wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 04:25 PM GMT

PDB ID	:	4EQV
Title	:	Structure of Saccharomyces cerevisiae invertase
Authors	:	Sainz-Polo, M.A.; Sanz-Aparicio, J.
Deposited on	:	2012-04-19
Resolution	:	3.40 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

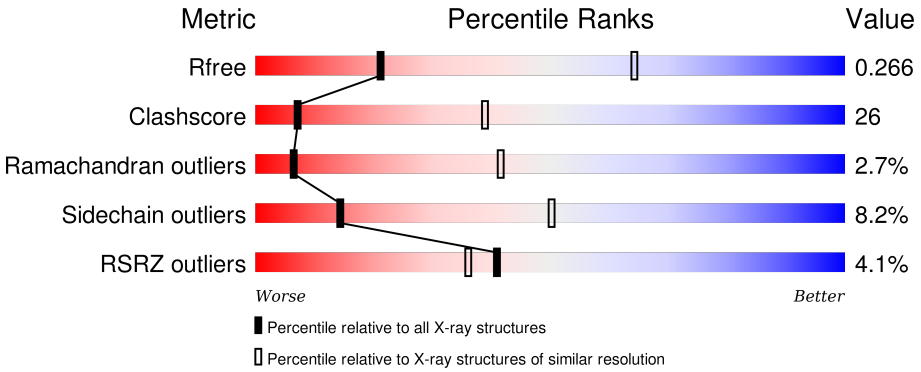
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.40 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



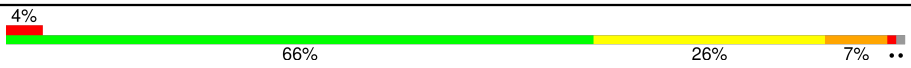
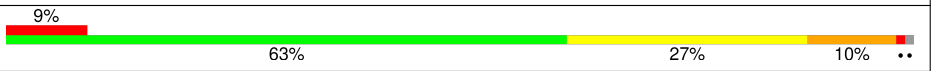
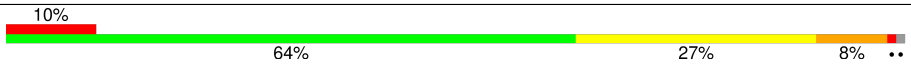
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1476 (3.50-3.30)
Clashscore	102246	1611 (3.50-3.30)
Ramachandran outliers	100387	1571 (3.50-3.30)
Sidechain outliers	100360	1571 (3.50-3.30)
RSRZ outliers	91569	1485 (3.50-3.30)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	512	71% 22% 6%
1	B	512	70% 24% 5%
1	C	512	2% 68% 25% 6%
1	D	512	3% 70% 22% 7%
1	E	512	4% 64% 27% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	512	
1	G	512	
1	H	512	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 33016 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Invertase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			
1	B	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			
1	C	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			
1	D	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			
1	E	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			
1	F	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			
1	G	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			
1	H	509	Total	C	N	O	S	0	0	0
			4124	2646	658	809	11			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	6	Total	O	0	0
			6	6		
2	B	2	Total	O	0	0
			2	2		
2	C	4	Total	O	0	0
			4	4		
2	D	2	Total	O	0	0
			2	2		
2	E	1	Total	O	0	0
			1	1		
2	F	4	Total	O	0	0
			4	4		

Continued on next page...

Continued from previous page...

