The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll

Andrew Gaskell, Susan Crennell and Garry Taylor*

School of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

Background: Sialidases, or neuraminidases, have been implicated in the pathogenesis of many diseases, but are also produced by many non-pathogenic bacteria. Bacterial sialidases are very variable in size, often possessing domains in addition to the catalytic domain. The sialidase from the non-pathogenic soil bacterium *Micromonospora viridifaciens* is secreted in two forms with molecular weights of 41 kDa or 68 kDa, depending on the nature of the carbohydrate used to induce expression.

Results: We report here the X-ray crystal structures of the 41 kDa and 68 kDa forms of the sialidase from *M. viridifaciens* at 1.8 Å and 2.5 Å resolution respectively. In addition, we report a complex of the 41 kDa form with an inhibitor at 2.0 Å resolution, and a complex of the 68 kDa form with galactose at 2.5 Å. The 41 kDa form shows the canonical sialidase β-propeller fold. The 68 kDa form possesses two additional domains, one with an immunoglobulin-like fold that serves as a linker to the second, which is homologous to the galactose-binding domain of a fungal galactose oxidase.

Conclusions: The presence of the additional carbohydrate-binding domain in the 68 kDa form of the bacterial sialidase reported here is a further example of a combination of carbohydrate binding and cleaving domains which we observed in the sialidase from *Vibrio cholerae*. This dual function may be common, not only to other bacterial and parasitic sialidases, but also to other secreted glycosidases involved in pathogenesis. The bacterium may have acquired both the immunoglobulin module and the galactose-binding module from eukaryotes, as the enzyme shows a remarkable similarity to a fungal galactose oxidase which possesses similar domains performing different functions and assembled in a different order.

Structure 15 November 1995, **3**:1197–1205

Key words: galactose binding, immunoglobulin fold, *Micromonospora viridifaciens*, neuraminidase, sialidase

Introduction

Sialidases, or neuraminidases, catalyze the removal of terminal sialic acid residues from various glycoconjugates, and play roles in pathogenesis, bacterial nutrition and cellular interactions [1]. The influenza virus neuraminidase, which plays a role in processing progeny virus particles from infected cells, is well characterized. The X-ray structures of neuraminidases from several influenza subtypes have been reported [2–7], and these structures have been used for the rational design of a novel inhibitor [8] which is undergoing clinical trials as a prophylactic against influenza. Bacterial sialidases have been implicated in many diseases, for example: gas gangrene (*Clostridium*); septicaemia (*Streptococcus*, *Pneumococcus*, *Bacteroides*, *Corynebacterium*); pneumonia (*Streptococcus*); peritonitis (*Clostridium*, *Bacteroides*); meningitis (*Streptococcus B*) and cholera (*Vibrio cholerae*) [1]. In the parasite *Trypanosoma cruzi*, a novel trans-sialidase has been characterized [9], which has been implicated in cell attachment and immune response evasion in the infective trypomastigote form of the parasite [10]. Many non-pathogenic bacteria also secrete sialidases The function of these enzymes is to scavenge host-cell sialic acid to use as a carbon and energy source. Such a nutritional role has also been found for some pathogenic bacteria. For example, although *V. cholerae* sialidase plays a defined role in pathogenesis by processing of higher-order gangliosides to G_{M1} receptors for cholera toxin, the bacterium also possesses the enzymes for metabolism of sialic acid.

Sequence analysis of bacterial sialidases shows them to be of diverse length with low sequence identity and molecular weights ranging from ~40–115 kDa. There are two motifs common among the sequences: one is the 'FRIP' motif (Phe-Arg-Ile-Pro, or variations on this theme), which occurs upstream of the second motif, the 'Asp-box' (most commonly Ser/Thr-X-Asp-[X]-Gly-X-Thr-Trp/Phe) which repeats three to five times along the sequences [11]. Our previous studies on two bacterial sialidases from *Salmonella typhimurium* and *V. cholerae* showed that the arginine of the FRIP motif represents one of the catalytic triad of arginines that binds the carboxylate group of sialic acid, and that the Asp-boxes are remote from the active site, in topologically equivalent locations. The structure of the 42 kDa sialidase from *S. typhimurium* shows the same six-bladed β-propeller seen in the influenza virus neuraminidase, despite a sequence identity of only 15% with the influenza enzyme [12]. Each blade of the propeller is a four-stranded antiparallel β-sheet, with the Asp-boxes forming the turn between the third and fourth strands of each sheet on the periphery of the propeller. The aspartate residues of each Asp-box are exposed to solvent and the aromatic group of the Asp-box tucks in between adjacent sheets. The function of these Asp-boxes is unknown: they may simply dictate the fold, although they are apparently absent from the influenza enzymes, or they may play a role in a protein secretory mechanism. The structure of the 83 kDa sialidase from *V. cholerae* revealed a central β-propeller catalytic domain flanked by two lectin-like domains, one

*Corresponding author.

 ISSN 0969-2126

at the N terminus and one inserted between the second and third blades of the propeller [13]. Although no carbohydrate-binding site has yet been identified in these lectin-like wings, it is known that active *V. cholerae* sialidase is able to bind to cells, and we have speculated that the lectin domains play a role in cell adhesion in the turbulent environment of the small intestine.

The sialidase from the soil bacterium *Micromonospora viridifaciens* was first purified from a bacterial culture filtrate [14]. When colominic acid was used as an inducer, a soluble 41 kDa sialidase was secreted. However, when milk casein was used as an inducer, a higher molecular weight (68–72 kDa) soluble sialidase was secreted [15]. Subsequent cloning and sequencing of the sialidase gene from *M. viridifaciens* and expression in *Streptomyces lividans*, revealed an enzyme with a molecular weight of 68 kDa which is easily digested to 52 kDa and 41 kDa forms, most likely through attack by *S. lividans* proteases [16]. All three forms, 41 kDa, 52 kDa and 68 kDa, have almost identical catalytic properties suggesting that the 41 kDa form contains the catalytic domain. The *M. viridifaciens* sialidase is a member of the family 33 glycosyl hydrolases, as are the sialidases from *S. typhimurium* and *V. cholerae*, in contrast to the influenza virus neuraminidase which belongs to family 34 [17,18].

Here we report the crystal structures of both the 41 kDa and 68 kDa forms of the enzyme. The latter exhibits a modular architecture which may be common to many other sialidases.

Results and discussion

We have determined the crystal structure of the fully active 41 kDa form of the sialidase from *M. viridifaciens* by the method of multiple isomorphous replacement. The structure has been refined to 1.82 Å resolution with a crystallographic R-value of 17.3%. A complex of the 41 kDa form of the enzyme with the inhibitor 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid (Neu5Ac2en, DANA) has been refined to 2.0 Å with an R-value of 15.9%. We have also determined the structure of the larger 68 kDa form of the sialidase by molecular replacement combined with careful weighted difference-Fourier analyses, and have refined the structure to 2.5 Å resolution with an R-value of 21.4%. A complex of the 68 kDa form with galactose has been refined to 2.5 Å, with an R-value of 22.8%.

