Published paper
Muller, A., Severi, E., Mulligan, C., Watts, A.G., Kelly, D.J., Wilson, K.S., Wilkinson, A.J. and Thomas, G.H. (2006) *Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus influenzae.* Journal of Biological Chemistry, 281 (31). pp. 22212-22222.

JBC Papers in Press. Published on May 15, 2006 as Manuscript M603463200
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M603463200

CONSERVATION OF STRUCTURE AND MECHANISM IN PRIMARY AND SECONDARY TRANSPORTERS EXEMPLIFIED BY SiaP, A SIALIC ACID-BINDING VIRULENCE FACTOR FROM *HAEMOPHILUS INFLUENZAE*.

Axel Müller[1], Emmanuele Severi[2], Christopher Mulligan[2], Andrew G Watts[1$], David J Kelly[3], Keith S Wilson[1], Anthony J Wilkinson[1*] & Gavin H Thomas[2*]

[1]Structural Biology Laboratory, Department of Chemistry and [2]Department of Biology, University of York, York YO10 5YW, UK,[3]Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield. S10 2TN, UK.
[$]Current address: Department of Pharmacy and Pharmacology, University of Bath, BA2 7AY, UK.

*Corresponding authors. Tel. +44 1904 328678. Email ght2@york.ac.uk; Tel. +44 1904 328261. Email ajw@ysbl.york.ac.uk;

Running title: Structure of a sialic acid binding protein, SiaP

Extracytoplasmic solute receptors (ESRs) are important components of solute uptake systems in bacteria, having been studied extensively as parts of ATP-binding cassette (ABC) transporters. Herein we report the first crystal structure of an ESR protein from a functionally characterised electrochemical ion gradient-dependent secondary transporter. This protein, SiaP, forms part of a tripartite ATP-independent periplasmic (TRAP) transporter specific for sialic acid in *Haemophilus influenzae*. Surprisingly, the structure reveals an overall topology similar to ABC ESR proteins, which is not apparent from the sequence, demonstrating that primary and secondary transporters can share a common structural component. The structure of SiaP in the presence of the sialic-acid analogue Neu5Ac2en reveals the ligand bound in a deep cavity with its carboxylate group forming a salt bridge with a highly conserved Arg residue. Sialic acid binding, which obeys simple bimolecular association kinetics as determined by stopped-flow fluorescence spectroscopy, is accompanied by domain closure about a hinge region and the kinking of an α-helix hinge component. The structure provides insight into the evolution, mechanism and substrate specificity of ESR-dependent secondary transporters that are widespread in prokaryotes.

The uptake of solutes into bacterial cells is critical for growth and survival in every environment and is catalysed by a variety of different protein-mediated transport systems. One common feature of uptake systems, including the well-studied ATP-binding cassette (ABC[1]) transporters is an extracytoplasmic solute receptor (ESR) (often also known as a periplasmic binding protein) which captures the substrate for the transporter and delivers it to the membrane subunits (1). Structures have been determined for many ESR proteins from ABC transporters and their mechanism of ligand binding is so well established that these proteins are now being used in a host of biotechnological applications (2).

The ABC systems are examples of primary active transporters, so called because they hydrolyse ATP directly to energise transport (3). These differ from another major grouping, the secondary active transporters, like *Escherichia coli* lactose permease, so-called because many of them use the membrane potential to energise uptake and direct ATP hydrolysis is not involved. The use of an ESR protein, which endows the transporter with high-affinity for its substrate, was for a long time believed to be exclusive to the ABC transporters, but biochemical studies of a C_4-dicarboxylate uptake system from *Rhodobacter capsulatus* led to the discovery of a novel family of ESR-dependent secondary transporters, the so called tripartite ATP-independent periplasmic (TRAP) transporters (4-8). These transporters contain two membrane protein components, the larger of which contains 12 predicted transmembrane (TM) helices. This subunit is a member of the ion transporter (IT) superfamily (6;9) and probably forms the translocation channel. The smaller membrane component of 4 TM helices has an unknown but essential function (7). Microbial genome sequencing has revealed that the TRAP transporters are widespread in the prokaryotic world (10) and known substrates now include sialic acid, ectoine, 2,3-diketo-

L-gulonate and pyruvate in addition to C_4-dicarboxylates (11-14). We have recently characterised the sialic acid TRAP transporter from the human pathogen *Haemophilus influenzae* and demonstrated that this transporter is essential for uptake of sialic acid (Neu5Ac) in this bacterium. Neu5Ac is an important host-acquired molecule that is used by the bacterium to modify its lipopolysaccharide (LPS) to make it appear as 'self' and evade the innate immune response (15). Deletion of the TRAP transporter results in loss of LPS sialylation and serum resistance in *H. influenzae* Rd (13), a phenotype also observed recently in non-typeable strains of *H. influenzae* (16) and in the related animal pathogen *Pasteurella multocida* (17). These were the first reports of TRAP transporters having a role in virulence and highlight the importance of a greater understanding of the function and mechanism of these systems in prokaryotes.

The sialic acid binding protein SiaP is a member of the DctP protein family, named after the first characterised TRAP ESR protein that binds C_4-dicarboxylates (4). This is the major family of ESR proteins found in TRAP transport systems (10). Given the potential importance of TRAP transporters in the biology of prokaryotes but the paucity of information on them, we solved the structure of SiaP at 1.7 Å in an unliganded form and also at 2.2 Å in complex with the sialic acid analogue, 2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid (Neu5Ac2en). Our study provides important new information on sialic acid transport and insight into the function and evolution of this novel family of ESR-dependent secondary transporters.

EXPERIMENTAL PROCEDURES

Expression and purification of SiaP

The SiaP protein was purified from *E. coli* using a modification of the methods described in Severi *et al*. (13). Cells of *E. coli* BL21(DE3) pLysS pGTY3 were grown in 5 ml LB (Lennox broth) for 6 hours, washed in M9 minimal medium (18), and used to inoculate 50 ml M9 minimal medium for overnight growth at 37 ºC. This was used to inoculate 1 l of M9 minimal medium at 25 ºC. Cells were allowed to grow to an OD_{650} of 0.3-0.4 before inducing expression with 1 mM IPTG followed by overnight incubation. Cells were washed in ice-cold 50 mM Tris-HCl pH 8, incubated in SET buffer (0.5 M sucrose, 5 mM EDTA in 50 mM Tris-HCl pH 8, 600 µg/ml lysozyme) for 1 hour at 30 ºC and the periplasmic fraction was then clarified by centrifugation and dialysed against 50 mM Tris-HCl pH 8 containing 1.5 M $(NH_4)_2SO_4$. SiaP was purified by FPLC using a hydrophobic interaction column followed by size exclusion chromatography using a G75 sepharose column as described previously (13). Protein concentration was determined from the absorbance at 280 nm using a molar absorption coefficient for SiaP of 23840 M^{-1} cm^{-1}. The correct cleavage of the signal peptide (first 23 amino acids) and the absence of pre-bound Neu5Ac was confirmed by electrospray mass spectrometry (13). For preparation of the selenomethionine (SeMet) derivative of SiaP, the protein was expressed from a 1 l culture as described (19) and purified to ~ 95% homogeneity using a single anion exchange step (MonoQ).

Crystallisation

For crystallisation, SiaP was concentrated to 30 mg/ml in 20 mM Tris-HCl pH 8, 150 mM NaCl in the presence or absence of 5 mM Neu5Ac2en and 5 mM zinc acetate. Crystallisation experiments utilised the vapour diffusion method and a MOSQUITO nanolitre dispensing robot to set up sitting drops. Three crystal forms were analysed. Form 1 crystals belonging to space group $P2_12_12$ were grown from drops made up of 150 nl of SeMet-substituted SiaP and 150 nl of 100 mM Tris-HCl, pH 8.0, 20 % PEG 6000 and 10 mM zinc acetate. Form 2 crystals of native SiaP belonging to space group *I222* grew under identical conditions. Form 3 crystals belonging to space group *C2* were grown from drops made up from 150 nl of 100 mM Tris-HCl pH 8.5, 0.2 M magnesium chloride and 25 % PEG 3350. Even though SiaP is not zinc-dependent, no crystals appeared in the absence of this metal.

Data collection and structure solution

Three wavelength data were collected from the Form 1 crystals together with single wavelength data from the Form 2 and 3 crystals at the ESRF, Grenoble, on beamline BM14 (Table 1). The Form 1 SeMet crystals diffracted to 2.6 Å with high values of R_{merge} in the outer ranges. Although the data beyond 2.9 Å were very weak, they proved to be essential for successful phasing. Prior to data collection, it had been expected that the native and SeMet crystals would be isomorphous and that the isomorphous and anomalous components could be combined in the phasing procedure. Unfortunately this proved not to be so.

