

SCIENTIFIC REPORTS

OPEN

Pathological macromolecular crystallographic data affected by twinning, partial-disorder and exhibiting multiple lattices for testing of data processing and refinement tools

Ivan Campeotto^{1,2,3}, Andrey Lebedev⁴, Antoine M. M. Schreurs⁵, Loes M. J. Kroon-Batenburg⁵, Edward Lowe², Simon E. V. Phillips^{1,4}, Garib N. Murshudov⁶ & Arwen R. Pearson^{1,7}

Twinning is a crystal growth anomaly, which has posed a challenge in macromolecular crystallography (MX) since the earliest days. Many approaches have been used to treat twinned data in order to extract structural information. However, in most cases it is usually simpler to rescreen for new crystallization conditions that yield an untwinned crystal form or, if possible, collect data from non-twinned parts of the crystal. Here, we report 11 structures of engineered variants of the *E. coli* enzyme N-acetylneuraminic lyase which, despite twinning and incommensurate modulation, have been successfully indexed, solved and deposited. These structures span a resolution range of 1.45–2.30 Å, which is unusually high for datasets presenting such lattice disorders in MX and therefore these data provide an excellent test set for improving and challenging MX data processing programs.

Twinning is a crystal growth anomaly or lattice disorder in which the crystal is composed of separate domains of differing orientations¹. Twinning has posed a challenge in macromolecular crystallography since the earliest days^{2,3} and multiple computational approaches have been developed in order to treat twinned data in order to extract structural information. Several exhaustive reviews are available that discuss twinning and the methods to address it in detail^{1,4–7} nevertheless, for clarity, we give here a brief description of this phenomenon. Twinning is characterised by the twin law (a set symmetry operators, which relate the different orientations of the domains); and the twin fractions, α_i , that characterise the relative volumes of the twinning domains. There are several types of twinning: merohedral twinning (when the twin operators are a subset of the exact rotational symmetry of the lattice); pseudo-merohedral twinning (when the twin operators approximate the rotational symmetry of the lattice); and non-merohedral twinning or epitaxial twinning (when the twin operators have the rotational symmetry of a sublattice in three or fewer dimensions). In this paper we present examples of pseudo-merohedral twinning.

¹Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK. ²Biochemistry Department, Oxford University, South Parks Road, Oxford, OX1 1HY, UK. ³Leicester Institute of Structural and Chemical Biology, University of Leicester, Lancaster Road, Leicester, LE1 7RH, UK. ⁴Research Complex at Harwell (RCaH), Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxford, OX11 0FA, UK. ⁵Department of Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands. ⁶Structural Studies Division, MRC-LMB, Francis Crick Avenue, Cambridge, CB2 0QH, UK. ⁷Hamburg Centre for Ultrafast Imaging, Institute of Nanostructure and Solid State Physics, Universität Hamburg, CFEL, Luruper Chaussee 149, 22761, Hamburg, Germany. Correspondence and requests for materials should be addressed to I.C. (email: ivan.campeotto@leicester.ac.uk) or A.R.P. (email: arwen.pearson@cfel.de)

Received: 23 January 2018

Accepted: 19 September 2018

Published online: 05 October 2018

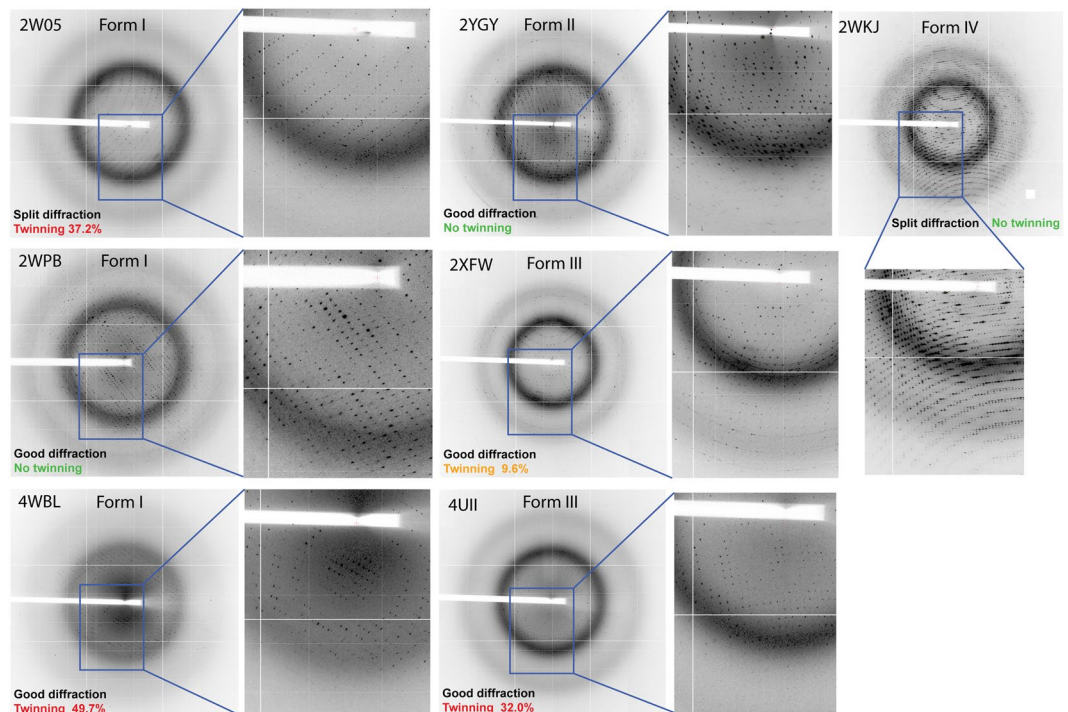


Figure 1. Diffraction pattern typologies observed for the crystal forms I, II, III and IV of NAL.

When there are two twin domains, and the twin operator is a 2-fold rotation, the twinning is called hemihedral twinning. When the twin domains are sufficiently large, the diffracted waves from these domains do not interfere (or interference is negligible, depending on the coherence radius of the beam and twin domain sizes) and the observed intensities are simply the weighted sum of the intensities from each of the individual domains⁸. If the twin fraction α approaches 0.5 the diffraction pattern acquires an additional symmetry, imposed by the twinning operator, which may lead to erroneous indexing in a higher symmetry space group. If the twin fraction equals 0.5, the crystal is perfectly twinned and the intensity measurements cannot be deconvoluted. If the twin fraction is <0.5 it is possible to deconvolute the data in order to recover the untwinned intensities¹. However, errors in the deconvoluted intensities increase proportionally and can become a very large fraction of the intensities as the twin fraction approaches 0.5. Twinning can thus hamper crystal structure determination at all stages, from indexing, to data reduction, phase determination and refinement.

