Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9852–9856, November 1993
Biochemistry

Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase

SUSAN J. CRENNELL*, ELSPETH F. GARMAN†, W. GRAEME LAVER‡, ERIC R. VIMR§, AND GARRY L. TAYLOR*¶

*Department of Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom; †Laboratory of Molecular Biophysics, University of Oxford, Rex Richards Building, South Parks Road, Oxford OX1 3QU, United Kingdom; ‡The John Curtin School of Medical Research, The Australian National University, GPO Box 334, Canberra, ACT 2601, Australia; and §Department of Microbiology and Department of Veterinary Pathobiology, University of Illinois at Urbana–Champaign, 2001 South Lincoln Avenue, Urbana, IL 61801

Communicated by David Phillips, May 28, 1993

ABSTRACT **Sialidases (EC 3.2.1.18 or neuraminidases) remove sialic acid from sialoglycoconjugates, are widely distributed in nature, and have been implicated in the pathogenesis of many diseases. The three-dimensional structure of influenza virus sialidase is known, and we now report the three-dimensional structure of a bacterial sialidase, from *Salmonella typhimurium* LT2, at 2.0-Å resolution and the structure of its complex with the inhibitor 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid at 2.2-Å resolution. The viral enzyme is a tetramer; the bacterial enzyme, a monomer. Although the monomers are of similar size (≈380 residues), the sequence similarity is low (≈15%). The viral enzyme contains at least eight disulfide bridges, conserved in all strains, and binds Ca^{2+}, which enhances activity; the bacterial enzyme contains one disulfide and does not bind Ca^{2+}. Comparison of the two structures shows a remarkable similarity both in the general fold and in the spatial arrangement of the catalytic residues. However, an rms fit of 3.1 Å between 264 C_α atoms of the *S. typhimurium* enzyme and those from an influenza A virus reflects some major differences in the fold. In common with the viral enzyme, the bacterial enzyme active site consists of an arginine triad, a hydrophobic pocket, and a key tyrosine and glutamic acid, but differences in the interactions with the O4 and glycerol groups of the inhibitor reflect differing kinetics and substrate preferences of the two enzymes. The repeating "Asp-box" motifs observed among the nonviral sialidase sequences occur at topologically equivalent positions on the outside of the structure. Implications of the structure for the catalytic mechanism, evolution, and secretion of the enzyme are discussed.**

Sialidase was originally identified as a "receptor-destroying enzyme" in extracts of *Vibrio cholerae* because of its ability to release influenza virus from the surface of erythrocytes. Sialidases have been found in viruses, bacteria, trypanosomes, and mammalian cells (1, 2). There is evidence for two families of the bacterial enzyme, distinguished by a requirement for a divalent metal ion for maximal activity. Those not requiring metal have molecular weights of ≈42,000 and share some sequence similarity—e.g., the sialidases of *Clostridium perfringens*, *Clostridium sordelli*, *Salmonella typhimurium*, and *Micromonospora viridifaciens*. These enzymes also share sequence similarity with the N-terminal domain of the membrane-bound sialidase of *Trypanosoma cruzi* (3). However, the sialidase from *V. cholerae*, which has been cloned (4), sequenced (5), and crystallized (6), has a molecular weight of 82,000 and requires a metal ion. The influenza virus enzymes possess a calcium-binding site, and although calcium is not essential for activity, it does enhance activity (7). Within the nonviral enzymes there is evidence for a conserved sequence motif (Ser/Thr-Xaa-Asp-[Xaa]-Gly-Xaa-Thr-Trp/Phe), or "Asp box," which repeats three to five times along the sequences (8).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Comparison of Bacterial and Viral Structures

The only sialidase three-dimensional (3-D) structure previously known was that of the influenza virus enzyme (9–11). The influenza virus sialidase forms tetramers on the virus surface, which remain as tetramers when released from the virus by Pronase, whereas other sialidases investigated to date may be monomers (12). The viral monomer has a molecular weight similar to the small sialidases, requires a divalent metal ion for maximal activity, and does not generally possess the Asp-box motifs.

We have determined the crystal structure of a sialidase from *S. typhimurium* by the method of multiple isomorphous replacement (MIR). The structure has been refined to 2.0-Å resolution with a crystallographic *R* value of 0.189.‖ The statistics are presented in Table 1, and the experimental details are given in its legend.