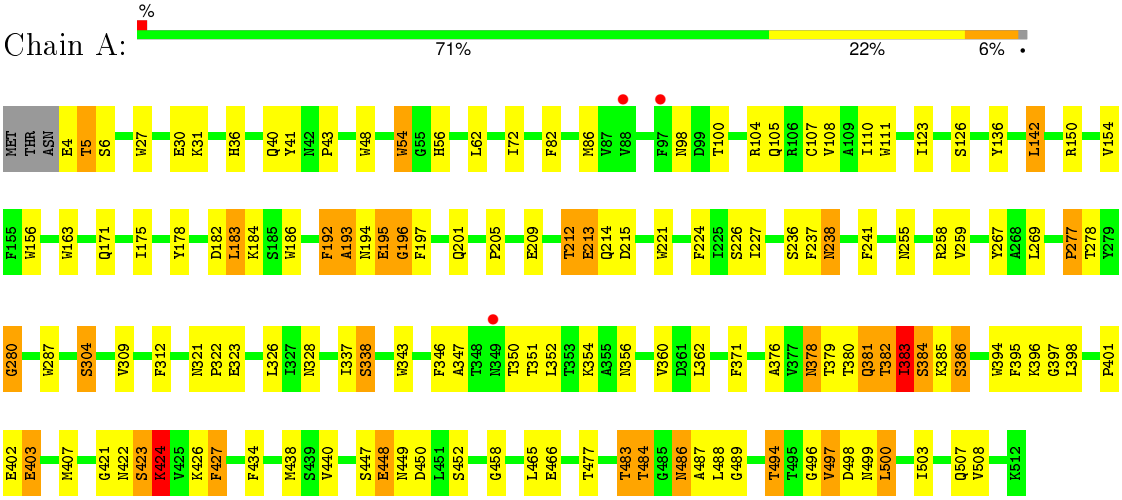
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	2	Total O 2 2	0	0
2	H	3	Total O 3 3	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

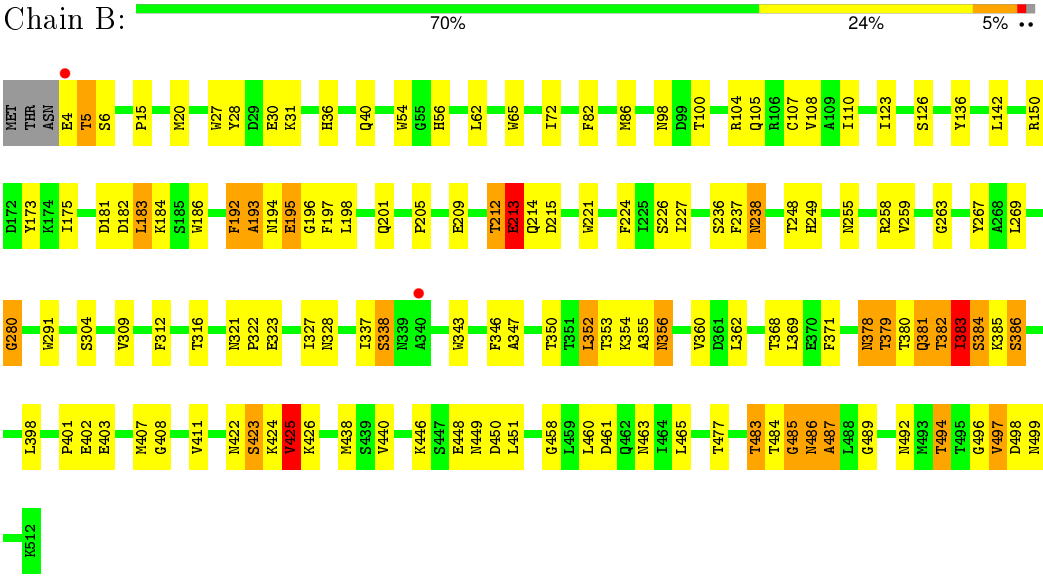
- Molecule 1: Invertase 2

Chain A:



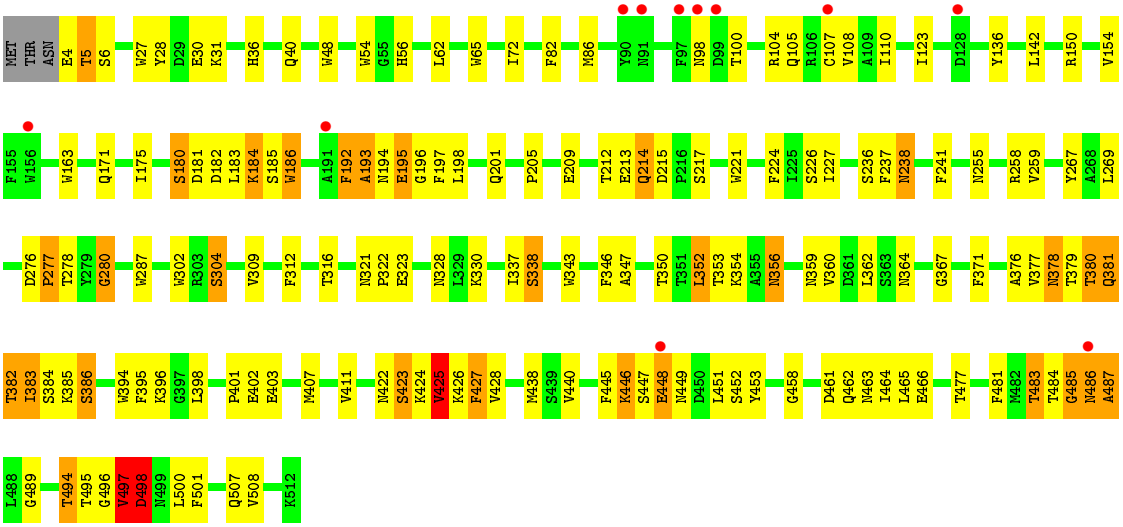
- Molecule 1: Invertase 2

Chain B:



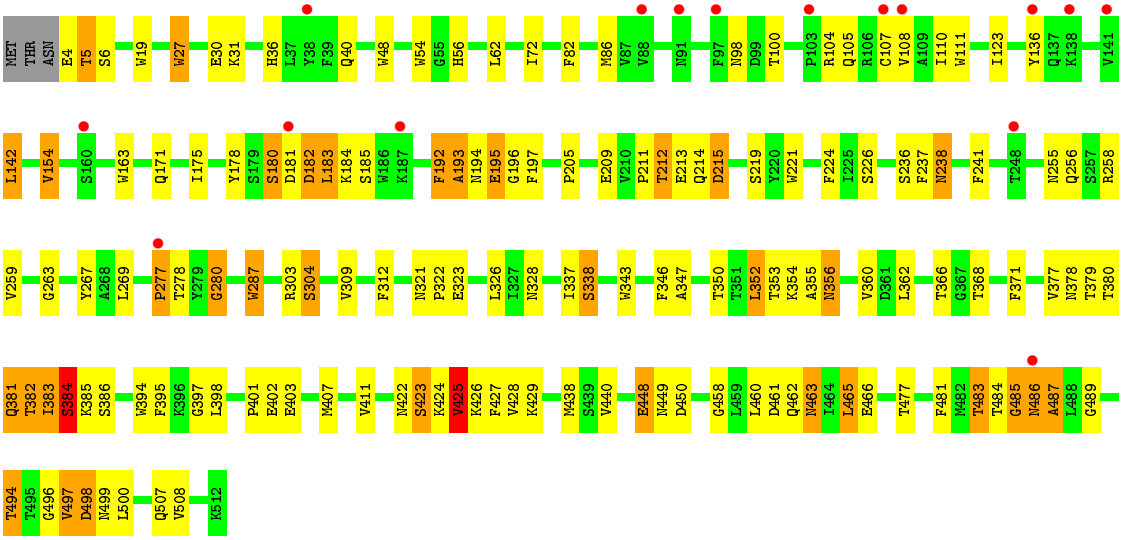
- Molecule 1: Invertase 2

Chain C:



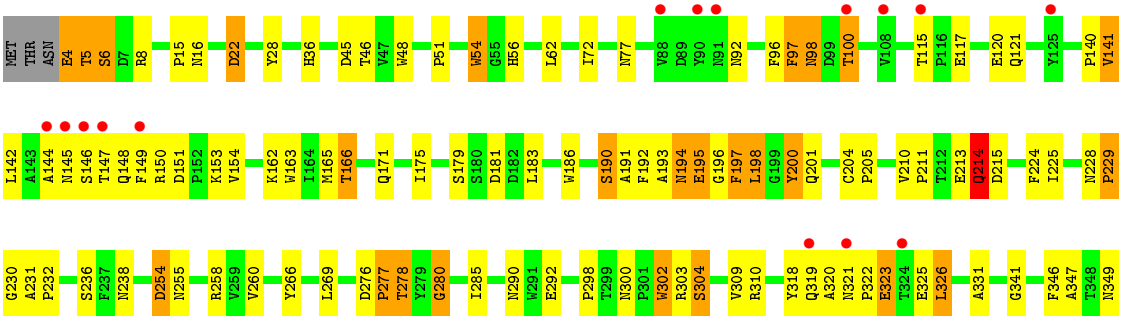
• Molecule 1: Invertase 2

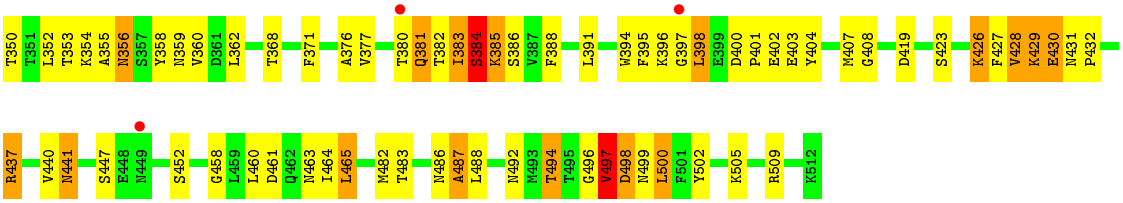
Chain D:



• Molecule 1: Invertase 2

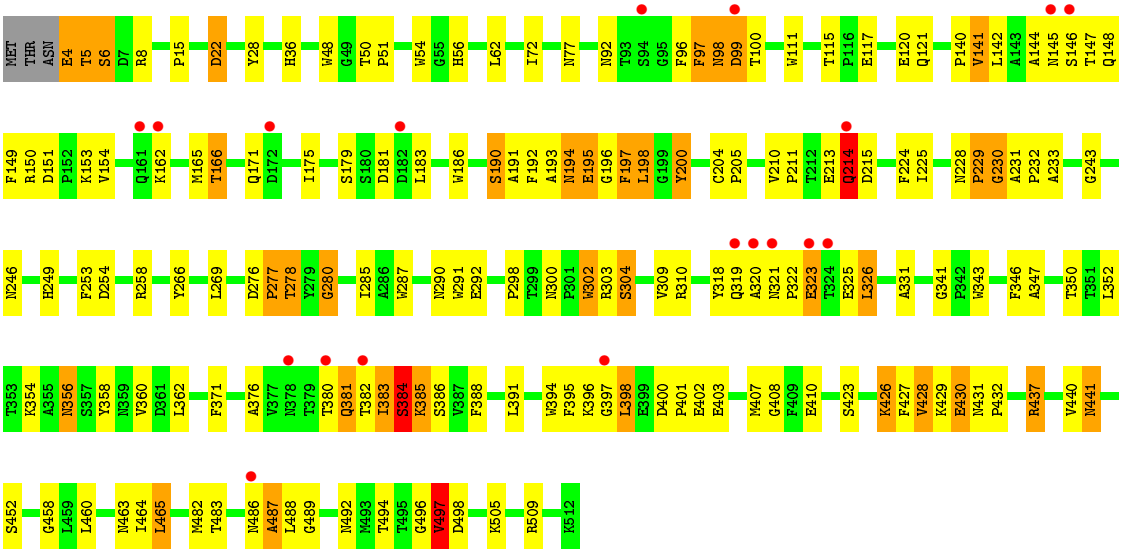
Chain E:





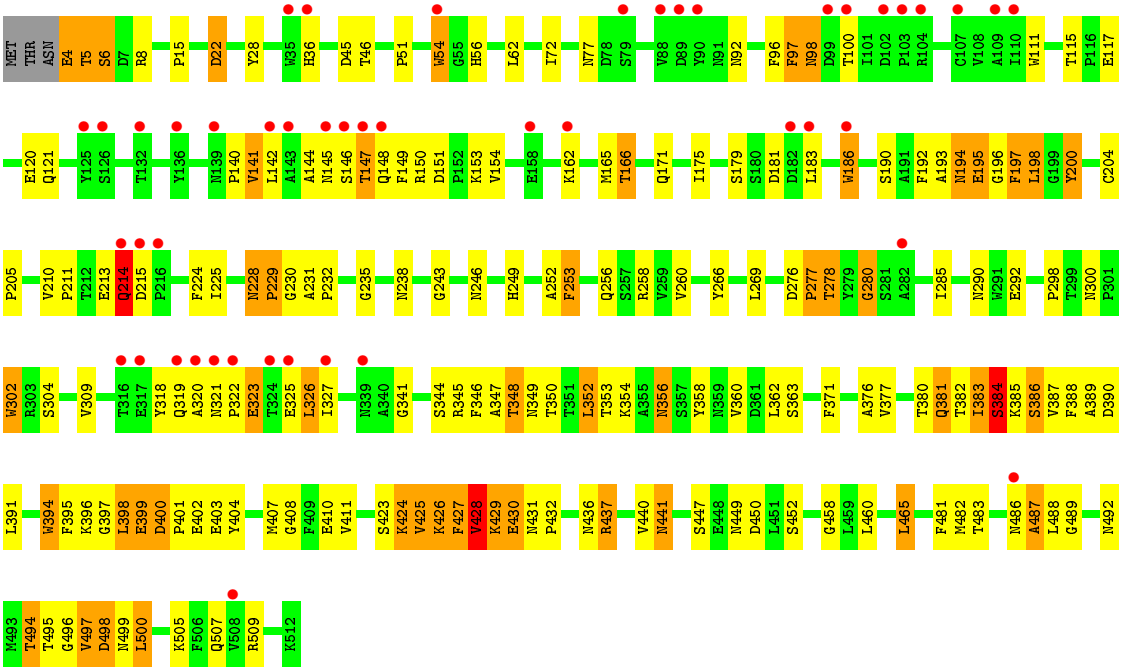
• Molecule 1: Invertase 2

Chain F:



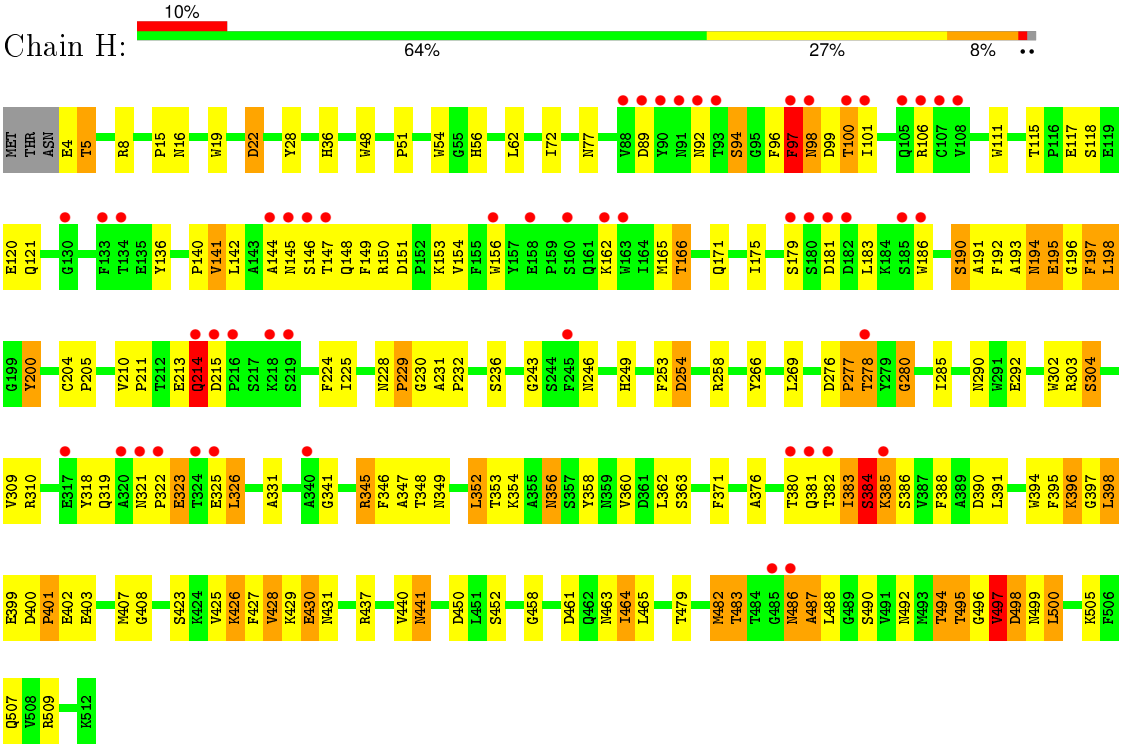
• Molecule 1: Invertase 2

Chain G:



• Molecule 1: Invertase 2

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	268.66Å 268.66Å 224.45Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	56.31 – 3.40 56.31 – 3.40	Depositor EDS
% Data completeness (in resolution range)	98.8 (56.31-3.40) 99.3 (56.31-3.40)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	2.92 (at 3.40Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.221 , 0.239 0.247 , 0.266	Depositor DCC
R_{free} test set	6376 reflections (5.28%)	DCC
Wilson B-factor (Å^2)	57.1	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 49.0	EDS
Estimated twinning fraction	0.730 for H, K, L 0.270 for -H, -K, L 0.000 for -h,-k,l	Xtriage
Reported twinning fraction	0.730 for H, K, L 0.270 for -H, -K, L	Depositor
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	1 of 127516 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	33016	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 21.37 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 7.1077e-03.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.60	6/4252 (0.1%)	0.61	0/5798
1	B	0.58	2/4252 (0.0%)	0.62	2/5798 (0.0%)
1	C	0.56	4/4252 (0.1%)	0.59	1/5798 (0.0%)
1	D	0.57	7/4252 (0.2%)	0.60	2/5798 (0.0%)
1	E	0.55	5/4252 (0.1%)	0.59	0/5798
1	F	0.55	6/4252 (0.1%)	0.59	0/5798
1	G	0.58	7/4252 (0.2%)	0.59	1/5798 (0.0%)
1	H	0.60	6/4252 (0.1%)	0.59	0/5798
All	All	0.57	43/34016 (0.1%)	0.60	6/46384 (0.0%)

The worst 5 of 43 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	403	GLU	CD-OE1	10.25	1.36	1.25
1	H	385	LYS	CD-CE	9.87	1.75	1.51
1	H	385	LYS	CE-NZ	-9.31	1.25	1.49
1	G	4	GLU	CD-OE1	-7.78	1.17	1.25
1	G	4	GLU	CD-OE2	-7.65	1.17	1.25

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	352	LEU	CA-CB-CG	5.84	128.74	115.30
1	D	352	LEU	CA-CB-CG	5.65	128.31	115.30
1	C	498	ASP	CB-CG-OD1	5.17	122.96	118.30
1	G	228	ASN	C-N-CD	-5.08	109.43	120.60
1	B	213	GLU	N-CA-C	-5.05	97.35	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4124	0	3831	176	0
1	B	4124	0	3831	156	0
1	C	4124	0	3831	187	0
1	D	4124	0	3831	178	0
1	E	4124	0	3831	243	0
1	F	4124	0	3831	232	0
1	G	4124	0	3831	282	0
1	H	4124	0	3831	271	0
2	A	6	0	0	1	0
2	B	2	0	0	0	0
2	C	4	0	0	1	0
2	D	2	0	0	0	0
2	E	1	0	0	0	0
2	F	4	0	0	0	0
2	G	2	0	0	0	0
2	H	3	0	0	0	0
All	All	33016	0	30648	1670	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 26.