Protein architecture of the 41 kDa form

The enzymatically active 41 kDa fragment, with 358 residues, forms the canonical β-propeller fold (Fig. 1) as seen in the influenza virus neuraminidase and bacterial sialidases from *S. typhimurium* and *V. cholerae*. Superposition of the *M. viridifaciens* sialidase onto other sialidase structures reveals that it has greatest similarity to the catalytic domain of *V. cholerae* sialidase (333 Cα atoms superimpose with a root mean square deviation [rmsd] of 2.31 Å) with which it shares 26% sequence identity. The five Asp-boxes of *M. viridifaciens* sialidase occur in topologically equivalent positions as observed in the other two bacterial sialidases, namely at the turn between the third and fourth strands of each sheet (Fig. 1b). Four of the Asp-boxes have very similar conformations, maintained by a hydrogen-bonding

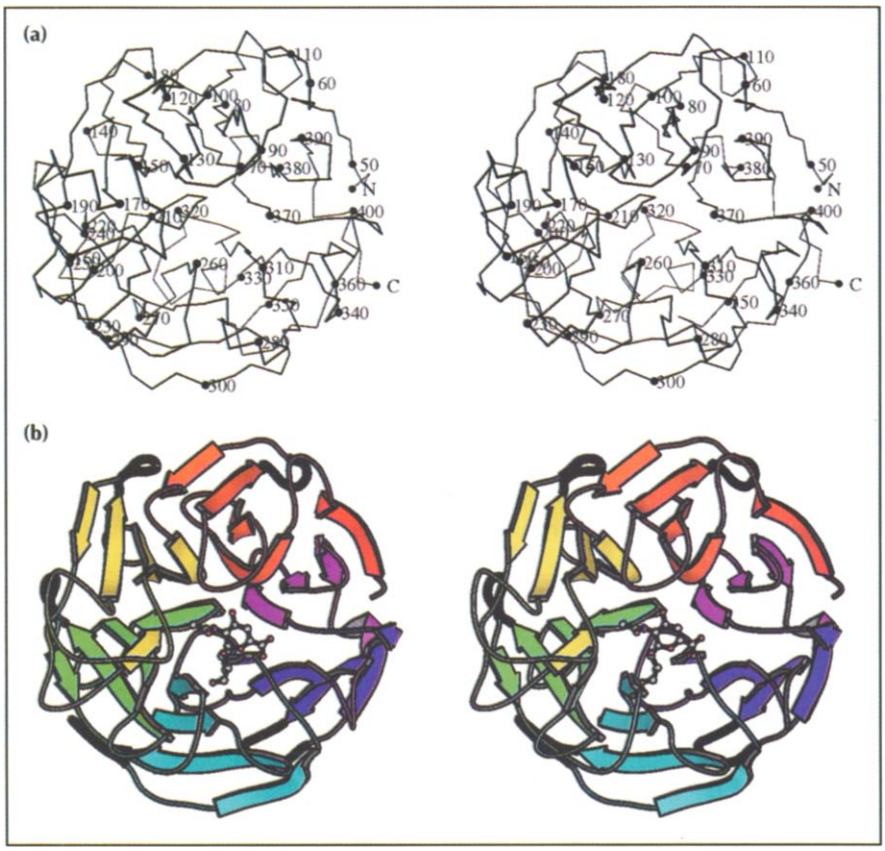

Fig. 1. β-propeller fold of the 41 kDa form of *M. viridifaciens* sialidase. **(a)** Stereo Cα trace of the 41 kDa form (residues 47–404) viewed from above the active site. **(b)** Stereoview, in the same orientation as (a), showing the six-bladed propeller structure with the inhibitor Neu5Ac2en bound. The location of the Asp-box motifs are highlighted as thick black turns. (Figure drawn using MOLSCRIPT [41].)

network between the conserved residues of the motif, and with the tryptophan or phenylalanine plugging into a hydrophobic pocket between the sheet to which the Asp-box belongs and the preceding sheet. The fifth Asp-box, residues 289–296, deviates from the others due to substitution of a tyrosine (Tyr296) for the tryptophan. This tyrosine is able to hydrogen bond to His294 resulting in a distortion of the loop from its normal conformation.

The Asp-box motif, therefore, appears simply to dictate the fold and integrity of the propeller, and may not be involved in a secretory mechanism as we proposed previously [12]. A similar motif was recently reported for a bacterial methanol dehydrogenase, dubbed the 'tryptophan docking motif', which is present in a topologically equivalent position in each sheet of an eight-bladed β-propeller structure [19]. However, its similarities to the sialidase Asp-box were not noted. This may suggest divergence from some ancestral four-stranded sheet, which through gene duplication formed six-bladed or eight-bladed propeller structures with different functions. Such a scheme for the evolution of these propeller folds has been discussed with reference to the *Drosophila kelch* motif observed in seven-bladed propeller structures [20]. Bork and Doolittle [20] argue that the the seven-bladed fungal galactose and fungal/prokaryotic glyoxal oxidases, as well as assorted proteins containing *kelch* motifs, have evolved from some early precursor which also diverged to form the six-bladed sialidases/neuraminidases and the eight-bladed methanol and methylamine dehydrogenases.

Active site of the 41 kDa form

The residues involved in the binding of substrate were identified by soaking a crystal of the 41 kDa *M. viridifaciens* sialidase in the inhibitor Neu5Ac2en, which inhibits most sialidases at micromolar concentrations. As observed in other sialidases, the active site is apparently rigid because little movement of the protein occurs upon binding inhibitor. Interactions between the protein and the carboxylate end of the inhibitor are conserved in all sialidase structures studied to date, both bacterial sialidases and the influenza virus neuraminidase [12,13], and these are detailed below (see also Fig. 2 and Table 1).

