wwPDB X-ray Structure Validation Summary Report (i)

Oct 23, 2021 – 10:56 AM EDT

PDB ID	:	1ZEI
Title	:	CROSS-LINKED B28 ASP INSULIN
Authors	:	Whittingham, J.L.; Edwards, E.J.; Antson, A.A.; Clarkson, J.M.; Dodson, G.G.
Deposited on	:	1998-07-14
Resolution	:	1.90 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

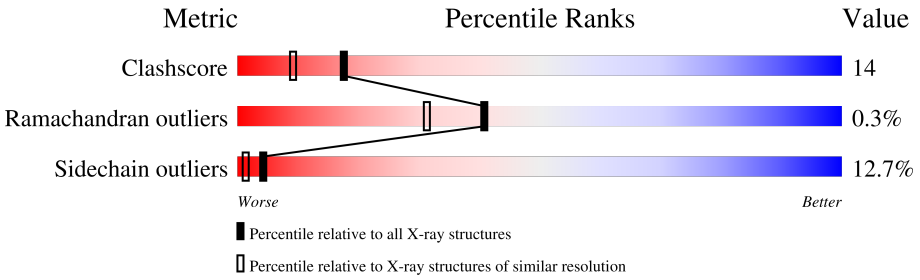
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 1.90 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))
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	53	45% 36% 15% ·
1	B	53	30% 53% 13% ·
1	C	53	40% 42% 17% ·
1	D	53	36% 45% 15% ·
1	E	53	40% 51% 8% ·
1	F	53	47% 43% 9%

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	CRS	B	56	-	X	-	-
2	CRS	D	54	-	X	-	-
2	CRS	E	55	-	X	-	-
2	CRS	F	55	-	X	-	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2806 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 INSULIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	53	Total	C	N	O	S	41	0	0
			417	264	68	79	6			
1	B	53	Total	C	N	O	S	8	1	0
			420	265	69	80	6			
1	C	53	Total	C	N	O	S	16	1	0
			421	266	68	81	6			
1	D	53	Total	C	N	O	S	16	0	0
			417	264	68	79	6			
1	E	53	Total	C	N	O	S	6	0	0
			417	264	68	79	6			
1	F	53	Total	C	N	O	S	7	1	0
			418	264	68	80	6			

There are 18 discrepancies between the modelled and reference sequences:

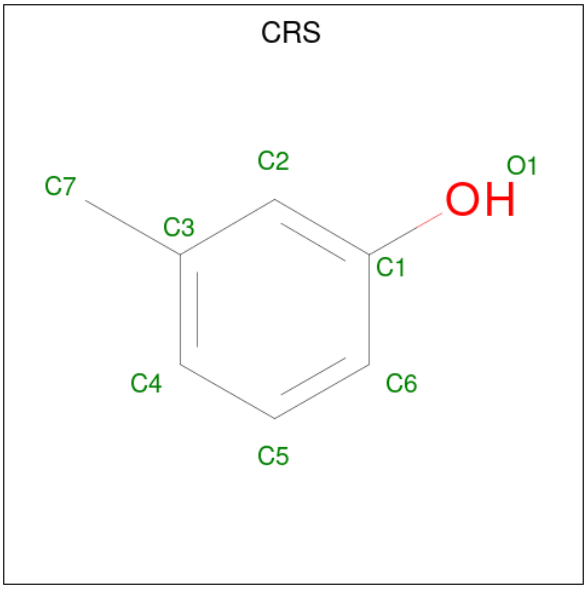
Chain	Residue	Modelled	Actual	Comment	Reference
A	28	ASP	PRO	engineered mutation	UNP P01315
A	31	ALA	-	insertion	UNP P01315
A	32	LYS	-	insertion	UNP P01315
B	28	ASP	PRO	engineered mutation	UNP P01315
B	31	ALA	-	insertion	UNP P01315
B	32	LYS	-	insertion	UNP P01315
C	28	ASP	PRO	engineered mutation	UNP P01315
C	31	ALA	-	insertion	UNP P01315
C	32	LYS	-	insertion	UNP P01315
D	28	ASP	PRO	engineered mutation	UNP P01315
D	31	ALA	-	insertion	UNP P01315
D	32	LYS	-	insertion	UNP P01315
E	28	ASP	PRO	engineered mutation	UNP P01315
E	31	ALA	-	insertion	UNP P01315
E	32	LYS	-	insertion	UNP P01315
F	28	ASP	PRO	engineered mutation	UNP P01315
F	31	ALA	-	insertion	UNP P01315

