



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2016 – 11:42 PM EDT

PDB ID : 5HP5  
Title : Srtucture of human peptidylarginine deiminase type I (PAD1)  
Authors : Unno, M.; Nagai, A.; Saijo, S.; Shimizu, N.; Kinjo, S.; Mashimo, R.; Kizawa, K.; Takahara, H.  
Deposited on : 2016-01-20  
Resolution : 3.20 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

<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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
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)	:	rb-20027939

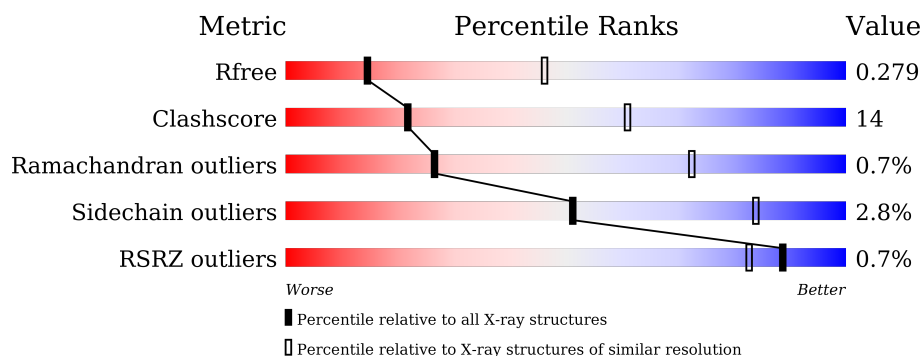
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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. 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	684	<div> <div></div> <div>66%</div> <div>29%</div> <div>..</div> </div>
1	B	684	<div> <div></div> <div>68%</div> <div>27%</div> <div>...</div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 10573 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 Protein-arginine deiminase type-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	669	Total	C	N	O	S	0	0	0
			5274	3361	903	980	30			
1	B	669	Total	C	N	O	S	0	0	0
			5274	3361	903	980	30			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	expression tag	UNP Q9ULC6
A	-19	GLY	-	expression tag	UNP Q9ULC6
A	-18	HIS	-	expression tag	UNP Q9ULC6
A	-17	HIS	-	expression tag	UNP Q9ULC6
A	-16	HIS	-	expression tag	UNP Q9ULC6
A	-15	HIS	-	expression tag	UNP Q9ULC6
A	-14	HIS	-	expression tag	UNP Q9ULC6
A	-13	HIS	-	expression tag	UNP Q9ULC6
A	-12	HIS	-	expression tag	UNP Q9ULC6
A	-11	HIS	-	expression tag	UNP Q9ULC6
A	-10	HIS	-	expression tag	UNP Q9ULC6
A	-9	HIS	-	expression tag	UNP Q9ULC6
A	-8	SER	-	expression tag	UNP Q9ULC6
A	-7	SER	-	expression tag	UNP Q9ULC6
A	-6	GLY	-	expression tag	UNP Q9ULC6
A	-5	HIS	-	expression tag	UNP Q9ULC6
A	-4	ILE	-	expression tag	UNP Q9ULC6
A	-3	GLU	-	expression tag	UNP Q9ULC6
A	-2	GLY	-	expression tag	UNP Q9ULC6
A	-1	ARG	-	expression tag	UNP Q9ULC6
A	0	HIS	-	expression tag	UNP Q9ULC6
B	-20	MET	-	expression tag	UNP Q9ULC6
B	-19	GLY	-	expression tag	UNP Q9ULC6
B	-18	HIS	-	expression tag	UNP Q9ULC6
B	-17	HIS	-	expression tag	UNP Q9ULC6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	HIS	-	expression tag	UNP Q9ULC6
B	-15	HIS	-	expression tag	UNP Q9ULC6
B	-14	HIS	-	expression tag	UNP Q9ULC6
B	-13	HIS	-	expression tag	UNP Q9ULC6
B	-12	HIS	-	expression tag	UNP Q9ULC6
B	-11	HIS	-	expression tag	UNP Q9ULC6
B	-10	HIS	-	expression tag	UNP Q9ULC6
B	-9	HIS	-	expression tag	UNP Q9ULC6
B	-8	SER	-	expression tag	UNP Q9ULC6
B	-7	SER	-	expression tag	UNP Q9ULC6
B	-6	GLY	-	expression tag	UNP Q9ULC6
B	-5	HIS	-	expression tag	UNP Q9ULC6
B	-4	ILE	-	expression tag	UNP Q9ULC6
B	-3	GLU	-	expression tag	UNP Q9ULC6
B	-2	GLY	-	expression tag	UNP Q9ULC6
B	-1	ARG	-	expression tag	UNP Q9ULC6
B	0	HIS	-	expression tag	UNP Q9ULC6