Since the intensities of twin related reflections are correlated, twinning reduces the information content of the data. In the limit case of perfect merohedral twinning that reduction is equivalent to a reduction in the resolution limit by a factor of 1.26. An additional complication is that the statistical properties of the data from twinned and untwinned crystals are different and therefore overall statistics describing model quality such as the $R_{\text{factor}}/R_{\text{free}}$ must be interpreted with extra care⁹. In particular, the gap between R_{factor} and R_{free} values as well as their individual values needs to be monitored during refinement. If refinement using the twin option leads to an increase of the gap between R_{factor} and R_{free} , this indicates a serious problem with the refinement protocol and data handling.

Another type of deviation from perfect periodicity in a crystal, is crystal modulation, in which the content of asymmetric unit is not perfectly replicated by the lattice operations and which can occur with a period commensurate or incommensurate with the lattice periodicity. As result of crystal modulation, primary Bragg reflections are flanked by off-lattice satellite reflections¹⁰. The direction and magnitude of such satellite reflections is described by an additional vector \mathbf{q} , which needs to be added to the reciprocal space vector \mathbf{H} to define a 4-dimensional reciprocal space vector. Although incommensurate crystals have been reported rarely in macromolecular protein crystallography^{11,12}, the EVAL software suite can index and process such data^{10,13}, and in silico simulations of modulated structure have been performed¹⁴.

In this report we present 11 diffraction data sets, in multiple space groups, from the *E. coli* enzyme *N*-acetyl-neuraminic acid lyase (NAL), which present twin lattices and incommensurate modulation. NAL is a tetramer in solution, that crystallises in low salt conditions¹⁵ to give four different crystal forms, three in space group $P2_1$ and one in $P2_12_12_1$ ¹⁵⁻¹⁷. Interestingly, the three crystal forms in space group $P2_1$ were not related to each other, two of them were twinned and shared the same twinning operator, which made the monoclinic cells a pseudo-orthorhombic cell.

Two of the crystal forms are reported here for the first time, and some were pseudo-merohedrally twinned with the additional complication of incommensurate modulation. Although they could all be solved by molecular replacement, they could not be refined satisfactorily using standard protocols. However, with improvements in REFMAC5, one of the software packages for macromolecular structure refinement available from ccp4 suite 7.0¹⁸, with direct contribution from the presented test cases, we were able to refine models satisfactorily against

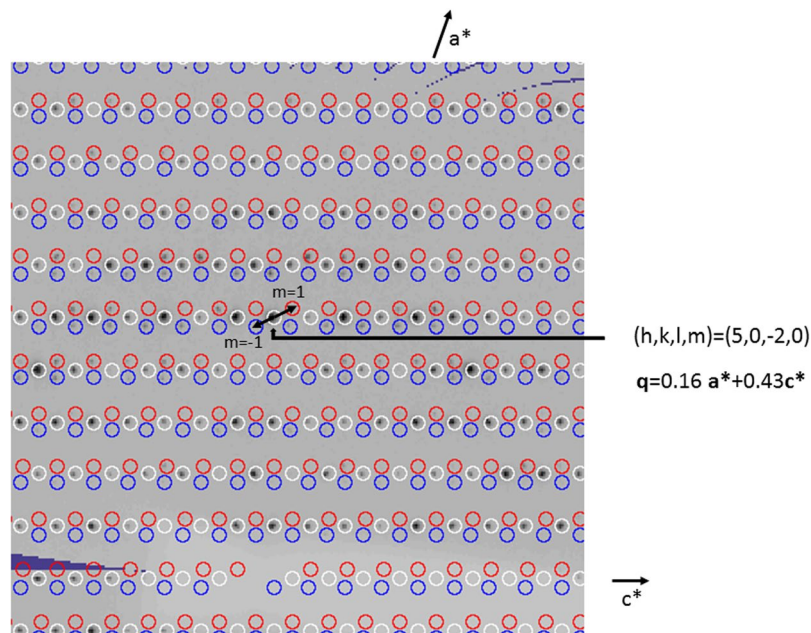


Figure 2. Precession reconstruction using of reciprocal space slice $h0l$ of 2WNN (with *Precession* in the EVAL suite). The main lattice is coloured white. Satellite reflections with $m = 1$ and -1 are coloured red and blue, respectively. Satellites of $(5, 0, -2)$ are indicated by arrows.

all 11 datasets. Due to the varied diffraction data pathologies (pseudo-merohedral twinning with α up to 0.497 as well as crystal modulation) we believe these data form a useful test set for the development of macromolecular crystallographic data processing and structure refinement software and therefore we made them available to the community through the public repository Zenodo (public links in Data Records).

Results and Discussion

Data processing. NAL crystallised in four different crystal forms from the same crystallisation conditions, and in the same drops (100 mM Tris-HCl pH 8.0–8.2, 200 mM Na acetate, 18–22% w/v PEG3350). It was not possible to discriminate between the four crystal forms of NAL solely by inspection of the crystal morphology (Fig. S1). The most commonly obtained crystal form (I) belonged to space group $P2_1$ with unit cell parameters $a = 55 \text{ \AA}$, $b = 142 \text{ \AA}$, $c = 84 \text{ \AA}$, $\beta = 109^\circ$ (decimals are omitted due to variability between datasets); followed by crystal forms II ($a = 84 \text{ \AA}$, $b = 95 \text{ \AA}$, $c = 91 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 116^\circ$, $\gamma = 90^\circ$) and III ($a = 78 \text{ \AA}$, $b = 108 \text{ \AA}$, $c = 148 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 116^\circ$, $\gamma = 90^\circ$), both in space group $P2_1$, and crystal form IV ($a = 78 \text{ \AA}$, $b = 116 \text{ \AA}$, $c = 84 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$) in space group $P2_12_12_1$ (Table S1).

The diffraction patterns occasionally showed spot splitting in all four crystal forms of NAL and it was not possible to predict the successful indexing and scaling outcome based on the observed diffraction quality alone (Fig. 1). The main and satellite reflections are clearly distinct and the main lattice could be indexed separately while satellite reflections were ignored by MOSFLM¹⁹ (i.e. PDB 2WNN, 2WPB & 2WKJ, Fig. 1).

