Full wwPDB X-ray Structure Validation Report (i)

Aug 28, 2023 – 06:24 AM EDT

PDB ID	:	3JSD
Title	:	Insulin's biosynthesis and activity have opposing structural requirements: a new factor in neonatal diabetes mellitus
Authors	:	Weiss, M.A.; Wan, Z.L.; Dodson, E.J.; Liu, M.; Xu, B.; Hua, Q.X.; Turkenburg, M.; Whittingham, J.; Nakagawa, S.H.; Huang, K.; Hu, S.Q.; Jia, W.H.; Wang, S.H.; Brange, J.; Whittaker, J.; Arvan, P.; Katsoyannis, P.G.; Dodson, G.G.
Deposited on	:	2009-09-10
Resolution	:	2.50 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:

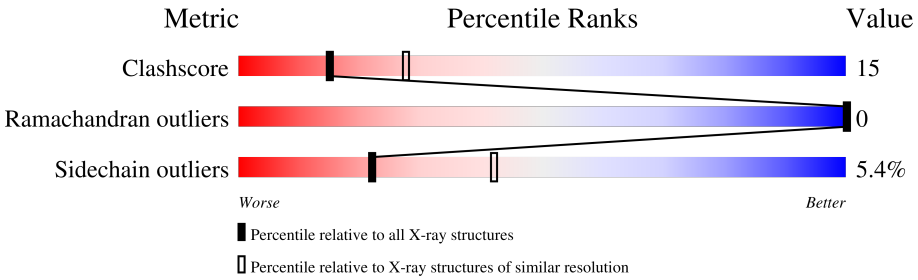
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

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 2.50 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	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)

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	21	67% 29% 5%
1	C	21	71% 24% 5%
2	B	30	87% 13%
2	D	30	83% 13% .

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	DAL	D	8	-	-	X	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 841 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
1	A	21	Total	C	N	O	S	0	0	0
			163	99	25	35	4			
1	C	21	Total	C	N	O	S	0	0	0
			163	99	25	35	4			

- Molecule 2 is a protein called Insulin B chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	30	Total	C	N	O	S	0	0	0
			243	159	40	42	2			
2	D	30	Total	C	N	O	S	0	0	0
			243	159	40	42	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	8	DAL	GLY	engineered mutation	UNP P01308
D	8	DAL	GLY	engineered mutation	UNP P01308

- 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	D	1	Total	Zn	0	0
			1	1		

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

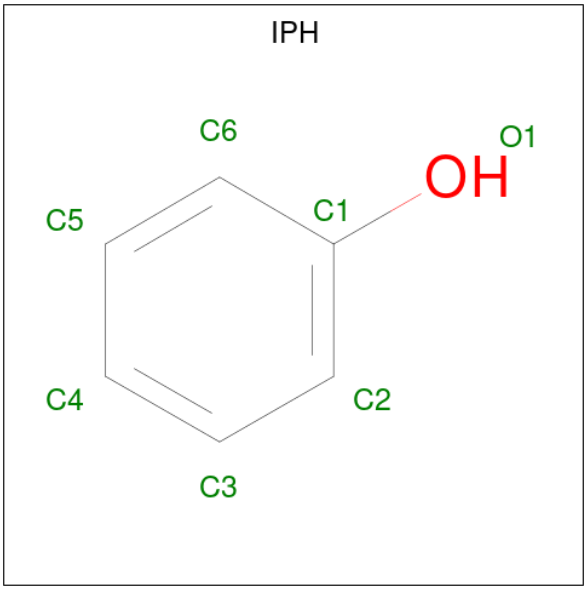
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Cl	0	0
			1	1		

Continued on next page...

Continued from previous page...

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

- Molecule 5 is PHENOL (three-letter code: IPH) (formula: C_6H_6O).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	1	Total C O 7 6 1	0	0

- Molecule 6 is water.

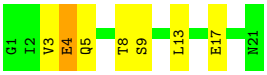
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	4	Total O 4 4	0	0
6	B	9	Total O 9 9	0	0
6	D	5	Total O 5 5	0	0

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.