A schematic view of the enzyme is shown in Fig. 1. The enzyme is mainly β-sheet with two small α-helical segments, with a shallow active site crevice, identified crystallographically by the soaking of DANA into crystals (Fig. 3), and on the opposite side it has a deep cleft extending ≈15 Å into the structure. This cleft proved effective in allowing differential binding of two Hg derivatives used in the phasing: *p*CMB with its bulky aromatic group was unable to penetrate the cleft, whereas $HgCl_2$ was able to reach cysteines deep inside the cleft. The fold topology is identical to that found in the influenza virus sialidases and consists of six four-stranded antiparallel β-sheets arranged as the blades of a propeller around an axis passing through the active site. However, there are major differences in the lengths of the β-strands and the loops between the sheets.

Unlike the *S. typhimurium* structure, the viral enzyme has a C-terminal extension which is involved in maintaining the tetramer through interactions with the first and second sheets of adjacent monomers. A sequence alignment based on structures of the bacterial enzyme and a viral enzyme is given in Fig. 2. In all influenza A and B sialidases sequenced to date, there are eight totally conserved disulfide bridges, which brace the sheets (25). In the *S. typhimurium* structure,

Abbreviations: 3-D, three-dimensional; MIR, multiple isomorphous replacement; *p*CMB, *p*-chloromercuribenzoate; DANA, 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid; NANA, *N*-acetylneuraminic acid.

¶To whom reprint requests should be addressed.

‖The atomic coordinates have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (entry codes 1SIL and 1SIM).

Table 1. X-ray data collection and phasing statistics

	Native	*p*CMB	Hg1	Hg4	DANA
Observations	31,846	50,741	21,381	46,320	27,589
Unique reflections	23,401	10,534	6,524	10,397	17,449
Completeness, %	70.2	93.6	94.1	70.5	80.4
Resolution, Å	2.0	2.7	3.2	2.6	2.2
R_I, %	6.5	4.4	6.9	5.4	5.8
No. of heavy atom sites		1 (A)	2 (A, B)	2 (A, C)	
Phasing power		1.1	1.6	1.2	
Cullis R		0.48	0.41	0.46	
Figure of merit	0.61 for 10,431 reflections				

$R_I = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$; phasing power = rms f_H/lack of closure; Cullis R = lack of closure/isomorphous difference; lack of closure $E = \Sigma||F_{PH} \pm F_P| - f_H|$, [rms $E = (\Sigma||F_{PH} \pm F_P| - f_H|^2/n)^{1/2}$]; rms $f_H = (\Sigma f_H^2/n)^{1/2}$; I = diffraction intensity; f_H = calculated heavy atom structure factor amplitude; F_P = structure factor amplitude of native crystals; F_{PH} = structure factor amplitude of derivative crystals; and sums are over all reflections n.

Sialidase from *S. typhimurium* LT2 crystallizes in the orthorhombic system, space group $P2_12_12_1$, with the unit cell a = 47.4 Å, b = 82.8 Å, and c = 92.4 Å, and with one molecule in the asymmetric unit (6). The structure was solved by using three mercury derivatives: a crystal soaked in 1 mM *p*-chloromercuribenzoate (*p*CMB) for 3 hr and two $HgCl_2$ derivatives, crystals from different crystallization batches known as Hg1 and Hg4, soaked in 1 mM $HgCl_2$ for 3 hr. All three derivatives showed substitution at one major site (A), while the two $HgCl_2$ derivatives had separate secondary Hg sites (B and C). X-ray diffraction data were collected on a Siemens area detector mounted on a rotating anode x-ray source operating at 40 kV and 80 mA. XDS was used for data processing and reduction (13), with the CCP4 suite of crystallographic programs being used for subsequent derivative processing (14). Initial heavy atom occupancies were derived and refined in Patterson space by using the program VECREF (15). The program MLPHARE (16) was then used for further refinement of the heavy atom parameters and refinement of phases derived from them. The MIR map calculated to 2.7 Å by using these phases was improved by using the solvent flattening method of Wang (17) as implemented by Leslie (18), the solvent content being calculated as 42%. A skeletonized representation of the electron density map was computed, then manually edited in the graphics package *O* (19), to give a polyalanine starting model. The sequence (20) was aligned to the refined polyalanine model by using the known heavy atom positions and disulfide bond site and the model was manually rebuilt in *O*. The published sequence was found to be incorrect for the last 50 residues; the sequence given in Fig. 2 is correct. Cycles of simulated annealing refinement (21), model building, and, finally, restrained least squares in TNT (22) using the native data to 2.0 Å gave the current model, which contains residues 2–382 (residue 1, Met, having been excised by *Escherichia coli* during expression) and has an R factor of 18.9% for all reflections between 6 and 2.0 Å with reasonable geometry (rms deviation from ideality in bond lengths is 0.017 Å; in bond angles it is 3.38°) ($R = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$). Inhibitor crystals were prepared by soaking crystals 1–3 days with 10 mM 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid (DANA) in the phosphate buffer used in the crystallization. The complex structure was refined in a similar way as the native structure, and it has an R factor of 16.9% for all reflections between 6 and 2.2 Å (rms bond and angle derivations, 0.017 Å and 3.35°). The complex structure contains 141 water molecules.