Closer inspection of the diffraction pattern of four of the seven datasets in crystal form I with DIALLS viewer²⁰ revealed two lattices but also some extra reflections, which did not belong to either lattices (Fig. S2). MOSFLM successfully indexed the main lattice in all cases (Table 1), but we decided to further investigate whether these extra reflections in the diffraction pattern could be caused by crystal modulation, as they appeared to be occurring in a periodic manner.

All the datasets in crystal form I were therefore indexed with Dirax²¹ to determine whether incommensurate modulation was present. This was indeed the case for four of the seven datasets, three of which were deposited in the PDB: 2WNN, 2WNNQ, 2W05, whilst one, called Y137A, was not, due to unsatisfactory statistics. In those cases, reflections could be indexed and assigned either to the main lattice or to the satellite reflections with order $m = -1$ or 1 (see 2WNN as example in Fig. 2). No evidence of splitting of the main lattice was found, implying that the pseudo-merohedral twinned lattices almost exactly overlap. The data were processed with Eval¹⁰ and scaled with SADABS²² in $2/m$ point group symmetry. The resulting statics are shown in Table 2.

All the modulated structures appeared also to be partially twinned (Tables 1 and 2). We speculate that the lack of modulation in 2WNNZ and 4BWL is probably due to the larger unit cell axis a , which is large enough not to be incommensurate. With the $P2_1$ indexing choice, POINTLESS initially assigned the space group $C222_1$ but reflections belonging to one of the 2-fold axes were much stronger than the others (data not shown), which is consistent with pseudo-merohedral twinning in $P2_1$, and indeed with this choice the structures could be easily solved.

However, in all the crystal forms, space group attribution was difficult or sometimes impossible and the choice of the point group was made based on the R_{meas} values²³. Weak molecular replacement solutions could also be obtained in multiple space groups. As a general rule, whenever only a single lattice with no incommensurate

Datasets	Res	Crystal form and cell parameters	Obliquity* (ω)	Twinning fraction*	Twin on	Twin off	PDB code	Diamond Station
Wild type apo	2.20 Å	P2 ₁ crystal form I a = 54.8 b = 142.2 c = 84.2 $\alpha = 90.00 \beta = 108.97 \gamma = 90.00$	0.019	0.372	R _{factor} = 0.200 R _{free} = 0.267	R _{factor} = 0.251 R _{free} = 0.319	2W05	I02
Wild type pyruvate complex	1.65 Å	P2 ₁ crystal form I a = 54.7 b = 142.5 c = 83.6 $\alpha = 90.00 \beta = 109.16 \gamma = 90.00$	0.070	0.334	R _{factor} = 0.201 R _{free} = 0.245	R _{factor} = 0.256 R _{free} = 0.293	2WNN	I03
E192N apo	1.80 Å	P2 ₁ crystal form I a = 54.6 b = 142.8 c = 84.3 $\alpha = 90.00 \beta = 108.8 \gamma = 90.00$	0.130	0.463	R _{factor} = 0.195 R _{free} = 0.244	R _{factor} = 0.272 R _{free} = 0.320	2WNNQ	I04
E192N pyruvate complex	1.80 Å	P2 ₁ crystal form I a = 56.9 b = 143.0 c = 83.9 $\alpha = 90.00 \beta = 109.8 \gamma = 90.00$	0.000	—	R _{factor} = 0.178 R _{free} = 0.209	R _{factor} = 0.187 R _{free} = 0.223	2WNNZ	I02
E192N + pyruvate + THB**	2.05 Å	P2 ₁ crystal form I a = 57.0 b = 143.7 c = 84.3 $\alpha = 90.00 \beta = 109.9 \gamma = 90.00$	0.130	—	R _{factor} = 0.192 R _{free} = 0.238	R _{factor} = 0.191 R _{free} = 0.242	2WNNP	I03
Y137A pyruvate complex	1.80 Å	P2 ₁ crystal form I a = 54.7 b = 142.2 c = 83.6 $\alpha = 90.0 \beta = 109.0 \gamma = 90.0$	0.119	0.149	R _{factor} = 0.287 R _{free} = 0.331	R _{factor} = 0.296 R _{free} = 0.357	n./a.***	I04
Y137A pyruvate, ManNAc and Neu5Ac complex	2.00 Å	P2 ₁ crystal form I a = 56.1 b = 143.5 c = 83.6 $\alpha = 90.0 \beta = 109.6 \gamma = 90.0$	0.094	0.497	R _{factor} = 0.183 R _{free} = 0.236	R _{factor} = 0.265 R _{free} = 0.321	4BWL	I02
Wild type apo	1.90 Å	P2 ₁ crystal form II a = 84.3 b = 95.9 c = 91.4 $\alpha = 90.00 \beta = 115.33 \gamma = 90.00$	2.10	—	R _{factor} = 0.198 R _{free} = 0.226	R _{factor} = 0.197 R _{free} = 0.225	2YGY	I02
E192N/Y137F pyruvate complex	1.80 Å	P2 ₁ crystal form III a = 78.0 b = 116.7 c = 83.7 $\alpha = 90.0 \beta = 118.06 \gamma = 90.00$	0.290	0.328	R _{factor} = 0.156 R _{free} = 0.183	R _{factor} = 0.206 R _{free} = 0.228	2YDZ	I02
E192N + pyruvate complex	1.85 Å	P2 ₁ crystal form III a = 78.1 b = 116.5 c = 83.7 $\alpha = 90.00 \beta = 116.5 \gamma = 90.00$	0.150	0.096	R _{factor} = 0.165 R _{free} = 0.186	R _{factor} = 0.174 R _{free} = 0.193	2XFW	I02
E192N + pyruvate	1.45 Å	P2 _{1,2,2,1} crystal form IV a = 78.3 b = 108 c = 148.3 $\alpha = \beta = \gamma = 90.00$	0.000	—	R _{factor} = 0.191 R _{free} = 0.201	R _{factor} = 0.188 R _{free} = 0.205	2WKJ	I04

Table 1. Improvement of refinement statistics upon applying the twin option in REFMAC. *As defined in Nespolo *et al.*,³⁵. **THB refers to the competitive inhibitor (2*R*,3*R*)-2,3,4-trihydroxy-*N,N*-dipropylbutanamide, as reported in Campeotto *et al.*¹⁶. ***Refinement statistics were not of enough quality for model and data deposition, although data analysis was beneficial for the discussion presented here and the raw images were deposited in the public Zenodo database.

modulation was present, indexing, data reduction and molecular replacement were possible, but the (non-twin) refinement stalled at R_{factor} and R_{free} values of 30–35% for all datasets (resolution range 1.45–2.3 Å, $\langle I/\sigma(I) \rangle$ cut off = 2.0) where we would expect R_{factor} values near or below 20% for well-behaved refinements.

