



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 8, 2018 – 08:18 PM EST

PDB ID : 5WDM
Title : An ultra-stable single-chain insulin analog resists thermal inactivation and exhibits biological signaling duration equivalent to the native protein
Authors : Yee, V.C.; Aldabbagh, K.; Peng, Y.
Deposited on : 2017-07-05
Resolution : 2.80 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
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)	:	rb-20030736

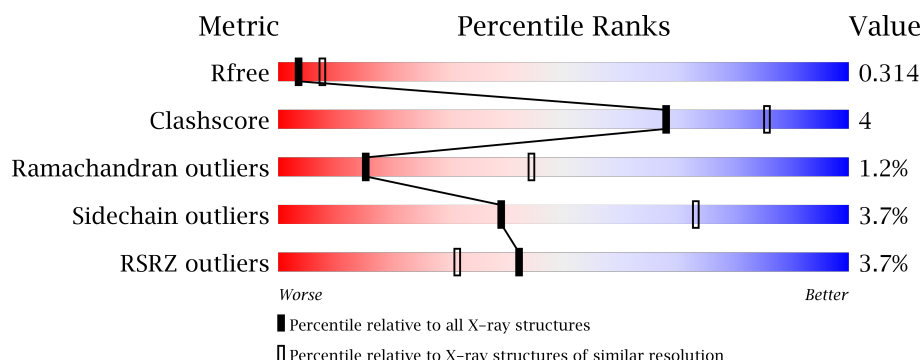
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (2.80-2.80)

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	57	<div> <div>7%</div> <div>63%</div> <div>14%</div> <div>21%</div> </div>
1	B	57	<div> <div>2%</div> <div>74%</div> <div>5%</div> <div>19%</div> </div>
1	C	57	<div> <div>70%</div> <div>5%</div> <div>23%</div> </div>
1	D	57	<div> <div>7%</div> <div>72%</div> <div>11%</div> <div>18%</div> </div>
1	E	57	<div> <div>2%</div> <div>68%</div> <div>12%</div> <div>18%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	57	 A horizontal bar chart showing the quality of chain F. The bar is divided into three segments: a green segment representing 72%, a yellow segment representing 2%, and a grey segment representing 23%. The green segment is the largest, followed by the grey segment, and the yellow segment is the smallest. The percentages are labeled below the bar: 72% under the green segment, 23% under the grey segment, and 2% under the yellow segment. There are two small black dots on the bar, one at the end of the green segment and one at the end of the yellow segment.

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 4114 atoms, of which 1959 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 Single-chain insulin analog.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	45	Total	C	H	N	O	S	0	0	0
			676	222	321	59	68	6			
1	B	46	Total	C	H	N	O	S	0	0	0
			690	227	328	60	69	6			
1	C	44	Total	C	H	N	O	S	0	0	0
			660	217	314	57	66	6			
1	D	47	Total	C	H	N	O	S	0	0	0
			714	233	341	64	70	6			
1	E	47	Total	C	H	N	O	S	0	0	0
			714	233	341	64	70	6			
1	F	44	Total	C	H	N	O	S	0	0	0
			660	217	314	57	66	6			