The programs SHELXC and SHELXD (20) readily found 14 of the 16 expected Se atoms in the asymmetric unit of the crystal using the combined MAD signal. However, the resulting 2.6 Å resolution map was difficult to interpret and simple application of the ARP/wARP suite (21) produced a model with a large number of disconnected peptides and scarcely any of the sequence docked into the density. The programme RESOLVE (22) produced a model consisting of about half of the protein backbone but still with very few side chains docked successfully. The breakthrough came via the use of the experimental phase probability distributions in terms of the Hendrickson-Lattman coefficients as restraints during the ARP/wARP-REFMAC rebuilding giving a model with more than 570 of the expected 612 residues and with most of the side chain correctly assigned. This model was used for molecular replacement with the Form 2 native data and MOLREP (23) subsequently provided an essentially complete model using ARP/wARP-REFMAC. The Form 2 crystal structure was in turn used as a search model in molecular replacement calculations with the Form 3 crystal data in the programme MOLREP leading to the identification of four molecules in the asymmetric unit. For three of these molecules A, C and D the maps were of satisfactory quality. It became apparent that relative domain movements had taken place in molecule B because the calculated maps satisfactorily covered only the amino terminal domain I. A mask was therefore applied to the 3.5 molecules which fitted the maps well, and further calculations using the program MOLREP using the carboxyl domain II as a search model completed molecular replacement. The model was refined by iterative cycles of REFMAC (24) interspersed with manual modelling in COOT (25). Refinement statistics for the Form 2 and Form 3 structures are given in Table 1. Coordinates and structure factors have been deposited with the Protein Data Bank (Unliganded structure, 2CEY, Neu5Ac2en structure, 2CEX).

Steady-state and stopped flow fluorescence spectroscopy

Steady-state protein fluorescence studies were performed as previously described (13) unless specifically outlined in the text. The K_d values were determined from at least 4 titrations, except that for dNeu5Ac which was determined from 3 titrations. Stopped-flow kinetic measurements were made using an Applied Photophysics sequential stopped-flow spectrofluorimeter (slit width = 1 nm) using an excitation wavelength of 280 nm, and monitoring the fluorescence emission above 305 nm (the emission maximum of SiaP occurs at 310 nm (13)). All reactions were performed using 1 μM SiaP (final concentration) at 20 ºC in 50 mM Tris-HCl pH 8 containing 100 mM NaCl. Neu5Ac binding to SiaP was monitored under pseudo-first-order conditions using at least a 4-fold excess of Neu5Ac over purified SiaP. One thousand data points were recorded over the course of each reaction, and at least six runs were averaged for each measurement. Kinetic traces were analysed using the Pro-K software supplied by Applied Photophysics Ltd. The reactions were rapid and monophasic and were fitted to a single-exponential consistent with a simple one-step equilibrium process (26;27).

$$P + L \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PL$$

The k_{obs} obtained by a fitting of the traces was plotted in SigmaPlot, from which the dependence of k_{obs} on Neu5Ac concentration was determined.

Examination of ligand binding to SiaP using mass-spectrometry was performed as described previously (13). The Neu5Ac derivatives used in this work were prepared as described previously for sialyl amide (28) and for 2-deoxy-β-*N*-acetylneuraminic acid (dNeu5Ac) (29).

Sequence analysis and bioinformatics

Sequences of TRAP ESR proteins and other components have been collected into the TRAP-DB database (www.trapdb.org) (Mulligan, Bryant and Thomas, unpublished) which contains sequences of over 1000 TRAP transporter proteins from bacteria and archaea. The sequences selected for a multiple sequence alignment were homologues of SiaP, a member of the DctP family of TRAP ESRs, that are encoded within operons containing the genes for the two membrane components of the transporter (either as separate genes or a single fused gene as in *siaQM* from *H. influenzae*). These 248 sequences were aligned using ClustalX and the percentage sequence conservation of the residues present in *H. influenzae* Rd SiaP was calculated in Excel after exporting the ClustalX alignment into BioEDIT.

RESULTS

The SiaP structure is a variation of a typical ESR fold

The structure of SiaP was solved to 1.7 Å spacing by MAD phasing of a SeMet derivative crystal (Form 1).

The structure of the Form 2 crystal, which diffracted to higher resolution, was then solved by molecular replacement (Table 1). The refined model contains all 306 residues of SiaP and 307 water molecules. SiaP has two α/β domains connected by three segments of the polypeptide and separated by a large cleft (Fig. 1A). Domain I, encompassing residues 1-124 and 213-252, contains a 5 stranded β-sheet against which are packed six α-helices. The strand order is β2-β1-β3-β10-β4, with strand β10 running anti-parallel to the other four strands (Fig. 2). Domain II contains residues 125-212 and 253-306 and has a 6 stranded β-sheet surrounded by 3 α-helices (Fig. 1A and Fig. 2). Here the sheet topology is β7-β6-β8-β5-β9-β11 with strand β5 running anti-parallel to the other strands. Residues 280-306 at the C-terminus of the molecule form a pair of α-helices which fold across the base of the molecule and pack against both domains. A striking feature of the structure is the long helix, α9, which spans the breadth of the molecule (Fig. 1).

A DALI search revealed the existence of a large number of structures with similarity to SiaP. The highest scoring match (Z = 8.6; with 112 of 309 Cα atoms aligning with a root mean square deviation (rmsΔ) of 2.9 Å) is the periplasmic glycine-betaine ESR protein from an ABC transporter (PDB 1R9L). Many other matches were found to other ABC ESR proteins, LysR-type transcription factors and eukaryotic glutamate receptors. It is immediately apparent from an examination of the domain topologies that SiaP is a type II ESR protein (Fig. 2). These are characterised by a domain dislocation of one of the β-strands in each of the β-strands (30).

The ligand-bound SiaP protein adopts a closed conformation.

To investigate the structural basis for ligand binding we grew crystals of SiaP in the presence of Neu5Ac and selected analogues (13). Analysis of a third crystal form (Form 3), grown in the presence of one of these sialic acid analogues revealed the presence of 4 molecules in the asymmetric unit. The ligand, Neu5Ac2en, was clearly defined in the electron density maps for molecule B following molecular replacement (Fig. 1B), however molecules A, C and D were unliganded. Superposition of 297 equivalent Cα atoms of molecules A, C and D by least squares minimisation methods gives pairwise root mean square positional deviations in the range 0.4-0.8 Å, similar to those seen when these structures are superposed on the coordinates from the Form 2 crystal (0.5-1.0 Å). However, superposition onto the liganded molecule B gave much larger deviations in the range of 2.5-3.1 Å. The superposition was much better when the individual domains were overlaid (0.4-0.9 Å). These comparisons suggest that a rigid body domain movement accompanies ligand binding, as is apparent in Fig. 1C. Quantitative comparison of these structures using DynDom (31) reveals that this movement can be described by a rotation of 25-31^{o} about a hinge that runs close to the peptide bonds connecting residues Thr127-Arg128, Ile211-Leu212 and Glu254-Lys255.

While the majority of the conformational changes can be attributed to the rigid body rotation about the hinge, a small region near the surface of the ligand binding cleft in Domain II (Ala186 to Tyr197) makes an additional movement beyond that caused by the hinge bending, which results in the reorientation of Phe170 to form a stacking interaction against the side of the sugar ring. Interestingly, the hinge bending observed in the ligand-bound form also results in the kinking of the α-helical component of the hinge.

The ligand-bound molecule B in the Form 3 crystal contains a single molecule of Neu5Ac2en which is consistent with the 1:1 stoichiometry of binding for Neu5Ac2en and Neu5Ac determined by electrospray mass spectrometry (13). Neu5Ac2en differs from the physiological ligand Neu5Ac in that it contains a C2 C3 double bond that introduces partial planarity into the sugar ring (its structure is drawn in Fig. 5). The ligand is bound in a pocket formed by the two domains and its carboxylate group forms a salt bridge to Arg147 and a polar interaction with Asn187, both in the C-terminal domain (Fig. 3A and B). A salt bridging interaction is also made with Arg127 which is in the hinge region. Unusually, the carboxylate in Neu5Ac2en is almost perpendicular to rather than planar with the ring. The glycerol group of Neu5Ac2en appears to form two hydrogen bonds to Glu67. There is an additional contact between Asn10 and the carbonyl oxygen of the *N*-acetyl group. The ligand is almost completely buried with only 32 Å^2 of its 435 Å^2 surface area accessible to the solvent (Fig. 1D).

SiaP binds Neu5Ac by a simple bimolecular association

We wished to determine the mechanism of binding of Neu5Ac by SiaP using pre-steady state kinetics and specifically test whether the data supported previous kinetic schemes proposed for other TRAP and ABC

ESR proteins. While ligand binding to a number of ESR proteins appears to follow monophasic kinetics, the mechanism of binding can be distinguished based on the dependence of the observed rate constant (k_{obs}) on the concentration of ligand. In ABC ESR proteins the linear dependence of k_{obs} on the concentration of ligand implies that ligand binding occurs by a single-step, bimolecular mechanism and that the ESR is predominantly in an open unliganded conformation before undergoing fast closure upon ligand binding (26;27;32). However, analysis of ligand binding to the TRAP ESR DctP revealed a kinetic behaviour not seen in ABC ESRs in that k_{obs} decreased in a hyperbolic manner with increasing concentrations of ligands (33;34). This unusual behaviour was interpreted as the consequence of a pre-isomerisation process of the protein from a closed unliganded conformation to an open unliganded form before ligand binding. Using stopped-flow fluorescence spectroscopy under pseudo-first order conditions, we observed an enhancement in fluorescence after addition of Neu5Ac which could be fitted to a single exponential equation (Fig. 4A). The observed rate constant (k_{obs}) increased linearly with the Neu5Ac concentration (Fig. 4B), which suggests that SiaP binds Neu5Ac using a similar mechanism to ABC-ESR proteins.

