



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 06:50 am BST

PDB ID : 5URU  
Title : Insulin with proline analog DhP at position B28 in the R6 state  
Authors : Lieblich, S.A.; Fang, K.Y.; Tirrell, D.A.  
Deposited on : 2017-02-13  
Resolution : 2.41 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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 : 2.11  
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) : 2.11

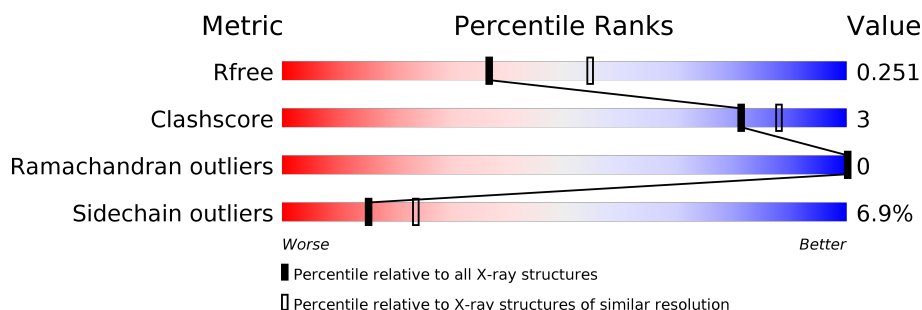
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.41 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	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)

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  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	21	90% 10%
1	C	21	95% 5%
1	E	21	95% 5%
1	G	21	95% 5%
2	B	30	77% 13% 7%
2	D	30	80% 7% 10%
2	F	30	73% 17% 10%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	H	30	 A horizontal bar chart showing the quality of chain H. The bar is divided into three segments: a long green segment representing 80%, a short yellow segment representing 10%, and a very short orange segment representing 7%. The segments are separated by thin black lines. The percentages are labeled below the bar: 80%, 10%, and 7%.

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 1462 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 Chain A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	21	Total	C	N	O	S	0	0	0
			151	91	25	31	4			
1	C	21	Total	C	N	O	S	0	0	0
			158	97	25	32	4			
1	E	21	Total	C	N	O	S	0	0	0
			139	84	23	28	4			
1	G	21	Total	C	N	O	S	0	0	0
			158	97	25	32	4			

- Molecule 2 is a protein called Insulin Chain B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	28	Total	C	N	O	S	0	0	0
			204	135	35	32	2			
2	D	27	Total	C	N	O	S	0	0	0
			206	135	35	34	2			
2	F	27	Total	C	N	O	S	0	0	0
			199	132	34	31	2			
2	H	28	Total	C	N	O	S	0	0	0
			205	135	34	34	2			

- Molecule 3 is PHENOL (three-letter code: IPH) (formula: C<sub>6</sub>H<sub>6</sub>O).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			7	6	1		
3	C	1	Total	C	O	0	0
			7	6	1		
3	E	1	Total	C	O	0	0
			7	6	1		
3	G	1	Total	C	O	0	0
			7	6	1		

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	H	1	Total	Zn	0	0
			1	1		
4	B	1	Total	Zn	0	0
			1	1		
4	D	1	Total	Zn	0	0
			1	1		
4	F	1	Total	Zn	0	0
			1	1		

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	1	Total	Cl	0	0
			1	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total 1	Cl 1	0	0
5	D	1	Total 1	Cl 1	0	0
5	F	1	Total 1	Cl 1	0	0


- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	1	Total 1	O 1	0	0
6	D	1	Total 1	O 1	0	0
6	E	1	Total 1	O 1	0	0
6	F	1	Total 1	O 1	0	0
6	H	2	Total 2	O 2	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.

- Molecule 1: Insulin Chain A

Chain A: 



- Molecule 1: Insulin Chain A

Chain C: 



- Molecule 1: Insulin Chain A

Chain E: 




- Molecule 1: Insulin Chain A

Chain G: 




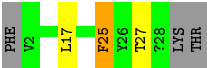
- Molecule 2: Insulin Chain B

Chain B: 



- Molecule 2: Insulin Chain B

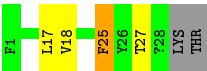
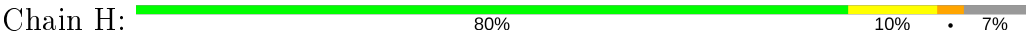
Chain D: 



- Molecule 2: Insulin Chain B



- Molecule 2: Insulin Chain B





## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.26 Å 79.26 Å 79.87 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	52.06 – 2.41 34.52 – 1.68	Depositor EDS
% Data completeness (in resolution range)	99.8 (52.06-2.41) 84.9 (34.52-1.68)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.18 (at 1.68 Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.187 , 0.246 0.200 , 0.251	Depositor DCC
$R_{free}$ test set	874 reflections (4.83%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	26.0	Xtriage
Anisotropy	0.120	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 58.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.071 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1462	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 62.32 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.1466e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, CL, IPH, 8LJ

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	0.87	0/151	0.88	0/204
1	C	0.87	0/159	0.82	0/215
1	E	0.79	0/139	0.74	0/189
1	G	0.79	0/159	0.87	0/215
2	B	1.29	3/202 (1.5%)	1.02	0/274
2	D	1.34	1/204 (0.5%)	1.19	1/276 (0.4%)
2	F	1.32	2/197 (1.0%)	0.96	0/267
2	H	1.22	1/203 (0.5%)	1.09	1/276 (0.4%)
All	All	1.12	7/1414 (0.5%)	0.97	2/1916 (0.1%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	25	PHE	CB-CG	-10.19	1.34	1.51
2	H	25	PHE	CB-CG	-8.40	1.37	1.51
2	F	25	PHE	CG-CD2	-6.67	1.28	1.38
2	F	25	PHE	CB-CG	-6.47	1.40	1.51
2	B	24	PHE	CG-CD1	-6.33	1.29	1.38