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	4	Total Ca 4 4	0	0
2	A	4	Total Ca 4 4	0	0

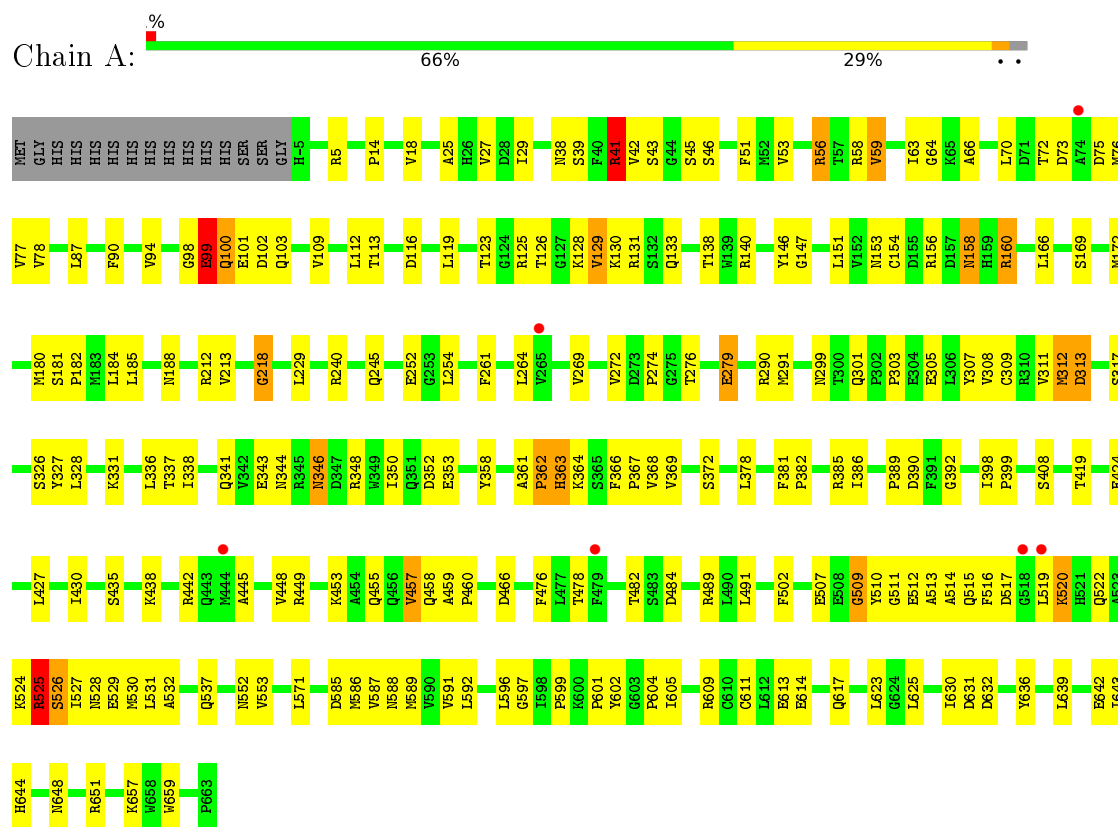
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	8	Total O 8 8	0	0
3	B	9	Total O 9 9	0	0

### 3 Residue-property plots

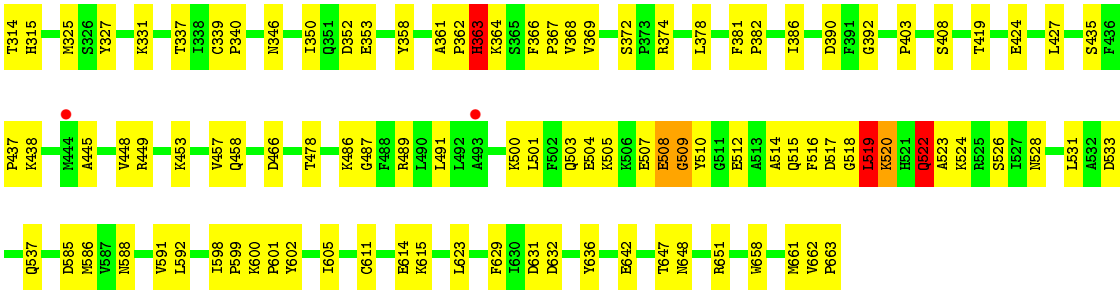
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Protein-arginine deiminase type-1



