



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 02:16 am GMT

PDB ID : 1K7W  
Title : Crystal Structure of S283A Duck Delta 2 Crystallin Mutant  
Authors : Sampaleanu, L.M.; Yu, B.; Howell, P.L.  
Deposited on : 2001-10-22  
Resolution : 1.96 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

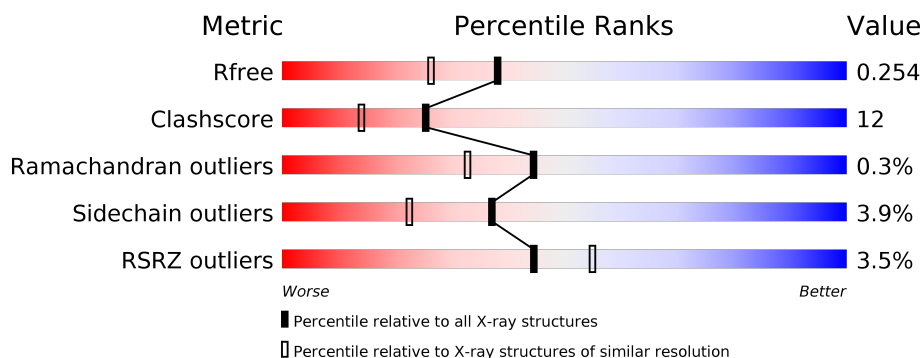
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.96 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))
$R_{free}$	100719	2004 (1.96-1.96)
Clashscore	112137	2136 (1.96-1.96)
Ramachandran outliers	110173	2117 (1.96-1.96)
Sidechain outliers	110143	2117 (1.96-1.96)
RSRZ outliers	101464	2018 (1.96-1.96)

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  $\geq 3$ , 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	468	<div> <div>4%</div> <div> <div></div> <div>72%</div> <div>23%</div> <div>• •</div> </div> </div>
1	B	468	<div> <div>5%</div> <div> <div></div> <div>71%</div> <div>22%</div> <div>• •</div> </div> </div>
1	C	468	<div> <div>2%</div> <div> <div></div> <div>74%</div> <div>20%</div> <div>• •</div> </div> </div>
1	D	468	<div> <div>2%</div> <div> <div></div> <div>72%</div> <div>21%</div> <div>• •</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	AS1	A	1004	X	-	-	X
2	AS1	B	1003	X	-	-	X
2	AS1	C	1001	X	-	-	X
2	AS1	D	1002	X	-	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 14719 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.

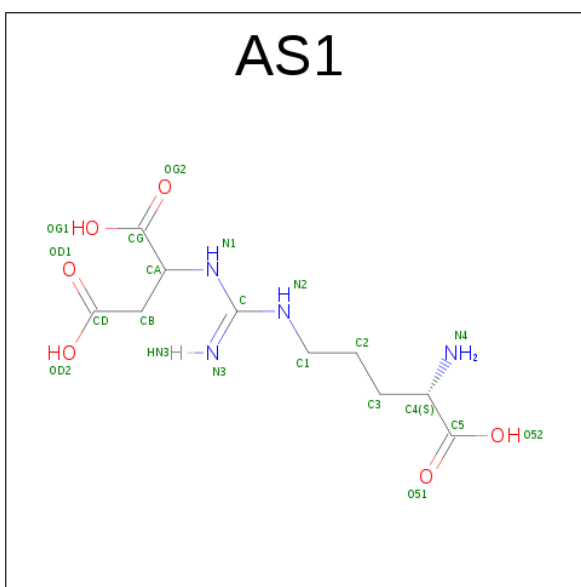
- Molecule 1 is a protein called delta 2 crystallin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	449	Total	C	N	O	S	0	0	0
			3485	2205	591	677	12			
1	B	448	Total	C	N	O	S	0	0	0
			3473	2198	589	674	12			
1	C	448	Total	C	N	O	S	0	0	0
			3482	2206	589	675	12			
1	D	450	Total	C	N	O	S	0	0	0
			3492	2212	591	677	12			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	283	ALA	SER	ENGINEERED	UNP P24058
A	443	GLY	ALA	CONFLICT	UNP P24058
B	283	ALA	SER	ENGINEERED	UNP P24058
B	443	GLY	ALA	CONFLICT	UNP P24058
C	283	ALA	SER	ENGINEERED	UNP P24058
C	443	GLY	ALA	CONFLICT	UNP P24058
D	283	ALA	SER	ENGINEERED	UNP P24058
D	443	GLY	ALA	CONFLICT	UNP P24058