Twinning analysis. H and L twinning tests, as implemented in TRUNCATE¹⁸, were used as diagnostic tools for twinning. In our experience the L-test prediction was more consistent with estimates of twinning fraction performed internally in REFMAC.

This is probably due to the fact that H and L tests are affected by experimental errors and lack discrimination power if one of the NCS-operation axes is parallel to twin-operation axis. However, the H-test requires for data to be merged in correct point group, and even then, in case of the NCS, it may seem to indicate partial twinning for data from single crystal. L-test is free from these two issues. For these reasons only the L-test is reported for the presented datasets (Fig. 3).

Micro-seeding techniques were employed in an attempt to avoid twinning by growing larger single crystals²⁴. However, twinning persisted, suggesting that it was likely to be a nucleation phenomenon, which was perpetuated when twinned seed crystals were used as nuclei. Diffraction data were collected at 100 K following flash cooling of crystals in cryo-protectants, which could have been a source of lattice disorder. Data collection at room temperature from multiple crystals, however, also showed both split diffraction and significant twinning (data not shown), indicating that the disorder pre-existed in the crystals. Ligand soaking experiments were similarly excluded as a cause of the twinning.

Refinement with the program REFMAC (versions 5.6 and 5.7) identified the twinning operator ($-h, -k, h + l$) for all the cases, in which twinning was detected. Twin refinement resulted in improved models with R_{factor} and R_{free} values ~18–20% (Table S2, data collection and final refinement statistics are summarised in Table 1). This improvement of the R_{factor} quality indices was accompanied by local improvements of the electron density maps, which became better defined and showed increased connectivity (Fig. 4). The best refined model for each crystal form was validated using ZANUDA¹⁸, which confirmed the space group assignment in all cases by transforming the individual space group into the lower symmetry space groups, followed by refinement of the corresponding models using REFMAC and selection of the model with highest symmetry from the ones with best refinement statistics.

Dataset ID	2WNN	2WNQ	2WNZ	2WO5	2WPB	4BWL	Y137A
Bravais	P	P	P	P	P	P	P
Pointgroup	2/m	2/m	2/m	2/m	2/m	2/m	2/m
Cell axes a,b,c (Å)	54.8, 142.8, 83.7	54.8, 142.1, 84.5	57.4, 143.0, 83.9	54.6, 141.9, 84.0	56.8, 143.5, 84.2	56.1, 143.5, 83.6	54.8, 142.4, 83.7
alpha (Å°)	90.00, 109.0, 90.00	90.00, 108.9, 90.00	90.00, 109.9, 90.00	90.00, 108.9, 90.00	90.00, 109.8, 90.00	90.00, 109.6, 90.00	90.00, 109.0, 90.00
qvx1*	0.16	0.14	—	0.18	—	—	0.22
qvy1*	0.00	0.00	—	0.00	—	—	0.00
qvz1*	0.43	0.42	—	0.42	—	—	0.42
Resolution (Å)	41.9-1.65	49.9-1.80	38.7-1.80	48.7-2.20	47.8-2.05	49.6-1.80	49.9-1.80
Rmerge	0.070 (0.482)	0.081 (0.643)	0.059 (0.445)	0.121 (0.619)	0.084 (0.367)	0.088 (1.043)	0.099 (0.751)
Rmeas	0.083 (0.585)	0.096 (0.769)	0.069 (0.543)	0.144 (0.737)	0.099 (0.429)	0.103 (1.217)	0.119 (0.899)
Rpim	0.044 (0.327)	0.052 (0.417)	0.036 (0.307)	0.077 (0.395)	0.052 (0.221)	0.053 (0.623)	0.064 (0.489)
<I/sigI>	10.6 (1.8)	8.5 (1.5)	12.3 (1.8)	6.8 (2.1)	9.0 (2.9)	7.4 (1.0)	7.8 (1.4)
Completeness (%)	92.0 (61.2)	97.9 (97.0)	99.4 (94.6)	99.9 (99.7)	98.5 (97.9)	98.7 (98.0)	99.8 (98.5)
Redundancy	3.5 (3.2)	3.4 (3.3)	3.6 (3.1)	3.4 (3.4)	3.6 (3.7)	3.7 (3.8)	3.5 (3.4)
Reflections	1358205 (75799)	1111355 (107314)	413549 (32220)	625336 (60979)	282230 (28483)	419064 (42200)	1132807 (108078)
Unique	401726 (26603)	331525 (32667)	116035 (11021)	183588 (18294)	78040 (7722)	113330 (11261)	335150 (33038)
Main reflections only							
Rmerge	0.051 (0.340)	0.053 (0.412)	—	0.077 (0.362)	—	—	0.061 (0.503)
Rmeas	0.061 (0.413)	0.064 (0.492)	—	0.092 (0.432)	—	—	0.073 (0.603)
Rpim	0.033 (0.232)	0.035 (0.267)	—	0.050 (0.232)	—	—	0.039 (0.329)
<I/sigI>	12.6 (2.8)	11.7 (2.5)	—	8.1 (3.0)	—	—	8.5 (1.9)
Completeness (%)	91.9 (61.3)	97.2 (96.6)	—	99.8 (99.2)	—	—	99.7 (98.3)
Redundancy	3.5 (3.2)	3.3 (3.4)	—	3.4 (3.4)	—	—	3.4 (3.4)
No. Reflections	447223 (25340)	354109 (35713)	—	203706 (20152)	—	—	373048 (35922)

Table 2. Analysis of the presence of crystal modulation in the structures belonging to crystal form I with EVAL15 package for modulated structures. Four of the seven datasets were modulated. Statistics may differ slightly from Table 1 due to the processing being performed using a different package. ***q** vector components.

Crystal packing analysis. Zanuda was used to expand the final refined models into space group P1 in order to compare packing in the different crystal forms. Inspection of the packing using the molecular graphics program COOT²⁵ highlighted how not only the inter-monomer contacts within the NAL tetramers were different, but also the inter-tetramer contacts in the crystal lattice (Fig. 5). We speculate that the likelihood of NAL of crystallising in any one of the four forms is determined by small differences in the interfaces between tetramer during nucleation and the early stages of crystal growth. This process is kinetically and thermodynamically difficult to control and attempts to select for a specific crystal form were hindered by the fact that all four forms were obtained in the same crystallisation drops and therefore from identical crystallisation conditions. Surface accessible areas and free energies of interaction were calculated using PISA (Table 2)²⁶. These did not show any significant differences in the strength of intra-tetramer interactions between the different crystal forms, consistent with our observation that all four crystal forms appeared in the same crystallisation drops.

