with four mechanically independent subunits. Brain Res. **1333**, 28-35

20 Compton, E. L., Karinou, E., Naismith, J. H., Gabel, F. and Javelle, A. (2011) Low resolution structure of a bacterial SLC26 transporter reveals dimeric stoichiometry and mobile intracellular domains. The Journal of biological chemistry. **286**, 27058-27067
21 Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. and de Gier, J. W. (2006) Optimization of membrane protein overexpression and purification using GFP fusions. Nat Methods. **3**, 303-313
22 Abulencia, A., Acosta, D., Adelman, J., Affolder, T., Akimoto, T., Albrow, M. G., Ambrose, D., Amerio, S., Amidei, D., Anastassov, A., Anikeev, K., Annovi, A., Antos, J., Aoki, M., Apollinari, G., Arguin, J. F., Arisawa, T., Artikov, A., Ashmanskas, W., Attal, A., Azfar, F., Azzi-Bacchetta, P., Azzurri, P., Bacchetta, N., Bachacou, H., Badgett, W., Barbaro-Galtieri, A., Barnes, V. E., Barnett, B. A., Baroiant, S., Bartsch, V., Bauer, G., Bedeschi, F., Behari, S., Belforte, S., Bellettini, G., Bellinger, J., Belloni, A., Ben Haim, E., Benjamin, D., Beretvas, A., Beringer, J., Berry, T., Bhatti, A., Binkley, M., Bisello, D., Blair, R. E., Blocker, C., Blumenfeld, B., Bocci, A., Bodek, A., Boisvert, V., Bolla, G., Bolshov, A., Bortoletto, D., Boudreau, J., Boveia, A., Brau, B., Bromberg, C., Brubaker, E., Budagov, J., Budd, H. S., Budd, S., Burkett, K., Busetto, G., Bussey, P., Byrum, K. L., Cabrera, S., Campanelli, M., Campbell, M., Canelli, F., Canepa, A., Carlsmith, D., Carosi, R., Carron, S., Casarsa, M., Castro, A., Catastini, P., Cauz, D., Cavalli-Sforza, M., Cerri, A., Cerrito, L., Chang, S. H., Chapman, J., Chen, Y. C., Chertok, M., Chiarelli, G., Chlachidze, G., Chlebana, F., Cho, I., Cho, K., Chokheli, D., Chou, J. P., Chu, P. H., Chuang, S. H., Chung, K., Chung, W. H., Chung, Y. S., Ciljak, M., Ciobanu, C. I., Ciocci, M. A., Clark, A., Clark, D., Coca, M., Compostella, G., Convery, M. E., Conway, J., Cooper, B., Copic, K., Cordelli, M., Cortiana, G., Crescioli, F., Cruz, A., Cuenca Almenar, C., Cuevas, J., Culbertson, R., Cyr, D., Daronco, S., D'Auria, S., D'Onofrio, M., Dagenhart, D., de Barbaro, P., De Cecco, S., Deisher, A., De Lentdecker, G., Dell'Orso, M., Delli Paoli, F., Demers, S., Demortier, L., Deng, J., Deninno, M., De Pedis, D., Derwent, P. F., Dionisi, C., Dittmann, J. R., DiTuro, P., Dorr, C., Donati, S., Donega, M., Dong, P., Donini, J., Dorigo, T., Dube, S., Ebina, K., Efron, J., Ehlers, J., Erbacher, R., Errede, D., Errede, S., Eusebi, R., Fang, H. C., Farrington, S., Fedorko, I., Fedorko, W. T., Feild, R. G., Feindt, M., Fernandez, J. P., Field, R., Flanagan, G., Flores-Castillo, L. R., Foland, A., Forrester, S., Foster, G. W., Franklin, M., Freeman, J. C., Furic, I., Gallinaro, M., Galyardt, J., Garcia, J. E., Garcia Sciveres, M., Garfinkel, A. F., Gay, C., Gerberich, H., Gerdes, D., Giagu, S., Giannetti, P., Gibson, A., Gibson, K., Ginsburg, C., Giokaris, N., Giolo, K., Giordani, M., Giromini, P., Giunta, M., Giurgiu, G., Glagolev, V., Glenzinski, D., Gold, M., Goldschmidt, N., Goldstein, J., Gomez, G., Gomez-Ceballos, G., Goncharov, M., Gonzalez, O., Gorelov, I., Goshaw, A. T., Gotra, Y., Goulianos, K., Gresele, A., Griffiths, M., Grinstein, S., Grosso-Pilcher, C., Grundler, U., Guimaraes da Costa, J., Gunay-Unalan, Z., Haber, C., Hahn, S. R., Hahn, K., Halkiadakis, E., Hamilton, A., Han, B. Y., Han, J. Y., Handler, R., Happacher, F., Hara, K., Hare, M., Harper, S., Harr, R. F., Harris, R. M., Hatakeyama, K., Hauser, J., Hays, C., Heijboer, A., Heinemann, B., Heinrich, J., Herndon, M., Hidas, D., Hill, C. S., Hirschbuehl, D., Hocker, A., Holloway, A., Hou, S., Houlden, M., Hsu, S. C., Huffman, B. T., Hughes, R. E., Huston, J., Incandela, J., Introzzi, G., Iori, M., Ishizawa, Y., Ivanov, A., Iyutin, B., James, E., Jang, D., Jayatilaka, B., Jeans, D., Jensen, H., Jeon, E. J., Jindariani, S., Jones, M., Joo, K. K., Jun, S. Y., Junk, T. R., Kamon, T., Kang, J., Karchin, P. E., Kato, Y., Kemp, Y., Kephart, R., Kerzel, U., Khotilovich, V., Kilminster, B., Kim, D. H., Kim, H. S., Kim, J. E., Kim, M. J., Kim, S. B., Kim, S. H., Kim, Y. K., Kirsch, L., Klimenko, S., Klute, M., Knuteson, B., Ko, B. R., Kobayashi, H., Kondo, K., Kong, D. J., Konigsberg, J., Korytov, A., Kotwal, A. V., Kovalev, A., Kraan, A., Kraus, J., Kravchenko, I., Kreps, M., Kroll, J., Krumnack, N., Kruse, M., Krutelyov, V., Kuhlmann, S. E., Kusakabe, Y., Kwang, S., Laasanen, A. T., Lai, S., Lami, S., Lammel, S., Lancaster, M., Lander, R. L., Lannon, K., Lath, A., Latino, G., Lazzizzera, I., LeCompte, T., Lee, J., Lee, Y. J., Lee, S. W., Lefevre, R., Leonardo, N., Leone, S., Levy, S., Lewis, J. D., Lin, C., Lin, C. S., Lindgren, M., Lipeles, E., Lister, A., Litvintsev, D. O., Liu, T., Lockyer, N. S., Loginov, A., Loreti, M., Loverre, P., Lu, R. S., Lucchesi, D., Lujan, P., Lukens, P., Lungu, G., Lyons, L., Lys, J., Lysak, R., Lytken, E., Mack, P., MacQueen, D., Madrak, R., Maeshima, K., Maki, T., Maksimovic, P., Malde, S., Manca, G., Margaroli, F., Marginean, R., Marino, C., Martin, A., Martin, V., Martinez, M., Maruyama, T., Mastrandrea, P., Matsunaga, H., Mattson, M. E., Mazini, R., Mazzanti, P., McFarland, K. S., McIntyre, P., McNulty, R., Mehta, A., Menzemer, S., Menzione, A., Merkel, P., Mesropian, C., Messina, A., von der Mey, M., Miao, T., Miladinovic, N., Miles, J., Miller, R., Miller, J. S., Mills, C., Milnik, M., Miquel, R., Mitra, A., Mitselmakher, G., Miyamoto, A., Moggi, N., Mohr, B.,