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	32	LYS	-	insertion	UNP P01315

- Molecule 2 is M-CRESOL (three-letter code: CRS) (formula: C_7H_8O).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 8 7 1	0	0
2	B	1	Total C O 8 7 1	0	0
2	C	1	Total C O 8 7 1	0	0
2	D	1	Total C O 8 7 1	0	0
2	E	1	Total C O 8 7 1	0	0
2	E	1	Total C O 8 7 1	0	0
2	F	1	Total C O 8 7 1	0	0
2	F	1	Total C O 8 7 1	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Zn 1 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	1	Total Zn 1 1	0	0

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Cl 1 1	0	0
4	C	1	Total Cl 1 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	25	Total O 25 25	0	0
5	B	34	Total O 34 34	0	0
5	C	40	Total O 40 40	0	0
5	D	36	Total O 36 36	0	0
5	E	49	Total O 49 49	0	0
5	F	44	Total O 44 44	0	0

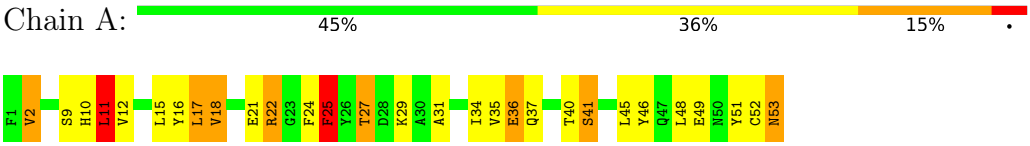
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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.

Note EDS was not executed.

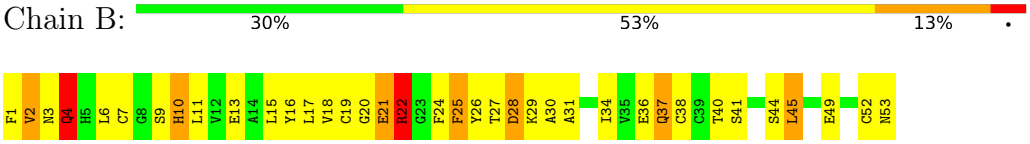
• Molecule 1: INSULIN

Chain A:



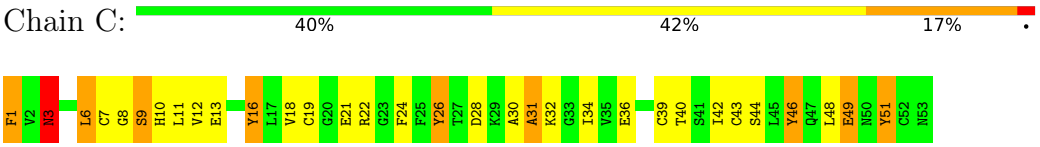
• Molecule 1: INSULIN

Chain B:



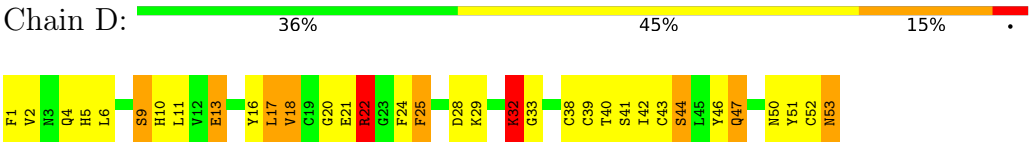
• Molecule 1: INSULIN

Chain C:



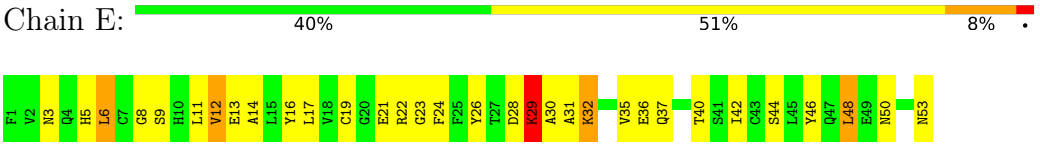
• Molecule 1: INSULIN

Chain D:



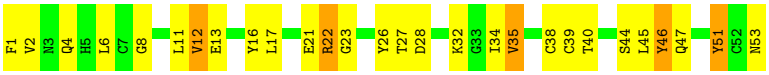
• Molecule 1: INSULIN

Chain E:



• Molecule 1: INSULIN

Chain F:


F1
V2
N3
Q4
H5
L6
C7
G8
L11
V12
E13
Y16
L17
E21
R22
G23
Y26
T27
D28
K32
G33
I34
V35
C38
C39
T40
S44
L45
Y46
Q47
Y51
C52
N53

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	53.95Å 64.77Å 48.91Å 90.00° 109.81° 90.00°	Depositor
Resolution (Å)	20.00 – 1.90	Depositor
% Data completeness (in resolution range)	87.8 (20.00-1.90)	Depositor
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	REFMAC	Depositor
R, R_{free}	0.176 , 0.232	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2806	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

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

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	1.39	2/425 (0.5%)	2.41	20/572 (3.5%)
1	B	1.71	3/434 (0.7%)	2.78	31/584 (5.3%)
1	C	1.66	10/434 (2.3%)	2.27	22/584 (3.8%)
1	D	1.90	9/425 (2.1%)	2.43	36/572 (6.3%)
1	E	2.03	6/425 (1.4%)	2.64	31/572 (5.4%)
1	F	1.69	4/431 (0.9%)	2.53	34/580 (5.9%)
All	All	1.74	34/2574 (1.3%)	2.52	174/3464 (5.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2
1	B	0	6
1	C	0	3
1	D	0	2
1	E	0	5
1	F	0	2
All	All	0	20

The worst 5 of 34 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	23	GLY	CA-C	9.84	1.67	1.51
1	F	46	TYR	CB-CG	8.25	1.64	1.51
1	E	9	SER	CB-OG	-7.93	1.31	1.42
1	D	38	CYS	CB-SG	7.47	1.95	1.82
1	E	53	ASN	C-OXT	7.44	1.37	1.23

The worst 5 of 174 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	22	ARG	NE-CZ-NH2	28.83	134.72	120.30
1	A	22	ARG	NE-CZ-NH2	-17.79	111.41	120.30
1	E	24	PHE	CB-CG-CD2	-16.47	109.27	120.80
1	B	22	ARG	NE-CZ-NH1	-15.54	112.53	120.30
1	A	22	ARG	NE-CZ-NH1	14.08	127.34	120.30

There are no chirality outliers.

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	11	LEU	Mainchain
1	A	18	VAL	Mainchain
1	B	2	VAL	Mainchain
1	B	4[A]	GLN	Mainchain
1	B	4[B]	GLN	Mainchain

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	417	0	391	21	0
1	B	420	0	392	17	0
1	C	421	0	394	13	0
1	D	417	0	392	17	2
1	E	417	0	392	8	0
1	F	418	0	393	6	2
2	A	8	0	8	2	0
2	B	8	0	8	2	0
2	C	8	0	8	0	0
2	D	8	0	8	0	0
2	E	16	0	16	0	0
2	F	16	0	16	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	25	0	0	1	0
5	B	34	0	0	2	0
5	C	40	0	0	2	0
5	D	36	0	0	4	0
5	E	49	0	0	3	0
5	F	44	0	0	0	0
All	All	2806	0	2418	66	2

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 14.