Future developments. The presented datasets were the result of an extensive screening at the data collection stage and of an extensive processing at the data reduction and data refinement stages with very low success rate (Fig. S3). The development of twin refinement in REFMAC, which at the time was only implemented in the experimental version of the program, allowed the determination of several apo- and ligand bound structures of NAL and the proposal of the first detailed mechanism of the enzyme reaction^{15,16}. Although twin refinement is currently included in REFMAC, the presented datasets are still a challenging test for current indexing and scaling programs, including iMosflm, LABELIT²⁷ and XDS²⁸, and they therefore offer an excellent opportunity for the development of these softwares.

Several improvements in MX software are still very desirable in the part of dealing with pathological data. This includes robust diagnostics and warning messages, automated space group assignment in at least obvious cases of twinning, and, importantly, robust integration of partially overlapping reflections and communication of all the necessary data and metadata to a refinement program. Crystal modulation was also detected only after structure deposition and although this had no effect on data processing in the presented cases, its diagnosis should be implemented to avoid reflection overlaps, which in severe cases can seriously hamper indexing, data reduction and ultimately phasing and satisfactory refinement.

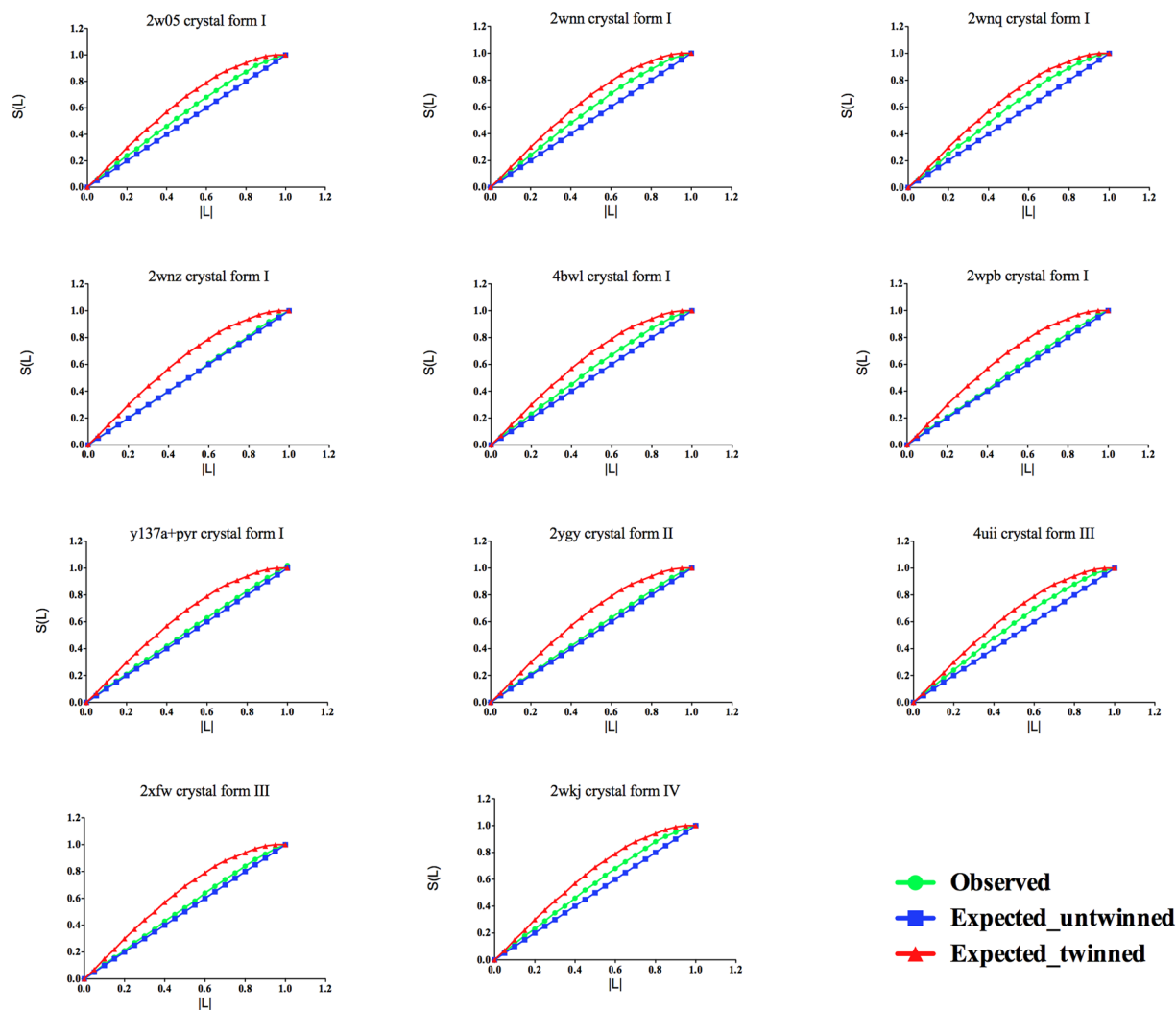


Figure 3. L-test analysis of the 11 NAL datasets reported here. For each crystal the crystal form is indicated. Cumulative intensity difference plot of the intensity difference of local pairs of intensities that are not twin-related $|L| \{L = [I(h\ 1) - I(h\ 2)]/[I(h\ 1) + I(h\ 2)]\}$ against the cumulative probability distribution $N(L)$ of the parameter L .

Methods

Data collection and structure solution. We have previously reported several structures of wild type NAL and engineered variants and NAL crystals were obtained as previously described^{15,16}. NAL crystals are plate-shaped and tend to grow in clusters and therefore micro-seeding experiments were required to obtain single large crystals. Crystal cryo-protection was achieved by serial transfer of the crystals through mother liquor containing 20% and then 25% v/v PEG 400, with 2 minutes soak time at each step. Eleven datasets were collected from single crystals at Diamond Light Source (beamlines I02, I03 and I04), at 100 K with a 1 s exposure and an oscillation of 0.5° per image and using a Q315 ADSC CCD detector. Data were processed using iMOSFLM and scaled and merged using SCALA²⁹.