Moore, R., Morello, M., Movilla Fernandez, P., Mulmenstadt, J., Mukherjee, A., Muller, T., Mumford, R., Murat, P., Nachtman, J., Naganoma, J., Nahn, S., Nakano, I., Napier, A., Naumov, D., Necula, V., Neu, C., Neubauer, M. S., Nielsen, J., Nigmanov, T., Nodulman, L., Norniella, O., Nurse, E., Ogawa, T., Oh, S. H., Oh, Y. D., Okusawa, T., Oldeman, R., Orava, R., Osterberg, K., Pagliarone, C., Palencia, E., Paoletti, R., Papadimitriou, V., Paramonov, A. A., Parks, B., Pashapour, S., Patrick, J., Pauletta, G., Paulini, M., Paus, C., Pellett, D. E., Penzo, A., Phillips, T. J., Piacentino, G., Piedra, J., Pinera, L., Pitts, K., Plager, C., Pondrom, L., Portell, X., Poukhov, O., Pounder, N., Prakoshyn, F., Pronko, A., Proudfoot, J., Ptohos, F., Punzi, G., Pursley, J., Rademacker, J., Rahaman, A., Rakitin, A., Rappoccio, S., Ratnikov, F., Reisert, B., Rekovic, V., van Remortel, N., Renton, P., Rescigno, M., Richter, S., Rimondi, F., Ristori, L., Robertson, W. J., Robson, A., Rodrigo, T., Rogers, E., Rolli, S., Roser, R., Rossi, M., Rossin, R., Rott, C., Ruiz, A., Russ, J., Rusu, V., Saarikko, H., Sabik, S., Safonov, A., Sakumoto, W. K., Salamanna, G., Salto, O., Saltzberg, D., Sanchez, C., Santi, L., Sarkar, S., Sartori, L., Sato, K., Savard, P., Savoy-Navarro, A., Scheidle, T., Schlabach, P., Schmidt, E. E., Schmidt, M. P., Schmitt, M., Schwarz, T., Scodellaro, L., Scott, A. L., Scribano, A., Scuri, F., Sedov, A., Seidel, S., Seiya, Y., Semenov, A., Sexton-Kennedy, L., Sfiligoi, I., Shapiro, M. D., Shears, T., Shepard, P. F., Sherman, D., Shimojima, M., Shochet, M., Shon, Y., Shreyber, I., Sidoti, A., Sinervo, P., Sisakyan, A., Sjolin, J., Skiba, A., Slaughter, A. J., Sliwa, K., Smith, J. R., Snider, F. D., Snihur, R., Soderberg, M., Soha, A., Somalwar, S., Sorin, V., Spalding, J., Spezziga, M., Spinella, F., Spreitzer, T., Squillacioti, P., Stanitzki, M., Staveris-Polykalas, A., St Denis, R., Stelzer, B., Stelzer-Chilton, O., Stentz, D., Strologas, J., Stuart, D., Suh, J. S., Sukhanov, A., Sumorok, K., Sun, H., Suzuki, T., Taffard, A., Takashima, R., Takeuchi, Y., Takikawa, K., Tanaka, M., Tanaka, R., Tanimoto, N., Tecchio, M., Teng, P. K., Terashi, K., Tether, S., Thom, J., Thompson, A. S., Thomson, E., Tipton, P., Tiwari, V., Tkaczyk, S., Toback, D., Tokar, S., Tollefson, K., Tomura, T., Tonelli, D., Tonnesmann, M., Torre, S., Torretta, D., Tourneur, S., Trischuk, W., Tsuchiya, R., Tsuno, S., Turini, N., Ukegawa, F., Unverhau, T., Uozumi, S., Usynin, D., Vaiciulis, A., Vallecorsa, S., Varganov, A., Vataga, E., Velev, G., Veramendi, G., Veszpremi, V., Vidal, R., Vila, I., Vilar, R., Vine, T., Vollrath, I., Volobouev, I., Volpi, G., Wurthwein, F., Wagner, P., Wagner, R. G., Wagner, R. L., Wagner, W., Wallny, R., Walter, T., Wan, Z., Wang, S. M., Warburton, A., Waschke, S., Waters, D., Wester, W. C., 3rd, Whitehouse, B., Whiteson, D., Wicklund, A. B., Wicklund, E., Williams, G., Williams, H. H., Wilson, P., Winer, B. L., Wittich, P., Wolbers, S., Wolfe, C., Wright, T., Wu, X., Wynne, S. M., Yagil, A., Yamamoto, K., Yamaoka, J., Yamashita, T., Yang, C., Yang, U. K., Yang, Y. C., Yao, W. M., Yeh, G. P., Yoh, J., Yorita, K., Yoshida, T., Yu, G. B., Yu, I., Yu, S. S., Yun, J. C., Zanello, L., Zanetti, A., Zaw, I., Zetti, F., Zhang, X., Zhou, J. and Zucchelli, S. (2006) Observation of B(s)0-->K+ K- and measurements of branching fractions of charmless two-body decays of B0 and B(s)0 mesons in pp collisions at square root of s = 1.96 TeV. Phys Rev Lett. **97**, 211802