From the gradient of the plot of k_{obs} versus ligand concentration in Fig. 4B, we calculated the value of k_1 for the process as $3.5 \pm 0.1 \times 10^7$ M^{-1} s^{-1}. We were not able to determine a k_{-1} from this plot as the intercept of the line was too close to zero and hence cannot be reliably inferred from a linear plot (35). However, we used steady-state fluorescence spectroscopy to calculate a K_d of 58 ± 5 nM (Fig. 4C) for Neu5Ac binding to the protein under identical conditions to those used in the pre-steady state analysis (20 ^oC in the presence of 100 mM NaCl). This is about 2-fold lower than the value we determined at 37 ^oC with no salt (138 ± 6 nM (data not shown) that is similar to the value of 120 ± 6 nM reported previously (13)). From this value of the K_d for Neu5Ac binding we were able to calculate the k_{-1} to be 2.03 s^{-1}. The k_1 and k_{-1} values determined for Neu5Ac binding to SiaP are in the range of those determined for cognate ligand binding to a number of ESR proteins from ABC transporters (26;27). The unusual binding mechanism observed for DctP has been observed with another TRAP ESR protein, RRC01191 from *Rhodobacter capsulatus*. However, the data from this study using SiaP and those using the *E. coli* TRAP ESR protein YiaO (14), suggest that this is not a conserved property of the TRAP ESRs.

The carboxylate group of Neu5Ac is essential for high-affinity binding to SiaP

The clear interaction between the carboxylate group of Neu5Ac2en and the side chains of Arg147/Arg127/Asn187 in the structure of SiaP suggests that the carboxylate is important for binding to SiaP. To probe the significance of this interaction we investigated ligand binding by a derivative of Neu5Ac in which the carboxylate was replaced by an amide (sialyl amide). This ligand bound weakly to SiaP, as judged by tyrosine fluorescence spectroscopy, with a K_d of 243 ± 28 μM, which is around 2000-fold higher than that for Neu5Ac (138 nM). The ligand binding properties of other variants of Neu5Ac have been described (13) (Fig. 5) where the *N*-acetyl group is altered or removed, a lactose group is added at C2, or the C2 position is dehydrated resulting in partial ring flattening (Neu5Ac2en). However, the change of the carboxylate for an amide gives by far the greatest decrease in affinity suggesting that this functional group of the ligand is the most important for binding to SiaP.

The differences between Neu5Ac and Neu5Ac2en are dehydration of C2 C3 and the partial flattening of the ring (See Fig. 5). To determine the contribution of the hydroxyl group at C2, we investigated the binding of 2-deoxy-β-*N*-acetylneuraminic acid (dNeu5Ac) by SiaP. dNeu5Ac retains the chair conformation of the ring seen in Neu5Ac but has lost the hydroxyl at C2 (Fig. 5). It binds with a K_d of 34 ± 2.5 μM, which is similar to that reported for Neu5Ac2en (20 ± 3.8 μM (13)) suggesting that the lower affinity of SiaP for Neu5Ac2en relative to Neu5Ac is primarily caused by the loss of the hydroxyl on C2. This suggests that the natural ligands for SiaP are sialic acids with a free hydroxyl group on C2 and not conjugated forms.

To investigate further the contributions of the carboxylate and *N*-acetyl groups, we tested the binding of 4-acetylaminocyclohexane carboxylic acid to SiaP. This molecule contains both a carboxylate and an *N*-acetyl group in analogous positions to the natural ligand Neu5Ac, and also adopts a chair conformation similar to Neu5Ac (Fig. 5). We were unable to detect binding of this compound to SiaP using either tyrosine fluorescence spectroscopy or electrospray mass spectroscopy (data not shown). Combined, these results suggest that the carboxylate and *N*-acetyl groups are essential, but not sufficient, for the high-affinity binding of Neu5Ac to SiaP.

Sequence analysis of the TRAP ESR proteins in light of the SiaP structure

Multiple sequence alignment of SiaP with its 7 most similar homologues, all of which are encoded in operons with genes for sialometabolism, revealed strong conservation of residues involved in coordination of the carboxylate of the Neu5Ac2en (Arg147, Asn187, Arg127) as well as Phe170 that stacks against the ligand (Fig. 6). The other residues in domain I that bind the glycerol moiety (Glu67 and Asp49) are conserved in 7 out of 8 sequences, however, Asn10 that bonds to the *N*-acetyl group of Neu5Ac2en is very poorly conserved, being replaced by either glutamine, valine or threonine in the other 7 sequences. This suggests that the position of the *N*-acetyl group of Neu5Ac2en in the structure is probably not exactly the same as in Neu5Ac or that this component of the interaction between the protein and ligand is not dependent on a conserved residue in this position. Indeed in other proteins for which structures of complexes with both Neu5Ac and Neu5Ac2en are known, the binding sites are identical but the conformation of the ligands is different (36;37). Given that the affinity of SiaP for Neu5Ac is 200 fold greater than for Neu5Ac2en it would follow that the analogue is presumably not bound in exactly the same conformation as the physiological ligand.

We next expanded the analysis to a wider set of TRAP ESR that are (i) homologous to DctP, and (ii) whose genes are located adjacent to genes for the membrane subunits of a TRAP transporter. This set contained 248 proteins that bind a range of different ligands. The analysis revealed that the most highly conserved residues fall in domain II of the protein (Fig. 7). The most highly conserved residue is Arg147 (present in 98 % of the sequences) that forms a salt bridge to the carboxylate group of Neu5Ac2en in SiaP. The region directly preceding Arg147 is the most highly conserved region in the family (Asp140 92 % conserved, Gly143 95 % conserved, Lys145 86 % conserved), suggesting that the correct positioning of Arg147 within β6 is critical for function of the TRAP ESRs. Additionally, highly conserved residues pack against this region from above (Gly162, 92% conserved) and below (Asp183 92 % conserved). None of the other residues implicated in coordinating the Neu5Ac2en are conserved to this extent across the whole TRAP ESR family. It should also be noted that there is a region of conserved charge on the surface of domain II that is unusual in that it is not seen with ABC ESRs. In domain I there are only two residues that are well conserved, both of which are glycines (Gly 34 90 % conserved and Gly 59 91 % conserved). These sit at turns in the domain following α-helices and probably play important roles in maintaining the overall structure of the domain.

A comparison of the residues conserved in the SiaP group and the larger alignment of all TRAP ESRs, revealed an additional region of SiaP that is very highly conserved within the sialic binding ESR cluster but not outside of this (Fig. 6). This is the α6 and η5 regions, which are adjacent to each other in the structure of domain II and form a face on the surface of the protein that could have a role in specific recognition of the membrane subunits of these particular transporters.

Discussion

The widespread occurrence of TRAP transporters in prokaryotes, including pathogens, suggests that they have important functions in the biology of these organisms and this paper provides the first structural information for a component of a functionally characterised TRAP transporter. One of the most unusual features of TRAP ESR proteins reported in the literature has come from kinetic data that suggest the protein predominates in a closed conformation even in the absence of ligand (33). This is different to all other ESR proteins for which binding data are available implying a mechanistic difference between TRAP ESR proteins and ABC ESR proteins. However, the kinetics of SiaP and *E. coli* YiaO (14), are similar to ABC ESR proteins in that in the absence of ligand they predominate in an open conformation. The unusual properties of DctP and RRC01191 suggest that a subset of the TRAP ESRs have evolved to adopt a closed conformation in the absence of ligand but that this is not an overriding feature of a TRAP ESR.

In accordance with ABC ESRs, we suggest that Neu5Ac binding to SiaP is initiated by the interaction of the carboxylate group of the ligand and the conserved Arg residue (Arg147 in SiaP) in domain II of the protein. In SiaP, the full coordination of the carboxylate also includes an interaction with Arg127 which is within the hinge region and the formation of this interaction could be involved in triggering the hinge bending of the protein, as has been proposed for the *E. coli* maltose binding protein (38). Once the hinge bending has occurred, the domain I interactions with the ligand can also form, keeping the ligand bound tightly. This mechanism is possible for many or all of the TRAP ESR proteins given the

conservation of the Arg residues and the presence of a carboxylate within all characterised ligands of TRAP transporters.

A unique feature of SiaP compared to ABC ESRs is the presence of a 'mixed hinge' consisting of two β-strands and an α-helix. The hinge region in typical ABC ESR proteins comprises 2 or 3 short β-strands, e.g. GlnH, while the more recently described structures of family 9 ESRs and siderophore-binding ESRs contain a single long inflexible α-helix (39-41). In SiaP the hinge α-helix is 35 amino acid residues in length, similar to the α-helical hinge in the family 9 ESRs, but it is positioned towards the C-terminus of the protein, rather than between the two domains as in the cluster 9 proteins. In the siderophore binding ESR proteins that contain a single α-helix hinge, there is only a relatively small movement of domains I and II upon ligand binding and the ligand sits in a shallow groove formed by the two domains rather than being deeply buried between the domains. However, in SiaP there is significant bending upon ligand binding which is similar to that seen in ESRs with hinges composed entirely of β-strands. To accommodate this hinge bending, a kink is induced in this α-helix in SiaP which will result in an altered surface of this region after ligand binding. This is expected to be an energetically unfavourable event and perhaps functions as a switch to hold the protein in either the open or closed conformation.