only one disulfide is observed, between residues 42 and 103 and linking the innermost β-strands of the first and second sheets.

The location of the Asp boxes at equivalent positions in the fold of the protein, at the turn between the third and fourth β-strands of the first four sheets, is shown in Fig. 1. The aromatic residues pack into a hydrophobic core that stabilizes the turn, with the aspartic residues pointing out into solvent. The turns have identical folds, with rms fits of 0.13–0.40 Å between the C_α atoms of the seven or eight amino acids of the four Asp boxes. They may be involved in secretion or protein folding, as they are on the surface, remote from the active site, and they do not appear in the viral enzyme.

Comparison of the Active Sites

The catalytic site of the influenza A and B virus sialidases has been characterized by binding of the inhibitor DANA or of the product *N*-acetylneuraminic acid (NANA), which is a weak inhibitor of the viral enzyme (11, 26, 28). Kinetic studies on the *S. typhimurium* enzyme have shown inhibition by DANA with a K_i similar to the viral enzyme but no inhibition by NANA, and a kinetic preference for sialyl α2→3 linkages over α2→6 similar to the influenza virus (29). The catalytic site shares several features with the viral enzyme (Fig. 4): (*i*) Three arginine residues (37, 246, and 309) stabilize the carboxylic acid group common to all natural sialic acid derivatives. (*ii*) A glutamic acid (361) stabilizes the position of one of the triad arginines. (*iii*) A tyrosine (342) approaches the sugar ring of the sialic acid from below; the tyrosine hydroxyl group is only 3.0 Å from the C1 and C2 carbons of DANA, a fact that lends support to its role in stabilizing a carbonium ion transition state intermediate (30). (*iv*) A hydrophobic pocket which accommodates the *N*-acetyl group of sialic acid is formed by Met-99, Trp-121, Trp-128, and Leu-175; in the viral enzymes it is formed by a tryptophan and an isoleucine. (*v*) A glutamate (231) donates a proton, a step that has been implicated in the viral enzyme mechanism (30).

Significant differences between the bacterial and viral active sites are as follows: (*i*) Interactions with sialic acid glycerol group. In the viral enzyme a glutamic residue forms two strong hydrogen bonds (2.5–2.9 Å) with O8 and O9 of the glycerol side group of DANA and NANA (11, 28), while the *S. typhimurium* enzyme has only one direct weak hydrogen bond, from Trp-128 to O9. The large difference in turnover number of 2700 s^{-1} for the *S. typhimurium* enzyme (29) and ≈10 s^{-1} for the viral enzyme (G. Air, personal communication) could be explained by the extra stabilization of the glycerol in the case of the viral enzyme, as it is not uncommon for product release to be the limiting factor in catalysis (31). This difference could also explain why a derivative of DANA with a bulky arylazide group at C9, in place of the hydroxyl, is bound by the bacterial enzymes (32). (*ii*) Interactions with sialic acid O4. Uniquely, the *S. typhimurium* enzyme has Asp-100 and Arg-56 making strong hydrogen bonds to the O4 of DANA (2.8 and 2.9 Å, respectively), plus a weak hydrogen bond from Asp-62 (3.2 Å); the viral enzyme just has the latter weak hydrogen bond from its corresponding aspartic residue. This difference explains the observation that *N*-acetyl-4-*O*-acetylneuraminyl-(2→3)-lactose is hydrolyzed by the viral enzyme but not by a bacterial enzyme (33): the bulky acetyl group inhibits binding to the bacterial enzyme, whereas in the viral enzyme there is room to accommodate this extra group, and substitution at this site has produced the most effective viral enzyme inhibitors to date (34).