23 Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. **76**, 2879-2886

24 Whitmore, L. and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers. **89**, 392-400

25 Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B. and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. **29**, 845-854

26 Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. **78**, 1950-1958

27 Cordomí, A., Caltabiano, G. and Pardo, L. (2012) Membrane protein simulations using AMBER force field and Berger lipid parameters. . J. Chem. Theory Comput. **8**, 948-958

28 Berendsen, H. J. C., Grigera, J. R. and Straatsma, T. P. (1987) The missing term in effective pair potentials. J. Phys. Chem. . **91**, 6269-6271

29 Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. R. H. J. and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. J. Phys. Chem. . **81**, 3684

30 Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity rescaling. J Chem Phys. **126**, 014101

31 Loland, C. J. (2014) The use of LeuT as a model in elucidating binding sites for substrates and inhibitors in neurotransmitter transporters. Biochimica et biophysica acta
32 Vandenberg, R. J. and Ryan, R. M. (2013) Mechanisms of glutamate transport. Physiological reviews. **93**, 1621-1657
33 Volkov, V. V. and Svergun, D. I. (2003) Uniqueness of *ab intio* shape determination in small-angle scattering. J. Appl.Cryst. **36**, 860-864
34 Bode, B. E., Margraf, D., Plackmeyer, J., Durner, G., Prisner, T. F. and Schiemann, O. (2007) Counting the monomers in nanometer-sized oligomers by pulsed electron-electron double resonance. J Am Chem Soc. **129**, 6736-6745
35 Soding, J., Biegert, A. and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. **33**, W244-248
36 Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. and Sali, A. (2006) Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics. **Chapter 5**, Unit 5 6
37 Hallworth, R. and Nichols, M. G. (2012) Prestin in HEK cells is an obligate tetramer. J Neurophysiol. **107**, 5-11
38 Zheng, J., Du, G. G., Anderson, C. T., Keller, J. P., Orem, A., Dallos, P. and Cheatham, M. (2006) Analysis of the oligomeric structure of the motor protein prestin. Journal of Biological Chemistry. **281**, 19916-19924
39 Sharma, A. K., Rigby, A. C. and Alper, S. L. (2011) STAS domain structure and function. Cell Physiol Biochem. **28**, 407-422
40 Vastermark, A. and Saier, M. H., Jr. (2014) Evolutionary relationship between 5+5 and 7+7 inverted repeat folds within the amino acid-polyamine-organocation superfamily. Proteins. **82**, 336-346

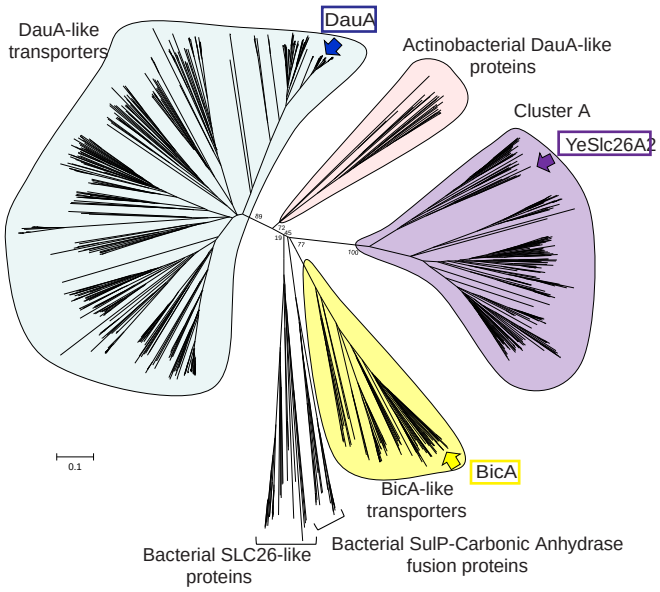

DauA-like
transporters
DauA
Actinobacterial DauA-like
proteins
Cluster A
YeSlc26A2
89
77
100
0.1
BicA
BicA-like
transporters
Bacterial SLC26-like
proteins
Bacterial SulP-Carbonic Anhydrase
fusion proteins