Firstly, three arginines (Arg68, Arg276 and Arg342) form interactions in a plane with the carboxylate group of Neu5Ac2en. Secondly, the position of the first arginine (Arg 68) is stabilized by a glutamic acid (Glu386) through a hydrogen bond between the Nε of the arginine and Oδ1 of the glutamic acid. Thirdly, a tyrosine (Tyr370) and a glutamic acid (Glu260) hydrogen bond with each other and sit beneath, and close to, the C1–C2 bond of the inhibitor. It has been suggested that this tyrosine and glutamic acid combination is involved in the stabilization of an oxycarbonium ion intermediate in the catalysis [21]. Fourthly, an aspartic acid (Asp92) sits above the sugar ring

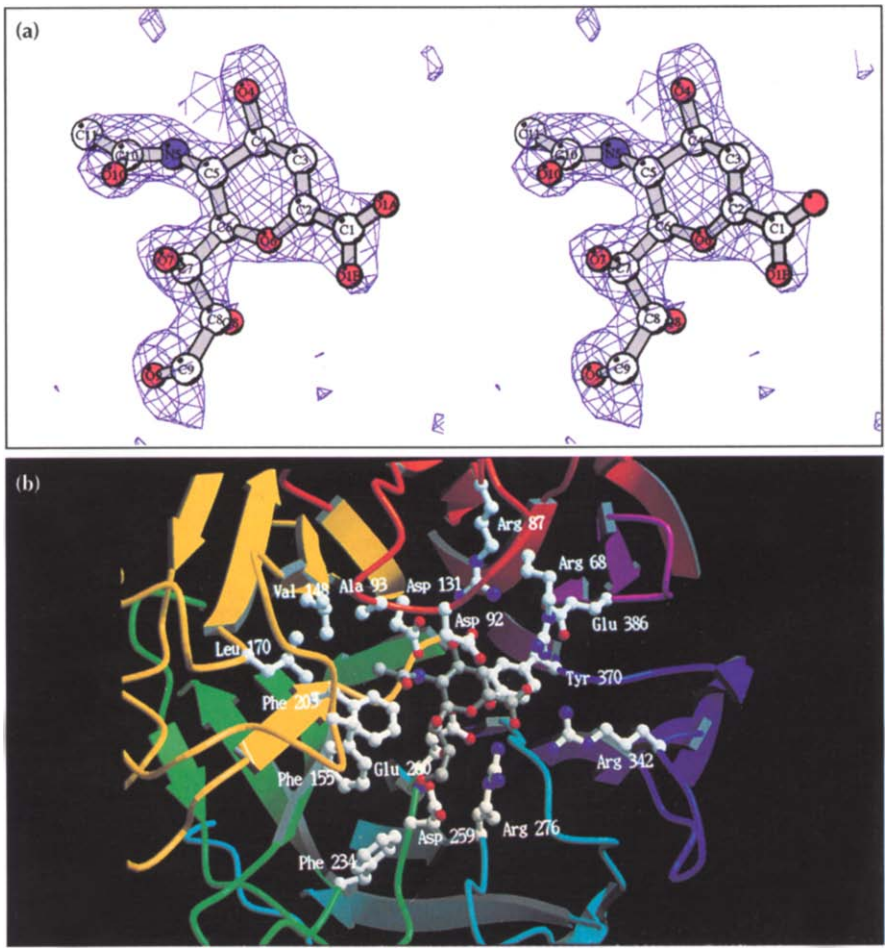

Fig. 2. Inhibitor bound to the 41 kDa form of the enzyme. **(a)** Stereoview of the 2.0 Å difference Fourier electron-density map for Neu5Ac2en, contoured at 3.5σ. The refined inhibitor is superimposed. **(b)** View of the active site of the 41 kDa form with Neu5Ac2en drawn with grey bonds. Residues which form hydrogen-bonding or van der Waals interactions with the inhibitor are shown. (Figure generated with RASTER-3D rendering software [42,43].)

Table 1. Hydrogen-bonding interactions between *M. viridifaciens* sialidase and the inhibitor Neu5Ac2en (DANA).

Residue	Atom	Interacting with DANA	Distance (Å)
Arg68	Nη1, Nη2	O1A	2.73, 2.91
Arg87	Nη2	O4	3.03
Asp92	Oδ1	O4	2.72
Asp131	Oδ1	N5	2.96
Asp131	Oδ2	O4	2.75
Asp259	Oδ1	O8	2.69
Asp259	Oδ2	O9	3.11
Arg276	Nη2	O1B	3.21
Arg342	Nη1	O1A	2.84
Arg342	Nη2	O1B	2.95

on a loop between the second and third strands of the first sheet. A suggested role for this residue in the influenza enzyme, is in the stabilization of a proton-donating water molecule involved in the hydrolysis, rather than as a direct proton donor to the glycosidic bond [21,22].

In contrast, interactions between the protein and other chemical groups on the inhibitor differ between the sialidases. In common with the sialidases from *S. typhimurium* and *V. cholerae*, *M. viridifaciens* sialidase has an arginine (Arg87) and an aspartic acid (Asp131) which interact with the O4 hydroxyl of the inhibitor. The absence of these interactions in the influenza virus neuraminidase produces a pocket close to O4. Structure-based drug design, in which the 4-hydroxyl on Neu5Ac2en has been replaced with a guanidinium group, fills this pocket and produces an inhibitor with close to picomolar binding affinity for the influenza enzyme [23]. Not surprisingly, the 4-guanidino compound is a much poorer inhibitor of bacterial sialidases. It is interesting to note that the 4-guanidino inhibitor exploits a conserved glutamic acid residue in the influenza enzyme which takes the position of the isoleucine in the bacterial/eukaryotic FRIP motif, at which position the influenza enzyme has a conserved REP motif (Arg-Glu-Pro). No functional role has been assigned to this glutamic acid in the influenza enzyme, and it raises the possibility that the highly efficient virus could mutate this residue without loss of function when subjected to the new evolutionary pressure of the designed inhibitor.

The *N*-acetyl group of Neu5Ac2en is accommodated in a hydrophobic pocket, as observed in the other sialidase structures. In the *M. viridifaciens* enzyme, the pocket is formed from residues Ala93, Val148, Phe155, Leu170 and Phe203. The O8 and O9 hydroxyls of the glycerol moiety of the inhibitor form hydrogen bonds with Asp259 which precedes the catalytic Glu260, in a similar fashion to the pair of adjacent glutamic acid residues seen in the influenza enzyme.

M. viridifaciens sialidase, in common with many bacterial sialidases, can cleave α(2→3), α(2→6) and α(2→8) linked sialic acids [14]. *S. typhimurium* sialidase is unusual in that it has a 260-fold kinetic preference for α(2→3) linked sialic acid over α(2→6), similar to the preference of the influenza enzyme [24]. In the structures of sialidases studied to date, no binding site has been observed for the carbohydrate group adjacent to the sialic acid moiety in natural substrates. Variations in substrate specificity appear to be due to differences in the lengths of the loops around the active site causing steric hindrance.

Protein architecture of the 68 kDa form

The 68 kDa form of the sialidase consists of 601 residues arranged as three domains: first, the 41 kDa β-propeller domain (residues 47–402); second, a β-sandwich linker domain (residues 403–505); and third, a β-sandwich galactose-binding domain (residues 506–647) (Fig. 3). The only significant structural change between the isolated 41 kDa form and the same domain in the 68 kDa form is in residues 402–404, which move to connect to the linker domain. Two adjacent glycine residues (Gly402, Gly403) between the 41 kDa domain and the linker may act as a pivot point for movement of the other two domains.