Discussion

A recent comparison of seven sialidase sequences from five different genera and two protozoan sialidase sequences from *T. cruzi* demonstrates homology (common descent) among small and large enzymes, consistent with the existence of a sialidase superfamily (12, 20). Together with the current structural analysis of the *S. typhimurium* sialidase active site, the results indicate that all sialidases should have similar catalytic mechanisms. Our current results strongly indicate that the influenza virus enzyme is part of the sialidase superfamily. Furthermore, the possibility that bacteria originally acquired the sialidase genetic information from eukaryotes is consistent with the phylogenetic distribution of sialic acids and sialidases (12, 20) and with the structural similar-

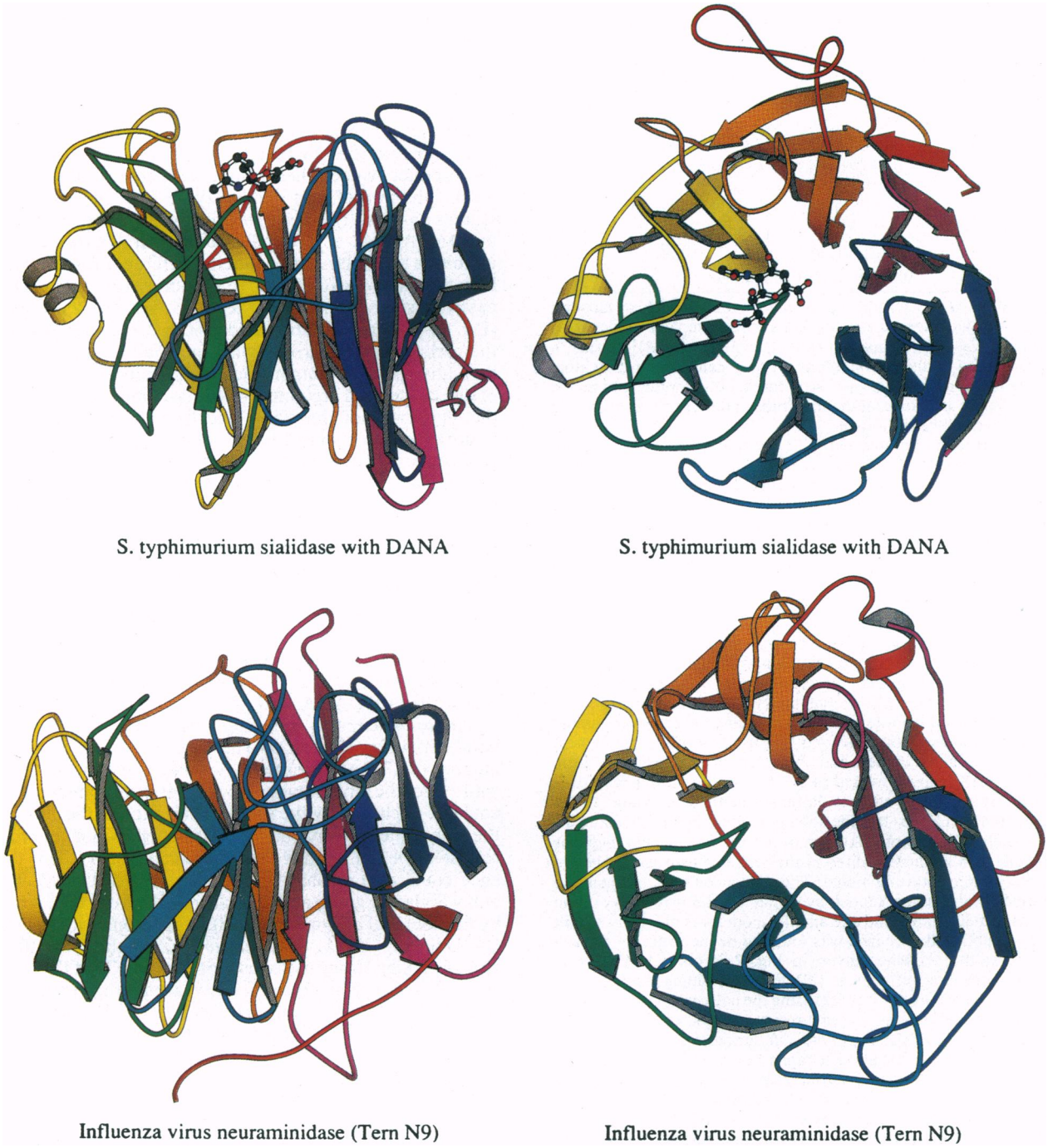

FIG. 1. Orthogonal views of the sialidase from *S. typhimurium* with bound inhibitor DANA (*Upper*) and the neuraminidase from tern influenza virus subtype N9 (*Lower*) produced by the MOLSCRIPT program (23), looking from the side (*Left*) and from above the active site (*Right*). The N9 structure is viewed after optimizing its fit to the bacterial structure by using the program SHP (24), which gave an rms fit of 3.1 Å for 264 C_α atoms. In both structures, the chain is colored from red at the N terminus to violet at the C terminus. (The sequences of the enzymes are aligned in Fig. 2.) Note that, in both enzymes, the sixth sheet is formed