Figure 1

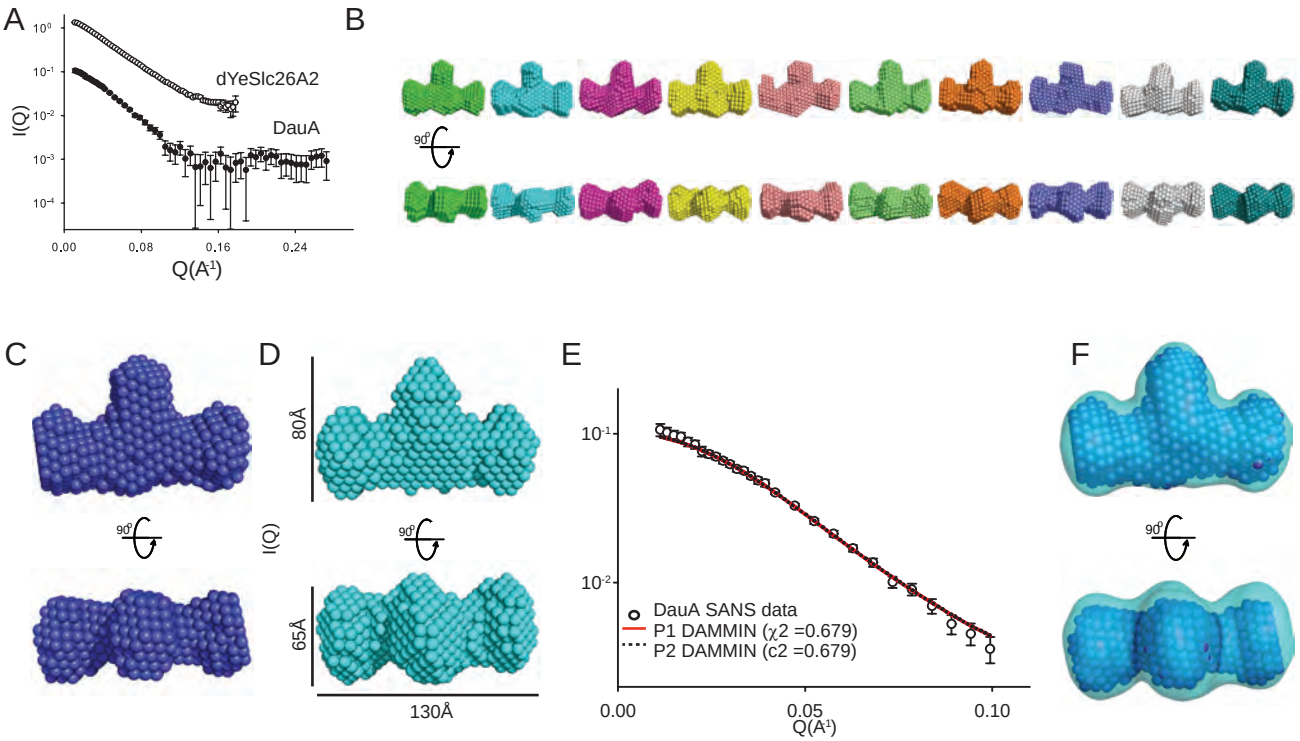

A
B
C
D
E
F
dYeSlc26A2
DauA
I(Q)
Q(A¹)
90°
80Å
65Å
130Å
DauA SANS data
P1 DAMMIN (χ2 =0.679)
P2 DAMMIN (c2 =0.679)

Figure 2

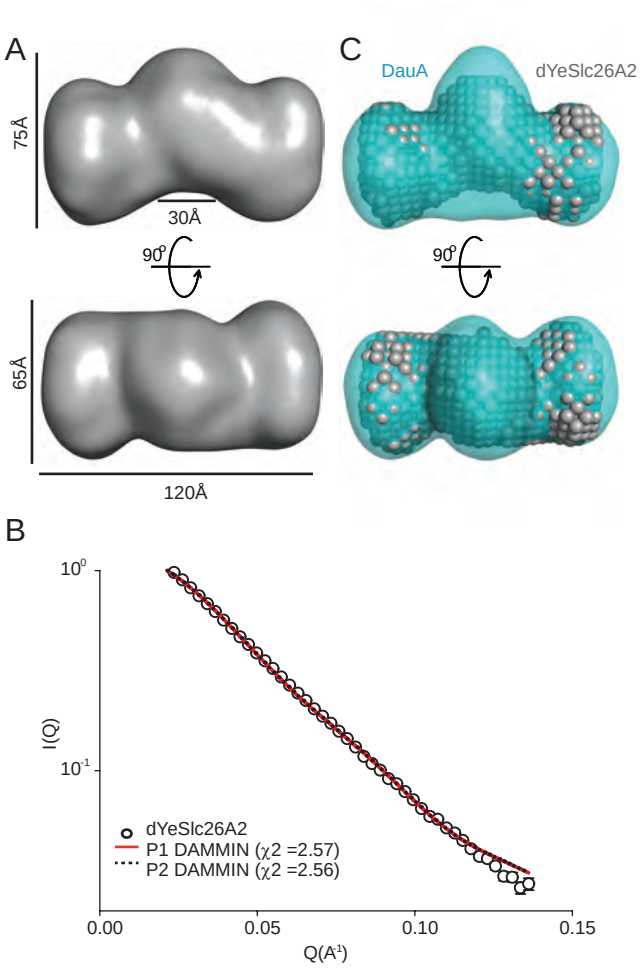
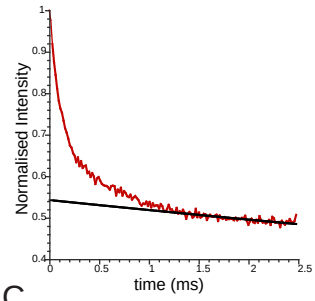

A
75Å
30Å
90°
65Å
120Å
C
DauA
dYeSlc26A2
90°
B
10^0
10^-1
I(Q)
dYeSlc26A2
P1 DAMMIN (χ2 =2.57)
P2 DAMMIN (χ2 =2.56)
0.00
0.05
0.10
0.15
Q(A^-1)

