



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:58 PM GMT

PDB ID : 4GSF
Title : The structure analysis of cysteine free insulin degrading enzyme (ide) with (s)-2-{2-[carboxymethyl-(3-phenyl-propionyl)-amino]-acetylamino}-3-(3h-imidazol-4-yl)-propionic acid methyl ester
Authors : Guo, Q.; Deprez-Poulain, R.; Deprez, B.; Tang, W.J.
Deposited on : 2012-08-27
Resolution : 2.70 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 ⓘ 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 (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

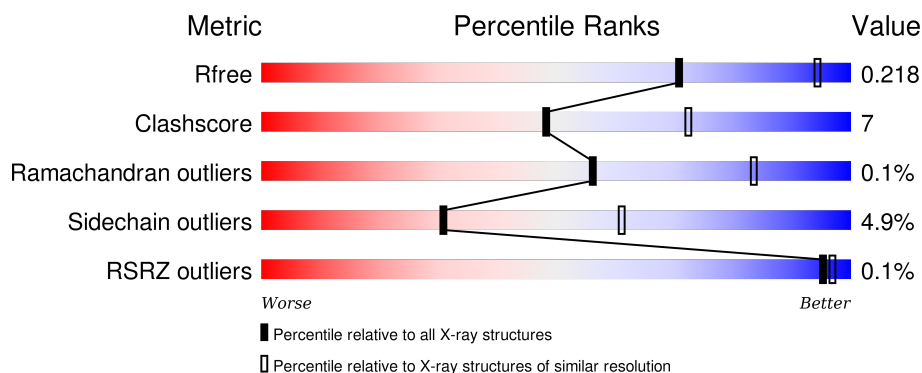
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION



The reported resolution of this entry is 2.70 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	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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	990	 80% 15% . .
1	B	990	 78% 17% . .

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	MGH	A	1101	-	-	-	X
2	MGH	B	1101	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15908 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-degrading enzyme.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	956	Total	C	N	O	S	0	0	0
			7768	5008	1307	1431	22			
1	B	954	Total	C	N	O	S	0	0	0
			7758	5002	1305	1429	22			

There are 52 discrepancies between the modelled and reference sequences:

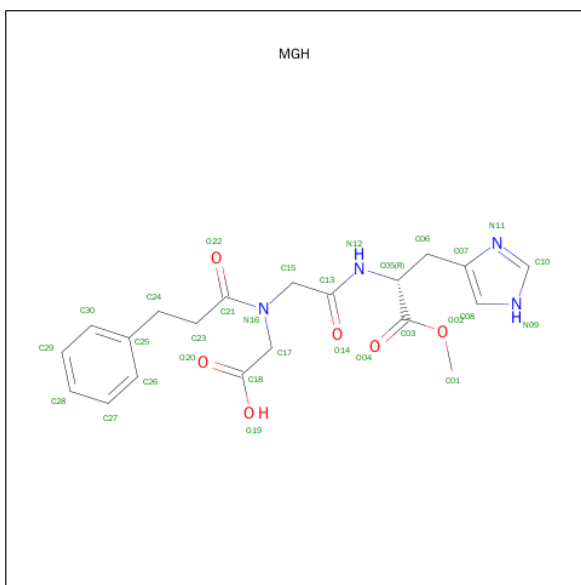
Chain	Residue	Modelled	Actual	Comment	Reference
A	30	MET	-	EXPRESSION TAG	UNP P14735
A	31	HIS	-	EXPRESSION TAG	UNP P14735
A	32	HIS	-	EXPRESSION TAG	UNP P14735
A	33	HIS	-	EXPRESSION TAG	UNP P14735
A	34	HIS	-	EXPRESSION TAG	UNP P14735
A	35	HIS	-	EXPRESSION TAG	UNP P14735
A	36	HIS	-	EXPRESSION TAG	UNP P14735
A	37	ALA	-	EXPRESSION TAG	UNP P14735
A	38	ALA	-	EXPRESSION TAG	UNP P14735
A	39	GLY	-	EXPRESSION TAG	UNP P14735
A	40	ILE	-	EXPRESSION TAG	UNP P14735
A	41	PRO	-	EXPRESSION TAG	UNP P14735
A	110	LEU	CYS	ENGINEERED MUTATION	UNP P14735
A	111	GLN	GLU	ENGINEERED MUTATION	UNP P14735
A	171	SER	CYS	ENGINEERED MUTATION	UNP P14735
A	178	ALA	CYS	ENGINEERED MUTATION	UNP P14735
A	257	VAL	CYS	ENGINEERED MUTATION	UNP P14735
A	414	LEU	CYS	ENGINEERED MUTATION	UNP P14735
A	573	ASN	CYS	ENGINEERED MUTATION	UNP P14735
A	590	SER	CYS	ENGINEERED MUTATION	UNP P14735
A	789	SER	CYS	ENGINEERED MUTATION	UNP P14735
A	812	ALA	CYS	ENGINEERED MUTATION	UNP P14735
A	819	ALA	CYS	ENGINEERED MUTATION	UNP P14735
A	904	SER	CYS	ENGINEERED MUTATION	UNP P14735
A	966	ASN	CYS	ENGINEERED MUTATION	UNP P14735