from an N-terminal β-strand and three C-terminal β-strands. Note that in the *Lower Right* view the viral tetramer is formed by a fourfold axis approximately through the bottom right of the picture and normal to the page.

ities between microbial sialidases demonstrated in this article.

The function(s) of bacterial sialidases is (are) not entirely clear. Nucleotide sequencing and biochemical analysis of *S. typhimurium nanH* and its encoded sialidase indicated a primarily intracellular enzyme location and the absence of a typical prokaryotic signal sequence (21, 29). When this *nanH* copy was subcloned in a sialidase-negative *V. cholerae* strain, which normally excretes its own sialidase (4), up to 25% of the total *S. typhimurium* sialidase activity was found in the culture medium compared with less than 2% of β-lactamase encoded by the vector (S. M. Steenbergen and E.R.V., unpublished data). It may be that the Asp boxes, which are not generally conserved in the viral sialidases, are involved with sialidase secretion and that these "signals" are not being efficiently recognized by *S. typhimurium*. Regardless of whether Asp boxes play any role in secretion, *S. typhimurium* LT2 is able to use sialyl-α2-3-lactose as a sole carbon and energy source, while a *nanH* mutant strain cannot (S. M. Steenbergen and E.R.V., unpublished data), consistent with a primarily nutritional function for most bacterial sialidases (20). Thus, any role of the enzyme in bacterial