In the case of the datasets of crystal form I, diffraction patterns were inspected with DIALS²⁰ for the presence of satellite reflection, indexed with Dirax²² and processed with EVAL15¹⁰. Scaling was performed with SADABS in 2/m point group symmetry. The results are shown in Table 2. For structure refinement only the main lattice reflections from MOSFLM were used, ignoring the weak satellite reflections.

In each dataset five percent of the reflections were excluded from the refinement and constituted the R_{free} set. A new R_{free} set was generated randomly for each new crystal form and then transferred to all datasets belonging to the same crystal form.

The first crystal structure obtained for each crystal form was solved by molecular replacement using PHASER³⁰ and 1NAL as a starting model³¹, while refinement against other datasets of the same crystal form started with 20 cycles of rigid body refinement (resolution range 10.0–6.0 Å) followed by 10 cycles of preliminary restrained refinement (whole resolution range) in REFMAC5.

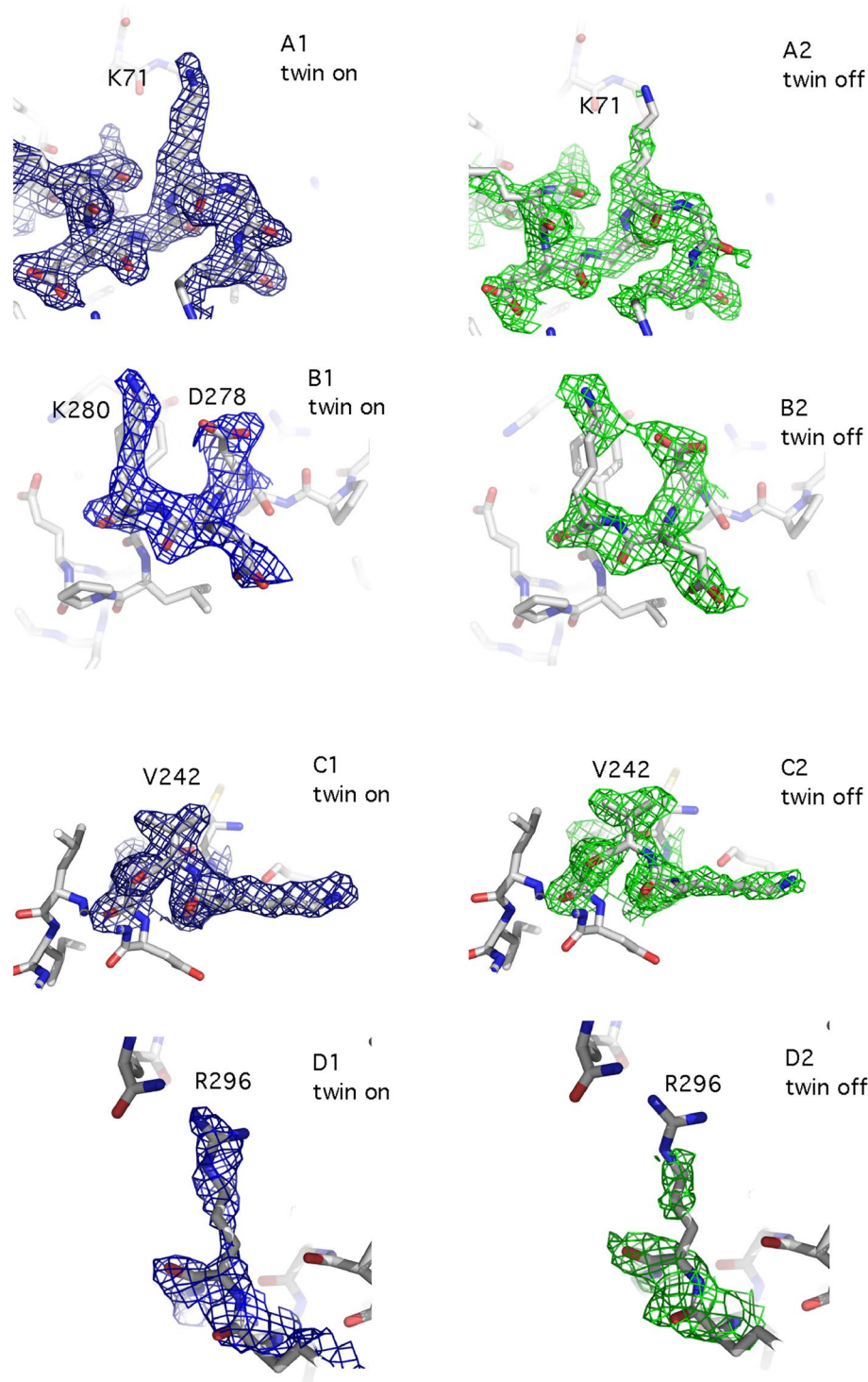


Figure 4. Comparison between equivalent portions of the electron density map before (A1, B1, C1, D1) and after (A2, B2, C2, D2) applying the twin option in REFMAC. The electron density maps refer to different regions of dataset 4BWL, which belongs to crystal form I and showed a twin fraction of almost 50%.

Refinement and Crystal packing analysis. Refinement was performed using REFMAC 5.6 or 5.7 (i.e. the latest version at the time of deposition or final refinement for each structure), with and without twin refinement, both for electron density calculations and evaluation of statistics. Refinement was performed with the same settings for all reported structures, *i.e.* 20 cycles per run (using the whole resolution range of the data), a weight matrix of 0.1³², with riding hydrogen atoms.

For all structures involved, regardless of whether the unit cell parameters allowed for twinning by merohedry or not, the refinement protocol was identical and included twin-refinement in the final refinement rounds³². If no

