



Full wwPDB X-ray Structure Validation Report ⓘ

Apr 8, 2019 – 10:40 AM EDT

PDB ID : 6C5C
Title : Crystal structure of Aro1p from *Candida albicans* SC5314 in complex with NADH
Authors : Michalska, K.; Evdokimova, E.; Di Leo, R.; Stogios, P.J.; Savchenko, A.; Joachimiak, A.; Satchell, K.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on : 2018-01-16
Resolution : 1.85 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix)	:	1.13
EDS	:	rb-20031633
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031633

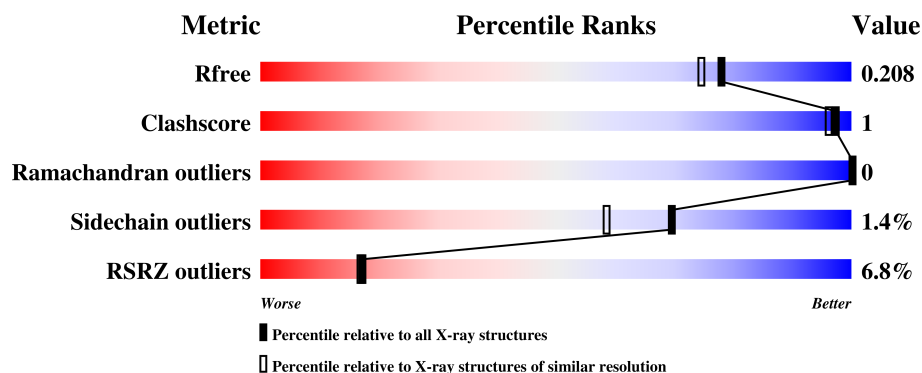
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.85 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}	111664	2111 (1.86-1.86)
Clashscore	122126	2258 (1.86-1.86)
Ramachandran outliers	120053	2234 (1.86-1.86)
Sidechain outliers	120020	2234 (1.86-1.86)
RSRZ outliers	108989	2075 (1.86-1.86)

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 $\leq 5\%$. 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	390	<div> <div>5%</div> <div>96%</div> <div>...</div> </div>
1	B	390	<div> <div>8%</div> <div>95%</div> <div>..</div> </div>

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 6478 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.

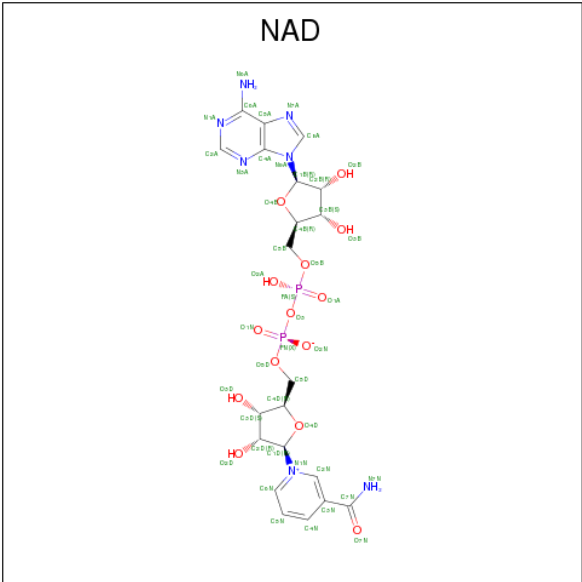
- Molecule 1 is a protein called 3-dehydroquinate synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	385	Total	C	N	O	S	0	2	0
			2983	1902	511	555	15			
1	B	383	Total	C	N	O	S	0	2	0
			2965	1891	507	552	15			

There are 6 discrepancies between the modelled and reference sequences:

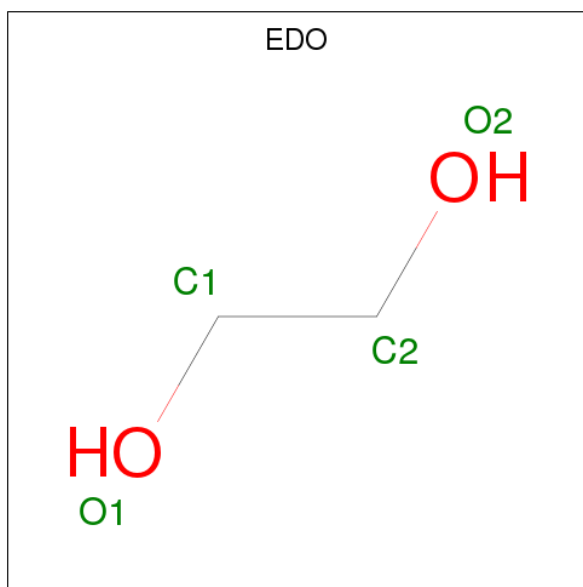
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q5AME2
A	-1	ASN	-	expression tag	UNP Q5AME2
A	0	ALA	-	expression tag	UNP Q5AME2
B	-2	SER	-	expression tag	UNP Q5AME2
B	-1	ASN	-	expression tag	UNP Q5AME2
B	0	ALA	-	expression tag	UNP Q5AME2

