Full wwPDB X-ray Structure Validation Report (i)

Aug 20, 2023 – 09:11 PM EDT

PDB ID	:	2HY6
Title	:	A seven-helix coiled coil
Authors	:	Liu, J.; Zheng, Q.; Deng, Y.; Cheng, C.S.; Kallenbach, N.R.; Lu, M.
Deposited on	:	2006-08-04
Resolution	:	1.25 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

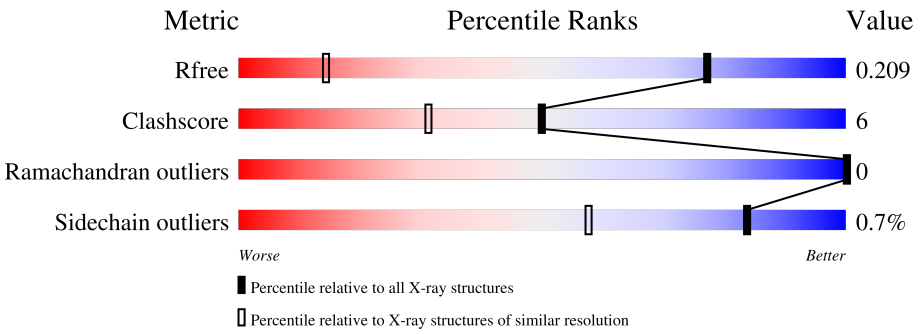
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 1.25 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}	130704	1023 (1.28-1.24)
Clashscore	141614	1060 (1.28-1.24)
Ramachandran outliers	138981	1029 (1.28-1.24)
Sidechain outliers	138945	1028 (1.28-1.24)

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%

Mol	Chain	Length	Quality of chain
1	A	34	88% • 9%
1	B	34	91% 9%
1	C	34	82% 9% 9%
1	D	34	94% 6%
1	E	34	82% 9% 9%
1	F	34	91% • 6%
1	G	34	85% • 12%

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	HEZ	A	202	-	-	X	-
2	HEZ	D	203	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1855 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 General control protein GCN4.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	A	31	Total C N O 228 142 41 45	0	1	0
1	B	31	Total C N O 226 141 41 44	0	0	0
1	C	31	Total C N O 233 145 41 47	0	2	0
1	D	32	Total C N O 240 148 45 47	0	1	0
1	E	31	Total C N O 228 142 41 45	0	1	0
1	F	32	Total C N O S 236 148 43 44 1	0	1	0
1	G	30	Total C N O 221 138 41 42	0	1	0

There are 70 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	LYS	-	cloning artifact	UNP P03069
A	3	VAL	-	cloning artifact	UNP P03069
A	7	ALA	GLU	engineered mutation	UNP P03069
A	9	ALA	LYS	engineered mutation	UNP P03069
A	14	ALA	LEU	engineered mutation	UNP P03069
A	16	ALA	LYS	engineered mutation	UNP P03069
A	21	ALA	GLU	engineered mutation	UNP P03069
A	23	ALA	GLU	engineered mutation	UNP P03069
A	28	ALA	LYS	engineered mutation	UNP P03069
A	30	ALA	LEU	engineered mutation	UNP P03069
B	2	LYS	-	cloning artifact	UNP P03069
B	3	VAL	-	cloning artifact	UNP P03069
B	7	ALA	GLU	engineered mutation	UNP P03069
B	9	ALA	LYS	engineered mutation	UNP P03069
B	14	ALA	LEU	engineered mutation	UNP P03069

Continued on next page...

Continued from previous page...