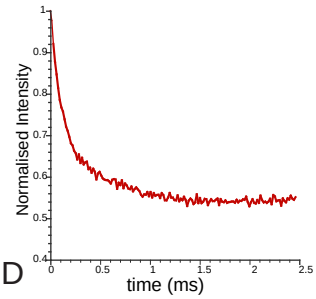
Figure 3

A

B



C



D

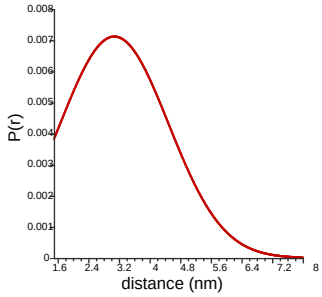


Figure 4

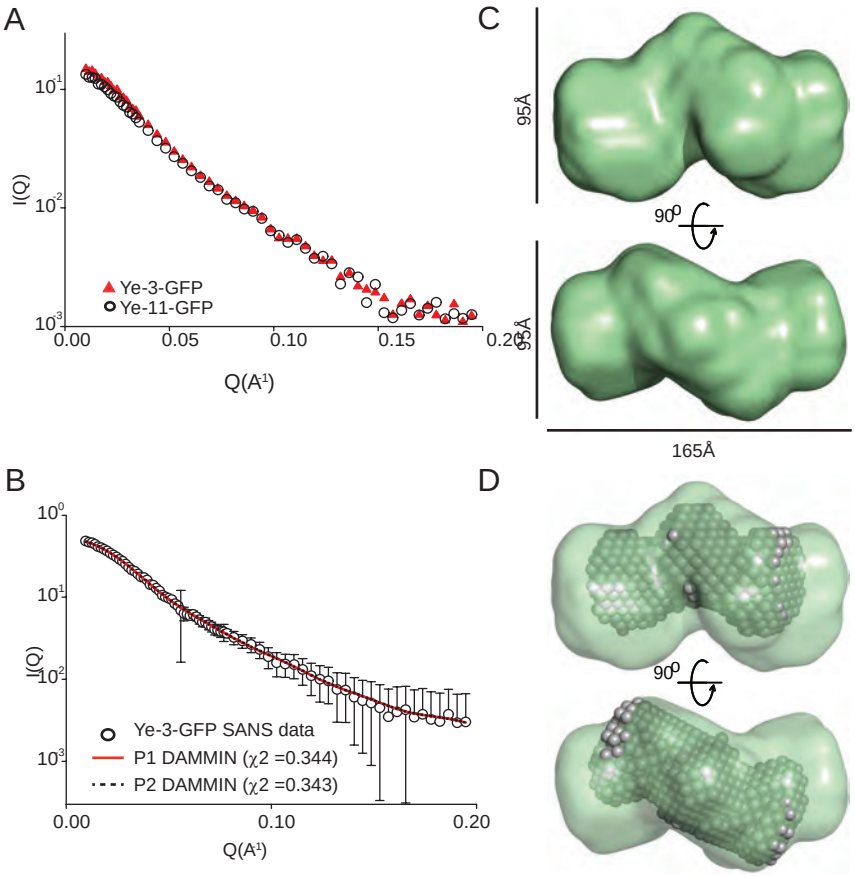

A
C
B
D
I(Q)
Q(A-1)
Ye-3-GFP
Ye-11-GFP
95Å
90°
165Å
Ye-3-GFP SANS data
P1 DAMMIN (χ2 =0.344)
P2 DAMMIN (χ2 =0.343)

