



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 05:11 AM GMT

PDB ID : 2PSG  
Title : REFINED STRUCTURE OF PORCINE PEPSINOGEN AT 1.8  
ANGSTROMS RESOLUTION  
Authors : James, M.N.G.; Sielecki, A.R.  
Deposited on : 1991-01-23  
Resolution : 1.80 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](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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

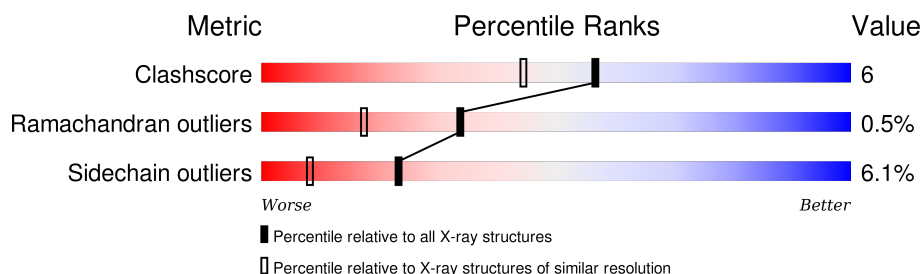
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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))
Clashscore	102246	5383 (1.80-1.80)
Ramachandran outliers	100387	5320 (1.80-1.80)
Sidechain outliers	100360	5319 (1.80-1.80)

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  $\geq 3$ , 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	370	 <div>69%26%. .</div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3031 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 PEPSINOGEN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	370	Total	C	N	O	P	S	0	1	0
			2794	1766	432	585	1	10			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	19P	ASP	ASN	CONFLICT	UNP P00791
A	?	-	ILE	DELETION	UNP P00791
A	263	ASP	ASN	CONFLICT	UNP P00791

- Molecule 2 is water.

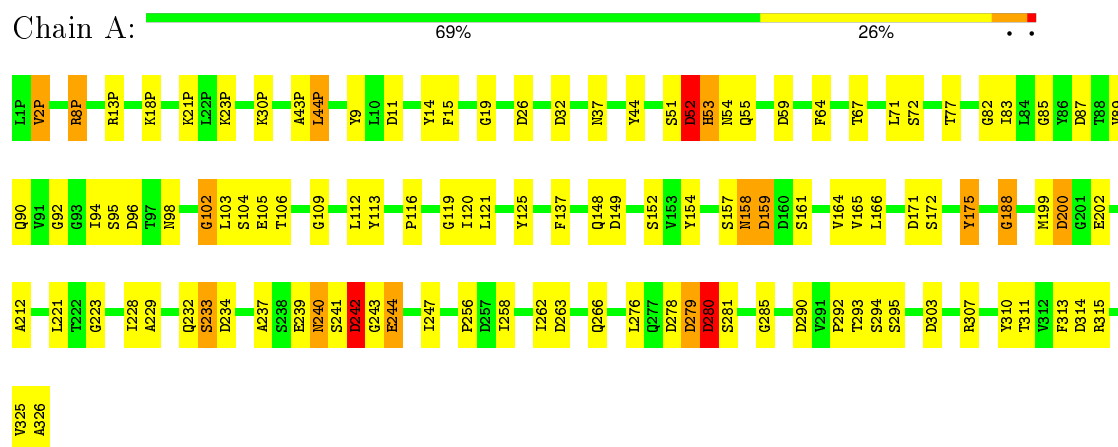
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	237	Total	O	0	0
			237	237		

### 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 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 ( $\text{RSRZ} > 2$ ). 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.

Note EDS was not executed.

#### • Molecule 1: PEPSINOGEN



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.78Å 43.41Å 88.58Å 90.00° 91.40° 90.00°	Depositor
Resolution (Å)	8.00 – 1.80	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-1.80)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, $R_{free}$	0.164 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3031	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	18.0	wwPDB-VP

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SEP

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	1.00	1/2850 (0.0%)	2.01	88/3889 (2.3%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	85	GLY	N-CA	5.14	1.53	1.46