- Molecule 2 is NICOTINAMIDE-ADENINE-DINUCLEOTIDE (three-letter code: NAD) (formula: C₂₁H₂₇N₇O₁₄P₂).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			44	21	7	14	2		
2	B	1	Total	C	N	O	P	0	0
			44	21	7	14	2		

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	A	1	Total	C	O	0	0
			4	2	2		
3	B	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			4	2	2		
3	B	1	Total	C	O	0	0
			4	2	2		
3	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Cl	0	0
			1	1		

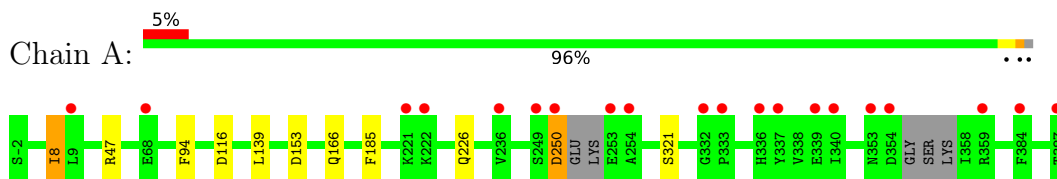
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	220	Total	O	0	0
			220	220		
5	B	169	Total	O	0	1
			169	169		

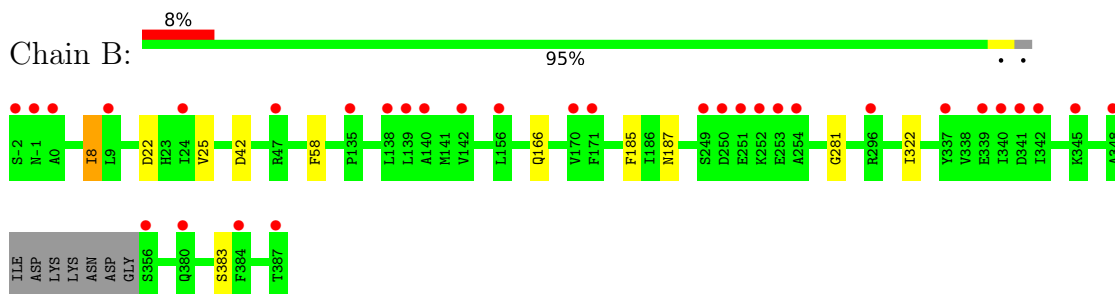
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: 3-dehydroquinate synthase



- Molecule 1: 3-dehydroquinate synthase



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	50.68Å 92.06Å 88.92Å 90.00° 97.57° 90.00°	Depositor
Resolution (Å)	30.00 – 1.85 29.51 – 1.84	Depositor EDS
% Data completeness (in resolution range)	99.3 (30.00-1.85) 98.9 (29.51-1.84)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.22 (at 1.84Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.176 , 0.203 0.184 , 0.208	Depositor DCC
R_{free} test set	2122 reflections (3.07%)	wwPDB-VP
Wilson B-factor (Å ²)	32.0	Xtriage
Anisotropy	0.406	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 44.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	6478	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.37% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, NAD, CL

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	0.70	0/3028	0.78	2/4088 (0.0%)
1	B	0.72	0/3011	0.79	1/4068 (0.0%)
All	All	0.71	0/6039	0.78	3/8156 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	153	ASP	CB-CG-OD2	-6.08	112.83	118.30
1	B	42	ASP	CB-CG-OD1	5.81	123.53	118.30
1	A	116	ASP	CB-CG-OD2	-5.01	113.79	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	2983	0	3097	6	0
1	B	2965	0	3077	5	0
2	A	44	0	26	0	0
2	B	44	0	26	0	0
3	A	36	0	53	3	0
3	B	16	0	24	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	1	0	0	0	0
5	A	220	0	0	1	0
5	B	169	0	0	0	0
All	All	6478	0	6303	11	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 1.