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	974	ALA	CYS	ENGINEERED MUTATION	UNP P14735
B	30	MET	-	EXPRESSION TAG	UNP P14735
B	31	HIS	-	EXPRESSION TAG	UNP P14735
B	32	HIS	-	EXPRESSION TAG	UNP P14735
B	33	HIS	-	EXPRESSION TAG	UNP P14735
B	34	HIS	-	EXPRESSION TAG	UNP P14735
B	35	HIS	-	EXPRESSION TAG	UNP P14735
B	36	HIS	-	EXPRESSION TAG	UNP P14735
B	37	ALA	-	EXPRESSION TAG	UNP P14735
B	38	ALA	-	EXPRESSION TAG	UNP P14735
B	39	GLY	-	EXPRESSION TAG	UNP P14735
B	40	ILE	-	EXPRESSION TAG	UNP P14735
B	41	PRO	-	EXPRESSION TAG	UNP P14735
B	110	LEU	CYS	ENGINEERED MUTATION	UNP P14735
B	111	GLN	GLU	ENGINEERED MUTATION	UNP P14735
B	171	SER	CYS	ENGINEERED MUTATION	UNP P14735
B	178	ALA	CYS	ENGINEERED MUTATION	UNP P14735
B	257	VAL	CYS	ENGINEERED MUTATION	UNP P14735
B	414	LEU	CYS	ENGINEERED MUTATION	UNP P14735
B	573	ASN	CYS	ENGINEERED MUTATION	UNP P14735
B	590	SER	CYS	ENGINEERED MUTATION	UNP P14735
B	789	SER	CYS	ENGINEERED MUTATION	UNP P14735
B	812	ALA	CYS	ENGINEERED MUTATION	UNP P14735
B	819	ALA	CYS	ENGINEERED MUTATION	UNP P14735
B	904	SER	CYS	ENGINEERED MUTATION	UNP P14735
B	966	ASN	CYS	ENGINEERED MUTATION	UNP P14735
B	974	ALA	CYS	ENGINEERED MUTATION	UNP P14735

- Molecule 2 is METHYL N-(CARBOXYMETHYL)-N-(3-PHENYLPROPANOYL)GLYCYL-D-HISTIDINATE (three-letter code: MGH) (formula: C₂₀H₂₄N₄O₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			30	20	4	6		
2	B	1	Total	C	N	O	0	0
			30	20	4	6		

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

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

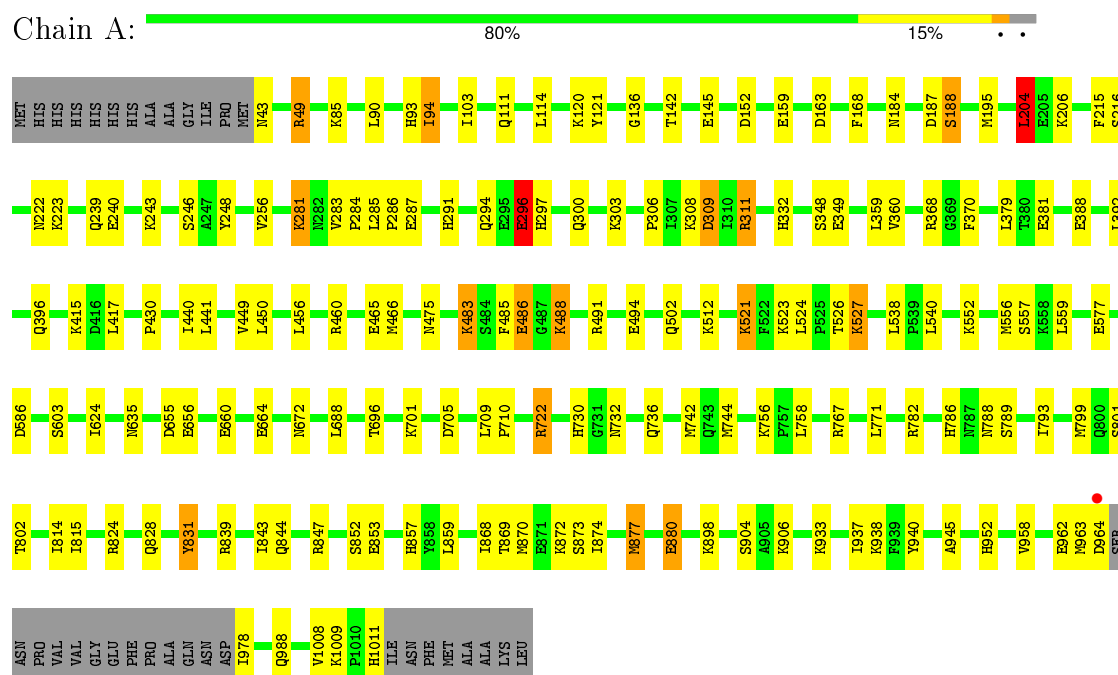
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	183	Total	O	0	0
			183	183		
4	B	137	Total	O	0	0
			137	137		