All (88) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	96	ASP	CB-CG-OD1	13.50	130.45	118.30
1	A	315	ARG	NE-CZ-NH1	12.61	126.60	120.30
1	A	13(P)	ARG	NE-CZ-NH2	11.24	125.92	120.30
1	A	313	PHE	CB-CG-CD2	-10.92	113.16	120.80
1	A	310	TYR	CB-CG-CD2	-10.14	114.92	121.00
1	A	154	TYR	CB-CG-CD1	-8.83	115.70	121.00
1	A	280	ASP	CB-CG-OD1	8.60	126.04	118.30
1	A	311	THR	O-C-N	8.48	136.26	122.70
1	A	148	GLN	O-C-N	8.30	135.99	122.70
1	A	9	TYR	CB-CG-CD1	-8.21	116.07	121.00
1	A	9	TYR	CB-CG-CD2	7.92	125.75	121.00
1	A	315	ARG	NE-CZ-NH2	-7.89	116.35	120.30
1	A	310	TYR	CB-CG-CD1	7.58	125.55	121.00
1	A	154	TYR	CB-CG-CD2	7.48	125.49	121.00
1	A	2(P)	VAL	O-C-N	7.40	134.55	122.70
1	A	326	ALA	N-CA-CB	7.35	120.39	110.10
1	A	26	ASP	CB-CG-OD2	-7.22	111.81	118.30
1	A	137	PHE	CB-CG-CD1	-7.16	115.79	120.80
1	A	303	ASP	CB-CG-OD2	-7.09	111.92	118.30
1	A	11	ASP	CB-CG-OD1	-6.89	112.09	118.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	13(P)	ARG	CD-NE-CZ	6.86	133.20	123.60
1	A	233	SER	CB-CA-C	6.84	123.10	110.10
1	A	263	ASP	CB-CG-OD2	-6.60	112.36	118.30
1	A	175	TYR	CB-CG-CD1	-6.53	117.08	121.00
1	A	89	VAL	CG1-CB-CG2	6.52	121.33	110.90
1	A	125	TYR	CB-CG-CD1	6.48	124.89	121.00
1	A	158	ASN	CA-CB-CG	6.39	127.46	113.40
1	A	202	GLU	O-C-N	6.37	132.89	122.70
1	A	113	TYR	CB-CG-CD1	-6.36	117.18	121.00
1	A	158	ASN	N-CA-CB	-6.31	99.25	110.60
1	A	165	VAL	N-CA-CB	6.18	125.09	111.50
1	A	137	PHE	CD1-CG-CD2	6.15	126.29	118.30
1	A	242	ASP	CB-CG-OD2	6.13	123.82	118.30
1	A	67	THR	N-CA-CB	6.08	121.84	110.30
1	A	44	TYR	CA-CB-CG	-6.04	101.92	113.40
1	A	307	ARG	NE-CZ-NH1	-6.02	117.29	120.30
1	A	149	ASP	CB-CG-OD1	6.01	123.71	118.30
1	A	313	PHE	CB-CG-CD1	5.99	124.99	120.80
1	A	244	GLU	N-CA-C	5.97	127.12	111.00
1	A	52	ASP	CB-CG-OD2	-5.95	112.95	118.30
1	A	171	ASP	CB-CG-OD2	5.94	123.65	118.30
1	A	102	GLY	CA-C-O	5.93	131.28	120.60
1	A	229	ALA	CB-CA-C	5.88	118.92	110.10
1	A	119	GLY	O-C-N	5.88	132.10	122.70
1	A	53	HIS	CA-CB-CG	-5.83	103.69	113.60
1	A	106	THR	CA-CB-CG2	5.80	120.52	112.40
1	A	59	ASP	CB-CG-OD1	5.79	123.51	118.30
1	A	54	ASN	O-C-N	-5.79	113.44	122.70
1	A	266	GLN	CA-CB-CG	-5.79	100.67	113.40
1	A	212	ALA	N-CA-CB	5.78	118.19	110.10
1	A	96	ASP	OD1-CG-OD2	-5.74	112.39	123.30
1	A	262	ILE	O-C-N	5.71	131.84	122.70
1	A	234	ASP	CB-CG-OD2	-5.69	113.18	118.30
1	A	44(P)	LEU	CB-CA-C	5.68	121.00	110.20
1	A	313	PHE	CD1-CE1-CZ	-5.62	113.36	120.10
1	A	161	SER	N-CA-C	-5.57	95.95	111.00
1	A	72	SER	N-CA-CB	5.55	118.82	110.50
1	A	303	ASP	OD1-CG-OD2	5.53	133.81	123.30
1	A	109	GLY	O-C-N	-5.51	113.89	122.70
1	A	221	LEU	CB-CG-CD2	-5.49	101.66	111.00
1	A	87	ASP	N-CA-CB	5.48	120.47	110.60
1	A	32	ASP	N-CA-CB	-5.48	100.74	110.60

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	98	ASN	O-C-N	5.47	131.45	122.70
1	A	152	SER	CB-CA-C	5.39	120.34	110.10
1	A	55	GLN	O-C-N	5.39	131.32	122.70
1	A	165	VAL	N-CA-C	-5.38	96.48	111.00
1	A	164	VAL	CB-CA-C	-5.36	101.22	111.40
1	A	285	GLY	O-C-N	-5.32	114.19	122.70
1	A	325	VAL	N-CA-CB	5.32	123.20	111.50
1	A	64	PHE	CB-CG-CD1	-5.31	117.08	120.80
1	A	105	GLU	O-C-N	5.31	131.20	122.70
1	A	149	ASP	O-C-N	5.31	131.19	122.70
1	A	105	GLU	OE1-CD-OE2	-5.29	116.95	123.30
1	A	19	GLY	N-CA-C	-5.27	99.92	113.10
1	A	37	ASN	CB-CG-OD1	-5.23	111.14	121.60
1	A	314	ASP	CB-CG-OD2	-5.21	113.61	118.30
1	A	120	ILE	O-C-N	5.21	131.03	122.70
1	A	266	GLN	CG-CD-OE1	-5.16	111.28	121.60
1	A	243	GLY	C-N-CA	5.16	134.59	121.70
1	A	171	ASP	OD1-CG-OD2	-5.15	113.52	123.30
1	A	14	TYR	O-C-N	5.12	130.89	122.70
1	A	290	ASP	CB-CG-OD2	-5.11	113.70	118.30
1	A	200	ASP	CB-CG-OD2	-5.11	113.70	118.30
1	A	188	GLY	N-CA-C	-5.10	100.36	113.10
1	A	8(P)	ARG	NE-CZ-NH2	5.07	122.83	120.30
1	A	223	GLY	O-C-N	5.05	130.71	121.10
1	A	171	ASP	CB-CG-OD1	5.04	122.83	118.30
1	A	161	SER	N-CA-CB	5.00	118.00	110.50