Other structures of proteins that bind sialic acid are known. A conserved arginine is a common theme amongst proteins which have diverse structures and biological functions (42-44). The most studied sialic acid binding proteins are the neuraminidases that contain a characteristic arginine triad that coordinates the carboxylate group (45-47). SiaP is similar in using a triad of residues to coordinate the carboxylate group but achieves this using two arginine residues and one asparagine, a conserved structural motif that appears to be important for high-affinity binding. Functionally SiaP is more similar to the Siglec molecules found on eukaryotic cell surfaces which bind sialic acids but do not modify them. Siglecs have general roles in adhesion and signalling (48) and bind sialic acid in a surface groove of a V-set immunoglobulin-like fold with only one face of the Neu5Ac being in contact with the protein. There are multiple interactions between the protein and ligand, including a salt bridge between the carboxylate and an invariant arginine (49). Similarly, the structure of the Neu5Ac-bound lectin domain of the *Vibrio cholerae* neuraminidase reveals the ligand bound in a shallow cleft such that only the anomeric oxygen and the O9 of the glycerol side chain are not involved in interactions with the protein (50). This domain binds Neu5Ac with a K_d of ~ 30 μM which is relatively low affinity compared to SiaP (13;50).

Following ligand binding, the ESR must interact with the membrane subunits of the transporter and given the similarity in structure and ligand binding mechanisms between SiaP and ABC ESRs there must be some similarity in how the ligand is 'delivered' to the membrane subunits. However, uptake via a TRAP transporter is not coupled to ATP binding and hydrolysis events, but rather to the symport of a coupling ion (there is evidence for both H^+ and Na^+ ions being the coupling ions for TRAP transporters) (8;10;54) and so it is likely that there will be differences in the mechanism by which ESR opening and ligand release is coupled to movement of the ligand across the membrane.

Sequence analysis of the sialic acid cluster of TRAP ESRs supports the hypothesis of a direct interaction between the ESR and the translocation pore, as in ABC transporters. Thus a particularly well conserved face of SiaP formed by the α6 and η5 region (Fig. 6) is in an analogous position on the surface of the domains as is seen in ABC ESRs like MBP. Importantly this region is not conserved in the larger alignment of TRAP ESRs, supporting the idea that it confers specificity of interaction with the cognate membrane subunit of the transporter.

The structure of SiaP also provides insight into the evolution of the TRAP transporters due to its structural similarity to an ancestral type-II ESR. This suggests that an ancestral type-II ESR was recruited to work with an ancestral secondary transporter of the IT superfamily and that over time its sequence diverged from ABC ESR proteins beyond the level of detection. During this divergence, the TRAP ESRs have added additional sequence to the ancestral type-II sequence including the α-helix that forms the 'mixed hinge', an extra β-strand in domain II and two extra helices which interact with the additional helices found in domain I.

While the DctP family of ESRs are used in the majority of TRAP transporters, we defined a different family of ESRs called the TAXI family (InterPro family IPR011852) which are found in a small number of uncharacterised TRAP transporters (10). Fortuitously, the structure of a protein that we believe

is a TAXI ESR has been solved as part of a structural genomics project, although this was not recognised by the authors (55). This ESR from *Thermus thermophilus* is encoded by a gene (*TTHA1157*) adjacent to the gene for a fused TRAP membrane subunit (*TTHA1158*) and therefore is very likely to be a genuine component of a TAXI-TRAP transporter. The structure revealed that the ESR bound glutamate/glutamine and that it is clearly a type-II ESR, although interestingly it binds the amino acid ligand using a completely different set of residues to the ABC-type ESRs like GlnH. Finally, the recent structure of the BugD protein from *Bordetella pertusis* provides additional support for the widespread nature of the type-II ESR fold for use with secondary transporters (56). This protein of unknown function is not encoded alongside genes for a transporter. However, it is homologous to BctA, a component of a tripartite tricarboxylate transporter (TTT), which form a second smaller family of ESR-dependent secondary transporters. Again, this structure has a type-II ESR fold but coordinates its ligand (aspartate) using an unusual set of interactions, in fact in this structure the carboxylate of the aspartate is coordinated solely by water molecules.

In summary, the structure of SiaP provides insight into a high-affinity binding site for sialic acid and in combination with bioinformatics reveals the importance of the Arg/carboxylate interaction in all TRAP transporters. The additional finding that SiaP is a type-II ESR supports the hypothesis that TRAP transporters have evolved from ancestral secondary transporters via the recruitment of an ancestral type-II ESR to specifically catalyse uptake of organic anions with high affinity and that this appears to have been a common feature of the evolution of the related TAXI TRAP and also the TTT families of ESR-dependent secondary transporters.

Acknowledgements

We would like to thank the ESRF, Grenoble, for excellent data collection facilities. The work described here was funded by the European Commission as SPINE, contract-no. QLG2-CT-2002-00988 under the RTD programme "Quality of Life and Management of Living Resources" and by a grant to GHT and DJK from the UK Biotechnology and Biological Science Research Council. We would also like to thank the participants of the SPINE meeting (York 2005) which greatly assisted in the solution of the structure and Prof. Colin Kleanthous for comments on the manuscript.

Reference List

1. Wilkinson, A. J. and Verschueren, K. H. G. (2003) Crystal structures of periplasmic solute-binding proteins in ABC transport complexes illuminate their function. In Holland, I. B., Cole, S. P. C., Kuchler, K., and Higgins, C. F., editors. *ABC Proteins: From Bacteria to Man*, Elsevier,

2. Dwyer, M. A. and Hellinga, H. W. (2004) *Curr.Opin.Struct.Biol* **14,** 495-504

3. Davidson, A. L. and Chen, J. (2004) *Annu.Rev.Biochem.* **73,** 241-268

4. Shaw, J. G., Hamblin, M. J., and Kelly, D. J. (1991) *Mol.Microbiol.* **5,** 3055-3062

5. Hamblin, M. J., Shaw, J. G., and Kelly, D. J. (1993) *Mol.Gen.Genet.* **237,** 215-224

6. Rabus, R., Jack, D. L., Kelly, D. J., and Saier, M. H., Jr. (1999) *Microbiology* **145 (Pt 12),** 3431-3445

7. Wyborn, N. R., Alderson, J., Andrews, S. C., and Kelly, D. J. (2001) *FEMS Microbiol.Lett.* **194,** 13-17

8. Forward, J. A., Behrendt, M. C., Wyborn, N. R., Cross, R., and Kelly, D. J. (1997) *J.Bacteriol.* **179,** 5482-5493

9. Prakash, S., Cooper, G., Singhi, S., and Saier, M. H., Jr. (2003) *Biochim.Biophys.Acta* **1618,** 79-92

10. Kelly, D. J. and Thomas, G. H. (2001) *FEMS Microbiol.Rev.* **25,** 405-424

11. Tetsch, L. and Kunte, H. J. (2002) *FEMS Microbiol.Lett.* **211,** 213-218

12. Grammann, K., Volke, A., and Kunte, H. J. (2002) *J.Bacteriol.* **184,** 3078-3085

13. Severi, E., Randle, G., Kivlin, P., Whitfield, K., Young, R., Moxon, R., Kelly, D., Hood, D., and Thomas, G. H. (2005) *Mol.Microbiol.* **58,** 1173-1185

14. Thomas, G. H., Southworth, T., Leon-Kempis, M. R., Leech, A., and Kelly, D. J. (2006) *Microbiology* **152,** 187-198

15. Hood, D. W., Makepeace, K., Deadman, M. E., Rest, R. F., Thibault, P., Martin, A., Richards, J. C., and Moxon, E. R. (1999) *Mol.Microbiol.* **33,** 679-692

16. Allen, S., Zaleski, A., Johnston, J. W., Gibson, B. W., and Apicella, M. A. (2005) *Infect.Immun.* **73,** 5291-5300

17. Steenbergen, S. M., Lichtensteiger, C. A., Caughlan, R., Garfinkle, J., Fuller, T. E., and Vimr, E. R. (2005) *Infect.Immun.* **73,** 1284-1294

18. Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) *J Bacteriol.* **119,** 736-747

19. Ducros, V. M., Lewis, R. J., Verma, C. S., Dodson, E. J., Leonard, G., Turkenburg, J. P., Murshudov, G. N., Wilkinson, A. J., and Brannigan, J. A. (2001) *J Mol.Biol* **306,** 759-771

20. Schneider, T. R. and Sheldrick, G. M. (2002) *Acta Crystallogr.D.Biol Crystallogr.* **58,** 1772-1779

21. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) *Acta Crystallogr.D.Biol Crystallogr.* **57,** 1445-1450

22. Terwilliger, T. C. (2003) *Acta Crystallogr.D.Biol Crystallogr.* **59,** 38-44

23. Vagin, A. and Teplyakov, A. (2000) *Acta Crystallogr.D.Biol Crystallogr.* **56 Pt 12,** 1622-1624

24. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr.D.Biol Crystallogr.* **53,** 240-255