- Molecule 2 is ARGININOSUCCINATE (three-letter code: AS1) (formula: C<sub>10</sub>H<sub>18</sub>N<sub>4</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	C	1	Total 20	C 10	N 4	O 6	0	0
2	D	1	Total 20	C 10	N 4	O 6	0	0
2	B	1	Total 20	C 10	N 4	O 6	0	0
2	A	1	Total 20	C 10	N 4	O 6	0	0

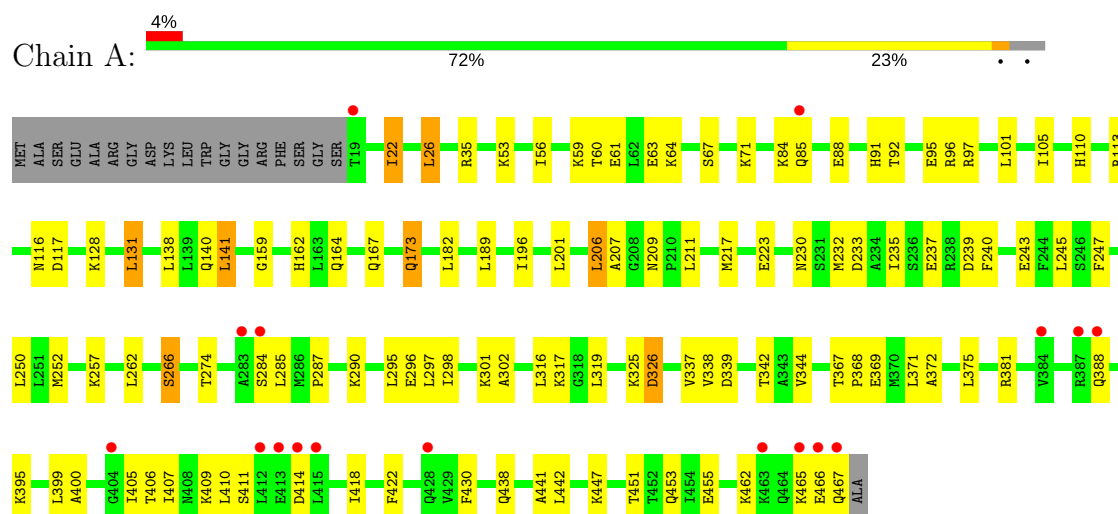
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	211	Total O 211 211	0	0
3	B	137	Total O 137 137	0	0
3	C	191	Total O 191 191	0	0
3	D	168	Total O 168 168	0	0

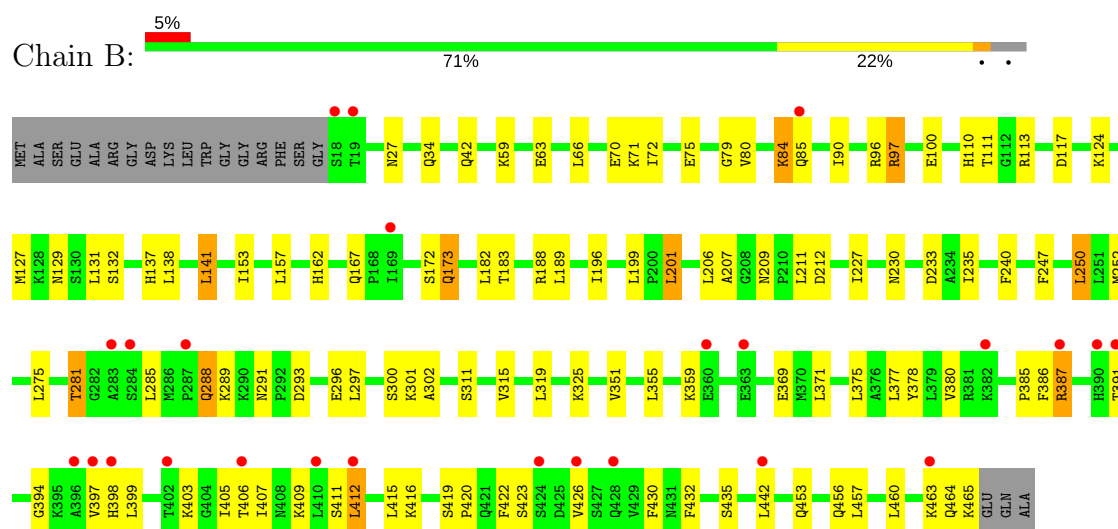
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