Figure 5

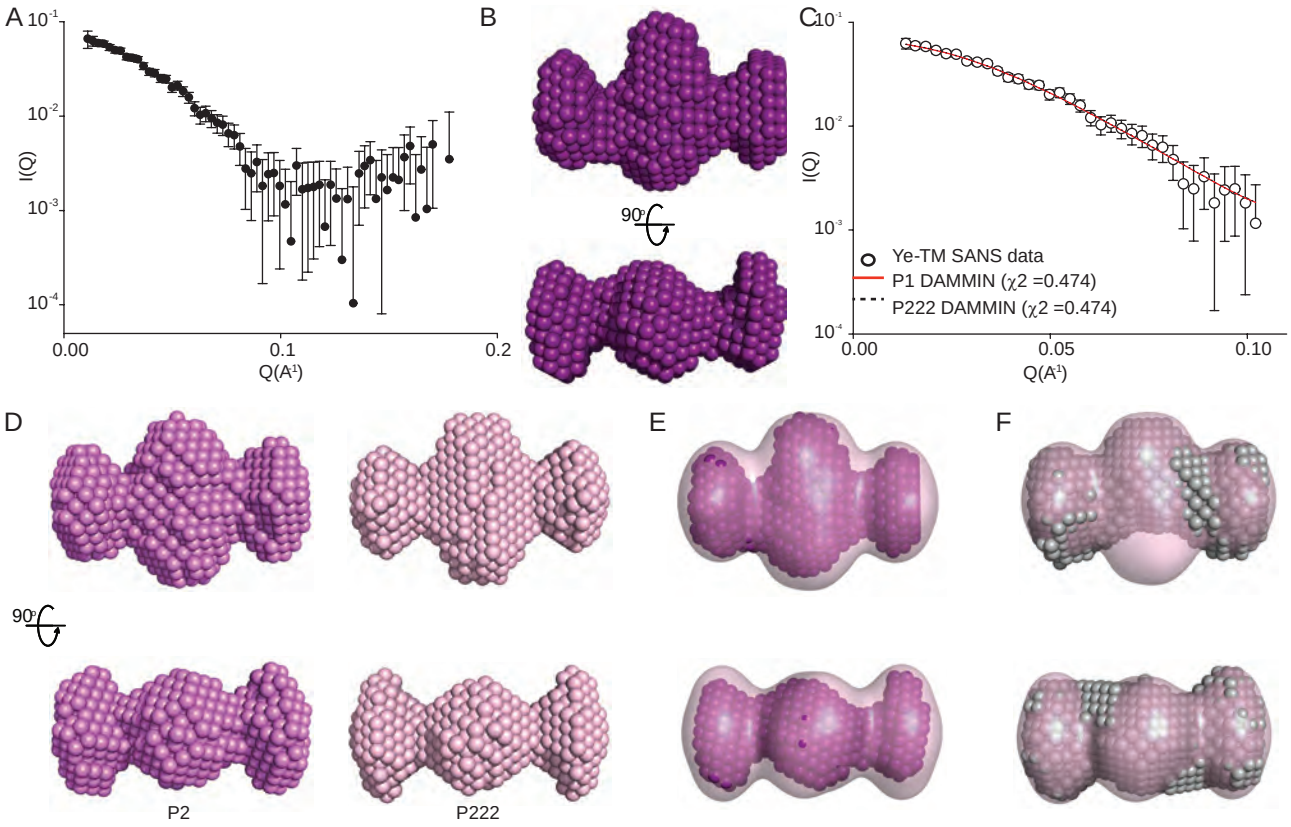

A
B
C
I(Q)
Q(A¹)
90°
Ye-TM SANS data
P1 DAMMIN (χ2 =0.474)
P222 DAMMIN (χ2 =0.474)
D
E
F
P2
P222

Figure 6

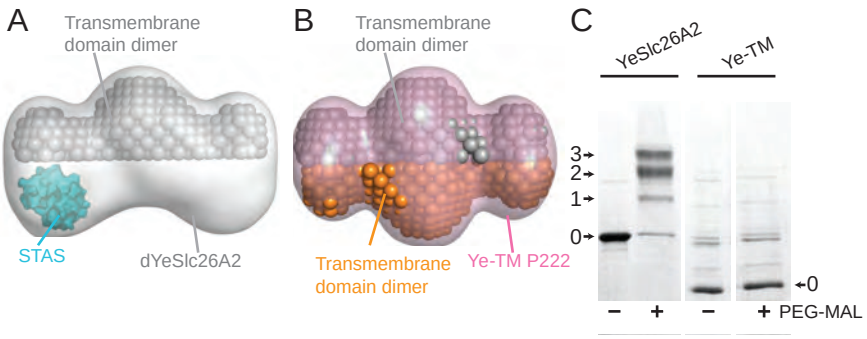
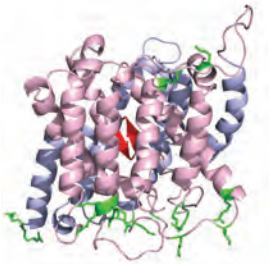

A
Transmembrane domain dimer
STAS
dYeSlc26A2
B
Transmembrane domain dimer
Transmembrane domain dimer
Ye-TM P222
C
YeSlc26A2
Ye-TM
3
2
1
0
0
− + − + PEG-MAL

Figure 7

A

B



C

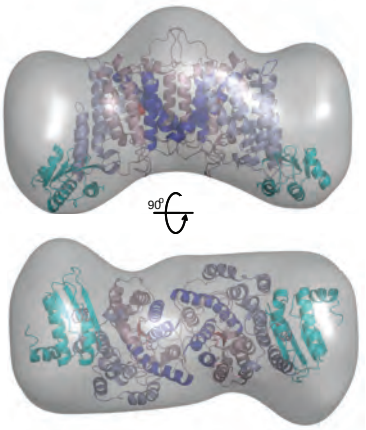


Figure 8

Sample	Rg (Å) / Guinier	Rg (Å) / GNOM
DauA	47.7 +/- 3.6	41.4
dYeSlc26A2	46.7 +/- 0.4	41.5
Ye-3-GFP	51.9 +/- 2.2	51.7
Ye-11-GFP	53.1 +/- 2.4	-
YeSTM	39.8 +/- 3.8	38.9

Figure S1. SANS characterisation of DauA.
Distance distribution P(R) used in the modelling of DauA *ab initio* envelopes.

Figure S2. SANS characterisation of dYeSlc26A2.
(A) Distance distribution P(R) used in the modelling of dYeSlc26A2 *ab initio* envelopes. (B) Averaged P1 dYeSlc26A2 model with no symmetry constraints. (C) Superimposition of P1 dYeSlc26A2 (beads) and P2 dYeSlc26A2 (outline) models.

Figure S3. Biochmical characterisation of Ye-D416C
(A) CW EPR spectra for Ye-D416C (B)Ye-D416C and YeSlc26A2 CD spectra.

Figure S4. SANS characterisation of Ye-3-GFP.
(A) Distance distribution P(R) used in the modelling of Ye-3-GFP *ab initio* envelopes. (B) Averaged P1 Ye-3-GFP model with no symmetry constraints. (C) Superimposition of P1 Ye-3-GFP (beads) and P2 Ye-3-GFP (outline) models.

Figure S5. SANS characterisation of Ye-TM
(A) Distance distribution P(R) used in the modelling of Ye-3-GFP *ab initio* envelopes. (B) Experimental Ye-TM SANS data and predicted scattering curves for typical DAMMIN Ye-TM models generated with no symmetry (P1) and P2 symmetry (P2) constraints.

Figure S6. Accessibility of YeSlc26A2 cysteine residues.
MAL-PEG labelling of YeSlc26A2 cysteineless (No cys) and single cysteine (C15, C51 and 272) mutants.

Figure S7. YeSlc26A2 and DauA molecular dynamics simulations.
100ns all-atom molecular dynamics simulations of homology models in a fully solvated, POPC bilayer. (A) Overlay of final YeSlc26A2 model after all-atom molecular dynamics simulations in fully solvated POPC bilayer (rainbow: N-terminus blue to C-terminus red) overlaid with UraA (grey; pdb: 3QE7). (B) Three independent RMSD trajectories for the YeSlc26A2 homology model over the course of molecular dynamic simulations. (C) Overlay of final DauA and YeSlc26A2 models. (D) Three independent RMSD trajectories for the DauA homology model over the course of the molecular dynamic simulations. (E) Final YeSlc26A2 model with conserved residues shown as sphere representation. The core domain is shown in light pink, the gate domain in light blue, the β-strand motif (red) and helices containing the signature motif (dark blue).

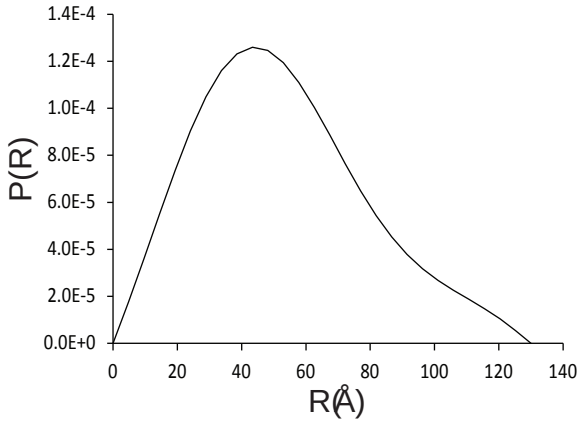

1.4E-4
1.2E-4
1.0E-4
8.0E-5
6.0E-5
4.0E-5
2.0E-5
0.0E+0
P(R)
0
20
40
60
80
100
120
140
R(Å)