Linker domain

The linker domain (residues 403–505) possesses an immunoglobulin (Ig) fold. The β-sandwich comprises four antiparallel strands over three antiparallel strands and is reminiscent of the constant domains of immunoglobulins [25] (Fig. 4a). In contrast to most Ig folds, an almost cylindrical seven-stranded β-barrel structure is formed, which is stabilized by the first strand (strand *a* in Ig nomenclature) forming hydrogen bonds with strands *b* and *g*. Strand *a* accomplishes this feat by having a proline halfway along (Pro411) which shifts the strand direction

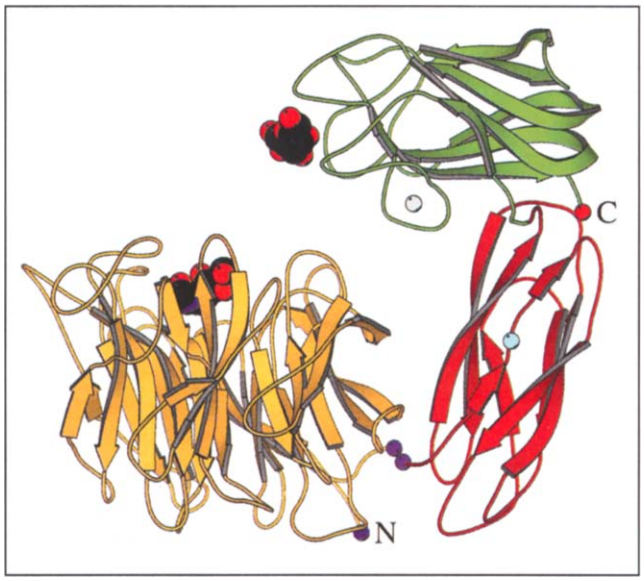

Fig. 3. View of the 68 kDa form of *M. viridifaciens* sialidase. The 41 kDa domain is shown in yellow, the linker in red and the galactose-binding domain in green. The Neu5Ac2en and galactose molecules are drawn in space-filling mode. The monovalent cation is drawn as a grey sphere, the two glycines (Gly402, Gly403), which may form a hinge, are drawn as purple spheres, and the proline residue in the linker domain, which causes the first strand to dogleg between the sheets, is drawn as a light blue sphere.

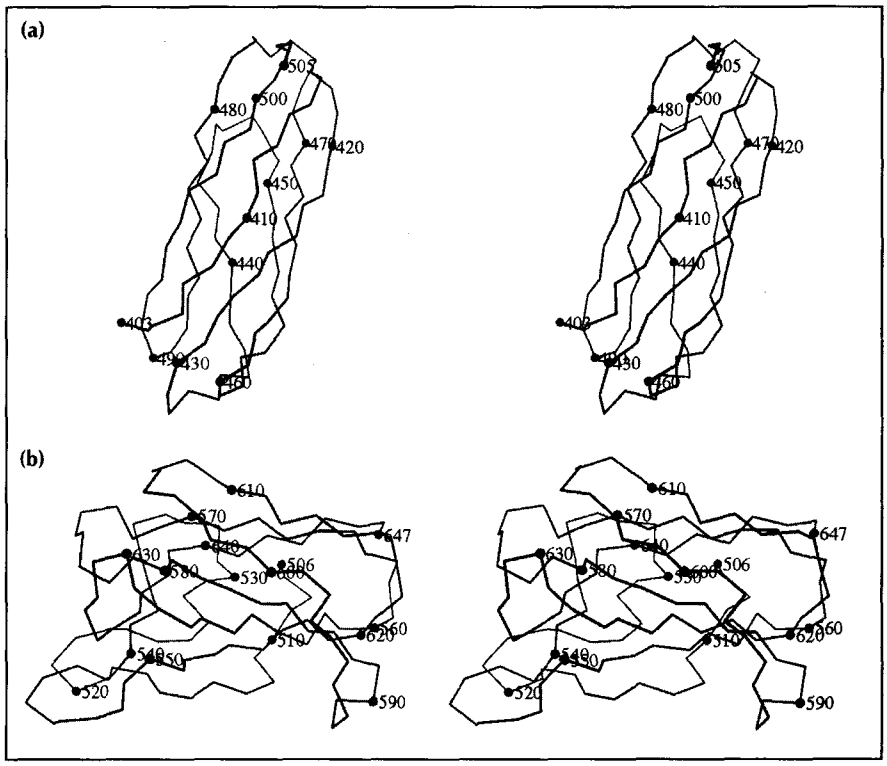

Fig. 4. Stereoview Cα traces of the additional domains in the 68 kDa form of the enzyme. **(a)** The immunoglobulin-like linker domain (residues 403–505). **(b)** The galactose-binding domain (residues 506–647).

from the *b* strand to the *g* strand on the opposite sheet. We compared the topology of the linker fold with the entire April 1995 release of the Brookhaven Protein Data Bank (PDB) using PROTEP [26]. This comparison revealed that the linker has greatest topological similarity to several light-chain variable regions of immunoglobulins (although they possess two additional strands), to the C-terminal domain of CD4, and to the N-terminal domain of the chaperone protein PapD (PDB code 3DPA). It is interesting to note that, in PapD, strand *a* also doglegs to form hydrogen bonds to both strands *b* and *g*, leading to a more cylindrical barrel similar to that in the sialidase linker.

The sialidase linker does not contain any disulphide bonds and its sequence does not conform to consensus patterns reported for Ig folds [25]. It therefore represents another variation on the Ig fold. With its cylindrical shape and well-packed hydrophobic core, its role in the sialidase appears to be that of a rigid linker between the other two domains.

Galactose-binding domain

The C-terminal domain of the *M. viridifaciens* sialidase (residues 506–647) shares 35.4% sequence identity with the N-terminal domain of a fungal galactose oxidase from *Dactylium dendroides*, which was found to possess a galactose-binding site [27]. Most of the residues identified as being involved in galactose binding or cation binding in galactose oxidase are conserved in this third sialidase domain. The two domains are very similar in structure and superimpose with an rmsd of 1.16 Å for 139 Cα atoms. The topology is that of a β-sandwich with a five-stranded antiparallel β-sheet packed on top of a three-stranded antiparallel β-sheet, and can be described as a jelly-roll (Fig. 4b). A cation, which maintains the conformation of a loop (residues 528–536), was