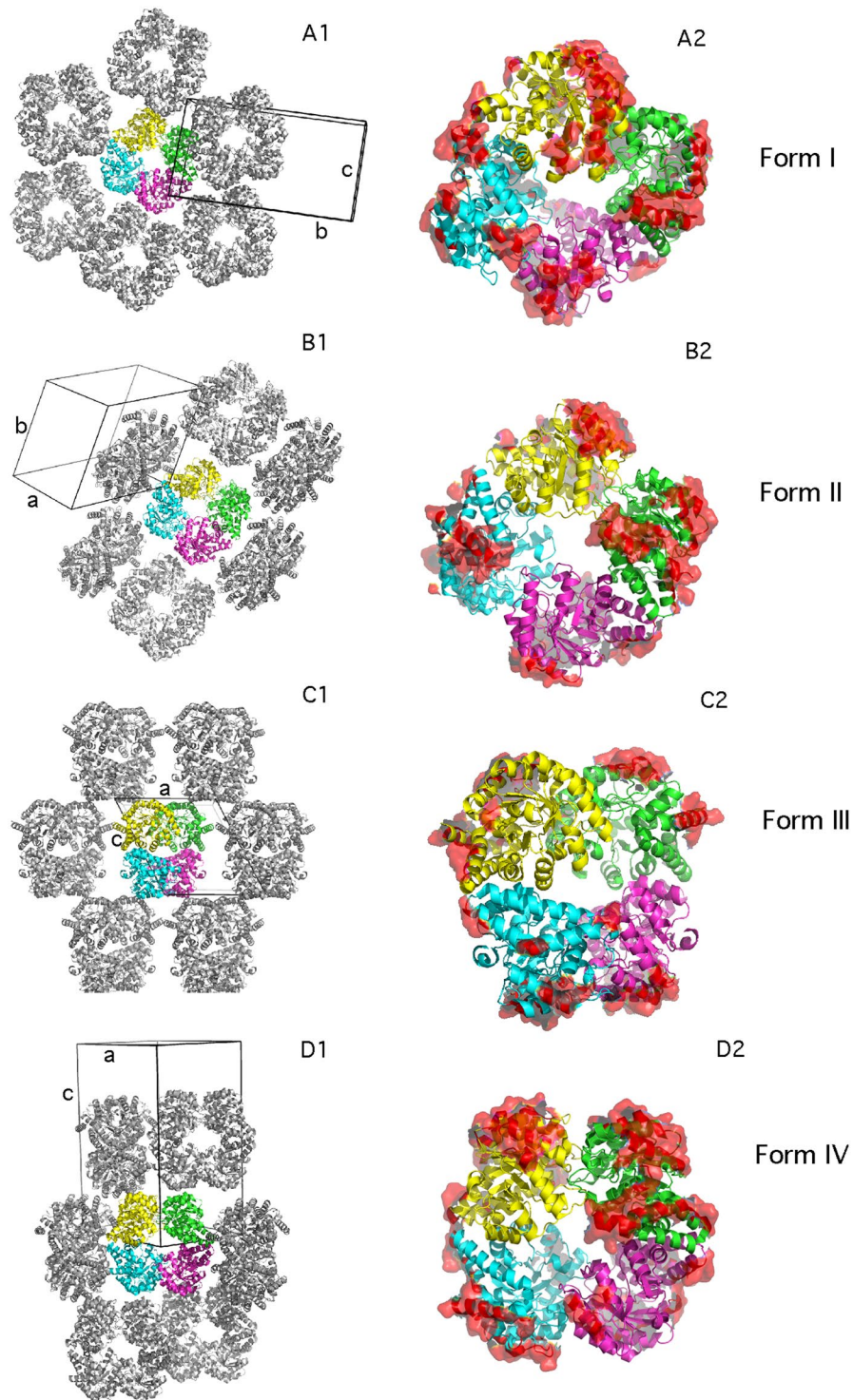


Figure 5. Crystal packing and crystal contacts in the four crystal forms of NAL. For each crystal form the crystal packing was inspected manually and the least overlapping orientation is presented as 2D layer (A1, B1, C1 and D1). The crystal contacts between tetramers in the given orientation are represented in more detail (A2, B2, C2, D2) as a red surface with the orientations of the tetramers kept the same as in the corresponding 2D layer. As examples of crystal forms I, II, III and IV, the structures of PDB code 2WNN, 2YGY, 2XFW and 2WKJ are represented respectively. Images were produced in PYMOL version 1.6.0.0.

twinning operations are present, the twin refinement option means that REFMAC uses approximation to the likelihood target rather than its exact version. Such usage therefore only makes sense for comparison of refinement results for twinned and untwinned crystals.

R_{factor} and R_{free} values were compared before and after twin refinement.

The values of the obliquity angle, which are a measure of pseudo-symmetry, were monitored and manual inspection of the diffraction pattern were performed with ADXV³³.

The concept of obliquity is a measure of the overlap of lattices on the individuals forming a twin and Friedel provided a formal mathematical description since the early day of crystallography³⁴. Briefly, the closest is the obliquity angle to zero, the more likely is the presence of merohedral twinning³⁵, as the two twin lattices tend to overlap. Values of obliquity close to zero are, however, only a possible indicator that twinning may be present but not a fixed rule, as some of the presented datasets highlight. For instance, in the case of crystal form I and III, the obliquity angle is small enough to allow twinning in some cases, whilst in crystal form II is too large for twinning to occur (Table 1).

Manual model building was performed in COOT. Zanuda was used to expand the unit cell of each crystal form into P1 for each crystal form and these were refined against the data processed in P1 in order to confirm the correctness of the space group assignment in each case.

In order to assess how the four crystal forms of NAL were related to each other, Csymmatch from CCP4 was used to bring all the P1-expanded structures to the same origin and NCONTACT¹⁸ was used to calculate inter-tetramer contacts. The input file from NCONTACT was used in PYMOL to visualise the contact surface between monomers. Surface accessibility areas and crystal contact energy were calculated using PISA²⁶ (Table S3).

Data Records

The datasets (raw diffraction images) discussed in this manuscript have been deposited in the publicly available database zenodo at, <https://doi.org/10.5281/zenodo.54568> and [10.5281/zenodo.1240503](https://doi.org/10.5281/zenodo.1240503). Structural models and processed structure factor data deposited in the PDB are available under the accession codes given in Table 1, with the exception of dataset Y137A, as the R factor indices were not satisfactory for PDB deposition.

