wwPDB X-ray Structure Validation Summary Report (i)

May 28, 2020 – 08:16 pm BST

PDB ID	:	1QJ0
Title	:	HUMAN INSULIN HEXAMERS WITH CHAIN B HIS MUTATED TO TYR
Authors	:	Tang, L.; Whittingham, J.L.; Verma, C.S.; Caves, L.S.D.; Dodson, G.G.
Deposited on	:	1999-06-18
Resolution	:	2.40 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:

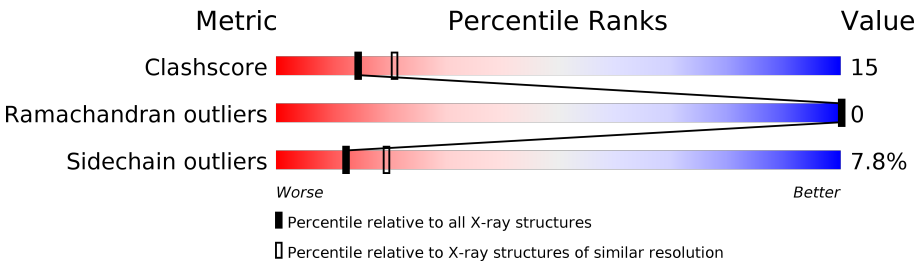
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.40 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	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)

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 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	21	48% 33% 10% 10%
1	C	21	62% 29% 10%
2	B	30	53% 23% 10% 7% 7%
2	D	30	50% 33% 13% •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 830 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 A CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	21	163	99	25	35	4	23	0	0
1	C	21	163	99	25	35	4	7	0	0

- Molecule 2 is a protein called INSULIN B CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	28	225	148	36	39	2	20	0	0
2	D	30	248	163	38	45	2	16	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	5	TYR	HIS	engineered mutation	UNP P01308
D	5	TYR	HIS	engineered mutation	UNP P01308

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

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

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

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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	4	4	4	0	0
			Total	O		
5	B	11	11	11	0	0
			Total	O		
5	C	3	3	3	0	0
			Total	O		
5	D	10	10	10	0	0

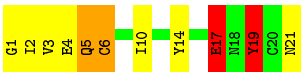
3 Residue-property plots (i)

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. 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 A CHAIN

Chain A: 48% 33% 10% 10%

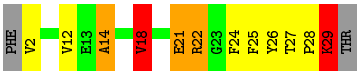


- Molecule 1: INSULIN A CHAIN

Chain C: 62% 29% 10%

- Molecule 2: INSULIN B CHAIN

Chain B: 53% 23% 10% 7% 7%



- Molecule 2: INSULIN B CHAIN

Chain D: 50% 33% 13%

4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	80.94Å 80.94Å 37.63Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	25.70 – 2.40	Depositor
% Data completeness (in resolution range)	97.0 (25.70-2.40)	Depositor
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, R_{free}	0.199 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	830	wwPDB-VP
Average B, all atoms ($Å^2$)	22.0	wwPDB-VP

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, CL

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	1.10	1/164 (0.6%)	1.86	7/220 (3.2%)
1	C	0.97	0/164	1.97	4/220 (1.8%)
2	B	1.46	2/231 (0.9%)	2.20	10/312 (3.2%)
2	D	0.89	0/260	2.13	15/350 (4.3%)
All	All	1.13	3/819 (0.4%)	2.07	36/1102 (3.3%)

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	4
1	C	0	3
2	B	1	2
2	D	0	1
All	All	1	10

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	21	GLU	CA-CB	-14.15	1.22	1.53
1	A	6	CYS	C-N	-9.22	1.12	1.34
2	B	29	LYS	CA-C	6.27	1.69	1.52

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	21	GLU	N-CA-CB	13.79	135.43	110.60
2	B	22	ARG	NE-CZ-NH1	-11.89	114.36	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	22	ARG	NE-CZ-NH2	-11.23	114.68	120.30
2	D	18	VAL	O-C-N	-10.44	105.99	122.70
1	A	5	GLN	N-CA-CB	-9.62	93.28	110.60

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	21	GLU	CA

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	17	GLU	Mainchain
1	A	19	TYR	Mainchain
1	A	5	GLN	Mainchain
1	A	6	CYS	Mainchain
2	B	21	GLU	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	163	0	148	6	0
1	C	163	0	149	2	0
2	B	225	0	215	14	0
2	D	248	0	236	8	0
3	B	1	0	0	0	0
3	D	1	0	0	0	0
4	B	1	0	0	0	0
5	A	4	0	0	0	0
5	B	11	0	0	3	0
5	C	3	0	0	0	0
5	D	10	0	0	0	1
All	All	830	0	748	22	1

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 15.

The worst 5 of 22 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:1:GLY:HA3	2:B:29:LYS:HE3	1.52	0.90
2:B:22:ARG:HD3	5:B:2009:HOH:O	1.81	0.81
2:B:12:VAL:HG12	2:D:12:VAL:HG23	1.65	0.77
2:B:12:VAL:CG1	2:D:12:VAL:HG23	2.23	0.67
2:D:17:LEU:O	2:D:17:LEU:HD12	1.95	0.66

All (1) 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)
5:D:2010:HOH:O	5:D:2010:HOH:O[3_555]	2.14	0.06

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	19/21 (90%)	18 (95%)	1 (5%)	0	100	100
1	C	19/21 (90%)	19 (100%)	0	0	100	100
2	B	26/30 (87%)	25 (96%)	1 (4%)	0	100	100
2	D	29/30 (97%)	28 (97%)	1 (3%)	0	100	100
All	All	93/102 (91%)	90 (97%)	3 (3%)	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	20/20 (100%)	19 (95%)	1 (5%)	24	40
1	C	20/20 (100%)	20 (100%)	0	100	100
2	B	24/26 (92%)	21 (88%)	3 (12%)	4	5
2	D	27/26 (104%)	24 (89%)	3 (11%)	6	8
All	All	91/92 (99%)	84 (92%)	7 (8%)	12	20

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	29	LYS
2	D	17	LEU
2	D	4	GLN
2	B	2	VAL
2	D	12	VAL

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

Mol	Chain	Res	Type
1	C	18	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	6:CYS	C	7:CYS	N	1.12

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.