25. Emsley, P. and Cowtan, K. (2004) *Acta Crystallogr.D.Biol Crystallogr.* **60,** 2126-2132

26. Miller, D. M., III, Olson, J. S., and Quiocho, F. A. (1980) *J Biol Chem* **255,** 2465-2471

27. Miller, D. M., III, Olson, J. S., Pflugrath, J. W., and Quiocho, F. A. (1983) *J Biol Chem* **258,** 13665-13672

28. Brossmer, R. and Holmquist, L. (1971) *Z.Physiol.Chem.* **352,** 1715-1719

29. Schmid, W., Christian, R., and Zbiral, E. (1988) *Tetrahedron Letters* **29,** 3643-3646

30. Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (2003) *Mol.Biol.Evol.* **20,** 267-277

31. Hayward, S. and Lee, R. A. (2002) *J Mol.Graph.Model.* **21,** 181-183

32. Ledvina, P. S., Tsai, A. L., Wang, Z., Koehl, E., and Quiocho, F. A. (1998) *Protein Sci.* **7,** 2550-2559

33. Walmsley, A. R., Shaw, J. G., and Kelly, D. J. (1992) *J.Biol.Chem.* **267,** 8064-8072

34. Walmsley, A. R., Shaw, J. G., and Kelly, D. J. (1992) *Biochemistry* **31,** 11175-11181

35. Wallis, R., Moore, G. R., James, R., and Kleanthous, C. (1995) *Biochemistry* **34,** 13743-13750

36. Lawrence, M. C., Borg, N. A., Streltsov, V. A., Pilling, P. A., Epa, V. C., Varghese, J. N., McKimm-Breschkin, J. L., and Colman, P. M. (2004) *J Mol.Biol* **335,** 1343-1357

37. Crennell, S., Takimoto, T., Portner, A., and Taylor, G. (2000) *Nat.Struct.Biol* **7,** 1068-1074

38. Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quiocho, F. A. (1992) *Biochemistry* **31,** 10657-10663

39. Lawrence, M. C., Pilling, P. A., Epa, V. C., Berry, A. M., Ogunniyi, A. D., and Paton, J. C. (1998) *Structure.* **6,** 1553-1561

40. Lee, Y. H., Deka, R. K., Norgard, M. V., Radolf, J. D., and Hasemann, C. A. (1999) *Nat.Struct.Biol.* **6,** 628-633

41. Clarke, T. E., Ku, S. Y., Dougan, D. R., Vogel, H. J., and Tari, L. W. (2000) *Nat.Struct.Biol.* **7,** 287-291

42. Angata, T. and Varki, A. (2002) *Chem Rev.* **102,** 439-469

43. Crocker, P. R. and Varki, A. (2001) *Trends Immunol.* **22,** 337-342

44. Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004) *Microbiol.Mol.Biol Rev.* **68,** 132-53, table

45. Taylor, G. (1996) *Curr.Opin.Struct.Biol* **6,** 830-837

46. Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R., and Taylor, G. L. (1993) *Proc.Natl.Acad.Sci.U.S.A* **90,** 9852-9856

47. Burmeister, W. P., Ruigrok, R. W., and Cusack, S. (1992) *EMBO J* **11,** 49-56

48. Crocker, P. R. (2002) *Curr.Opin.Struct.Biol* **12,** 609-615

49. May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R., and Jones, E. Y. (1998) *Mol.Cell* **1,** 719-728

50. Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., Kiefel, M. J., von Itzstein, M., and Taylor, G. (2004) *J Biol Chem* **279,** 40819-40826

51. Austermuhle, M. I., Hall, J. A., Klug, C. S., and Davidson, A. L. (2004) *J Biol Chem* **279,** 28243-28250

52. Chen, J., Lu, G., Lin, J., Davidson, A. L., and Quiocho, F. A. (2003) *Mol.Cell* **12,** 651-661

53. Lu, G., Westbrooks, J. M., Davidson, A. L., and Chen, J. (2005) *Proc.Natl.Acad.Sci.U.S.A* **102,** 17969-17974

54. Jacobs, M. H., van der, H. T., Driessen, A. J., and Konings, W. N. (1996) *Proc.Natl.Acad.Sci.U.S.A* **93,** 12786-12790

55. Takahashi, H., Inagaki, E., Kuroishi, C., and Tahirov, T. H. (2004) *Acta Crystallogr.D.Biol Crystallogr.* **60,** 1846-1854

56. Huvent, I., Belrhali, H., Antoine, R., Bompard, C., Locht, C., Jacob-Dubuisson, F., and Villeret, V. (2006) *J.Mol.Biol.* **356,** 1014-1026

57. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A., and Noble, M. (2004) *Acta Crystallogr.D.Biol Crystallogr.* **60,** 2288-2294

58. Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (1999) *J Mol.Biol* **286,** 279-290

Footnotes.

[1] Abbreviations used in this manuscript: tripartite ATP-independent periplasmic (TRAP) transporters, sialic acid or *N*-acetylneuraminic acid (Neu5Ac), 2,3-didehydro-2-deoxy-*N*-acetyl-neuraminic acid (Neu5Ac2en), 2-deoxy-β-*N*-acetylneuraminic acid (dNeu5Ac). ATP-binding cassette (ABC), extracytoplasmic solute receptor (ESR).

Figure legends.

Fig. 1. Overall structure of the unliganded (A) and Neu5Ac2en-bound (B) forms of SiaP and the superposition by least squares minimisation procedures applied to the positions of the Cα atoms of domain I (C) in an orientation to illustrate the domain closure around the ligand. The ribbon diagrams were drawn using CCP4MG (57) and the ligand is coloured by atom type. (D) Surface representation of the Neu5Ac2en structure viewed looking down into the binding cleft. The ligand is in blue and domain I is on the left while domain II is on the right.

Fig. 2. Schematic diagram of the topology of A) SiaP in comparison with B) an ancestral type-II ESR protein (58). The N-terminal domain is in light grey and the C-terminal domain is in black. Features which distinguish SiaP are in dark grey. Unfilled circles are 3_{10} helices. This diagram adopts the style of Fukami-Kobayashi (58) where the hinge β-strands are drawn as part of the β-sheets. The β4 and β5 schematic elements for SiaP actually form a single extended β-strand which extends across both domains but has been displayed as two elements in the schematic to be consistent with the numbering scheme for the type-II ESR.

Fig. 3. A) Stereo view of the electron density (contoured at 1.5σ) for Neu5Ac2en and residues involved in the coordination of this ligand in SiaP. B) Ligplot representation of the interactions between Neu5Ac2en and the protein.

Fig. 4. Investigation of the presteady-state kinetics of Neu5Ac binding to SiaP monitored using stopped-flow fluorescence spectroscopy. (A) Trace of SiaP (1 μM) pushed against buffer (flat line) and against 8 μM Neu5Ac. The binding data have been fit to a single exponential equation. (B) Plot of the k_{obs} of the association between SiaP (1 μM) and Neu5Ac versus the concentration of Neu5Ac under pseudo-first order conditions. The points on the graph are the averages from the three independent titrations. The k_1 was determined from the slope of the line of best fit and averaged from three independent sets of titrations. (C) Representative steady-state fluorescence titration of SiaP with Neu5Ac under identical conditions used in the pre-steady-state experiments.

Fig. 5. Chemical structures and K_d values of sialic acid (Neu5Ac) and related analogues used in this study and in Severi *et al*., (13). The abbreviations are those used in the text and also Neu5Gc, *N*-glycolylneuraminic acid, KDN, 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid.

Fig. 6. Multiple sequence alignment of *H. influenzae* SiaP and 7 related TRAP ESR proteins that are likely to form components of a sialic acid transporter. The genes for all 8 of these sequences are encoded in operons encoding sialometabolic genes. The regions indicated by β4 and β5 form a single extended β-strand that is part of both domains but is labelled as two β-strands for consistency with Fig. 2 and the text.

Fig. 7. Surface structures of SiaP overlaid by shading to indicate the percentage conservation of amino acid residues in the TRAP-ESR family. The view on the left is looking down into the binding cleft of the open unliganded form of SiaP. There are four shadings of blue in the figure, which from the darkest represent, 90 %, 80%, 50% and 25% conservation. The images are rotated by 180^o to illustrate the nature of the conserved surface charge visible on the domain II. Two aspartate residues that are highly conserved and surface exposed are indicated in red (D140 and D183 in SiaP, both around 92% conserved).

Fig. 1.