All (11) 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:A:166:GLN:HG3	3:A:405:EDO:H12	1.56	0.87
1:B:166:GLN:HG3	3:B:403:EDO:H12	1.62	0.81
1:A:250:ASP:N	1:A:250:ASP:OD1	2.38	0.57
1:A:8:ILE:HG12	3:A:405:EDO:H21	1.95	0.48
1:A:94:PHE:CD1	3:A:404:EDO:H21	2.50	0.47
1:B:187:ASN:OD1	1:B:281:GLY:N	2.49	0.45
1:B:322:ILE:HG23	3:B:405:EDO:H12	1.99	0.44
1:B:8:ILE:HG12	3:B:403:EDO:H21	2.00	0.44
1:B:25:VAL:HG22	1:B:58:PHE:CD1	2.54	0.42
1:A:139:LEU:C	1:A:139:LEU:HD13	2.40	0.42
1:A:47:ARG:NE	5:A:510:HOH:O	2.54	0.41

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	381/390 (98%)	372 (98%)	9 (2%)	0	100	100
1	B	381/390 (98%)	370 (97%)	11 (3%)	0	100	100
All	All	762/780 (98%)	742 (97%)	20 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	331/333 (99%)	326 (98%)	5 (2%)	67	56
1	B	329/333 (99%)	325 (99%)	4 (1%)	74	65
All	All	660/666 (99%)	651 (99%)	9 (1%)	69	58

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	8	ILE
1	A	185	PHE
1	A	226	GLN
1	A	250	ASP
1	A	321	SER
1	B	8	ILE
1	B	22	ASP
1	B	185	PHE
1	B	383	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	264	HIS
1	B	264	HIS

5.3.3 RNA ⓘ

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](#)

Of 16 ligands modelled in this entry, 1 is monoatomic - leaving 15 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	NAD	A	401	-	39,48,48	1.03	3 (7%)	44,73,73	1.54	6 (13%)
3	EDO	A	402	-	3,3,3	0.52	0	2,2,2	0.31	0
3	EDO	A	403	-	3,3,3	0.49	0	2,2,2	0.38	0
3	EDO	A	404	-	3,3,3	0.25	0	2,2,2	0.42	0
3	EDO	A	405	-	3,3,3	1.35	1 (33%)	2,2,2	0.76	0
3	EDO	A	406	-	3,3,3	0.47	0	2,2,2	0.24	0
3	EDO	A	407	-	3,3,3	0.35	0	2,2,2	0.36	0
3	EDO	A	408	-	3,3,3	0.44	0	2,2,2	0.44	0
3	EDO	A	409	-	3,3,3	0.47	0	2,2,2	0.52	0
3	EDO	A	410	-	3,3,3	0.54	0	2,2,2	0.18	0
2	NAD	B	401	-	39,48,48	0.96	1 (2%)	44,73,73	1.47	7 (15%)
3	EDO	B	402	-	3,3,3	0.38	0	2,2,2	0.60	0
3	EDO	B	403	-	3,3,3	0.58	0	2,2,2	0.72	0
3	EDO	B	404	-	3,3,3	0.55	0	2,2,2	0.34	0
3	EDO	B	405	-	3,3,3	0.35	0	2,2,2	0.57	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAD	A	401	-	-	0/22/62/62	0/5/5/5
3	EDO	A	402	-	-	0/1/1/1	0/0/0/0
3	EDO	A	403	-	-	0/1/1/1	0/0/0/0
3	EDO	A	404	-	-	0/1/1/1	0/0/0/0
3	EDO	A	405	-	-	0/1/1/1	0/0/0/0
3	EDO	A	406	-	-	0/1/1/1	0/0/0/0
3	EDO	A	407	-	-	0/1/1/1	0/0/0/0
3	EDO	A	408	-	-	0/1/1/1	0/0/0/0
3	EDO	A	409	-	-	0/1/1/1	0/0/0/0
3	EDO	A	410	-	-	0/1/1/1	0/0/0/0
2	NAD	B	401	-	-	0/22/62/62	0/5/5/5
3	EDO	B	402	-	-	0/1/1/1	0/0/0/0
3	EDO	B	403	-	-	0/1/1/1	0/0/0/0
3	EDO	B	404	-	-	0/1/1/1	0/0/0/0
3	EDO	B	405	-	-	0/1/1/1	0/0/0/0

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	405	EDO	O2-C2	-2.22	1.30	1.42
2	A	401	NAD	C5A-C4A	2.29	1.45	1.40
2	A	401	NAD	C2A-N3A	2.37	1.36	1.32
2	A	401	NAD	O4B-C1B	3.02	1.45	1.41
2	B	401	NAD	C5A-C4A	3.07	1.47	1.40