There are 150 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	29	GLU	LYS	conflict	UNP A6XGL2
A	?	-	ARG	deletion	UNP A6XGL2
A	?	-	ARG	deletion	UNP A6XGL2
A	?	-	GLU	deletion	UNP A6XGL2
A	?	-	ALA	deletion	UNP A6XGL2
A	?	-	GLU	deletion	UNP A6XGL2
A	?	-	ASP	deletion	UNP A6XGL2
A	?	-	LEU	deletion	UNP A6XGL2
A	?	-	GLN	deletion	UNP A6XGL2
A	?	-	GLY	deletion	UNP A6XGL2
A	?	-	SER	deletion	UNP A6XGL2
A	?	-	LEU	deletion	UNP A6XGL2
A	?	-	GLN	deletion	UNP A6XGL2
A	?	-	PRO	deletion	UNP A6XGL2
A	?	-	LEU	deletion	UNP A6XGL2
A	?	-	ALA	deletion	UNP A6XGL2
A	?	-	LEU	deletion	UNP A6XGL2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	GLU	deletion	UNP A6XGL2
A	31	GLU	GLY	conflict	UNP A6XGL2
A	32	GLU	SER	conflict	UNP A6XGL2
A	33	GLY	LEU	conflict	UNP A6XGL2
A	34	PRO	GLN	conflict	UNP A6XGL2
A	35	ARG	LYS	conflict	UNP A6XGL2
A	44	HIS	THR	conflict	UNP A6XGL2
A	50	GLU	TYR	conflict	UNP A6XGL2
B	29	GLU	LYS	conflict	UNP A6XGL2
B	?	-	ARG	deletion	UNP A6XGL2
B	?	-	ARG	deletion	UNP A6XGL2
B	?	-	GLU	deletion	UNP A6XGL2
B	?	-	ALA	deletion	UNP A6XGL2
B	?	-	GLU	deletion	UNP A6XGL2
B	?	-	ASP	deletion	UNP A6XGL2
B	?	-	LEU	deletion	UNP A6XGL2
B	?	-	GLN	deletion	UNP A6XGL2
B	?	-	GLY	deletion	UNP A6XGL2
B	?	-	SER	deletion	UNP A6XGL2
B	?	-	LEU	deletion	UNP A6XGL2
B	?	-	GLN	deletion	UNP A6XGL2
B	?	-	PRO	deletion	UNP A6XGL2
B	?	-	LEU	deletion	UNP A6XGL2
B	?	-	ALA	deletion	UNP A6XGL2
B	?	-	LEU	deletion	UNP A6XGL2
B	?	-	GLU	deletion	UNP A6XGL2
B	31	GLU	GLY	conflict	UNP A6XGL2
B	32	GLU	SER	conflict	UNP A6XGL2
B	33	GLY	LEU	conflict	UNP A6XGL2
B	34	PRO	GLN	conflict	UNP A6XGL2
B	35	ARG	LYS	conflict	UNP A6XGL2
B	44	HIS	THR	conflict	UNP A6XGL2
B	50	GLU	TYR	conflict	UNP A6XGL2
C	29	GLU	LYS	conflict	UNP A6XGL2
C	?	-	ARG	deletion	UNP A6XGL2
C	?	-	ARG	deletion	UNP A6XGL2
C	?	-	GLU	deletion	UNP A6XGL2
C	?	-	ALA	deletion	UNP A6XGL2
C	?	-	GLU	deletion	UNP A6XGL2
C	?	-	ASP	deletion	UNP A6XGL2
C	?	-	LEU	deletion	UNP A6XGL2
C	?	-	GLN	deletion	UNP A6XGL2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	?	-	GLY	deletion	UNP A6XGL2
C	?	-	SER	deletion	UNP A6XGL2
C	?	-	LEU	deletion	UNP A6XGL2
C	?	-	GLN	deletion	UNP A6XGL2
C	?	-	PRO	deletion	UNP A6XGL2
C	?	-	LEU	deletion	UNP A6XGL2
C	?	-	ALA	deletion	UNP A6XGL2
C	?	-	LEU	deletion	UNP A6XGL2
C	?	-	GLU	deletion	UNP A6XGL2
C	31	GLU	GLY	conflict	UNP A6XGL2
C	32	GLU	SER	conflict	UNP A6XGL2