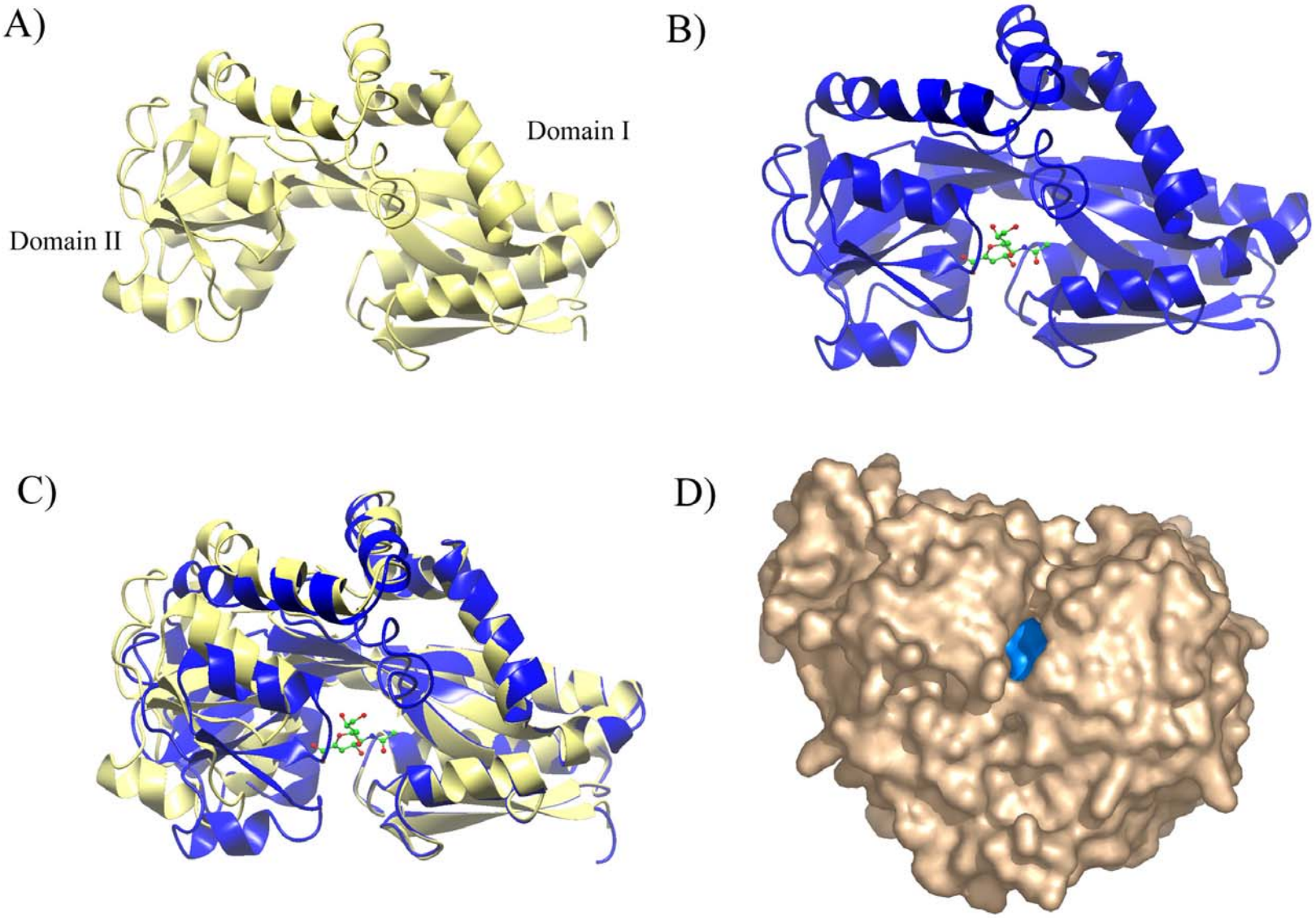

Fig. 2.