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	25	PHE	CB-CA-C	-10.24	89.92	110.40
2	H	25	PHE	CB-CA-C	-9.02	92.37	110.40

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	151	0	138	1	0
1	C	158	0	145	1	0
1	E	139	0	114	1	0
1	G	158	0	145	1	0
2	B	204	0	180	2	0
2	D	206	0	183	1	0
2	F	199	0	175	1	0
2	H	205	0	179	2	0
3	A	7	0	6	0	0
3	C	7	0	5	1	0
3	E	7	0	5	0	0
3	G	7	0	5	0	0
4	B	1	0	0	0	0
4	D	1	0	0	0	0
4	F	1	0	0	0	0
4	H	1	0	0	0	0
5	B	1	0	0	0	0
5	D	1	0	0	0	0
5	F	1	0	0	0	0
5	H	1	0	0	0	0
6	C	1	0	0	0	0
6	D	1	0	0	0	0
6	E	1	0	0	0	0
6	F	1	0	0	0	0
6	H	2	0	0	0	0
All	All	1462	0	1280	7	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 3.

The worst 5 of 7 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:B:25:PHE:CD1	2:B:25:PHE:N	2.69	0.59
2:H:25:PHE:CD1	2:H:25:PHE:N	2.80	0.49

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:G:13:LEU:HD22	2:H:18:VAL:HG22	1.98	0.45
1:C:6:CYS:O	3:C:101:IPH:O1	2.38	0.42
2:D:25:PHE:CD1	2:D:25:PHE:N	2.88	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	19/21 (90%)	19 (100%)	0	0	100	100
1	C	19/21 (90%)	19 (100%)	0	0	100	100
1	E	19/21 (90%)	18 (95%)	1 (5%)	0	100	100
1	G	19/21 (90%)	19 (100%)	0	0	100	100
2	B	26/30 (87%)	26 (100%)	0	0	100	100
2	D	25/30 (83%)	25 (100%)	0	0	100	100
2	F	25/30 (83%)	25 (100%)	0	0	100	100
2	H	26/30 (87%)	26 (100%)	0	0	100	100
All	All	178/204 (87%)	177 (99%)	1 (1%)	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	18/20 (90%)	17 (94%)	1 (6%)	21	32
1	C	19/20 (95%)	19 (100%)	0	100	100
1	E	14/20 (70%)	14 (100%)	0	100	100
1	G	19/20 (95%)	19 (100%)	0	100	100
2	B	18/25 (72%)	16 (89%)	2 (11%)	6	8
2	D	20/25 (80%)	18 (90%)	2 (10%)	7	10
2	F	18/25 (72%)	15 (83%)	3 (17%)	2	2
2	H	19/25 (76%)	17 (90%)	2 (10%)	7	9
All	All	145/180 (81%)	135 (93%)	10 (7%)	15	24

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	27	THR
2	F	6	LEU
2	F	27	THR
2	D	17	LEU
2	F	17	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues 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	8LJ	H	28	2	4,7,8	2.89	2 (50%)	3,8,10	1.95	1 (33%)
2	8LJ	B	28	2	4,7,8	3.31	3 (75%)	3,8,10	2.79	1 (33%)
2	8LJ	D	28	2	4,7,8	3.15	3 (75%)	3,8,10	2.44	1 (33%)
2	8LJ	F	28	2	4,7,8	3.11	3 (75%)	3,8,10	2.64	1 (33%)

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	8LJ	H	28	2	-	0/0/9/11	0/1/1/1
2	8LJ	B	28	2	-	0/0/9/11	0/1/1/1
2	8LJ	D	28	2	-	0/0/9/11	0/1/1/1
2	8LJ	F	28	2	-	0/0/9/11	0/1/1/1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	28	8LJ	CB-CG	-4.37	1.31	1.48
2	D	28	8LJ	CB-CG	-4.34	1.31	1.48
2	B	28	8LJ	CB-CG	-4.14	1.32	1.48
2	F	28	8LJ	CB-CG	-4.02	1.33	1.48
2	B	28	8LJ	CA-N	3.54	1.50	1.46

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	28	8LJ	CA-CB-CG	4.49	112.67	103.00
2	F	28	8LJ	CA-CB-CG	4.01	111.65	103.00
2	D	28	8LJ	CA-CB-CG	3.90	111.40	103.00
2	H	28	8LJ	CA-CB-CG	2.93	109.31	103.00

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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 12 ligands modelled in this entry, 8 are monoatomic - leaving 4 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	$\# Z  > 2$
3	IPH	E	101	-	7,7,7	0.45	0	8,8,8	0.36	0
3	IPH	G	101	-	7,7,7	0.47	0	8,8,8	0.51	0
3	IPH	A	101	-	7,7,7	0.47	0	8,8,8	0.46	0
3	IPH	C	101	-	7,7,7	0.66	0	8,8,8	0.30	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
3	IPH	E	101	-	-	-	0/1/1/1
3	IPH	G	101	-	-	-	0/1/1/1
3	IPH	A	101	-	-	-	0/1/1/1
3	IPH	C	101	-	-	-	0/1/1/1

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	101	IPH	1	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 ⓘ

### 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.