C	33	GLY	LEU	conflict	UNP A6XGL2
C	34	PRO	GLN	conflict	UNP A6XGL2
C	35	ARG	LYS	conflict	UNP A6XGL2
C	44	HIS	THR	conflict	UNP A6XGL2
C	50	GLU	TYR	conflict	UNP A6XGL2
D	29	GLU	LYS	conflict	UNP A6XGL2
D	?	-	ARG	deletion	UNP A6XGL2
D	?	-	ARG	deletion	UNP A6XGL2
D	?	-	GLU	deletion	UNP A6XGL2
D	?	-	ALA	deletion	UNP A6XGL2
D	?	-	GLU	deletion	UNP A6XGL2
D	?	-	ASP	deletion	UNP A6XGL2
D	?	-	LEU	deletion	UNP A6XGL2
D	?	-	GLN	deletion	UNP A6XGL2
D	?	-	GLY	deletion	UNP A6XGL2
D	?	-	SER	deletion	UNP A6XGL2
D	?	-	LEU	deletion	UNP A6XGL2
D	?	-	GLN	deletion	UNP A6XGL2
D	?	-	PRO	deletion	UNP A6XGL2
D	?	-	LEU	deletion	UNP A6XGL2
D	?	-	ALA	deletion	UNP A6XGL2
D	?	-	LEU	deletion	UNP A6XGL2
D	?	-	GLU	deletion	UNP A6XGL2
D	31	GLU	GLY	conflict	UNP A6XGL2
D	32	GLU	SER	conflict	UNP A6XGL2
D	33	GLY	LEU	conflict	UNP A6XGL2
D	34	PRO	GLN	conflict	UNP A6XGL2
D	35	ARG	LYS	conflict	UNP A6XGL2
D	44	HIS	THR	conflict	UNP A6XGL2
D	50	GLU	TYR	conflict	UNP A6XGL2
E	29	GLU	LYS	conflict	UNP A6XGL2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	?	-	ARG	deletion	UNP A6XGL2
E	?	-	ARG	deletion	UNP A6XGL2
E	?	-	GLU	deletion	UNP A6XGL2
E	?	-	ALA	deletion	UNP A6XGL2
E	?	-	GLU	deletion	UNP A6XGL2
E	?	-	ASP	deletion	UNP A6XGL2
E	?	-	LEU	deletion	UNP A6XGL2
E	?	-	GLN	deletion	UNP A6XGL2
E	?	-	GLY	deletion	UNP A6XGL2
E	?	-	SER	deletion	UNP A6XGL2
E	?	-	LEU	deletion	UNP A6XGL2
E	?	-	GLN	deletion	UNP A6XGL2
E	?	-	PRO	deletion	UNP A6XGL2
E	?	-	LEU	deletion	UNP A6XGL2
E	?	-	ALA	deletion	UNP A6XGL2
E	?	-	LEU	deletion	UNP A6XGL2
E	?	-	GLU	deletion	UNP A6XGL2
E	31	GLU	GLY	conflict	UNP A6XGL2
E	32	GLU	SER	conflict	UNP A6XGL2
E	33	GLY	LEU	conflict	UNP A6XGL2
E	34	PRO	GLN	conflict	UNP A6XGL2
E	35	ARG	LYS	conflict	UNP A6XGL2
E	44	HIS	THR	conflict	UNP A6XGL2
E	50	GLU	TYR	conflict	UNP A6XGL2
F	29	GLU	LYS	conflict	UNP A6XGL2
F	?	-	ARG	deletion	UNP A6XGL2
F	?	-	ARG	deletion	UNP A6XGL2
F	?	-	GLU	deletion	UNP A6XGL2
F	?	-	ALA	deletion	UNP A6XGL2
F	?	-	GLU	deletion	UNP A6XGL2
F	?	-	ASP	deletion	UNP A6XGL2
F	?	-	LEU	deletion	UNP A6XGL2
F	?	-	GLN	deletion	UNP A6XGL2
F	?	-	GLY	deletion	UNP A6XGL2
F	?	-	SER	deletion	UNP A6XGL2
F	?	-	LEU	deletion	UNP A6XGL2
F	?	-	GLN	deletion	UNP A6XGL2
F	?	-	PRO	deletion	UNP A6XGL2
F	?	-	LEU	deletion	UNP A6XGL2
F	?	-	ALA	deletion	UNP A6XGL2
F	?	-	LEU	deletion	UNP A6XGL2
F	?	-	GLU	deletion	UNP A6XGL2