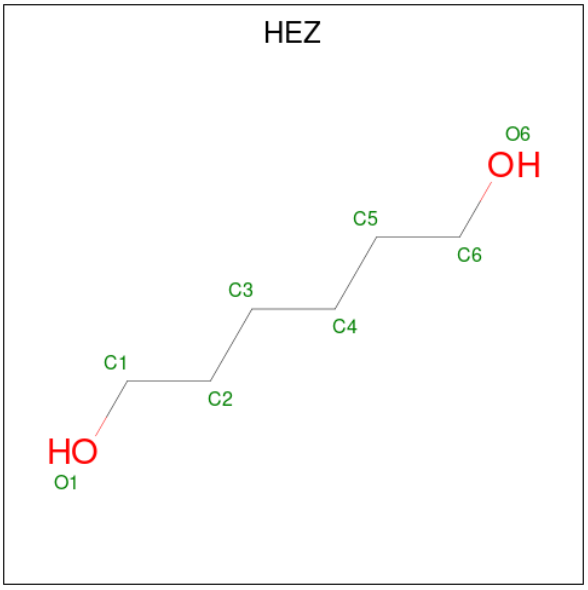
Chain	Residue	Modelled	Actual	Comment	Reference
B	16	ALA	LYS	engineered mutation	UNP P03069
B	21	ALA	GLU	engineered mutation	UNP P03069
B	23	ALA	GLU	engineered mutation	UNP P03069
B	28	ALA	LYS	engineered mutation	UNP P03069
B	30	ALA	LEU	engineered mutation	UNP P03069
C	2	LYS	-	cloning artifact	UNP P03069
C	3	VAL	-	cloning artifact	UNP P03069
C	7	ALA	GLU	engineered mutation	UNP P03069
C	9	ALA	LYS	engineered mutation	UNP P03069
C	14	ALA	LEU	engineered mutation	UNP P03069
C	16	ALA	LYS	engineered mutation	UNP P03069
C	21	ALA	GLU	engineered mutation	UNP P03069
C	23	ALA	GLU	engineered mutation	UNP P03069
C	28	ALA	LYS	engineered mutation	UNP P03069
C	30	ALA	LEU	engineered mutation	UNP P03069
D	2	LYS	-	cloning artifact	UNP P03069
D	3	VAL	-	cloning artifact	UNP P03069
D	7	ALA	GLU	engineered mutation	UNP P03069
D	9	ALA	LYS	engineered mutation	UNP P03069
D	14	ALA	LEU	engineered mutation	UNP P03069
D	16	ALA	LYS	engineered mutation	UNP P03069
D	21	ALA	GLU	engineered mutation	UNP P03069
D	23	ALA	GLU	engineered mutation	UNP P03069
D	28	ALA	LYS	engineered mutation	UNP P03069
D	30	ALA	LEU	engineered mutation	UNP P03069
E	2	LYS	-	cloning artifact	UNP P03069
E	3	VAL	-	cloning artifact	UNP P03069
E	7	ALA	GLU	engineered mutation	UNP P03069
E	9	ALA	LYS	engineered mutation	UNP P03069
E	14	ALA	LEU	engineered mutation	UNP P03069
E	16	ALA	LYS	engineered mutation	UNP P03069
E	21	ALA	GLU	engineered mutation	UNP P03069
E	23	ALA	GLU	engineered mutation	UNP P03069
E	28	ALA	LYS	engineered mutation	UNP P03069
E	30	ALA	LEU	engineered mutation	UNP P03069
F	2	LYS	-	cloning artifact	UNP P03069
F	3	VAL	-	cloning artifact	UNP P03069
F	7	ALA	GLU	engineered mutation	UNP P03069
F	9	ALA	LYS	engineered mutation	UNP P03069
F	14	ALA	LEU	engineered mutation	UNP P03069
F	16	ALA	LYS	engineered mutation	UNP P03069
F	21	ALA	GLU	engineered mutation	UNP P03069

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	23	ALA	GLU	engineered mutation	UNP P03069
F	28	ALA	LYS	engineered mutation	UNP P03069
F	30	ALA	LEU	engineered mutation	UNP P03069
G	2	LYS	-	cloning artifact	UNP P03069
G	3	VAL	-	cloning artifact	UNP P03069
G	7	ALA	GLU	engineered mutation	UNP P03069
G	9	ALA	LYS	engineered mutation	UNP P03069
G	14	ALA	LEU	engineered mutation	UNP P03069
G	16	ALA	LYS	engineered mutation	UNP P03069
G	21	ALA	GLU	engineered mutation	UNP P03069
G	23	ALA	GLU	engineered mutation	UNP P03069
G	28	ALA	LYS	engineered mutation	UNP P03069
G	30	ALA	LEU	engineered mutation	UNP P03069

- Molecule 2 is HEXANE-1,6-DIOL (three-letter code: HEZ) (formula: $C_6H_{14}O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	O		0	0
			8	6	2			
2	C	1	Total	C	O		0	0
			8	6	2			
2	D	1	Total	C	O		0	0
			8	6	2			
2	D	1	Total	C	O		0	0
			8	6	2			
2	E	1	Total	C	O		0	0
			8	6	2			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	F	1	8	6	2	0	0
			Total	C	O		
2	F	1	8	6	2	0	0
			Total	C	O		
2	G	1	8	6	2	0	0

- Molecule 3 is water.