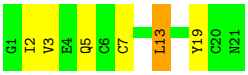
• Molecule 1: Insulin A chain

Chain A:



• Molecule 1: Insulin A chain

Chain C:

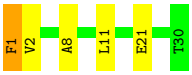


• Molecule 2: Insulin B chain

Chain B:

• Molecule 2: Insulin B chain

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	80.84Å 80.84Å 38.84Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	17.37 – 2.50 21.87 – 2.00	Depositor EDS
% Data completeness (in resolution range)	97.6 (17.37-2.50) 92.1 (21.87-2.00)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.30 (at 2.01Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.206 , 0.274 0.235 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	36.2	Xtriage
Anisotropy	0.393	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 54.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.039 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	841	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.66% 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

5.1 Standard geometry

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

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	0.51	0/164	0.60	0/220
1	C	0.39	0/164	0.61	0/220
2	B	0.51	0/244	0.75	0/327
2	D	0.51	0/244	0.63	0/327
All	All	0.49	0/816	0.66	0/1094

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	163	0	149	9	0
1	C	163	0	149	6	0
2	B	243	0	233	5	0
2	D	243	0	233	8	0
3	B	1	0	0	0	0
3	D	1	0	0	0	0
4	B	1	0	0	0	0
4	D	1	0	0	0	0
5	C	7	0	6	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	4	0	0	0	0
6	B	9	0	0	0	0
6	D	5	0	0	1	0
All	All	841	0	770	24	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 15.

All (24) 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:C:13:LEU:HD22	1:C:13:LEU:H	1.41	0.83
2:D:8:DAL:HB3	2:D:11:LEU:HB2	1.67	0.75
2:D:2:VAL:HG23	2:D:2:VAL:O	1.88	0.71
1:C:13:LEU:HD22	1:C:13:LEU:N	2.10	0.66
1:C:13:LEU:H	1:C:13:LEU:CD2	2.10	0.64
2:D:8:DAL:CB	2:D:11:LEU:HB2	2.27	0.63
2:D:8:DAL:CB	2:D:11:LEU:HD12	2.31	0.61
2:D:1:PHE:HB2	6:D:38:HOH:O	2.01	0.60
2:D:21:GLU:H	2:D:21:GLU:CD	2.06	0.59
1:A:17:GLU:HG2	2:B:18:VAL:HG11	1.85	0.57
1:A:13:LEU:O	1:A:17:GLU:HG3	2.07	0.55
1:A:4:GLU:HB2	2:B:30:THR:OXT	2.09	0.53
2:D:21:GLU:CD	2:D:21:GLU:N	2.65	0.50
2:D:8:DAL:HB2	2:D:11:LEU:HD12	1.94	0.49
2:B:29:LYS:O	2:B:30:THR:HB	2.12	0.49
1:A:3:VAL:HB	2:B:30:THR:HA	1.96	0.47
1:A:4:GLU:O	1:A:8:THR:HB	2.15	0.46
1:A:8:THR:O	1:A:8:THR:HG22	2.16	0.46
1:A:4:GLU:HA	1:A:8:THR:HB	1.98	0.44
1:C:2:ILE:HD12	1:C:19:TYR:CD2	2.53	0.44
1:A:17:GLU:HG2	2:B:18:VAL:CG1	2.47	0.44
1:C:3:VAL:O	1:C:7:CYS:HB2	2.19	0.43
1:A:5:GLN:O	1:A:9:SER:HB2	2.19	0.41
1:C:2:ILE:HD12	1:C:19:TYR:CG	2.57	0.41

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	19/21 (90%)	18 (95%)	1 (5%)	0	100	100
1	C	19/21 (90%)	18 (95%)	1 (5%)	0	100	100
2	B	27/30 (90%)	24 (89%)	3 (11%)	0	100	100
2	D	27/30 (90%)	24 (89%)	3 (11%)	0	100	100
All	All	92/102 (90%)	84 (91%)	8 (9%)	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	46
1	C	20/20 (100%)	18 (90%)	2 (10%)	7	15
2	B	26/26 (100%)	25 (96%)	1 (4%)	33	58
2	D	26/26 (100%)	25 (96%)	1 (4%)	33	58
All	All	92/92 (100%)	87 (95%)	5 (5%)	22	42

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	4	GLU
2	B	3	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	5	GLN
1	C	13	LEU
2	D	1	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	18	ASN
2	B	3	ASN
1	C	21	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

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.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
5	IPH	C	200	-	7,7,7	2.06	3 (42%)	8,8,8	1.75	2 (25%)

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
5	IPH	C	200	-	-	-	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	C	200	IPH	C5-C6	2.88	1.45	1.38
5	C	200	IPH	C2-C1	2.71	1.44	1.38
5	C	200	IPH	C3-C2	2.64	1.44	1.38

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	C	200	IPH	C6-C1-C2	3.43	125.54	119.77
5	C	200	IPH	C3-C2-C1	-2.43	115.88	119.31

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

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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

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

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

6.3 Carbohydrates

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

6.4 Ligands

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

6.5 Other polymers

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