Citation for published version

Müller, Axel and Severi, Emmanuele and Mulligan, Christopher and Watts, Andrew G and Kelly, David J and Wilson, Keith S and Wilkinson, Anthony J and Thomas, Gavin H  (2006) Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus influenzae. The Journal of biological chemistry, DOI

https://doi.org/10.1074/jbc.M603463200

Link to record in KAR

http://kar.kent.ac.uk/61635/

Document Version

Publisher pdf

Copyright & reuse
Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research
The version in the Kent Academic Repository may differ from the final published version. Users are advised to check http://kar.kent.ac.uk for the status of the paper. Users should always cite the published version of record.

Enquiries
For any further enquiries regarding the licence status of this document, please contact: researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at http://kar.kent.ac.uk/contact.html
Conservation of Structure and Mechanism in Primary and Secondary Transporters Exemplified by SiaP, a Sialic Acid Binding Virulence Factor from *Haemophilus influenzae*

Axel Müller†, Emmanuele Severi§, Christopher Mulligan‡, Andrew G. Watts†, David J. Kelly§, Keith S. Wilson‡, Anthony J. Wilkinson‡2, and Gavin H. Thomas§3

From the †Structural Biology Laboratory, Department of Chemistry and §Department of Biology, University of York, York YO10 5YW, United Kingdom and ‡Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield, S10 2TN, United Kingdom

Extractocytoplasmic solute receptors (ESRs) are important components of solute uptake systems in bacteria, having been studied extensively as parts of ATP binding cassette transporters. Herein we report the first crystal structure of an ESR protein from a functionally characterized electrochemical ion gradient-dependent secondary transporter. This protein, SiaP, forms part of a tripartite ATP-independent periplasmic transporter specific for sialic acid in *Haemophilus influenzae*. Surprisingly, the structure reveals an overall topology similar to ATP binding cassette 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 2,3-didehydro-2-deoxy-N-acetyleneuraminic acid 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 component. The structure provides insight into the evolution, mechanism, and substrate specificity of ESR-dependent secondary transporters that are widespread in prokaryotes.

The work described here was funded by the European Commission as Structural Proteomics in Europe (SPINE), Contract QLG2-CT-2002-00988 under the Research Technological Development (RTD) program “Quality of Life and Management of Living Resources” and by a grant from the United Kingdom Biotechnology and Biological Science Research Council (to G. H. T. and D. J. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby indicated.

The atomic coordinates and structure factors (code 2CEY (unliganded structure) and 2CEX (Neu5Ac2en structure)) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

1 Current address: Dept. of Pharmacy and Pharmacology, University of Bath, BA2 7AY, UK.

2 To whom correspondence may be addressed. Tel.: 44-1904-328261; E-mail: ajw@ysbl.york.ac.uk.

3 To whom correspondence may be addressed. Tel.: 44-1904-328268; E-mail: ght2@york.ac.uk.

4 The abbreviations used are: ABC, ATP binding cassette; TRAP transporter, tripartite ATP-independent periplasmic transporter; Neu5Ac, sialic acid, N-acetyleneuraminic acid; Neu5Ac2en, 2,3-didehydro-2-deoxy-N-acetylneuraminic acid; dNeu5Ac, 2-deoxy-β-N-acetyleneuraminic acid; ESR, extracytoplasmic solute receptor; SiaP, sialic acid-binding protein; SeMet, selenomethionine.
lipopolysaccharide to make it appear as “self” and evade the innate immune response (15). Deletion of the TRAP transporter results in loss of lipopolysaccharide 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 characterized 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-acetylenamidic 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 of LB (Lennox broth) for 6 h, washed in M9 minimal medium (18), and used to inoculate 50 ml M9 minimal medium for overnight growth at 37 °C. This was used to inoculate 1 liter of M9 minimal medium at 25 °C. Cells were allowed to grow to an *A$_{600}$* of 0.3–0.4 before inducing expression with 1 mm isopropyl-1-thio-β-d-galactopyranoside 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 mmm EDTA in 50 mm Tris-HCl, pH 8, 600 µg/ml lysozyme) for 1 h at 30 °C, and the periplasmic fraction was then clarified by centrifugation and dialyzed against 50 mm Tris-HCl, pH 8, containing 1.5 m (NH$_4$)$_2$SO$_4$. SiaP was purified by fast protein liquid chromatography 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 were confirmed by electrospray mass spectrometry (13). For preparation of the selenomethionine (SeMet) derivative of SiaP, the protein was expressed from a 1-liter culture as described (19) and purified to ~95% homogeneity using a single anion exchange step (Mono Q).

**Crystallization**—For crystallization, 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. Crystallization experiments utilized the vapor diffusion method and a MOSQUITO nanoliter dispensing robot to set up sitting drops. Three crystal forms were analyzed. Form 1 crystals belonging to space group P2$_1$2$_1$2 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% polyethylene glycol 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% polyethylene glycol 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 European Synchrotron Radiation Facility, 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. Before 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 program 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 program 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 N-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 carbohydrate domain II as a search model completed molecular replacement. The model was refined by iterative cycles of REFMAC (24) interspersed with manual modeling 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
ESR proteins and other components have been collected into previously for sialyl amide (28) and for dNeu5Ac (29). Kinetic analysis was performed as described previously (13). The Neu5Ac concentration was determined.