The worst 5 of 66 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:D:22:ARG:HD2	5:D:77:HOH:O	1.48	1.14
1:A:45:LEU:HD21	1:D:17:LEU:CD1	1.90	1.02
1:A:45:LEU:HD21	1:D:17:LEU:HD11	1.51	0.92
1:A:2:VAL:CG1	1:A:2:VAL:O	2.25	0.84
1:D:44:SER:H	1:D:47:GLN:HE21	1.30	0.80

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:29:LYS:NZ	1:F:53:ASN:O[2_455]	1.63	0.57
1:D:29:LYS:NZ	1:F:22:ARG:NH2[2_455]	1.87	0.33

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	51/53 (96%)	49 (96%)	2 (4%)	0	100 100
1	B	52/53 (98%)	49 (94%)	3 (6%)	0	100 100
1	C	52/53 (98%)	50 (96%)	1 (2%)	1 (2%)	8 1
1	D	51/53 (96%)	47 (92%)	4 (8%)	0	100 100
1	E	51/53 (96%)	51 (100%)	0	0	100 100
1	F	52/53 (98%)	52 (100%)	0	0	100 100
All	All	309/318 (97%)	298 (96%)	10 (3%)	1 (0%)	41 31

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	32	LYS

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	46/46 (100%)	37 (80%)	9 (20%)	1 0
1	B	47/46 (102%)	40 (85%)	7 (15%)	3 1
1	C	47/46 (102%)	43 (92%)	4 (8%)	10 4
1	D	46/46 (100%)	38 (83%)	8 (17%)	2 0
1	E	46/46 (100%)	41 (89%)	5 (11%)	6 2
1	F	47/46 (102%)	44 (94%)	3 (6%)	17 8
All	All	279/276 (101%)	243 (87%)	36 (13%)	4 1

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	3	ASN
1	F	46	TYR
1	E	6	LEU
1	E	48	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	22	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	53	ASN
1	F	5	HIS
1	F	53	ASN
1	F	47	GLN
1	C	3	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 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	CRS	F	55	-	8,8,8	2.03	3 (37%)	10,10,10	8.50	8 (80%)
2	CRS	E	55	-	8,8,8	1.83	3 (37%)	10,10,10	3.79	8 (80%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	CRS	E	54	-	8,8,8	1.44	1 (12%)	10,10,10	6.98	6 (60%)
2	CRS	C	56	-	8,8,8	1.70	2 (25%)	10,10,10	1.70	2 (20%)
2	CRS	A	54	-	8,8,8	1.02	0	10,10,10	2.71	6 (60%)
2	CRS	B	56	-	8,8,8	1.85	3 (37%)	10,10,10	3.85	5 (50%)
2	CRS	D	54	-	8,8,8	3.20	5 (62%)	10,10,10	8.34	5 (50%)
2	CRS	F	54	-	8,8,8	2.00	2 (25%)	10,10,10	2.23	3 (30%)

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	CRS	F	55	-	-	-	0/1/1/1
2	CRS	E	55	-	-	-	0/1/1/1
2	CRS	E	54	-	-	-	0/1/1/1
2	CRS	C	56	-	-	-	0/1/1/1
2	CRS	A	54	-	-	-	0/1/1/1
2	CRS	B	56	-	-	-	0/1/1/1
2	CRS	D	54	-	-	-	0/1/1/1
2	CRS	F	54	-	-	-	0/1/1/1

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	54	CRS	C5-C6	6.15	1.51	1.38
2	D	54	CRS	O1-C1	4.21	1.46	1.37
2	F	55	CRS	C2-C1	3.93	1.45	1.39
2	B	56	CRS	C5-C6	3.66	1.46	1.38
2	E	55	CRS	O1-C1	3.57	1.45	1.37

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	54	CRS	C1-C2-C3	-21.90	107.75	120.07
2	F	55	CRS	C1-C2-C3	-21.15	108.17	120.07
2	E	54	CRS	C1-C2-C3	-19.41	109.14	120.07
2	F	55	CRS	C6-C5-C4	-10.11	105.90	120.25
2	B	56	CRS	C1-C2-C3	-9.62	114.66	120.07

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	54	CRS	2	0
2	B	56	CRS	2	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	27:THR	C	28:ASP	N	1.18

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.