Continued on next page...

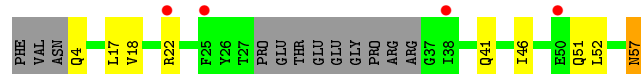
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	31	GLU	GLY	conflict	UNP A6XGL2
F	32	GLU	SER	conflict	UNP A6XGL2
F	33	GLY	LEU	conflict	UNP A6XGL2
F	34	PRO	GLN	conflict	UNP A6XGL2
F	35	ARG	LYS	conflict	UNP A6XGL2
F	44	HIS	THR	conflict	UNP A6XGL2
F	50	GLU	TYR	conflict	UNP A6XGL2

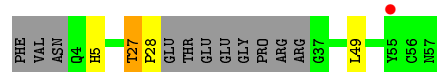
3 Residue-property plots [i](#)

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.

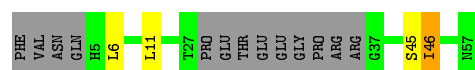
- Molecule 1: Single-chain insulin analog



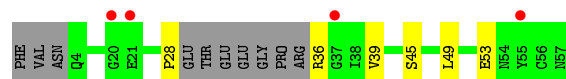
- Molecule 1: Single-chain insulin analog



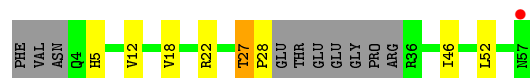
- Molecule 1: Single-chain insulin analog



- Molecule 1: Single-chain insulin analog

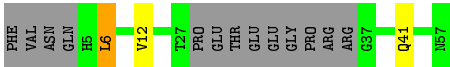


- Molecule 1: Single-chain insulin analog



- Molecule 1: Single-chain insulin analog





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	43.38Å 85.77Å 45.70Å 90.00° 110.89° 90.00°	Depositor
Resolution (Å)	38.22 – 2.80 38.22 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.2 (38.22-2.80) 99.2 (38.22-2.80)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.84 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.251 , 0.314 0.251 , 0.314	Depositor DCC
R_{free} test set	743 reflections (9.66%)	DCC
Wilson B-factor (Å ²)	63.1	Xtriage
Anisotropy	0.281	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 70.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	4114	wwPDB-VP
Average B, all atoms (Å ²)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.44% 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

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	0.36	0/361	0.50	0/484
1	B	0.36	0/369	0.57	0/496
1	C	0.39	0/352	0.57	0/472
1	D	0.39	0/380	0.58	0/510
1	E	0.34	0/380	0.56	0/510
1	F	0.37	0/352	0.52	0/472
All	All	0.37	0/2194	0.55	0/2944

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

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	355	321	321	6	0
1	B	362	328	328	2	0
1	C	346	314	313	3	0
1	D	373	341	341	2	0
1	E	373	341	341	5	0
1	F	346	314	313	2	0
All	All	2155	1959	1957	18	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 4.

The worst 5 of 18 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:18:VAL:O	1:A:22:ARG:NH2	2.30	0.64
1:D:49:LEU:O	1:D:53:GLU:N	2.28	0.63
1:A:41:GLN:NE2	1:A:51:GLN:OE1	2.32	0.62
1:E:18:VAL:O	1:E:22:ARG:NH2	2.33	0.62
1:A:4:GLN:HB3	1:A:46:ILE:HG23	1.88	0.56

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	41/57 (72%)	41 (100%)	0	0	100	100
1	B	42/57 (74%)	39 (93%)	1 (2%)	2 (5%)	2	8
1	C	40/57 (70%)	40 (100%)	0	0	100	100
1	D	43/57 (75%)	43 (100%)	0	0	100	100
1	E	43/57 (75%)	42 (98%)	0	1 (2%)	7	25
1	F	40/57 (70%)	40 (100%)	0	0	100	100
All	All	249/342 (73%)	245 (98%)	1 (0%)	3 (1%)	15	44

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	27	THR
1	B	5	HIS
1	B	49	LEU

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	40/51 (78%)	39 (98%)	1 (2%)	53	84
1	B	41/51 (80%)	40 (98%)	1 (2%)	54	85
1	C	39/51 (76%)	37 (95%)	2 (5%)	28	61
1	D	42/51 (82%)	40 (95%)	2 (5%)	30	63
1	E	42/51 (82%)	41 (98%)	1 (2%)	54	85
1	F	39/51 (76%)	37 (95%)	2 (5%)	28	61
All	All	243/306 (79%)	234 (96%)	9 (4%)	39	73

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	36	ARG
1	F	41	GLN
1	E	52	LEU
1	C	45	SER
1	D	45	SER

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

Mol	Chain	Res	Type
1	E	5	HIS

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 [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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	45/57 (78%)	0.77	4 (8%) 10 5	38, 83, 100, 113	0
1	B	46/57 (80%)	0.53	1 (2%) 62 52	38, 82, 106, 117	0
1	C	44/57 (77%)	0.13	0 100 100	39, 62, 79, 87	0
1	D	47/57 (82%)	0.57	4 (8%) 11 6	36, 70, 89, 108	0
1	E	47/57 (82%)	0.48	1 (2%) 64 54	52, 86, 103, 121	0
1	F	44/57 (77%)	0.27	0 100 100	34, 55, 79, 84	0
All	All	273/342 (79%)	0.46	10 (3%) 42 31	34, 69, 103, 121	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	57	ASN	4.0
1	A	25	PHE	3.7
1	A	22	ARG	3.3
1	D	20	GLY	3.0
1	D	21	GLU	2.4

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.