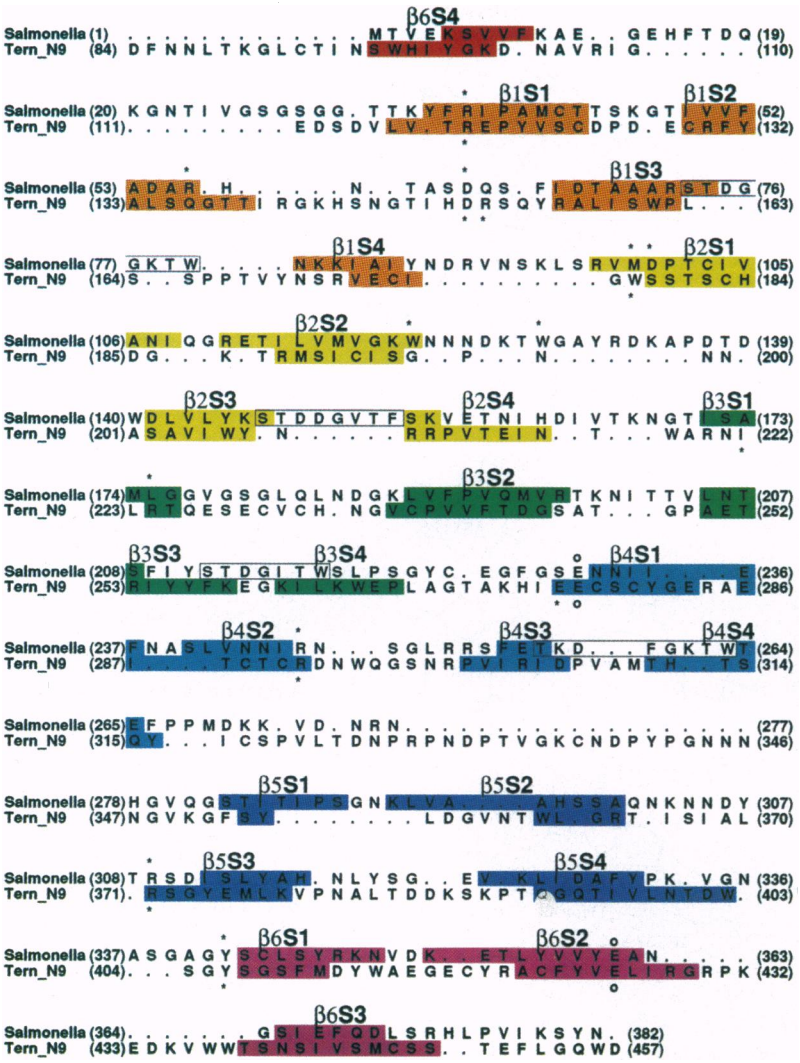

FIG. 2. An alignment of the *S. typhimurium* and tern influenza virus N9 sequences based on topologically equivalent residues. Dots indicate gaps introduced for optimal alignment. The β-strands follow the nomenclature of Colman (25); *βiSj* refers to the *j*th strand in the *i*th sheet with the innermost strand being strand 1, and the residues are color coded as in Fig. 1. Numbers at the start and end of lines refer to the residue numbering for the two sequences. Residues that interact with DANA are marked with an *; residues in both structures that stabilize active-site residues are marked with a °. The N9 residues involved in inhibitor interactions are taken from Bossart-Whitaker *et al.* (26). The Asp boxes are boxed in the bacterial sequence. The picture was produced with the aid of ALSCRIPT (27).

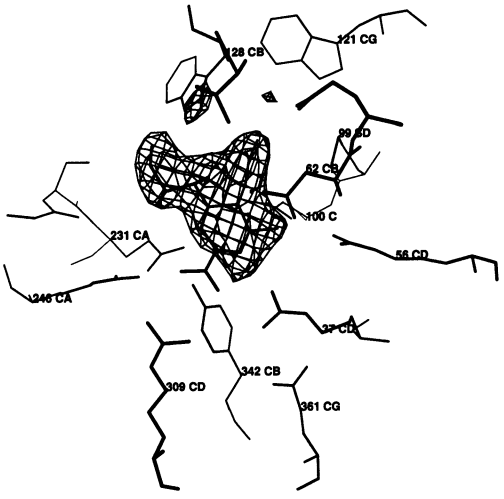

FIG. 3. A 2.2-Å ($F_{DANA} - F_{native}$) $\exp^{i\phi_{calc}}$ difference electron density map, contoured at 4σ (F_{DANA}, derivative structure factors; F_{native}, native structure factors; ϕ_{calc}, phases derived from the refined 2.0-Å native model). The atomic coordinates displayed are from the refined complex at 2.2 Å. Note the absence of electron density for the carboxyl group of DANA; this is explained by the fact that this site is occupied by a phosphate group in the native structure. The C_α atoms of 10 active-site residues with and without DANA have an rms fit of 0.16 Å. The only significant movements of side chains within the active site are those of Trp-121 and Trp-128, which move slightly towards the *N*-acetyl group of DANA. CA, C_α; CB, C_β; etc.

pathogenesis may be a secondary consequence of this nutritional function. Knowledge of bacterial sialidase structure–function will be central to the design of compounds for the treatment and prevention of diseases caused by sialidase-positive microbes, and it should help to provide insight into a variety of noninfectious diseases of humans that involve derangements in sialic acid catabolism.

We thank Gillian Air, Tony Corfield, Robert Eisenthal, Robert Webster, and Ian Williams for useful discussions. This work was supported by a grant from the Wellcome Trust in the United Kingdom (G.L.T. and S.J.C.). E.R.V. was supported, in part, by a grant from National Institute of Allergy and Infectious Diseases (AI-23039).