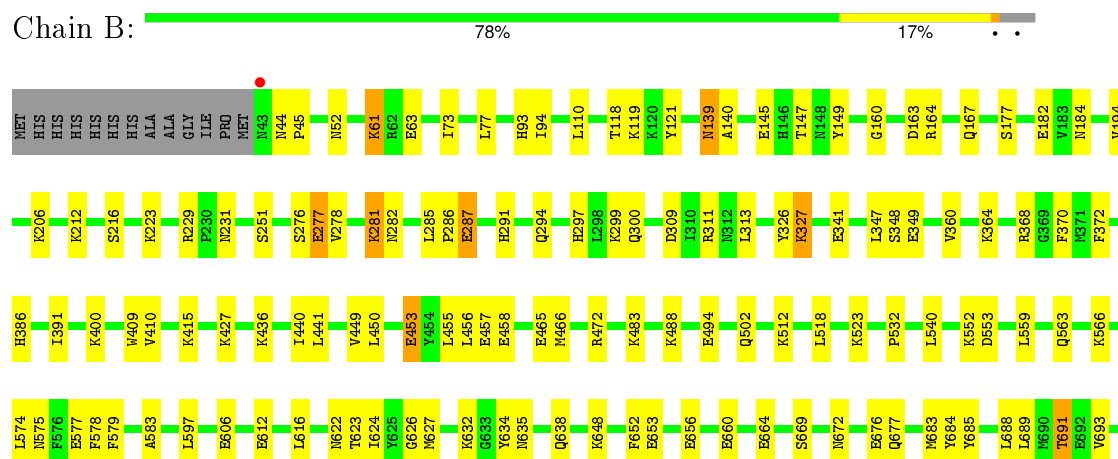
3 Residue-property plots

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: Insulin-degrading enzyme



- Molecule 1: Insulin-degrading enzyme



T696	K697	E702	T708	L709	P710	L725	E726	A727	L728	L729	N732	L733	M742	T748	E768	M776	E792	T797	D798	L810	M841	G842	L843	G844	G845	L846	R847	S852	E853	K854	L859	L867	E876	E880	K896	P897	K898	S913	Q914				
K933	I936	I937	K941	R951	H952	K953	V956	A960	M963	D964	SER	ASN	PRO	VAL	VAL	GLY	GLU	PHE	PRO	ALA	ALA	GLN	ASN	ASP	ILE	ASN	L980	P989	Q993	H994	M995	V1008	K1009	P1010	H1011	ILE	ASN	PHE	MET	ALA	ALA	LYS	LEU

4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α , β , γ	263.18 Å 263.18 Å 90.56 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 2.70 49.74 – 2.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (50.00-2.70) 100.0 (49.74-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.41 (at 2.69 Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.174 , 0.220 0.172 , 0.218	Depositor DCC
R_{free} test set	4911 reflections (5.25%)	DCC
Wilson B-factor (Å ²)	45.7	Xtriage
Anisotropy	0.028	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 40.2	EDS
Estimated twinning fraction	0.026 for h,-h-k,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 98494 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	15908	wwPDB-VP
Average B, all atoms (Å ²)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.78% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: ZN, MGH

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	1.12	14/7963 (0.2%)	0.98	10/10779 (0.1%)
1	B	1.10	16/7953 (0.2%)	0.95	7/10765 (0.1%)
All	All	1.11	30/15916 (0.2%)	0.97	17/21544 (0.1%)