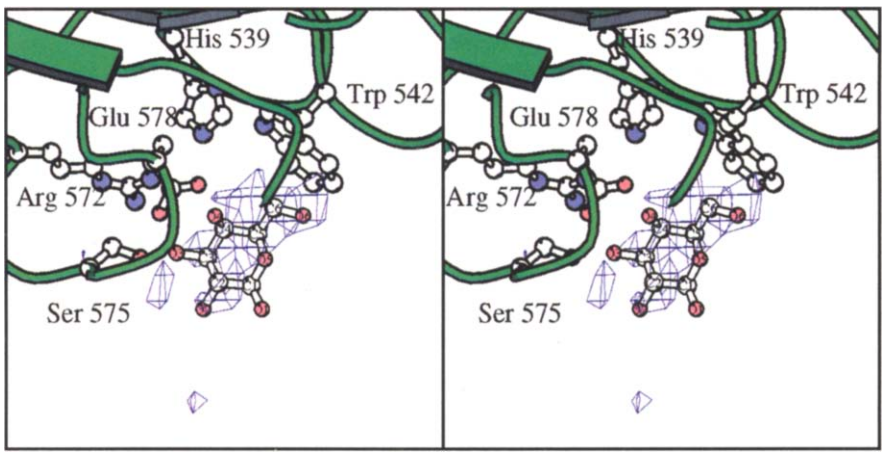

Fig. 5. Stereoview of the 2.5 Å difference Fourier electron-density map showing the binding of galactose and its environment in the 68 kDa form of *M. viridifaciens* sialidase. Map contoured at 3σ.

suggested to be Na^+ or a similar monovalent cation in the structure of galactose oxidase. The same ligands coordinate the ion in both *M. viridifaciens* sialidase and galactose oxidase, in a pentagonal bipyramidal configuration.

The location of the galactose-binding site was confirmed by soaking a crystal of the 68 kDa form in ~50 mM β-D-galactose for 4 h (Fig. 5). The major interactions between galactose and the protein are: hydrogen bonding of the O3 and O4 hydroxyls of galactose to Arg572; hydrogen bonding of O4 to Oε1 of Glu578, and a weak interaction between O4 and Nε2 of His539; a weak interaction with Oγ of Ser575; and finally, van der Waals interactions between C6 and Trp542. The galactose-binding site is located ~30 Å directly above the active site of the 41 kDa domain.

Biological implications

Sialidases have been implicated in the pathogenesis of many diseases, but are also secreted by several non-pathogenic bacteria for which they play a nutritional role. Bacterial sialidases range in size from 40 kDa to 115 kDa with little sequence identity between them apart from two small sequence motifs, one of which repeats up to five times along the sequence. The soil bacterium *Micromonospora viridifaciens* secretes two forms of a sialidase, with molecular weights of 41 kDa, and 68 kDa, depending on the substrate used to induce expression.

The 68 kDa form of the sialidase from *M. viridifaciens* encompasses the 41 kDa catalytic, six-bladed β-propeller domain and has two additional domains downstream of it. One of these domains, with the topology of an immunoglobulin module, appears to serve solely as a rigid linker arm to the last, galactose-binding, domain. This final domain is homologous to the galactose-binding domain found in a fungal galactose oxidase. Intriguingly, the *M. viridifaciens* sialidase represents a variation on a theme seen in the fungal enzyme, suggesting that they may be related in more ways than one. The catalytic domain of galactose oxidase has a seven-bladed β-propeller fold, with the galactose-binding domain situated at its N terminus, and a distorted immunoglobulin module at its C terminus. It is conceivable that *M. viridifaciens* has acquired the immunoglobulin and galactose-binding modules from eukaryotes such as fungi or animals. This bacterial sialidase and galactose oxidase may also share a common ancestor in a primitive four-stranded β-sheet, which, through multiple gene duplication events (as many as six or seven times) and functional divergence, has provided two domains with very different catalytic functions.

***M. viridifaciens* secretes the 41 kDa form of the enzyme when colominic acid (poly α2→8 *N*-acetylneuraminic acid) is used as the inducer, but secretes the 68 kDa form when milk casein is used. It may be that sialic acid on the casein glycoproteins together with galactose, either on the glycoproteins or the lactose 'contaminant' often found with casein, is inducing the larger form of the enzyme. The galactose-binding domain may anchor the enzyme to appropriate substrates allowing the catalytic domain to do its work. Alternatively it may work in concert with the catalytic domain in helping to feed appropriate oligosaccharides to the active site. Such concerted activity may require the relative movement of the catalytic and galactose-binding domains which appears possible around a flexible pivot point created by a pair of consecutive glycine residues.**

Our earlier discovery of two lectin-like domains in the sialidase from *Vibrio cholerae*, together with the findings presented here on the *M. viridifaciens* sialidase, suggest that this combination of catalytic and carbohydrate domains may be common to other members of the sialidase family. These carbohydrate-binding domains may play roles in cell adhesion or targeting specific substrates, and in the case of pathogenic bacteria may provide additional targets for therapeutic intervention.

Materials and methods

41 kDa form

Crystals of the 41 kDa form of *M. viridifaciens* sialidase were grown by the hanging drop vapour diffusion method from an 8% polyethylene glycol (PEG) 3350 solution at pH 5.0 or 5.5 as described previously [28]. The enzyme has its maximal activity at pH 5.0 [14]. The enzyme crystallizes in space group $P2_12_12_1$ with unit cell a=48.14 Å, b=82.73 Å and c=84.75 Å. There is one enzyme molecule in the asymmetric unit giving an estimated solvent content of 40%. All data were collected at room temperature (~18°) on a Siemens area detector mounted on a rotating anode X-ray source. The programs XDS and XSCALE were used for all data processing and reduction [29]. Inhibitor complex data were collected from a crystal soaked for 3 h in 5 mM Neu5Ac2en (DANA). The unit cell of the inhibitor-soaked crystals is a=48.00 Å, b=82.86 Å and c=84.94 Å.

Initial attempts were made at molecular replacement using the structure of *S. typhimurium* sialidase, with which *M. viridifaciens* sialidase shares 20% sequence identity and with the catalytic domain of *V. cholerae* sialidase, but these proved fruitless. The structure was solved using three mediocre derivatives. Derivatives were identified through Patterson maps and cross-phase difference Fourier maps using the CCP4 package [30], and the program MLPHARE [31] was used for refinement of the heavy-atom parameters and the phases derived from them (Table 2). A 3.0 Å multiple isomorphous replacement (MIR) map was used together with a solvent-flattened map, determined with the Wang method in the implementation of Leslie [32], in the interpretation of the structure. In spite of a relatively poor MIR map, many β-strands were evident. Interpretation was aided by a 3.0 Å difference Fourier map calculated from the inhibitor complex data. Even with the poor MIR phases used to calculate this difference Fourier, the asymmetric shape of the inhibitor was evident and this aided the positioning

Table 2. X-ray data collection and phasing statistics.