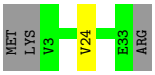
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	26	26	26	0	0
			Total	O		
3	B	22	22	22	0	0
			Total	O		
3	C	22	22	22	0	0
			Total	O		
3	D	28	28	28	0	0
			Total	O		
3	E	29	29	29	0	0
			Total	O		
3	F	22	22	22	0	0
			Total	O		
3	G	30	30	30	0	0

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.

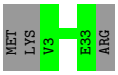
• Molecule 1: General control protein GCN4

Chain A:



• Molecule 1: General control protein GCN4

Chain B:

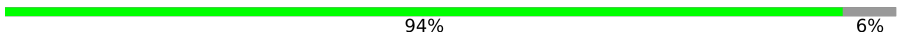


• Molecule 1: General control protein GCN4

Chain C:

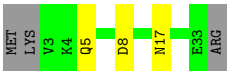
• Molecule 1: General control protein GCN4

Chain D:



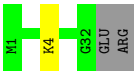
• Molecule 1: General control protein GCN4

Chain E:



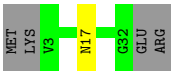
• Molecule 1: General control protein GCN4

Chain F:



• Molecule 1: General control protein GCN4

Chain G:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	31.49Å 35.30Å 48.98Å 86.54° 104.28° 99.95°	Depositor
Resolution (Å)	47.50 – 1.25 23.73 – 1.25	Depositor EDS
% Data completeness (in resolution range)	91.8 (47.50-1.25) 91.7 (23.73-1.25)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.09 (at 1.25Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.181 , 0.209 0.184 , 0.209	Depositor DCC
R_{free} test set	2596 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	12.1	Xtriage
Anisotropy	0.525	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 58.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	1855	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.96% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: HEZ

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	0.41	0/233	0.52	0/315
1	B	0.43	0/227	0.59	0/307
1	C	0.56	0/242	0.58	0/327
1	D	0.48	0/245	0.54	0/329
1	E	0.43	0/233	0.57	0/315
1	F	0.42	0/241	0.56	0/324
1	G	0.51	0/226	0.50	0/306
All	All	0.47	0/1647	0.55	0/2223

There are no bond length outliers.

There are no bond angle outliers.

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	228	0	230	1	0
1	B	226	0	229	0	0
1	C	233	0	232	5	0
1	D	240	0	243	0	0
1	E	228	0	230	4	0
1	F	236	0	249	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	221	0	225	1	0
2	A	8	0	14	9	0
2	C	8	0	14	2	0
2	D	16	0	28	10	0
2	E	8	0	14	3	0
2	F	16	0	28	3	0
2	G	8	0	14	0	0
3	A	26	0	0	0	0
3	B	22	0	0	0	0
3	C	22	0	0	2	0
3	D	28	0	0	0	0
3	E	29	0	0	0	0
3	F	22	0	0	0	0
3	G	30	0	0	0	0
All	All	1855	0	1750	20	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 6.

All (20) 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)
2:A:202:HEZ:H12	2:F:201:HEZ:H11	1.36	1.06
1:C:17:ASN:HD21	2:D:203:HEZ:H11	1.44	0.81
1:C:17:ASN:ND2	2:D:203:HEZ:H11	1.97	0.80
2:C:206:HEZ:H11	3:C:220:HOH:O	1.89	0.71
2:A:202:HEZ:C1	2:F:201:HEZ:H11	2.21	0.67
1:E:5:GLN:HG3	2:E:208:HEZ:H51	1.79	0.63
2:A:202:HEZ:H52	2:D:203:HEZ:H12	1.80	0.62
2:A:202:HEZ:O6	2:D:203:HEZ:H12	2.03	0.58
1:C:19:HIS:HE1	3:C:211:HOH:O	1.87	0.57
2:A:202:HEZ:C6	2:D:203:HEZ:H12	2.35	0.56
1:E:8:ASP:OD2	2:E:208:HEZ:H52	2.06	0.55
2:A:202:HEZ:H52	2:D:203:HEZ:H31	1.90	0.53
2:D:203:HEZ:H52	1:E:17:ASN:HD21	1.72	0.53
1:C:17:ASN:HD21	2:D:203:HEZ:C1	2.19	0.50
1:A:24:VAL:HG11	1:G:17[A]:ASN:OD1	2.12	0.49
2:A:202:HEZ:H52	2:D:203:HEZ:C1	2.44	0.48
2:A:202:HEZ:H12	2:F:201:HEZ:C1	2.26	0.48
2:A:202:HEZ:C5	2:D:203:HEZ:H12	2.42	0.47
1:C:4:LYS:NZ	2:C:206:HEZ:O1	2.45	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:5:GLN:HG3	2:E:208:HEZ:C5	2.48	0.42