**TABLE 1**

| Data collection at BM14 | Form I | Form II | Form III |
|------------------------|--------|---------|----------|
| Wavelength (Å) | 0.97907 | 0.97921 | 0.97624 |
| Resolution range (Å)/highest resolution shell | 50.0–2.63/2.72–2.63 | 50.0–2.70/2.80–2.70 | 50.0–1.70/1.76–1.70 |
| Space group | P2₁,2,2 | P2₁,2,2 | P2₁,2,2 |
| Unit-cell parameters (Å) | | |
| a = 40.68 Å | 1222 | 1222 | 1222 |
| b = 103.36 Å | a = 131.45 | a = 131.45 | a = 131.45 |
| c = 199.10 Å | b = 103.67 | b = 103.67 | b = 103.67 |
| Number of unique reflections, overall/outside shell | 28,863/2,562 | 26,676/2,434 | 28,271/2,192 |
| Completeness (%), overall/outside shell | 98.5/88.4 | 992/93.2 | 964.7/5.6 |
| Redundancy, overall/outside shell | 6.5/5.1 | 6.5/4.9 | 6.0/4.1 |
| I/(σ(I)), overall/outside shell | 10.1/1.5 | 10.7/1.0 | 8.8/0.8 |
| R_{merge} (%), overall/outside shell | 14.6/70.8 | 14.1/84.8 | 17.0/97.4 |

| Refinement and model statistics | |
| R-factor/R-free | 0.19/0.24 | 0.20/0.28 |
| Reflections (working/free) | 38,727/2,091 | 46,872/2,390 |
| Outer shell R-factor/R-free | 50.0/70.0 | 36.4/57.3 |
| Molecules/asymmetric unit | 1 | 4 |
| Number of protein non hydrogen atoms | 2711 | 10040 |
| Number of Zn²⁺ atoms | 2.5 | 6 |
| Number of water molecules | 307 | 467 |

| Root mean square deviation from target | |
| Bond lengths (Å) | 0.023 | 0.006 |
| Bond angles (°) | 1.944 | 0.899 |
| Average B-factor (Å²) | 30.9 | 27.7 |
| Ramachandran plot | 93.5/5.8/0.7/0 | 91.8/8.0/0.2/0 |

| \( R = \frac{\|F_{o}\| - |F_{c}|}{\|F_{o}\|} \) where \( F_{o} \) and \( F_{c} \) are the observed and calculated structure factor amplitudes, respectively. \( R_{merge} = \frac{\sum_{i,j} \sum_{k} |F_{ijkl} - F_{ijkl}^{\text{obs}}|}{\sum_{i,j} \sum_{k} |F_{ijkl}|} \) where \( F_{ijkl} \) are the observed and calculated structure factor amplitudes, respectively.

**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 diffraction 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 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 analyzed 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).

\[
p + L \rightleftharpoons PL \quad (\text{Eq. 1})
\]

The \( k_{\text{obs}} \) obtained by a fitting of the traces was plotted in Sigmaplot, from which the dependence of \( k_{\text{obs}} \) 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 dNeu5Ac (29).

Sequence Analysis and Bioinformatics—Sequences of TRAP ESR proteins and other components have been collected into the TRAP-DB data base, which contains sequences of >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.

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 diffraction 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 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 analyzed 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).

\[
p + L \rightleftharpoons PL \quad (\text{Eq. 1})
\]

The \( k_{\text{obs}} \) obtained by a fitting of the traces was plotted in Sigmaplot, from which the dependence of \( k_{\text{obs}} \) 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 dNeu5Ac (29).

Sequence Analysis and Bioinformatics—Sequences of TRAP ESR proteins and other components have been collected into
form a pair of α-helices that 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 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 characterized by a domain dislocation of one of the β-strands in each of the β-sheets (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 four molecules in the asymmetric unit. The ligand, Neu5Ac2en, was clearly defined in the electron density maps for molecule B after 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 minimization methods gives pairwise root mean
The ligand is almost completely buried, with only 32 Å between Asn-10 and the carbonyl oxygen of the hydrogen bonds to Glu-67. There is an additional contact the ring. The glycerol group of Neu5Ac2en appears to form two 127, which is in the hinge region. Unusually, the carboxylate in the dependence of the observed rate constant (k_{obs}) can be distinguished based on 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 behavior not seen in ABC ESRs in that k_{obs} decreased in a hyperbolic manner with increasing concentrations of ligands (33, 34). This unusual behavior was interpreted as the consequence of a pre-isomerization 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 the addition of Neu5Ac that 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 ± 0.1 × 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_{f} 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 °C in the presence of 100 mM NaCl). This is about 2-fold lower than the value we determined at 37 °C with no salt (138 ± 6 nm (data not shown), which is similar to the value of 120 ± 6 nm reported previously (13)). From this value of the K_{f} 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 R. 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 Arg-147/Arg-127/Asn-187 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_{f} 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 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_{f} 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 carboxylate group.
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 further investigate the contributions of the carboxylate and N-acetyl groups, we tested the binding of 4-acetylaminoxyphane 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

FIGURE 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.
Figure 4. Investigation of the presteady-state kinetics of Neu5Ac binding to SiaP monitored using stopped-flow fluorescence spectroscopy. A, trace SiaP (1 μM) pushed against buffer (flat line) and against 8 μM Neu5Ac. The binding data have been fit to a single exponential equation. The points on the graph are the averages from the three independent titrations. 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 k_{obs} 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.