Figure S1

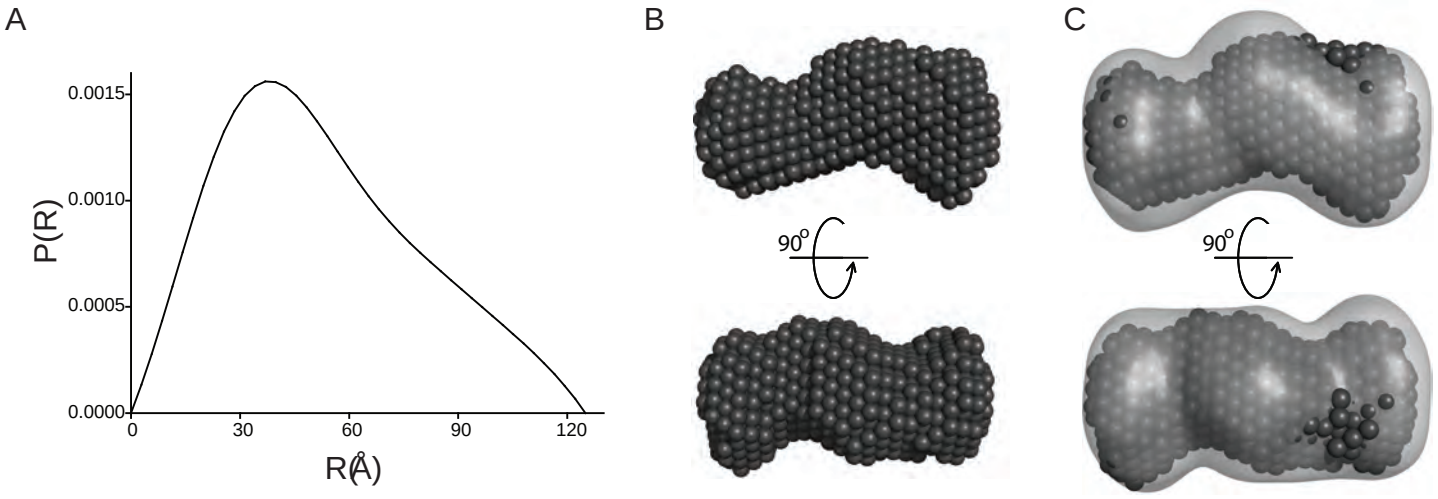
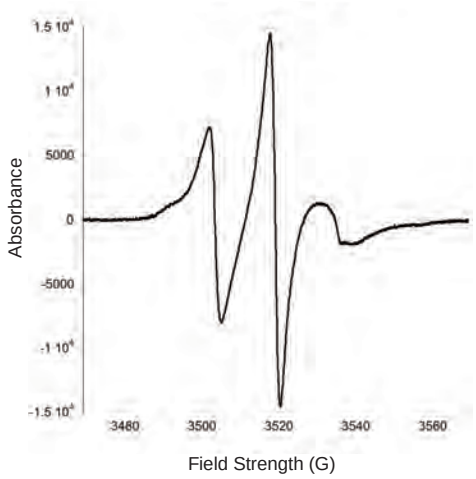


Figure S2

A


1.5 10⁴
1 10⁴
5000
0
-5000
-1 10⁴
-1.5 10⁴
Absorbance
3480
3500
3520
3540
3560
Field Strength (G)

B

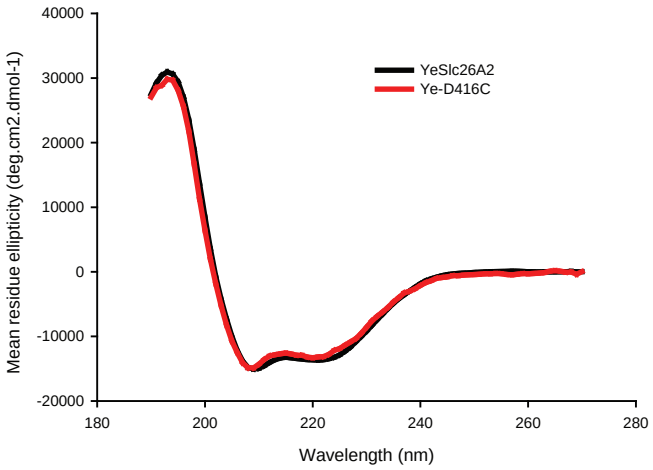

40000
30000
20000
10000
0
-10000
-20000
Mean residue ellipticity (deg.cm2.dmol-1)
YeSlc26A2
Ye-D416C
180
200
220
240
260
280
Wavelength (nm)