Derivative	41 kDa Native	41 kDa $CdCl_2$	41 kDa $La(ac)_3$	41 kDa KI_3	41 kDa Inhibitor	68 kDa Native	68 kDa with Galactose
No. of observations	168296	24180	24521	24577	93040	49584	35106
No. of unique reflections	29469	6327	6665	6814	20366	20112	16297
Resolution (Å)	1.82	3.0	3.0	3.0	2.0	2.5	2.5
Overall completeness (>1σ) (%)	93.4	85.3	86.7	87.9	83.0	79.2	59.3
Overall RI (%)*	5.5	5.6	8.2	7.9	6.3	3.9	5.4
Completeness (>1σ) (%)	71.7	57.8	58.3	58.0	67.4	76.8	49.3
in top shell	(2–1.82 Å)	(3.2–3 Å)	(3.2–3 Å)	(3.2–3 Å)	(2.5–2 Å)	(2.7–2.5 Å)	(2.7–2.5 Å)
R_I (%)* in top shell	9.4	8.2	14.4	13.9	11.2	9.2	11.8
Heavy-atom concentration (mM)		2	2	0.2			
Soaking time (days)		2	3	10			
Heavy-atom sites		C, D	A	B			
Phasing power† (acentric, centric)		0.8, 0.6	0.7, 0.4	0.4, 0.3			
Cullis R‡ (acentric, centric)		0.42, 0.74	0.62, 0.86	0.50, 0.93			

Figure of merit is 0.40 for all 6231 reflections between 20–3 Å.

*$R_I=\Sigma|I-\langle I\rangle|/\Sigma\langle I\rangle$, where I=diffraction density. †Phasing power=rms f_H/E, where f_H=calculated heavy-atom structure-factor amplitude and E=lack of closure=$\Sigma||F_{PH}\pm F_P|-f_H|$ (F_{PH}=structure-factor amplitude of derivative crystals and F_P=structure-factor amplitude of native crystals). ‡Cullis R=lack of closure/isomorphous difference. Sums over all reflections n.

of the homologous *S. typhimurium* sialidase–inhibitor complex model onto the *M. viridifaciens* sialidase MIR map. Skeletonized representations of both the MIR and solvent-flattened maps were computed and manually edited using the graphics package O [33] to give a continuous Cα trace. Conserved active-site residues in the two bacterial sialidases and the location of the Asp-boxes served as landmarks in the interpretation.

The model was refined in X-PLOR [34] initially using data to 3.0 Å. Cycles of simulated annealing refinement were followed by manual rebuilding with O using experimental, SIGMAA-weighted $2F_o-F_c$ [35] and omit maps. Data to 1.82 Å were gradually included, water molecules were identified and the whole refined to give the final model (Table 3), which contains residues 47–404 with 270 water molecules and has an R-value of 0.173 for all data between 6.0 Å and 1.82 Å. Residues 1–37 code a signal peptide, with the N terminus of the 41 kDa product reported to start at residue 43 [16]. However, no electron density is observed for residues 43–46. The C terminus of the 41 kDa product has not been determined chemically, but there is no interpretable density beyond residue 404. The final model has no residues in disallowed regions of the Ramachandran plot, and a 3D-1D plot showed that all the residues are in chemically sensible environments [36]. A Luzatti plot gave an estimated error in atomic coordinates of 0.15–0.20 Å.

A difference Fourier map was calculated with coefficients $(F_{Neu5Ac2en}-F_{Native})\exp i\phi_{calc}$ with phases calculated from the 1.82 Å model after water molecules had been removed from the active site, and clearly revealed the conformation of the inhibitor (Fig. 2a). The protein–inhibitor complex was refined using X-PLOR to give a final model which includes 265 water molecules with an R-value of 0.159 for all data from 6.0–2.0 Å (Table 3).

68 kDa form

The 68 kDa form was successfully derived from molecular replacement and careful difference Fourier maps, even with data that were initially incomplete to 3.5 Å. The progress of this structure solution is therefore given in some detail.

The 68 kDa form of *M. viridifaciens* sialidase was purified by the same method as the 41 kDa protein [28]. Crystals were obtained by microseeding followed by macroseeding using the hanging drop vapour diffusion method. All wells contained 0.1 M Tris buffer pH 8.4 and 8% PEG 8000 as precipitant. The initial drop contained 3 μl of protein at a concentration of 12 mg ml^{-1}, and 3 μl of well solution. A 3 μl volume of a small crystal suspension obtained from the initial drop was then dissolved in 100 μl of the well solution and aspirated violently. Next, 1 μl of this solution was added to 3 μl of protein and 3 μl of new well solution.

Table 3. Model refinement of the 41 kDa and 68 kDa forms.

Model	41 kDa Native	41 kDa Inhibitor complex	68 kDa Native	68 kDa Galactose complex
Number of non-hydrogen protein atoms	2701	2721	4536	4548
Number of water molecules/ions	270/0	265/0	0/1	0/1
Resolution range (Å)	6.0–1.82	6.0–2.0	6.0–2.5	6.0–2.5
Number of reflections	28814	20015	18632	15102
R-factor for all data (%)	17.3	15.9	21.4	22.8
Rmsd in bond lengths (Å)	0.016	0.020	0.016	0.012
Rmsd in bond angles (°)	2.57	2.74	2.57	1.96
Average temperature factors ($Å^2$)				
main-chain/side-chain all domains	14.7/21.8	14.8/24.3	17.7/19.0	
main-chain/side-chain domain 1	–	–	17.3/17.6	
main-chain/side-chain domain 2	–	–	19.4/23.6	
main-chain/side-chain domain 3	–	–	17.3/19.4	
inhibitor/galactose	–	12.9	–	50.5

Finally, the larger crystals obtained from this trial were washed in successive buffers containing increasing concentrations of PEG 8000 up to 8% and macroseeded in a solution containing 1 μl of protein and 3 μl of well solution. Crystals with dimensions of up to 0.4×0.1×0.1 mm^3 were obtained.

The crystals belong to space group $P2_1$, with unit cell a=51.1 Å, b=117.1 Å, c=60.0 Å and β=95.6°, giving a calculated solvent content of 53%. Crystals diffracted to only 4 Å on the in-house rotating anode source, and a low resolution set of data was collected and processed using XDS. Molecular replacement using the 41 kDa model was carried out with the AMoRE package [37] which gave a clear solution for the rotation and translation functions. Rigid-body refinement of the model gave a crystallographic R-value of 0.415 using data from 10–6 Å. A 6 Å SIGMAA-weighted $2F_o–F_c$ difference Fourier map, calculated using phases from the partial model, clearly showed additional electron density for the extra 27 kDa of protein which is located downstream of the 41 kDa fragment.