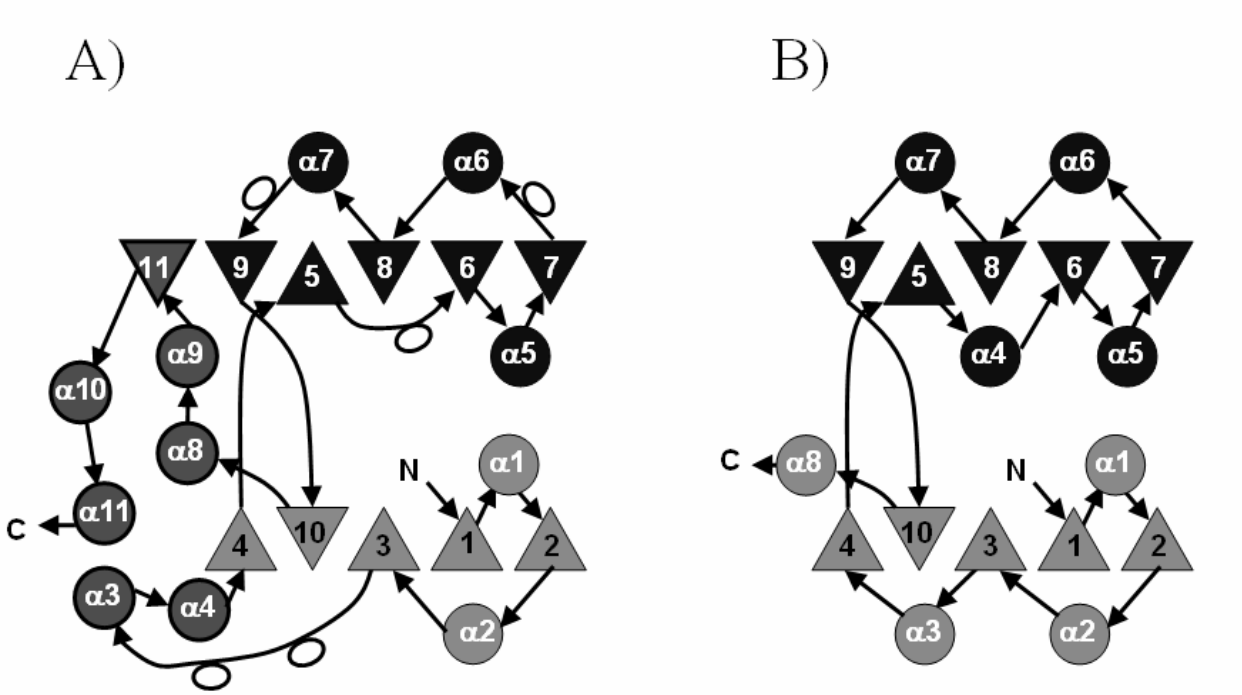

Fig. 3

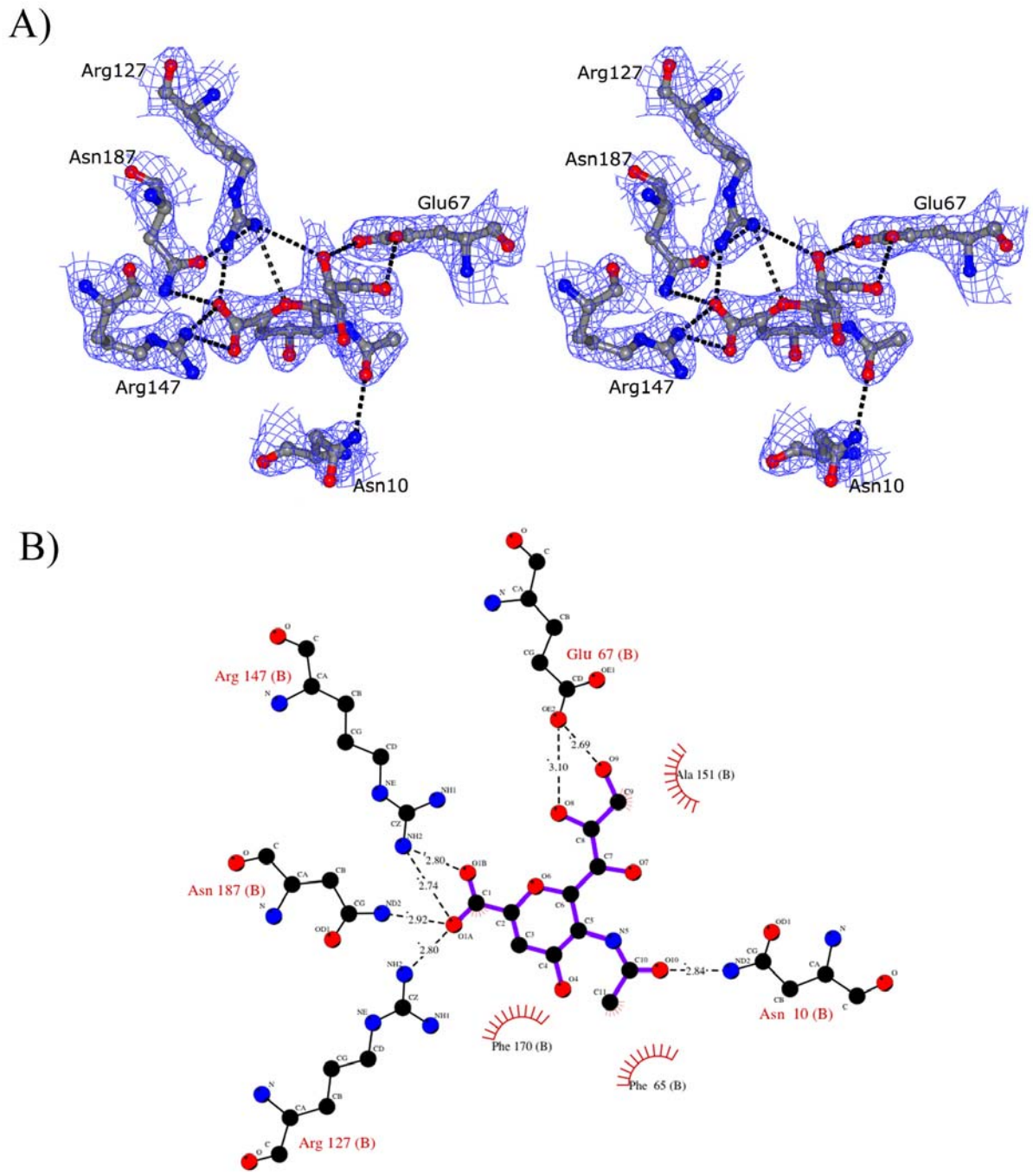

Fig. 4

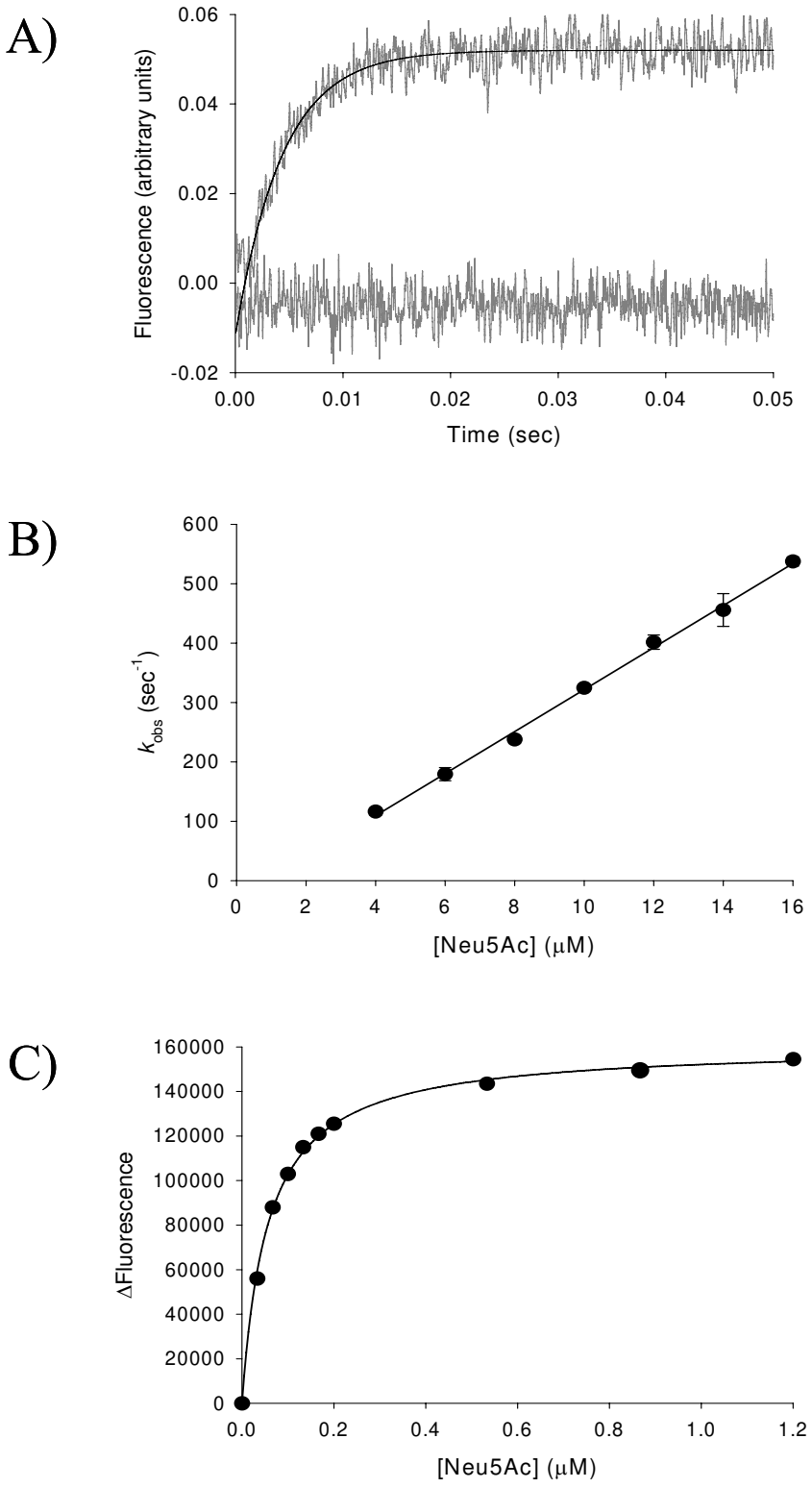

Fig. 5

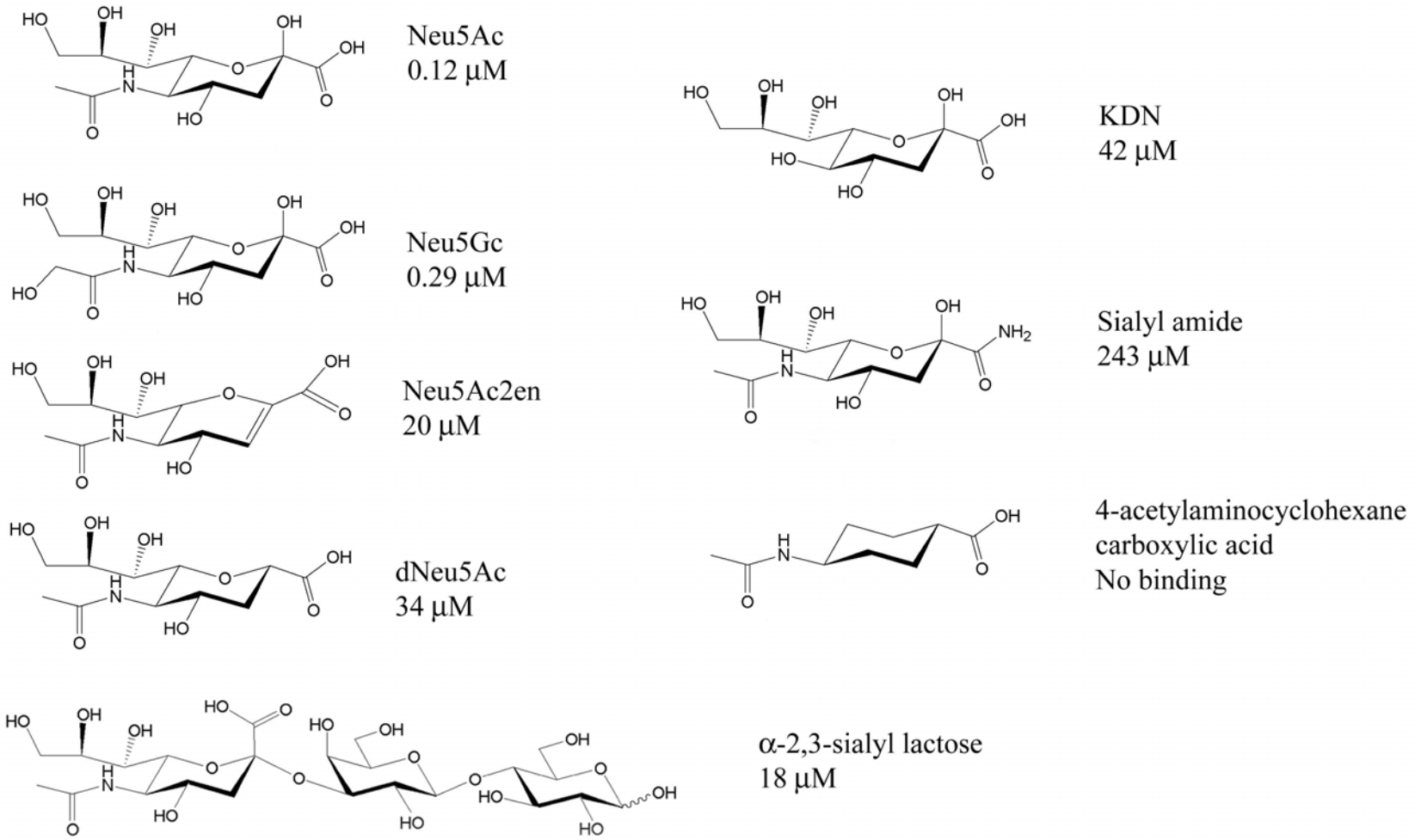

Fig. 6

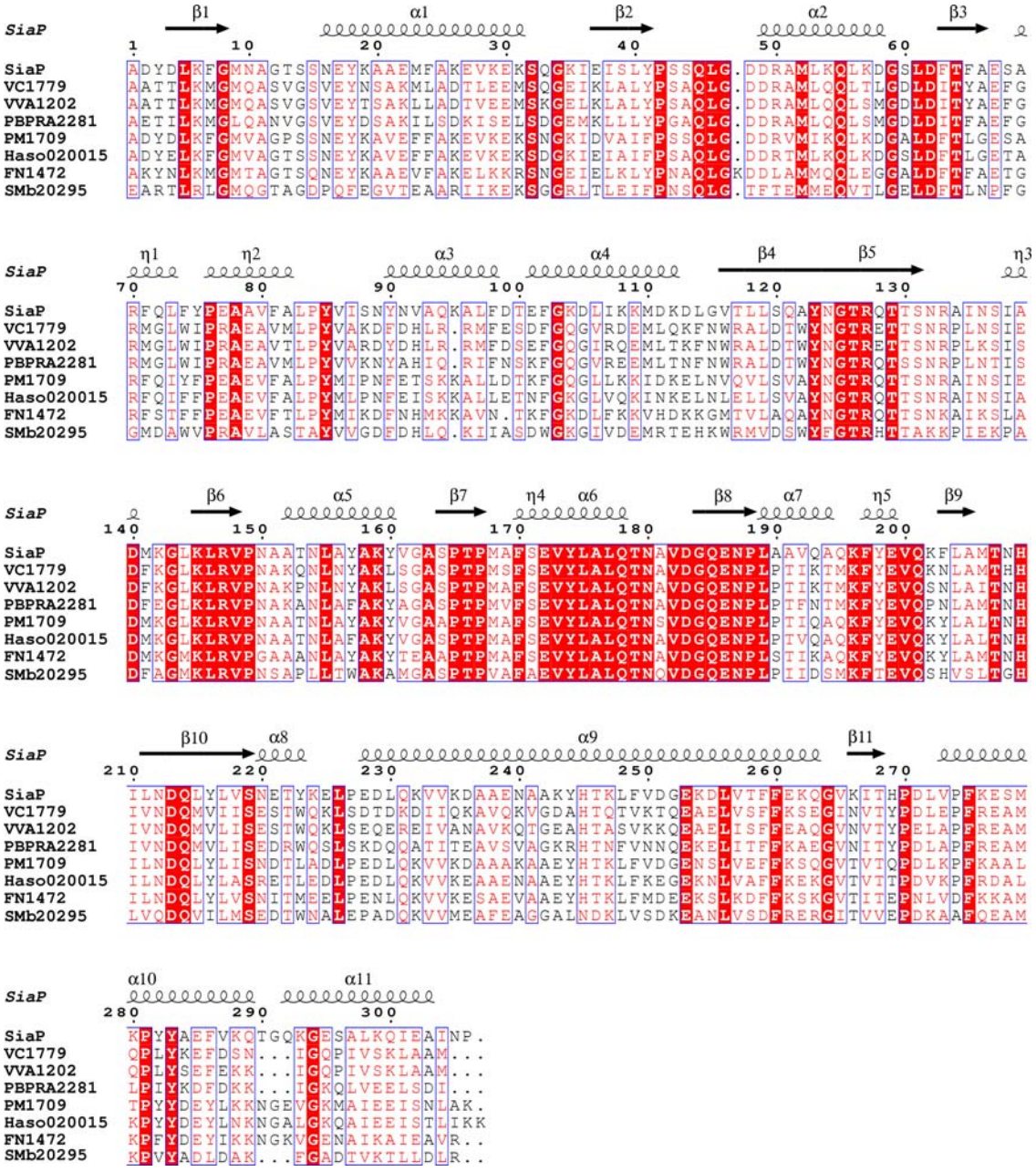

Fig. 7

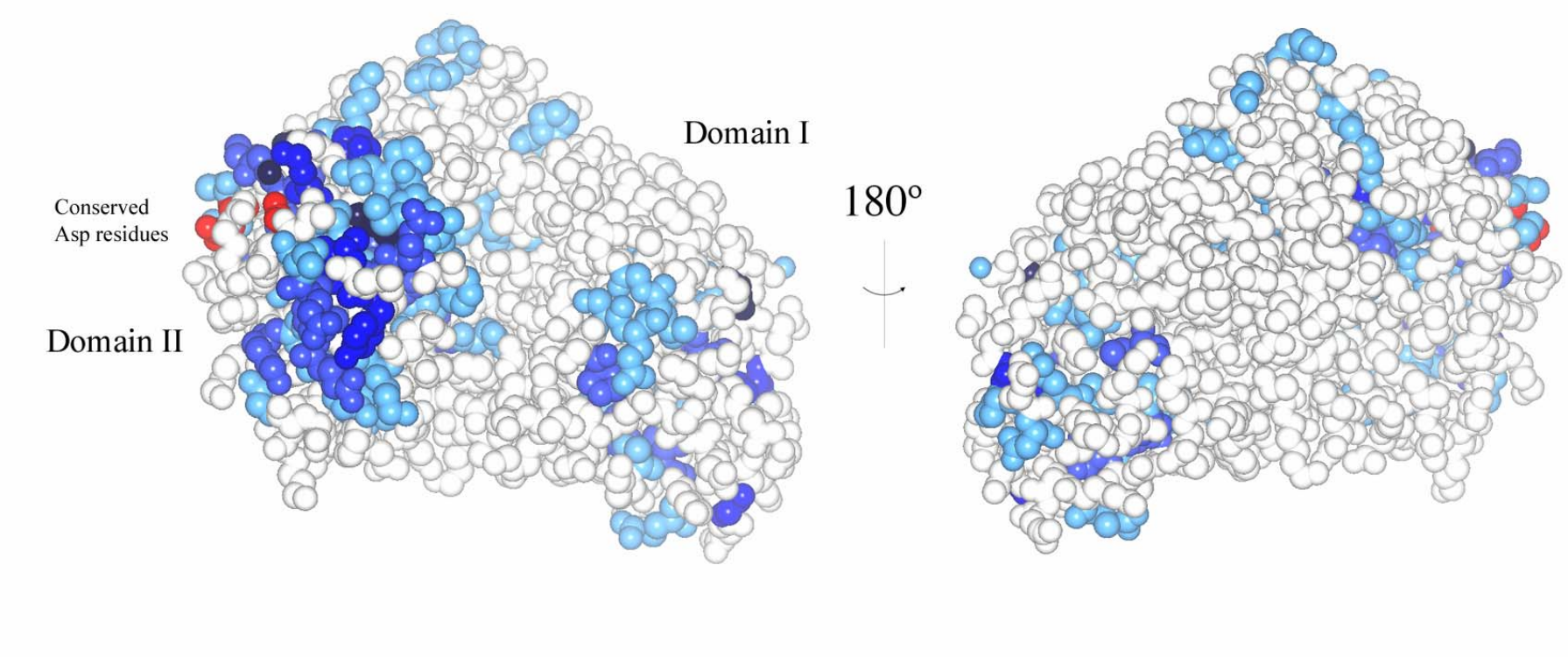

Table 1.

	Form I			Form II	Form III
Data collection at BM14					
Wavelength (Å)	0.97907	0.97921	0.90777	0.97624	0.97624
Resolution range (Å) / Highest resolution shell	50.0–2.63 / 2.72–2.63	50.0–2.70 / 2.80–2.70	50.0–2.63 / 2.72–2.63	50.0–1.70 / 1.76–1.70	50.0–2.20 / 2.28–2.20
Space group	$P2_12_12$			I222	C2
Unit-cell parameters (Å)	a=46.08 b=103.36 c=199.10	a=45.78 b=103.46 c=198.77	a=46.13 b=103.67 c=199.40	a=46.76 b=102.51 c=202.65	a=131.45 b=88.70 c=115.91 β=105.33°
Number of unique reflections, overall/ outer shell	28863 / 2562	26676 / 2434	28271 / 2192	41133 / 460	49350 / 826
Completeness (%), overall/ outer shell	98.5/88.4	99.2/93.2	96.4/75.6	77.4/8.8	75.7/12.7
Redundancy, overall/ outer shell	6.6/5.1	6.5/4.9	6.0/4.1	5.1/1.1	3.2/1.3
$I/\sigma(I)$, overall /outer shell	10.1/1.5	10.7/1.0	8.8/0.8	12.20/1.33	10.7/1.1