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	NAD	O7N-C7N-N7N	-3.87	117.09	122.60
2	A	401	NAD	N3A-C2A-N1A	-3.72	122.68	128.68
2	B	401	NAD	N3A-C2A-N1A	-3.33	123.32	128.68
2	B	401	NAD	C4A-C5A-N7A	-2.89	106.38	109.40
2	A	401	NAD	C1B-N9A-C4A	-2.87	121.68	126.64
2	B	401	NAD	C1B-N9A-C4A	-2.75	121.89	126.64
2	A	401	NAD	N6A-C6A-N1A	2.39	123.54	118.57
2	B	401	NAD	C3N-C2N-N1N	2.42	122.86	120.40
2	A	401	NAD	C3N-C2N-N1N	2.58	123.03	120.40
2	B	401	NAD	O2N-PN-O1N	2.74	125.97	112.21
2	B	401	NAD	C4B-O4B-C1B	3.71	113.69	109.83
2	B	401	NAD	C3N-C7N-N7N	3.77	122.04	117.75
2	A	401	NAD	C3N-C7N-N7N	4.84	123.26	117.75

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	404	EDO	1	0
3	A	405	EDO	2	0
3	B	403	EDO	2	0
3	B	405	EDO	1	0

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, 95th 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	385/390 (98%)	0.15	20 (5%) 27 26	23, 38, 60, 83	0
1	B	383/390 (98%)	0.31	32 (8%) 11 11	24, 41, 67, 80	0
All	All	768/780 (98%)	0.23	52 (6%) 17 17	23, 39, 65, 83	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	254	ALA	6.5
1	B	356	SER	6.2
1	B	252	LYS	5.7
1	A	353	ASN	5.6
1	B	254	ALA	5.2
1	B	0	ALA	5.2
1	A	250	ASP	5.2
1	B	-1	ASN	5.0
1	A	354	ASP	4.3
1	A	253	GLU	4.2
1	B	348	ALA	4.1
1	B	387	THR	4.0
1	B	342	ILE	3.9
1	B	253	GLU	3.9
1	B	251	GLU	3.9
1	B	142	VAL	3.7
1	A	221	LYS	3.5
1	A	222	LYS	3.4
1	B	345	LYS	3.4
1	A	337	TYR	3.4
1	B	337	TYR	3.3
1	B	-2	SER	3.3
1	B	250	ASP	3.2
1	A	336	HIS	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	387	THR	2.9
1	B	249	SER	2.8
1	B	47	ARG	2.7
1	A	384	PHE	2.7
1	B	138	LEU	2.7
1	B	341	ASP	2.6
1	B	380	GLN	2.6
1	A	332	GLY	2.5
1	A	9	LEU	2.5
1	A	68	GLU	2.4
1	A	333	PRO	2.4
1	B	140	ALA	2.4
1	B	135	PRO	2.3
1	B	339	GLU	2.3
1	B	340	ILE	2.3
1	B	139	LEU	2.3
1	B	171	PHE	2.2
1	A	249	SER	2.2
1	A	339	GLU	2.2
1	B	170	VAL	2.2
1	B	384	PHE	2.1
1	A	236	VAL	2.1
1	A	359	ARG	2.1
1	B	296[A]	ARG	2.1
1	B	156	LEU	2.1
1	B	9	LEU	2.1
1	B	24	ILE	2.0
1	A	340	ILE	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
3	EDO	A	409	4/4	0.84	0.24	57,57,57,57	0
3	EDO	A	410	4/4	0.84	0.15	52,56,58,61	0
3	EDO	B	405	4/4	0.87	0.18	51,51,51,53	0
3	EDO	A	406	4/4	0.87	0.26	52,56,56,62	0
3	EDO	A	404	4/4	0.89	0.16	35,37,37,48	0
3	EDO	A	403	4/4	0.90	0.17	37,41,43,44	0
3	EDO	A	407	4/4	0.90	0.23	41,45,46,48	0
3	EDO	B	404	4/4	0.91	0.20	49,54,54,55	0
3	EDO	A	405	4/4	0.91	0.20	20,24,27,35	0
3	EDO	A	408	4/4	0.92	0.26	60,60,61,63	0
3	EDO	B	402	4/4	0.92	0.16	48,51,51,53	0
2	NAD	B	401	44/44	0.95	0.11	30,35,42,43	0
3	EDO	B	403	4/4	0.96	0.21	26,29,31,44	0
3	EDO	A	402	4/4	0.96	0.12	41,46,47,47	0
2	NAD	A	401	44/44	0.97	0.09	26,29,32,40	0
4	CL	A	411	1/1	0.98	0.07	39,39,39,39	0

6.5 Other polymers [i](#)

There are no such residues in this entry.