- Molecule 1: Protein-arginine deiminase type-1





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.81Å 90.81Å 373.05Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.73 – 3.20 48.77 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.73-3.20) 99.9 (48.77-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.02 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.213 , 0.278 0.216 , 0.279	Depositor DCC
$R_{free}$ test set	1440 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	85.2	Xtriage
Anisotropy	0.297	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 30.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.477 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	10573	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.66% of the height of the origin peak. No significant pseudotranslation is detected.*

<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: CA

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	0.33	0/5408	0.65	5/7336 (0.1%)
1	B	0.32	0/5408	0.65	6/7336 (0.1%)
All	All	0.32	0/10816	0.65	11/14672 (0.1%)

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	11
1	B	0	12
All	All	0	23

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	522	GLN	CA-CB-CG	9.52	134.34	113.40
1	B	218	GLY	N-CA-C	-8.46	91.96	113.10
1	A	218	GLY	N-CA-C	-7.55	94.22	113.10
1	B	509	GLY	N-CA-C	7.46	131.74	113.10
1	A	41	ARG	NE-CZ-NH2	-6.28	117.16	120.30

There are no chirality outliers.

5 of 23 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	126	THR	Peptide

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	130	LYS	Peptide
1	A	133	GLN	Peptide
1	A	218	GLY	Peptide
1	A	312	MET	Peptide

## 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	5274	0	5174	162	0
1	B	5274	0	5174	130	1
2	A	4	0	0	0	0
2	B	4	0	0	0	0
3	A	8	0	0	2	0
3	B	9	0	0	1	0
All	All	10573	0	10348	292	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 14.

The worst 5 of 292 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:B:99:GLU:HA	1:B:101:GLU:HG2	1.31	1.10
1:B:101:GLU:HB2	1:B:102:ASP:HA	1.41	1.02
1:A:56:ARG:HH12	1:A:59:VAL:HG22	1.28	0.95
1:A:312:MET:SD	1:A:313:ASP:HA	2.10	0.92
1:A:515:GLN:N	1:A:525:ARG:O	2.09	0.84

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)
1:B:99:GLU:OE2	1:B:212:ARG:NH1[5_664]	2.12	0.08

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	667/684 (98%)	630 (94%)	33 (5%)	4 (1%)	30	75
1	B	667/684 (98%)	628 (94%)	34 (5%)	5 (1%)	26	72
All	All	1334/1368 (98%)	1258 (94%)	67 (5%)	9 (1%)	26	72

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	458	GLN
1	A	510	TYR
1	A	457	VAL
1	B	100	GLN
1	B	520	LYS

### 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	581/601 (97%)	562 (97%)	19 (3%)	45	81
1	B	581/601 (97%)	568 (98%)	13 (2%)	60	87
All	All	1162/1202 (97%)	1130 (97%)	32 (3%)	51	84

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	507	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	525	ARG
1	B	363	HIS
1	A	524	LYS
1	A	526	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	30	HIS
1	B	528	ASN
1	B	346	ASN
1	A	341	GLN
1	B	341	GLN

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 8 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](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	669/684 (97%)	0.12	6 (0%)	85 78	53, 78, 115, 134	0
1	B	669/684 (97%)	0.09	3 (0%)	93 90	53, 79, 113, 140	0
All	All	1338/1368 (97%)	0.11	9 (0%)	89 83	53, 78, 115, 140	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	444	MET	5.0
1	A	444	MET	4.9
1	A	518	GLY	4.6
1	B	106	GLY	2.7
1	A	479	PHE	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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
2	CA	B	701	1/1	0.98	0.22	-0.05	68,68,68,68	0
2	CA	A	702	1/1	0.92	0.18	-0.25	64,64,64,64	0
2	CA	A	701	1/1	0.97	0.18	-0.63	64,64,64,64	0
2	CA	B	703	1/1	1.00	0.17	-1.16	59,59,59,59	0
2	CA	A	703	1/1	0.99	0.14	-1.50	65,65,65,65	0
2	CA	B	702	1/1	0.95	0.15	-1.74	60,60,60,60	0
2	CA	A	704	1/1	0.98	0.12	-1.87	80,80,80,80	0
2	CA	B	704	1/1	0.98	0.11	-2.25	82,82,82,82	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.