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	30/34 (88%)	30 (100%)	0	0	100	100
1	B	29/34 (85%)	29 (100%)	0	0	100	100
1	C	31/34 (91%)	31 (100%)	0	0	100	100
1	D	31/34 (91%)	31 (100%)	0	0	100	100
1	E	30/34 (88%)	30 (100%)	0	0	100	100
1	F	31/34 (91%)	31 (100%)	0	0	100	100
1	G	29/34 (85%)	29 (100%)	0	0	100	100
All	All	211/238 (89%)	211 (100%)	0	0	100	100

There are no Ramachandran outliers to report.

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	22/24 (92%)	22 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	21/24 (88%)	21 (100%)	0	100	100
1	C	23/24 (96%)	23 (100%)	0	100	100
1	D	23/24 (96%)	23 (100%)	0	100	100
1	E	22/24 (92%)	22 (100%)	0	100	100
1	F	23/24 (96%)	22 (96%)	1 (4%)	29	2
1	G	21/24 (88%)	21 (100%)	0	100	100
All	All	155/168 (92%)	154 (99%)	1 (1%)	84	62

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	4	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	19	HIS

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)

8 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	HEZ	C	206	-	7,7,7	0.28	0	6,6,6	0.73	0
2	HEZ	D	203	-	7,7,7	0.38	0	6,6,6	0.40	0
2	HEZ	D	207	-	7,7,7	0.30	0	6,6,6	0.50	0
2	HEZ	F	205	-	7,7,7	0.30	0	6,6,6	0.51	0
2	HEZ	E	208	-	7,7,7	0.32	0	6,6,6	0.50	0
2	HEZ	A	202	-	7,7,7	0.42	0	6,6,6	0.79	0
2	HEZ	G	204	-	7,7,7	0.29	0	6,6,6	0.52	0
2	HEZ	F	201	-	7,7,7	0.32	0	6,6,6	0.52	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	HEZ	C	206	-	-	0/5/5/5	-
2	HEZ	D	203	-	-	3/5/5/5	-
2	HEZ	D	207	-	-	0/5/5/5	-
2	HEZ	F	205	-	-	4/5/5/5	-
2	HEZ	E	208	-	-	3/5/5/5	-
2	HEZ	A	202	-	-	3/5/5/5	-
2	HEZ	G	204	-	-	3/5/5/5	-
2	HEZ	F	201	-	-	2/5/5/5	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (18) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	F	205	HEZ	C2-C3-C4-C5
2	F	201	HEZ	C3-C4-C5-C6
2	E	208	HEZ	C3-C4-C5-C6
2	G	204	HEZ	C1-C2-C3-C4
2	A	202	HEZ	C3-C4-C5-C6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	F	201	HEZ	C1-C2-C3-C4
2	F	205	HEZ	C3-C4-C5-C6
2	A	202	HEZ	C1-C2-C3-C4
2	D	203	HEZ	O1-C1-C2-C3
2	F	205	HEZ	O1-C1-C2-C3
2	G	204	HEZ	C2-C3-C4-C5
2	E	208	HEZ	C2-C3-C4-C5
2	D	203	HEZ	C2-C3-C4-C5
2	F	205	HEZ	C4-C5-C6-O6
2	A	202	HEZ	C2-C3-C4-C5
2	G	204	HEZ	O1-C1-C2-C3
2	D	203	HEZ	C3-C4-C5-C6
2	E	208	HEZ	C4-C5-C6-O6

There are no ring outliers.

5 monomers are involved in 18 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	206	HEZ	2	0
2	D	203	HEZ	10	0
2	E	208	HEZ	3	0
2	A	202	HEZ	9	0
2	F	201	HEZ	3	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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.