The worst 5 of 30 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	204	LEU	CG-CD2	9.44	1.86	1.51
1	B	494	GLU	CG-CD	8.52	1.64	1.51
1	B	465	GLU	CG-CD	8.46	1.64	1.51
1	B	494	GLU	CB-CG	7.01	1.65	1.52
1	A	577	GLU	CG-CD	7.00	1.62	1.51

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	847	ARG	NE-CZ-NH2	-8.91	115.84	120.30
1	B	229	ARG	NE-CZ-NH2	8.31	124.46	120.30
1	A	767	ARG	NE-CZ-NH1	7.47	124.04	120.30
1	A	204	LEU	CB-CG-CD2	7.38	123.54	111.00
1	A	311	ARG	NE-CZ-NH2	-6.43	117.09	120.30

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	7768	0	7676	113	0
1	B	7758	0	7672	114	0
2	A	30	0	23	7	0
2	B	30	0	23	4	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	183	0	0	21	0
4	B	137	0	0	8	0
All	All	15908	0	15394	227	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 7.

The worst 5 of 227 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:204:LEU:CG	1:A:204:LEU:CD2	1.86	1.51
1:A:705:ASP:CB	4:A:1270:HOH:O	1.84	1.22
1:B:386:HIS:HE1	4:B:1215:HOH:O	1.49	0.94
1:B:309:ASP:H	1:B:672:ASN:HD21	1.15	0.94
1:A:978:ILE:N	4:A:1339:HOH:O	2.00	0.92

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	952/990 (96%)	915 (96%)	36 (4%)	1 (0%)	56 83
1	B	950/990 (96%)	911 (96%)	38 (4%)	1 (0%)	56 83

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1902/1980 (96%)	1826 (96%)	74 (4%)	2 (0%)	56 83

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	1010	PRO
1	A	103	ILE

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	837/879 (95%)	798 (95%)	39 (5%)	32 63
1	B	837/879 (95%)	794 (95%)	43 (5%)	29 59
All	All	1674/1758 (95%)	1592 (95%)	82 (5%)	31 61

5 of 82 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1009	LYS
1	B	139	ASN
1	B	841	ASN
1	A	1011	HIS
1	B	63	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	129	GLN
1	B	231	ASN
1	B	821	ASN
1	B	139	ASN
1	B	232	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](#)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	MGH	A	1101	-	24,31,31	2.00	3 (12%)	29,40,40	3.10	10 (34%)
2	MGH	B	1101	-	24,31,31	1.97	5 (20%)	29,40,40	2.36	8 (27%)

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	MGH	A	1101	-	-	0/29/31/31	0/2/2/2
2	MGH	B	1101	-	-	0/29/31/31	0/2/2/2

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1101	MGH	C26-C25	2.07	1.43	1.38

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1101	MGH	C30-C25	2.14	1.43	1.38
2	B	1101	MGH	C23-C21	2.24	1.56	1.51
2	A	1101	MGH	C15-N16	3.26	1.49	1.45
2	A	1101	MGH	C21-N16	5.09	1.46	1.35

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1101	MGH	C17-N16-C15	-8.10	104.29	116.54
2	A	1101	MGH	O02-C03-O04	-5.94	111.53	123.79
2	A	1101	MGH	C17-N16-C21	-5.36	108.86	121.74
2	B	1101	MGH	C17-N16-C15	-4.29	110.06	116.54
2	B	1101	MGH	O22-C21-N16	-3.96	113.58	122.08

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1101	MGH	7	0
2	B	1101	MGH	4	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, 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	956/990 (96%)	-0.43	1 (0%) 95 97	19, 35, 51, 75	0
1	B	954/990 (96%)	-0.39	1 (0%) 95 97	25, 39, 55, 82	0
All	All	1910/1980 (96%)	-0.41	2 (0%) 95 97	19, 37, 53, 82	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	43	ASN	2.6
1	A	964	ASP	2.6

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, 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	LLDF	B-factors(Å ²)	Q<0.9
2	MGH	B	1101	30/30	0.89	0.25	9.23	59,69,105,105	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
2	MGH	A	1101	30/30	0.85	0.28	7.15	49,78,97,97	0
3	ZN	B	1102	1/1	0.98	0.14	-	49,49,49,49	0
3	ZN	A	1102	1/1	0.98	0.13	-	44,44,44,44	0

6.5 Other polymers [i](#)

There are no such residues in this entry.