1. Corfield, T. (1992) *Glycobiology* **2,** 509–521.
2. Corfield, A. P., Lambre, C. R., Michalski, J.-C. & Schauer, R. (1992) *Conferences Philippe Laudat 1991*, Institut National de la Sante et de la Recherche Medicale (INSERM, Paris), pp. 113–134.
3. Pereira, M. E. A., Mejia, J. S., Ortega-Barria, E., Matzilevich, D. & Prioli, R. P. (1991) *J. Exp. Med.* **174,** 179–191.
4. Vimr, E. R., Lawrisuk, L., Galen, J. & Kaper, J. B. (1988) *J. Bacteriol.* **170,** 1495–1504.
5. Galen, J. E., Ketley, J. M., Fasano, A., Richardson, S. H., Wasserman, S. S. & Kaper, J. B. (1992) *Infect. Immun.* **60,** 406–415.
6. Taylor, G. L., Vimr, E., Garman, E. & Laver, G. (1992) *J. Mol. Biol.* **226,** 1287–1290.
7. Chong, A. K. J., Pegg, M. S. & von Itzstein, M. (1991) *Biochim. Biophys. Acta* **1077,** 65–71.

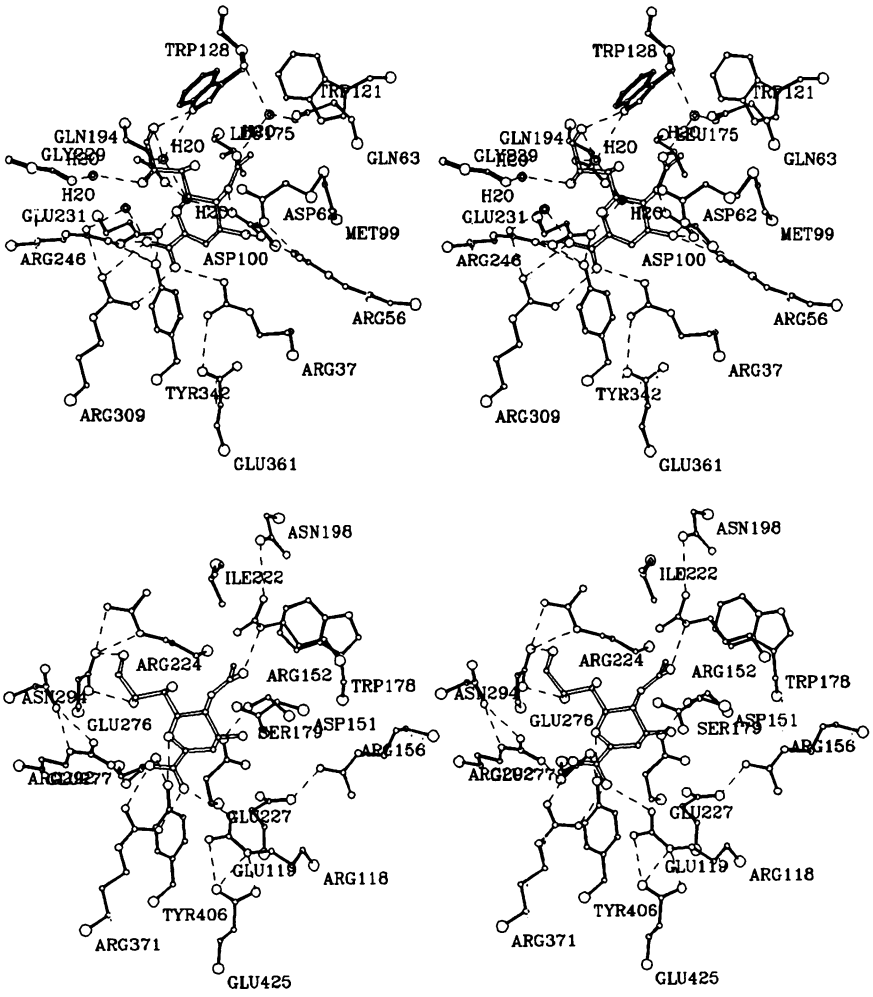

FIG. 4. Stereoviews of the active site. (*Upper*) *S. typhimurium* sialidase with DANA. (*Lower*) Tern influenza virus N9 sialidase with DANA (26). Hydrogen bonds (<3.2 Å) are drawn as broken lines. Water molecules (H2O) in *Upper* are drawn as two concentric circles. Both sites are viewed from the same direction, looking into the active site. The N9 structure has been optimally fitted to the bacterial structure as described in the legend of Fig. 1. The C_α coordinates of seven residues common to both sites have an rms fit of 1.01 Å (Arg-37/118, Asp-62/151, Glu-231/277, Arg-246/292, Arg-309/371, Tyr-342/406, and Glu-361/425).