At this point a partial native data set was collected at the DESY synchrotron, Hamburg, on station X31 at a wavelength of 0.993 Å, where diffraction to 2.8 Å was observed. Only 54° of data were collected from one crystal at room temperature in 1° frames which were processed using XDS. Synchrotron data were merged with the in-house data using XSCALE to give a reasonable working set of data to 3.5 Å, which was only 57% complete ($I>\sigma[I]$) with an R_I of 12.7%. Because of the paucity of the data at this stage, the free R-factor [38] was employed throughout the building and refinement procedures using X-PLOR, with 5% of the reflections as the test set. The 41 kDa domain was kept fixed throughout all subsequent refinement and the B-factors from the 1.82 Å refined 41 kDa structure were maintained. The 41 kDa domain was rigid-body refined in X-PLOR using data from 8–3.5 Å, giving an R-value of 0.425 (R_{free}=0.445). A SIGMAA-weighted $2F_o–F_c$ difference map was calculated using phases from this rigid-body refined partial model, which was subjected to eight cycles of solvent-flattening using the Leslie implementation of the Wang procedure but with a SIGMAA-weighted difference Fourier being back-transformed on each cycle. The resulting solvent-flattened map showed distinct electron density for the extra 27 kDa of protein, which appeared to form two domains (Fig. 6). A FastA [39] search of the SWISSPROT database with the 27 kDa sequence revealed a 34.6% sequence identity within a 127 amino acid region of galactose oxidase from *Dactylium dendroides*, the structure of which has been elucidated [27]. The N-terminal galactose-binding domain of galactose oxidase was moved manually in O as a rigid body into the *M. viridifaciens* 3.5 Å difference electron-density map, and was found to fit remarkably well. The 41 kDa domain and galactose-binding domain (with B-factors set to 15 $Å^2$) mutated to the *M. viridifaciens* sequence, were refined in X-PLOR as rigid bodies. The galactose-binding domain was then subjected to 200 cycles of positional refinement. Further SIGMAA-weighted and solvent-flattened difference Fourier maps allowed tentative tracing of the intervening domain and adjustments to the galactose-binding domain. Positional refinement of the galactose-binding domain and of a polyalanine model for the linker domain was carried out. After several rounds of rebuilding, all atoms of the additional domains of the 68 kDa form were subjected to positional refinement, with B-factors fixed at 15 $Å^2$, the 41 kDa domain still remaining fixed, to give a final R-value of 0.249 (R_{free}=0.297).

At this point a more complete synchrotron data set to higher resolution was collected in Hamburg on station X31, on a crystal which diffracted to 2.5 Å. A total of 101° of data were collected at 4°C in 1° steps. The data were integrated using the DENZO package (Table 2) [40]. This new crystal, of dimensions 0.4×0.2×0.2 mm^3 had grown slowly over two months using 4% PEG 4000 as precipitant, with an initial protein concentration of 12 mg ml^{-1} in 0.1 M Tris-HCl buffer at pH 8.2. The 3.5 Å model minus the linker domain was used as the starting model for simulated annealing refinement in X-PLOR against this new data (6.0–2.5 Å with a 2σ cutoff) to give an R-value of 0.307 (R_{free}=0.423). A SIGMAA-weighted $2F_o–F_c$ map confirmed the connectivity of the linker domain, but allowed rebuilding of some regions which were in doubt. After several rounds of simulated annealing refinement, rebuilding and finally isotropic B-factor refinement, the final R-value was 0.214 (R_{free}=0.340) (Table 3). The final model consists of 601 residues (47–647), and one cation arbitrarily assigned as a sodium. There are only two residues in the disallowed region of the Ramachandran plot, Val421 in the linker domain and Ala545 in the galactose-binding domain, and from a 3D-1D plot all residues are in acceptable environments. A Luzatti plot gave an estimated error in coordinates of 0.30 Å.

The previously exposed native crystal was soaked in ~50 mM β-D-galactose for 4 h, and 76° of data collected again on station X31 (Table 2). A difference Fourier map calculated with coefficients $(F_{Galactose}–F_{Native})\exp i\phi_{calc}$ with phases calculated from the 2.5 Å model revealed electron density into which a galactose molecule could be fitted. There was no apparent movement of the protein side chains. A single round of positional refinement was carried out to produce a final model with an R-value of 0.228 for all data between 6.0 Å and 2.5 Å (Table 3).

Atomic coordinates have been deposited in the Brookhaven Protein Data Bank.

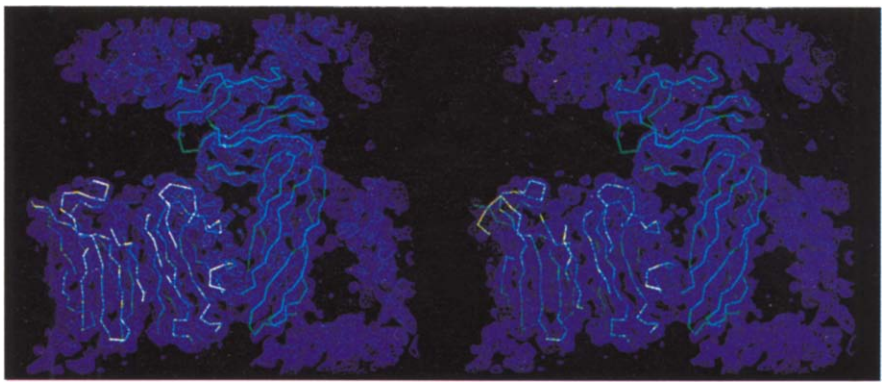

Fig. 6. Additional electron density for the 68 kDa form of the sialidase revealed in a difference Fourier. Stereoview of the 3.5 Å SIGMAA-weighted, solvent-flattened $2F_o–F_c$ difference Fourier map based on phases calculated from a molecular replacement solution of the 41 kDa domain only. The map is contoured at 1σ. The Cα trace of the final 68 kDa model is superimposed with the 41 kDa domain in yellow and the additional 27 kDa domains in blue.

Acknowledgements: We thank Dr G Laver for stimulating our interest in this project, Drs G Laver, M Hasegawa and T Uwajima for gift of the enzyme, Dr T Corfield for useful discussions, Drs J Rossjohn and R Russell for help in data collection at Hamburg and the staff of the EMBL Outstation, Hamburg for their assistance. Dr P Artymiuk is thanked for carrying out structural comparisons using PROTEP, which proved both useful in establishing the topology of the linker domain, and in revealing its homology to members of the immunoglobulin family of folds. This work was supported by a Wellcome Trust Project Grant (035997) to GLT and a BBSRC studentship to AG, and the European Union through its support of the work at EMBL Hamburg through the HCMP Access to Large Installations Project, Contract Number CHGE-CT93-0400.