Figure S3

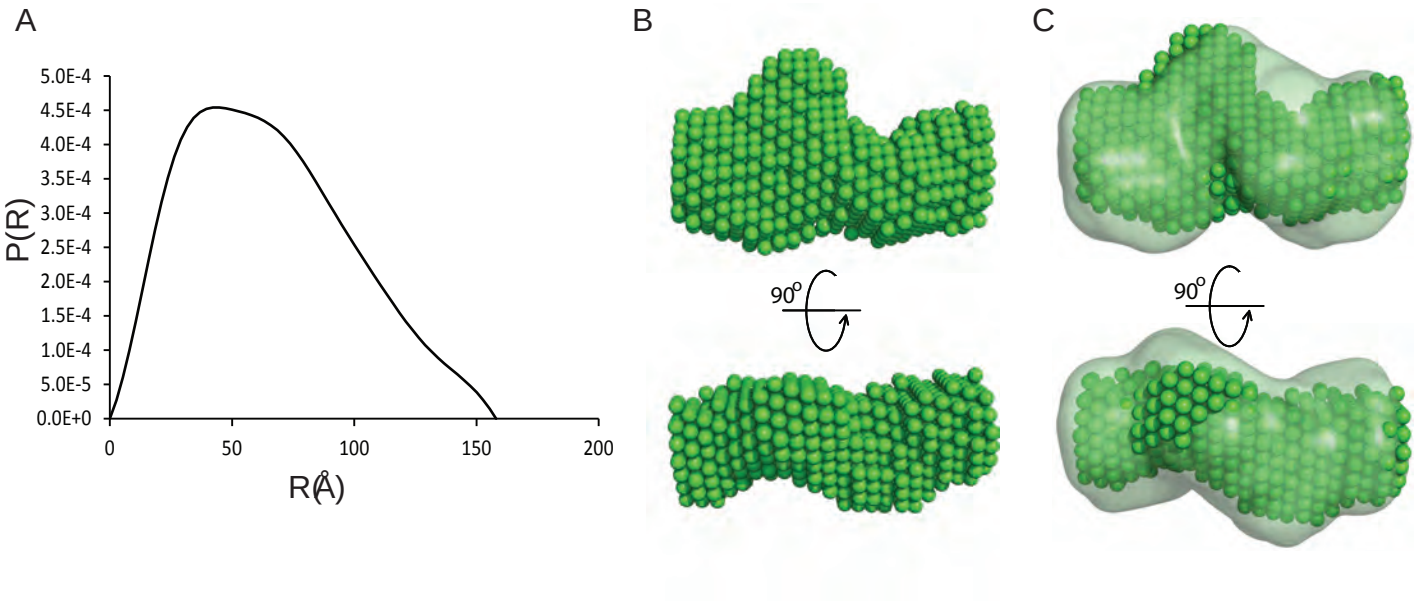
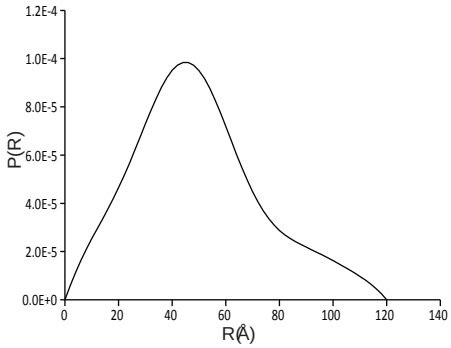

A
5.0E-4
4.5E-4
4.0E-4
3.5E-4
3.0E-4
2.5E-4
2.0E-4
1.5E-4
1.0E-4
5.0E-5
0.0E+0
P(R)
0
50
100
150
200
R(Å)
B
90°
C
90°

Figure S4

A


1.2E-4
1.0E-4
8.0E-5
6.0E-5
4.0E-5
2.0E-5
0.0E+0
P(R)
0
20
40
60
80
100
120
140
R(Å)

B

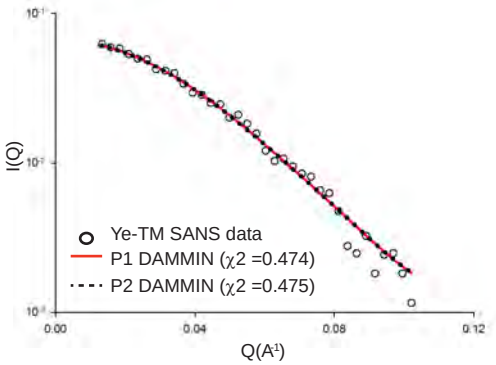

I(Q)
Ye-TM SANS data
P1 DAMMIN (χ2 =0.474)
P2 DAMMIN (χ2 =0.475)
0.00
0.04
0.08
0.12
Q(A-1)

Figure S5

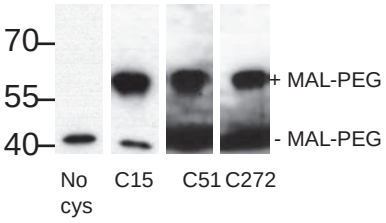
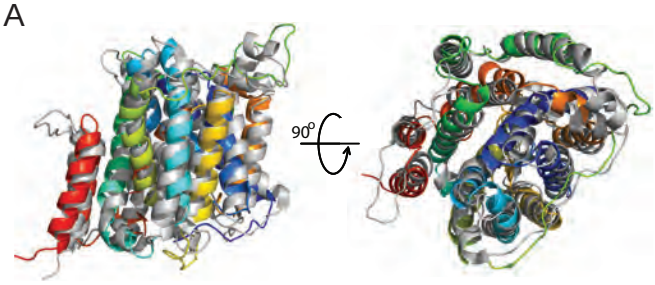
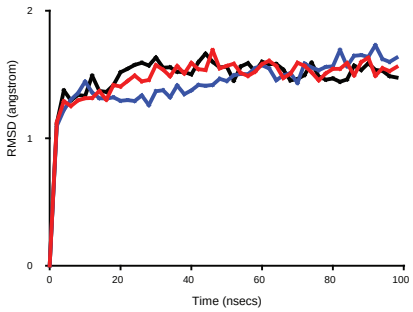


Figure S6

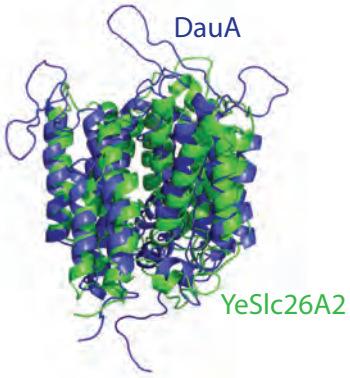
A



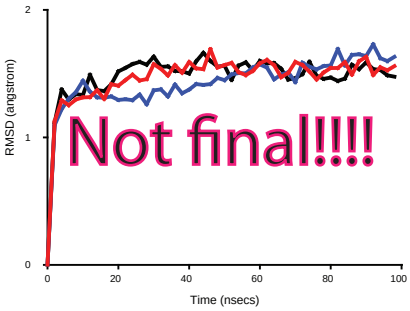
B



C



D



E

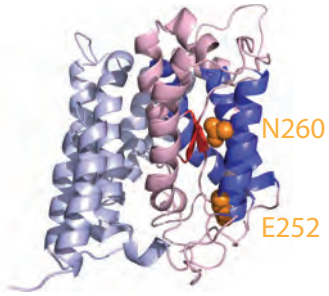


Figure S7