8. Roggentin, P., Rothe, B., Kaper, J. B., Galen, J., Lawrisuk, L., Vimr, E. R. & Schauer, R. (1989) *Glycoconjugate J.* **6**, 349–353.
9. Tulip, W. R., Varghese, J. N., Baker, A. T., van Donkelaar, A., Laver, W. G., Webster, R. G. & Colman, P. M. (1991) *J. Mol. Biol.* **221**, 487–497.
10. Varghese, J. N. & Colman, P. M. (1991) *J. Mol. Biol.* **221**, 473–486.
11. Burmeister, W. P., Ruigrok, R. W. H. & Cusack, S. (1992) *EMBO J.* **11**, 49–56.
12. Roggentin, P., Schauer, R., Hoyer, L. L. & Vimr, E. R. (1993) *Mol. Microbiol.*, in press.
13. Kabsch, W. J. (1988) *J. Appl. Crystallogr.* **21**, 916–924.
14. CCP4 (1979) *The SERC (U.K.) Collaborative Computing Project No. 4: A Suite of Programs for Protein Crystallography* (Daresbury Lab., Warrington, U.K.).
15. Tickle, I. J. (1991) in *Isomorphous Replacement and Anomalous Scattering: Proceedings of CCP4 Study Weekend*, eds. Wolf, W., Evans, P. R. & Leslie, A. G. W. (SERC Daresbury Lab., U.K.), pp. 87–95.
16. Otwinoski, Z. (1991) in *Isomorphous Replacement and Anomalous Scattering: Proceedings of CCP4 Study Weekend*, eds. Wolf, W., Evans, P. R. & Leslie, A. G. W. (SERC Daresbury Lab., U.K.), pp. 80–86.
17. Wang, B. C. (1985) *Methods Enzymol.* **115**, 90–112.
18. Leslie, A. G. W. (1989) in *Improving Protein Phases: Proceedings of CCP4 Study Weekend*, eds. Bailey, S., Dodson, E. & Phillips, S. (SERC Daresbury Lab., U.K.), pp. 13–24.
19. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119.
20. Hoyer, L. L., Hamilton, A. C., Steenbergen, S. M. & Vimr, E. R. (1992) *Mol. Microbiol.* **6**, 873–884.
21. Brunger, A. T. (1988) *J. Mol. Biol.* **203**, 803–816.
22. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987) *Acta Crystallogr. Sect. A* **43**, 489–501.
23. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
24. Stuart, D. I. (1979) Ph.D. thesis (Univ. of Bristol, U.K.).
25. Colman, P. M. (1989) in *The Influenza Viruses*, ed. Krug, R. M. (Plenum, New York), pp. 175–218.
26. Bossart-Whitaker, P., Carson, M., Babu, Y. S., Smith, C. D., Laver, W. G. & Air, G. M. (1993) *J. Mol. Biol.*, in press.
27. Barton, G. J. (1993) *Protein Eng.* **6**, 37–40.
28. Varghese, J. N., McKimm-Breschkin, J. B., Caldwell, J. B., Krott, A. A. & Colman, P. M. (1992) *Proteins Struct. Funct. Genet.* **14**, 327–332.
29. Hoyer, L. L., Roggentin, P., Schauer, R. & Vimr, E. R. (1991) *J. Biochem.* **110**, 462–467.
30. Burmeister, W. P., Henrissat, B., Bosso, C., Cusack, S. & Ruigrok, R. W. H. (1993) *Structure* **1**, 19–26.
31. Fersht, A. (1985) *Enzyme Structure and Mechanism* (Freeman, New York), 2nd Ed.
32. Warner, T. G., Harris, R., McDowell, R. & Vimr, E. R. (1992) *Biochem. J.* **285**, 957–964.
33. Messer, M. (1974) *Biochem. J.* **139**, 415–420.
34. von Itzstein, M. L., Wu, W.-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L., White, H. F., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M. & Penn, C. R. (1993) *Nature (London)* **363**, 418–426.