References

- Yeates, T. O. Detecting and overcoming crystal twinning. *Methods in enzymology* **276**, 344–358 (1997).
- Bernal, J. D., Fankuchen, I. & Perutz, M. F. A X-ray study of chymotrypsin and hemoglobin. *Nature*, 523–524 (1938).
- Blow, D. M., Rossmann, M. G. & Jeffery, B. A. The Arrangement of Alpha-Chymotrypsin Molecules in the Monoclinic Crystal Form. *Journal of molecular biology* **8**, 65–78 (1964).
- Dauter, Z. Twinned crystals and anomalous phasing. *Acta crystallographica. Section D, Biological crystallography* **59**, 2004–2016 (2003).
- Helliwell, J. R. Macromolecular crystal twinning, lattice disorders and multiple crystals. *Crystallography reviews* **14**, 189–250 (2008).
- Parsons, S. Introduction to twinning. *Acta crystallographica. Section D, Biological crystallography* **59**, 1995–2003 (2003).
- Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D. Surprises and pitfalls arising from (pseudo) symmetry. *Acta crystallographica. Section D, Biological crystallography* **64**, 99–107, <https://doi.org/10.1107/S090744490705531X> (2008).
- Grainger, C. T. Pseudo-merohedral twinning. The treatment of overlapped data. *Acta Crystallographica Section A* **A25**, 427–434 (1969).
- Murshudov, G. N. Some statistical properties of the crystallographic reliability index Rfactor: Effect of twinning. *Applied and Computational Mathematics*, 250–261 (2011).
- Porta, J., Lovelace, J. J., Schreurs, A. M., Kroon-Batenburg, L. M. & Borgstahl, G. E. Processing incommensurately modulated protein diffraction data with Eval15. *Acta crystallographica. Section D, Biological crystallography* **67**, 628–638, <https://doi.org/10.1107/S0907444911017884> (2011).
- Lovelace, J. J. *et al.* Protein crystals can be incommensurately modulated. *Journal of applied crystallography* **41**, 600–605 (2008).
- Rupp, B., Marshak, D. R. & Parkin, S. Crystallization and preliminary X-ray analysis of two new crystal forms of calmodulin. *Acta crystallographica. Section D, Biological crystallography* **52**, 411–413, <https://doi.org/10.1107/S0907444995011826> (1996).
- Schreurs, M. M. A., Xian, X. & Kroon-Batenburg, J. M. L. EVAL15: a diffraction data integration method based on ab initio predicted profile. *J. Appl. Cryst.* **43**, 70–82 (2010).
- Lovelace, J. J., Simone, P. D., Petricek, V. & Borgstahl, G. E. Simulation of modulated protein crystal structure and diffraction data in a supercell and in superspace. *Acta crystallographica. Section D, Biological crystallography* **69**, 1062–1072, <https://doi.org/10.1107/S0907444913004630> (2013).
- Campeotto, I. *et al.* Structure of an Escherichia coli N-acetyl-D-neuraminic acid lyase mutant, E192N, in complex with pyruvate at 1.45 angstrom resolution. *Acta crystallographica. Section F, Structural biology and crystallization communications* **65**, 1088–1090, <https://doi.org/10.1107/S1744309109037403> (2009).
- Campeotto, I. *et al.* Structural insights into substrate specificity in variants of N-acetylneuraminic Acid lyase produced by directed evolution. *Journal of molecular biology* **404**, 56–69, <https://doi.org/10.1016/j.jmb.2010.08.008> (2010).
- Daniels, A. D. *et al.* Reaction mechanism of N-acetylneuraminic acid lyase revealed by a combination of crystallography, QM/MM simulation, and mutagenesis. *ACS Chem Biol* **9**, 1025–1032, <https://doi.org/10.1021/cb500067z> (2014).
- Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235–242, <https://doi.org/10.1107/S0907444910045749> (2011).
- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* **67**, 271–281, <https://doi.org/10.1107/S0907444910048675> (2011).
- Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85–97, <https://doi.org/10.1107/S2059798317017235> (2018).
- Duisenberg, A. J. M. Indexing in Single-Crystal Diffractometry with an Obstinate List of Reflections. *J. Appl. Cryst.* **25**, 92–96 (1992).
- Krause, L., Herbst-Irmer, R., Sheldrick, G. M. & Stalke, D. Comparison of silver and molybdenum microfocus X-ray sources for single-crystal structure determination. *Journal of applied crystallography* **48**, 3–10, <https://doi.org/10.1107/S1600576714022985> (2015).
- Roversi, P., Blanc, E., Johnson, S. & Lea, S. M. Tetartohedral twinning could happen to you too. *Acta crystallographica. Section D, Biological crystallography* **68**, 418–424, <https://doi.org/10.1107/S0907444912006737> (2012).
- Bergfors, T. Seeds to crystals. *J Struct Biol* **142**, 66–76, [https://doi.org/10.1016/S1047-8477\(03\)00039-X](https://doi.org/10.1016/S1047-8477(03)00039-X) (2003).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486–501, <https://doi.org/10.1107/S0907444910007493> (2010).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774–797, <https://doi.org/10.1016/j.jmb.2007.05.022> (2007).

27. Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. Robust indexing for automatic data collection. *Journal of applied crystallography* **37**, 399–409, <https://doi.org/10.1107/S0021889804005874> (2004).
28. Kabsch, W. Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125–132, <https://doi.org/10.1107/S0907444909047337> (2010).
29. Evans, P. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography* **62**, 72–82, <https://doi.org/10.1107/S0907444905036693> (2006).
30. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658–674, <https://doi.org/10.1107/S0021889807021206> (2007).
31. Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G. & Colman, P. M. The three-dimensional structure of N-acetylneuraminidase from *Escherichia coli*. *Structure* **2**, 361–369 (1994).
32. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355–367, <https://doi.org/10.1107/S0907444911001314> (2011).
33. Arvai, A. ADXV - a program to display X-ray diffraction images, <http://www.scripps.edu/~arvai/adxv.html>. (2012).
34. Donnay, J. D. H. a. D., G. International Tables for X-ray Crystallography. *Birmingham: Kynoch Press III* (1959).
35. Nespolo, M. & Ferraris, G. Overlooked problems in manifold twins: twin misfit in zero-obliquity TLQS twinning and twin index calculation. *Acta crystallographica. Section A, Foundations of crystallography* **63**, 278–286, <https://doi.org/10.1107/S0108767307012135> (2007).

Acknowledgements

ARP was supported by a RCUK Academic Research Fellowship, and currently by the German Federal Excellence Cluster “Hamburg Centre for Ultrafast Imaging”. AL is supported by CCP4, GNM is supported by MRC grant (no MC_US_A025_0104) and IC was supported by the Wellcome Trust PhD programme, “The Molecular Basis of Biological Mechanisms”. We thank the MX beamline staff at Diamond Light Source for assistance with data collection and Diamond Light Source for access to beamlines i02, i03 and i04 (proposal number mx 302) for diffraction data collection.

Author Contributions

I.C. purified the protein, grew crystals, collected and processed the data, solved and deposited the structures and wrote the paper. A.L. provided advice for the refinement of the structures. L.K.-B. and A.S. analysed the datasets affected by crystal modulation, whilst E.L. provided advice for DIALS. S.E.V.P. supervised the research with A.R.P. A.R.P. also supervised data analysis, structure validation and deposition, and wrote the paper with I.C., G.N.M. developed some novel features of REFMAC for dealing with twinned data using some of the presented structures. All the authors contributed to the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-32962-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018