*R*merge (%), overall/ outer shell	14.6/70.8	14.1/84.8	17.0/97.4	11.4/96.5	9.7/56.0
Refinement and model statistics					
R-factor[a]/ *R*-free[b]				0.19/0.24	0.20/0.28
Reflections (working/free)				38727/2091	46872/2390
Outer shell *R*-factor / *R*-free[b]				50.0/70.0	36.4/57.3
Molecules/ asymmetric unit				1	4
Number of protein non hydrogen atoms				2711	10040
Number of Zn^{2+} atoms				2.5	6
Number of water molecules				307	467
R.m.s. deviation from target[c]					
Bond lengths (Å)				0.023	0.006
Bond angles (°)				1.944	0.899
Average *B*-factor (Å²)				30.6	27.7
Ramachandran plot[d]				93.5/5.8/0.7/0	91.8/8.0/0.2/0

[a]*R*-factor = $\sum||F_O| - |F_C||/\sum|F_O|$ where F_O and F_C are the observed and calculated structure factor amplitudes, respectively.

[b]*R*-free is the *R*-factor calculated with 5 % of the reflections chosen at random and omitted from refinement.

[c]Root-mean-square deviation of bond lengths and bond angles from ideal geometry.

[d]Percentage of residues in most-favoured/additionally allowed/generously allowed/disallowed regions of the Ramachandran plot, according to PROCHECK.