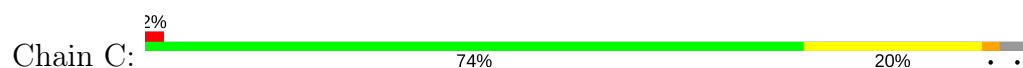
#### • Molecule 1: delta 2 crystallin

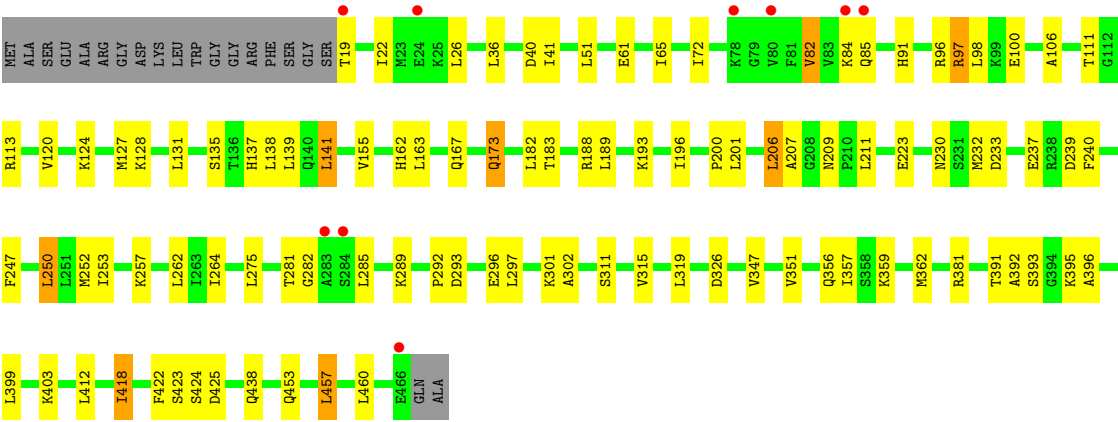


#### • Molecule 1: delta 2 crystallin

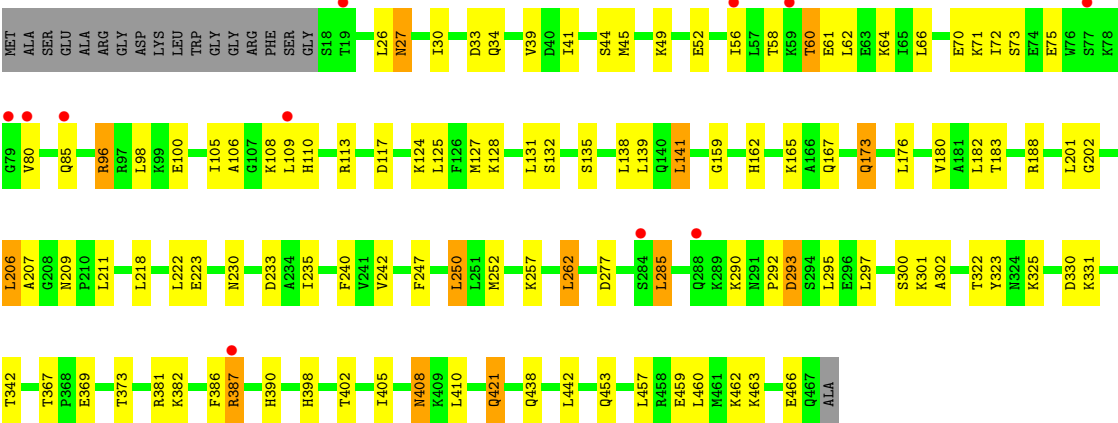


#### • Molecule 1: delta 2 crystallin





● Molecule 1: delta 2 crystallin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.76Å 98.64Å 106.15Å 90.00° 101.30° 90.00°	Depositor
Resolution (Å)	19.81 – 1.96 19.81 – 1.96	Depositor EDS
% Data completeness (in resolution range)	92.0 (19.81-1.96) 92.0 (19.81-1.96)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.65 (at 1.96Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.207 , 0.253 0.207 , 0.254	Depositor DCC
$R_{free}$ test set	12440 reflections (9.95%)	DCC
Wilson B-factor (Å <sup>2</sup> )	22.6	Xtriage
Anisotropy	0.204	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 57.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	14719	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.39% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: AS1

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	0.49	1/3529 (0.0%)	0.64	1/4759 (0.0%)
1	B	0.43	0/3517	0.63	0/4743
1	C	0.47	0/3527	0.65	0/4756
1	D	0.45	0/3537	0.63	1/4770 (0.0%)
All	All	0.46	1/14110 (0.0%)	0.64	2/19028 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	243	GLU	CB-CG	-5.20	1.42	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	159	GLY	N-CA-C	-5.18	100.15	113.10
1	D	159	GLY	N-CA-C	-5.15	100.23	113.10

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	3485	0	3604	100	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	3473	0	3595	94	0
1	C	3482	0	3603	81	0
1	D	3492	0	3607	101	0
2	A	20	0	14	0	0
2	B	20	0	14	1	0
2	C	20	0	14	0	0
2	D	20	0	14	1	0
3	A	211	0	0	11	0
3	B	137	0	0	5	0
3	C	191	0	0	8	0
3	D	168	0	0	6	0
All	All	14719	0	14465	343	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 12.