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 seven 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 (Arg-147, Asn-187, Arg-127) as well as Phe-170 that stacks against the ligand (Fig. 6). The other residues in domain I that bind the glycerol moiety (Glu-67 and Asp-49) are conserved in 7 of 8 sequences; however, Asn-10, which 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 ESRs 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 Arg-147 (present in 98% of the sequences) that forms a salt bridge to the carboxylate group of Neu5Ac2en in SiaP. The region directly preceding Arg-147 is the most highly conserved region in the family (Asp-140 is 92% conserved, Gly-143 is 95% conserved, and Lys-145 is 86% conserved), suggesting that the correct positioning of Arg-147 within β6 is critical for function of the TRAP ESRs. Additionally, highly conserved residues pack against this region from above (Gly-162, 92% conserved) and below (Asp-183, 92% conserved). None of the other residues implicated in coordinating the Neu5Ac2en is 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 is 90% conserved, and Gly-59 is 91% conserved). These sit at turns in the domain after 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 γ5 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 characterized 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 from 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 (Arg-147 in SiaP) in domain II of the protein. In SiaP, the full coordination of the carboxylate also includes an interaction with Arg-127, 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 (MBP) (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 characterized ligands of TRAP transporters.

A unique feature of SiaP compared with 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, whereas 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 toward 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 that 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 that will result in an altered surface of this region after ligand binding. This is expected to be an energetically unfavorable 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 among proteins that 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
Sialic acid binding Ig-like lectin (Siglec) molecules found on eukaryotic cell surfaces that bind sialic acids but do not modify them. Siglecs have general roles in adhesion and signaling (48) and bind sialic acid in a surface groove of a V-set immunoglobulin-like fold with only one face of the Neu5Ac 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 $\sim 30 \mu M$, which is relatively low affinity compared with SiaP (13, 50).

After 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 regions (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 ion transporter 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 that interact with the additional helices found in domain I.

Although 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) that are found in a small number of uncharacterized 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 recognized 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 pertussis provides additional support for the widespread nature of the type II ESR fold for use with secondary transporters (51). 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, which forms a second 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 catalyze 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 tripartite tricarboxylate transporter families of ESR-dependent secondary transporters.

Acknowledgments—We thank the European Synchrotron Radiation Facility, Grenoble, for excellent data collection facilities. We also thank the participants of the SPINE meeting (York, July 11–15, 2005), which greatly assisted in the solution of the structure and Prof. Colin Kleanthous for comments on the manuscript.

REFERENCES
1. Wilkinson, A. J., and Verschueren, K. H. G. (2003) in ABC Proteins: From Bacteria to Man (Holland, I. B., Cole, S. P. C., Kuchler, K., and Higgins, C. F., eds) pp. 187–207, Academic Press, London
2. Dwyer, M. A., and Hellinga, H. W. (2004) Cur. Opin. Struct. Biol 14, 22221
Structure of a Sialic Acid-binding Protein, SiaP

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, 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., Singh, S., and Saier, M. H., Jr. (2003) *Infect. Immun.* 71, 679–692

10. Allen, S., Zaleski, A., Johnston, I. W., Gibson, B. W., and Apicella, M. A. (2005) *Infect. Immun.* 73, 5291–5300

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

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

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

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

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

16. Tetsch, L., and Kunte, H. J. (2002) *Mol. Microbiol.* 45, 1173–1185

17. May, A. P., Robinson, R. C., Vinson, M., Crocker, P. R., and Jones, E. Y. (1999) *Acta Crystallogr.* 55, 830–837

18. Moustafa, I., Connaris, H., Taylor, M., Zaitsev, V., Wilson, J. C., and Paton, J. C. (1998) *Structure* 6, 1553–1561

19. Martin, A., Richards, J. C., and Moxon, E. R. (1999) *Mol. Microbiol.* 34, 3521–3530

20. Miller, D. M., III, Olson, J. S., and Quiocho, F. A. (1980) *Science* 218, 628–633

21.patients, I., Belrhal, H., Antoine, R., Bompard, C., Locht, C., and Paton, J. C. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 719–728

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

23. Brossmer, R., and Holmequist, L. (1971) *Hoppe-Seyler’s Z. Physiol. Chem.* 352, 1715–1719

24. Schmid, W., Christian, R., and Zbiral, E. (1988) *Tetrahedron Lett.* 29, 3643–3646

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

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

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 Holmequist, L. (1971) *Hoppe-Seyler’s Z. Physiol. Chem.* 352, 1715–1719

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