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	2794	0	2649	33	0
2	A	237	0	0	1	0
All	All	3031	0	2649	33	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 6.

All (33) 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:90:GLN:HE22	1:A:95:SER:HA	1.48	0.77
1:A:90:GLN:NE2	1:A:95:SER:HA	2.09	0.67
1:A:199:MET:CE	1:A:258:ILE:HB	2.33	0.59
1:A:239:GLU:HA	1:A:244:GLU:O	2.03	0.58
1:A:199:MET:HE3	1:A:258:ILE:HB	1.86	0.57
1:A:240:ASN:HD22	1:A:240:ASN:N	2.06	0.53
1:A:158:ASN:O	1:A:159:ASP:OD1	2.28	0.52
1:A:278:ASP:O	1:A:279:ASP:C	2.49	0.51
1:A:237:ALA:HA	1:A:247:ILE:HG13	1.93	0.51
1:A:240:ASN:ND2	1:A:244:GLU:HB2	2.26	0.50
1:A:21(P):LYS:HE2	2:A:485:HOH:O	2.11	0.50
1:A:292:PRO:O	1:A:295:SER:HB3	2.13	0.49
1:A:71:LEU:HD22	1:A:102:GLY:HA3	1.93	0.49
1:A:2(P):VAL:HG13	1:A:92:GLY:O	2.13	0.48
1:A:8(P):ARG:HH11	1:A:159:ASP:H	1.61	0.48
1:A:240:ASN:HD22	1:A:240:ASN:H	1.60	0.48
1:A:121:LEU:HD23	1:A:121:LEU:C	2.34	0.47
1:A:157:SER:O	1:A:158:ASN:HB3	2.14	0.47
1:A:15:PHE:CZ	1:A:116:PRO:HG2	2.50	0.47
1:A:90:GLN:HE22	1:A:95:SER:CA	2.22	0.46
1:A:228:ILE:O	1:A:232:GLN:HG2	2.16	0.45
1:A:159:ASP:CG	1:A:159:ASP:O	2.55	0.45
1:A:8(P):ARG:HD3	1:A:159:ASP:CB	2.47	0.45
1:A:8(P):ARG:HD3	1:A:159:ASP:HB2	1.98	0.45
1:A:172:SER:HA	1:A:175:TYR:CE1	2.51	0.45
1:A:276:LEU:HD23	1:A:276:LEU:N	2.32	0.44
1:A:82:GLY:HA3	1:A:103:LEU:O	2.18	0.43
1:A:43(P):ALA:HB3	1:A:77:THR:HA	2.00	0.43
1:A:279:ASP:OD1	1:A:280:ASP:OD1	2.38	0.42
1:A:53:HIS:NE2	1:A:112:LEU:O	2.49	0.42
1:A:240:ASN:O	1:A:242:ASP:O	2.37	0.41
1:A:51:SER:O	1:A:52:ASP:HB2	2.21	0.41
1:A:94:ILE:HD13	1:A:94:ILE:HG21	1.86	0.40

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	368/370 (100%)	360 (98%)	6 (2%)	2 (0%)	34 17

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	279	ASP
1	A	188	GLY

### 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	314/313 (100%)	295 (94%)	19 (6%)	23 8

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	18(P)	LYS
1	A	23(P)	LYS
1	A	30(P)	LYS
1	A	44(P)	LEU
1	A	52	ASP
1	A	83	ILE
1	A	104	SER
1	A	159	ASP
1	A	166	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	200	ASP
1	A	233	SER
1	A	240	ASN
1	A	241	SER
1	A	242	ASP
1	A	256	PRO
1	A	280	ASP
1	A	281	SER
1	A	293	THR
1	A	294	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	90	GLN
1	A	99	GLN
1	A	143	GLN
1	A	240	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	$\# Z  > 2$
1	SEP	A	68	1	8,9,10	1.03	1 (12%)	8,12,14	3.38	3 (37%)

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
1	SEP	A	68	1	-	0/6/8/10	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	68	SEP	P-OG	-2.29	1.52	1.60

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	68	SEP	O2P-P-OG	-5.26	91.41	106.56
1	A	68	SEP	O-C-CA	-2.02	120.22	125.49
1	A	68	SEP	OG-P-O1P	7.36	125.88	107.14

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 [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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

EDS was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section will therefore be empty.