The worst 5 of 343 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:C:183:THR:HG21	1:C:460:LEU:HD13	1.29	1.06
1:C:285:LEU:HD21	1:D:390:HIS:NE2	1.71	1.06
1:B:423:SER:O	1:B:426:VAL:HG23	1.71	0.91
1:D:183:THR:HG21	1:D:460:LEU:HD13	1.51	0.90
1:D:201:LEU:HD22	1:D:218:LEU:HD13	1.56	0.86

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	447/468 (96%)	439 (98%)	7 (2%)	1 (0%)	51 41

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	446/468 (95%)	434 (97%)	10 (2%)	2 (0%)	38	25
1	C	446/468 (95%)	439 (98%)	5 (1%)	2 (0%)	38	25
1	D	448/468 (96%)	438 (98%)	9 (2%)	1 (0%)	51	41
All	All	1787/1872 (96%)	1750 (98%)	31 (2%)	6 (0%)	44	33

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	206	LEU
1	A	206	LEU
1	B	206	LEU
1	C	206	LEU
1	B	464	GLN

### 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	391/404 (97%)	377 (96%)	14 (4%)	40	26
1	B	390/404 (96%)	373 (96%)	17 (4%)	33	18
1	C	391/404 (97%)	377 (96%)	14 (4%)	40	26
1	D	391/404 (97%)	375 (96%)	16 (4%)	35	21
All	All	1563/1616 (97%)	1502 (96%)	61 (4%)	37	22

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	387	ARG
1	C	82	VAL
1	D	293	ASP
1	B	412	LEU
1	B	457	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	230	ASN
1	C	85	GLN
1	D	408	ASN
1	B	291	ASN
1	B	438	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

### 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

4 ligands are modelled in this entry.

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	AS1	A	1004	-	7,19,19	3.22	3 (42%)	8,24,24	6.61	6 (75%)
2	AS1	B	1003	-	7,19,19	3.33	3 (42%)	8,24,24	6.73	6 (75%)
2	AS1	C	1001	-	7,19,19	3.22	3 (42%)	8,24,24	6.69	6 (75%)
2	AS1	D	1002	-	7,19,19	3.11	3 (42%)	8,24,24	6.76	6 (75%)

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	AS1	A	1004	-	1/1/6/8	0/13/23/23	0/0/0/0
2	AS1	B	1003	-	1/1/6/8	0/13/23/23	0/0/0/0
2	AS1	C	1001	-	1/1/6/8	0/13/23/23	0/0/0/0
2	AS1	D	1002	-	1/1/6/8	0/13/23/23	0/0/0/0

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	1002	AS1	CA-N1	3.35	1.51	1.46
2	A	1004	AS1	CA-N1	3.53	1.51	1.46
2	C	1001	AS1	CA-N1	3.55	1.51	1.46
2	B	1003	AS1	CA-N1	3.83	1.52	1.46
2	C	1001	AS1	C-N1	4.78	1.45	1.34

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1001	AS1	CD-CB-CA	-3.95	105.86	114.30
2	D	1002	AS1	CD-CB-CA	-3.86	106.05	114.30
2	A	1004	AS1	CD-CB-CA	-3.75	106.28	114.30
2	B	1003	AS1	CD-CB-CA	-3.38	107.08	114.30
2	A	1004	AS1	C3-C2-C1	4.67	126.05	112.11

All (4) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	D	1002	AS1	CA
2	B	1003	AS1	CA
2	C	1001	AS1	CA
2	A	1004	AS1	CA

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1003	AS1	1	0
2	D	1002	AS1	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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	449/468 (95%)	-0.01	17 (3%) 41 51	15, 25, 53, 71	0
1	B	448/468 (95%)	0.19	25 (5%) 25 35	16, 31, 53, 66	0
1	C	448/468 (95%)	-0.08	9 (2%) 65 75	13, 25, 45, 62	0
1	D	450/468 (96%)	0.05	11 (2%) 59 69	15, 28, 50, 62	0
All	All	1795/1872 (95%)	0.04	62 (3%) 44 55	13, 27, 51, 71	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	80	VAL	4.9
1	B	18	SER	4.6
1	A	19	THR	4.5
1	A	467	GLN	4.1
1	B	402	THR	3.8

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	AS1	A	1004	20/20	0.86	0.18	5.46	17,40,60,60	0
2	AS1	B	1003	20/20	0.87	0.17	3.50	18,39,61,62	0
2	AS1	D	1002	20/20	0.81	0.20	2.86	22,48,67,67	0
2	AS1	C	1001	20/20	0.87	0.18	2.59	17,40,57,57	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.