References

1. Corfield, T. (1992). Bacterial sialidases — roles in pathogenicity and nutrition. *Glycobiology* **2**, 509–521.
2. Varghese, J.N., Laver, W.G. & Colman, P.M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature* **303**, 35–40.
3. Varghese, J.N. & Colman, P.M. (1991). 3-Dimensional structure of the influenza virus A/Tokyo/3/67 at 2.2 Å resolution. *J. Mol. Biol.* **221**, 473–486.
4. Tulip, W.R., *et al.*, & Colman, P.M. (1991). Refined atomic structures of N9 subtype influenza virus neuraminidase and escape mutants. *J. Mol. Biol.* **221**, 487–497.
5. Burmeister, W.P., Ruigrok, R.W.H. & Cusack, S. (1992). The 2.2 Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. *EMBO J.* **11**, 49–56.
6. Bossart-Whitaker, P., Carson, M., Babu, Y.S., Smith, C.D., Laver, W.G. & Air, G.M. (1993). Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy-2,3-dehydro-*N*-acetyl neuraminic acid. *J. Mol. Biol.* **232**, 1069–1083.
7. Janakiraman, M.N., White, C.L., Laver, W.G., Air, G.M. & Luo, M. (1994). Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8 Å resolution: implications for the catalytic mechanism. *Biochemistry* **33**, 8172–8179.
8. von Itzstein, M.L., *et al.*, & Penn, C.R. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363**, 418–426.
9. Pereira, M.E.A., Mejia, J.S., Ortega-Barria, E., Matzilevich, D. & Prioli, R.P. (1991). The *Trypanosoma cruzi* neuraminidase contains sequences similar to bacterial neuraminidases, YWTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin. *J. Exp. Med.* **174**, 179–191.
10. Schenkman, S. & Eichiniger, D. (1993). *Trypanosoma cruzi* trans-sialidase and cell invasion. *Parasitol. Today* **9**, 218–222.
11. Roggentin, P., *et al.*, & Schauer, R. (1989). Conserved sequences in bacterial and viral sialidases. *Glycoconjugate J.* **6**, 349–353.
12. Crennell, S.J., Garman, E.F., Laver, W.G., Vimr, E.R. & Taylor, G.L. (1993). Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. *Proc. Natl. Acad. Sci. USA* **90**, 9852–9856.
13. Crennell, S.J., Garman, E.F., Laver, W.G., Vimr, E.R. & Taylor, G.L. (1994). Crystal structure of a *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure* **2**, 535–544.
14. Aisaka, K. & Uwajima, T. (1987). Production of neuraminidase by *Micromonospora viridifaciens*. *FEMS Microbiol. Lett.* **44**, 289–291.
15. Aisaka, K., Igarashi, A. & Uwajima, T. (1991). Purification, crystallization and characterisation of neuraminidase from *Micromonospora viridifaciens*. *Agricultural Biol. Chem.* **55**, 997–1004.
16. Sakurada, K., Ohta, T. & Hasegawa, M. (1992). Cloning, expression and characterisation of the *Micromonospora viridifaciens* neuraminidase gene in *Streptomyces lividans*. *J. Bacteriol.* **174**, 6896–6903.
17. Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.
18. Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788.
19. Ghosh, M., Anthony, C., Harlos, K., Goodwin, M.G. & Blake, C. (1995). The refined structure of the quinoprotein methanol dehydrogenase from *Methylobacterium extorquens* at 1.94 Å. *Structure* **3**, 177–187.
20. Bork, P. & Doolittle, R.F. (1994). *Drosophila kelch* motif is derived from a common enzyme fold. *J. Mol. Biol.* **236**, 1277–1282.
21. Burmeister, W.P., Henrissat, B., Bosso, C., Cusack, S. & Ruigrok, R.W.H. (1993). Influenza B virus neuraminidase can synthesize its own inhibitor. *Structure* **1**, 19–26.
22. Chong, A.K.J., Pegg, M.S., Taylor, N.R. & von Itzstein, M. (1992). Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.* **207**, 335–343.
23. Holzer, C.T., von Itzstein, M., Jin, B., Pegg, M.S., Stewart, W.P. & Wu, W.-Y. (1993). Inhibition of sialidases from viral, bacterial and mammalian sources by analogues of 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid modified at the C-4 position. *Glycoconjugate J.* **10**, 40–44.
24. Hoyer, L.L., Roggentin, P., Schauer, R. & Vimr, E.R. (1991). Purification and properties of cloned *Salmonella typhimurium* LT2 sialidase with virus-typical kinetic preference for sialyl $\alpha 2 \rightarrow 3$ linkages. *J. Biochem.* **110**, 462–467.
25. Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold: structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320.
26. Grindley, H.M., Artymiuk, P.J., Rice, D.W. & Willett, P. (1993). Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J. Mol. Biol.* **229**, 707–721.
27. Ito, N., Phillips, S.E.V., Yadav, K.D.S. & Knowles, P.F. (1994). Crystal structure of a free-radical enzyme, galactose-oxidase. *J. Mol. Biol.* **238**, 794–814.
28. Taylor, G., Dineley, E., Glowka, M. & Laver, G. (1992). Crystallization and preliminary crystallographic study of neuraminidase from *Micromonospora viridifaciens*. *J. Mol. Biol.* **225**, 1135–1136.
29. Kabsch, W.J. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown cell constants. *J. Appl. Cryst.* **26**, 795–800.
30. Collaborative Computing Project, Number 4 (1994). The CCP4 Suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
31. Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering: Proceedings of CCP4 Study Weekend.* (Wolf, W., Evans, P.R. & Leslie, A.G.W., eds), pp. 80–86, SERC Daresbury Laboratory, Warrington, UK.
32. Leslie, A.G.W. (1989). Solvent flattening versus additional derivative data in the improvement of phases. In *Improving Protein Phases: Proceedings of CCP4 Study Weekend.* (Bailey, S., Dodson, E. & Phillips, S., eds), pp. 13–24, SERC Daresbury Laboratory, Warrington, UK.
33. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
34. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458–460.
35. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149.
36. Luthy, R., Bowie, J.U. & Eisenberg, D. (1992). Assessment of protein models with 3-dimensional profiles. *Nature* **356**, 83–85.
37. Navaza, J. (1992). AMoRE: an automated package for molecular replacement. *Acta Cryst. A* **50**, 157–163.
38. Brünger, A.T. (1992). The free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–474.
39. Pearson, W.R. & Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
40. Otwinowski, Z. (1993). Oscillation data reduction program. In *Data Collection and Processing: Proceedings of CCP4 Study Weekend.* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
41. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
42. Bacon, D.J. & Anderson, W.F. (1988). A fast algorithm for rendering space-filling molecule pictures. (Abstract of paper presented at the Seventh Annual Meeting of the Molecular Graphics Society). *J. Mol. Graphics* **6**, 219–220.
43. Merritt, E.A. & Murphy, M.E.P. (1994). Raster 3D Version 2.0. A program for photorealistic molecular graphics. *Acta Cryst. D* **50**, 869–873.

Received: **27 Jul 1995**; revisions requested: **16 Aug 1995**; revisions received: **29 Aug 1995**. Accepted: **31 Aug 1995**.