The worst 5 of 1670 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:H:385:LYS:CE	1:H:385:LYS:CD	1.75	1.58
1:E:228:ASN:ND2	1:E:229:PRO:HB3	1.35	1.40
1:F:228:ASN:ND2	1:F:229:PRO:HB3	1.35	1.37
1:H:486:ASN:HA	1:H:487:ALA:CB	1.41	1.35
1:H:97:PHE:CE2	1:H:106:ARG:HB3	1.62	1.34

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	507/512 (99%)	452 (89%)	44 (9%)	11 (2%)	8 46
1	B	507/512 (99%)	448 (88%)	46 (9%)	13 (3%)	7 42
1	C	507/512 (99%)	448 (88%)	47 (9%)	12 (2%)	7 44
1	D	507/512 (99%)	446 (88%)	48 (10%)	13 (3%)	7 42
1	E	507/512 (99%)	446 (88%)	46 (9%)	15 (3%)	5 39
1	F	507/512 (99%)	447 (88%)	45 (9%)	15 (3%)	5 39
1	G	507/512 (99%)	452 (89%)	41 (8%)	14 (3%)	6 41
1	H	507/512 (99%)	448 (88%)	44 (9%)	15 (3%)	5 39
All	All	4056/4096 (99%)	3587 (88%)	361 (9%)	108 (3%)	6 41

5 of 108 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	423	SER
1	A	424	LYS
1	D	423	SER
1	E	229	PRO
1	E	487	ALA

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	450/453 (99%)	416 (92%)	34 (8%)	16 54

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	450/453 (99%)	421 (94%)	29 (6%)	22 62
1	C	450/453 (99%)	416 (92%)	34 (8%)	16 54
1	D	450/453 (99%)	420 (93%)	30 (7%)	20 60
1	E	450/453 (99%)	409 (91%)	41 (9%)	12 44
1	F	450/453 (99%)	411 (91%)	39 (9%)	13 47
1	G	450/453 (99%)	403 (90%)	47 (10%)	9 37
1	H	450/453 (99%)	409 (91%)	41 (9%)	12 44
All	All	3600/3624 (99%)	3305 (92%)	295 (8%)	14 50

5 of 295 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	166	THR
1	E	498	ASP
1	H	278	THR
1	E	194	ASN
1	E	383	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 101 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	40	GLN
1	E	148	GLN
1	H	171	GLN
1	D	194	ASN
1	D	319	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	509/512 (99%)	0.01	3 (0%) 90 86	29, 37, 51, 70	0
1	B	509/512 (99%)	-0.04	2 (0%) 93 91	29, 38, 49, 72	0
1	C	509/512 (99%)	0.16	11 (2%) 65 60	38, 50, 66, 79	0
1	D	509/512 (99%)	0.16	16 (3%) 52 48	37, 51, 68, 78	0
1	E	509/512 (99%)	0.20	18 (3%) 48 42	33, 47, 71, 114	0
1	F	509/512 (99%)	0.25	19 (3%) 45 40	31, 48, 75, 132	0
1	G	509/512 (99%)	0.60	47 (9%) 11 11	39, 59, 89, 130	0
1	H	509/512 (99%)	0.57	52 (10%) 9 9	40, 62, 93, 140	0
All	All	4072/4096 (99%)	0.24	168 (4%) 41 36	29, 48, 81, 140	0

The worst 5 of 168 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	146	SER	5.7
1	F	324	THR	5.5
1	H	145	ASN	4.8
1	H	162	LYS	4.4
1	H	107	CYS	4.4

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.