



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 08:32 am BST

PDB ID : 1Y1U  
Title : Structure of unphosphorylated STAT5a  
Authors : Neculai, D.; Neculai, A.M.; Verrier, S.; Straub, K.; Klumpp, K.; Pfitzner, E.;  
Becker, S.  
Deposited on : 2004-11-19  
Resolution : 3.21 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
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

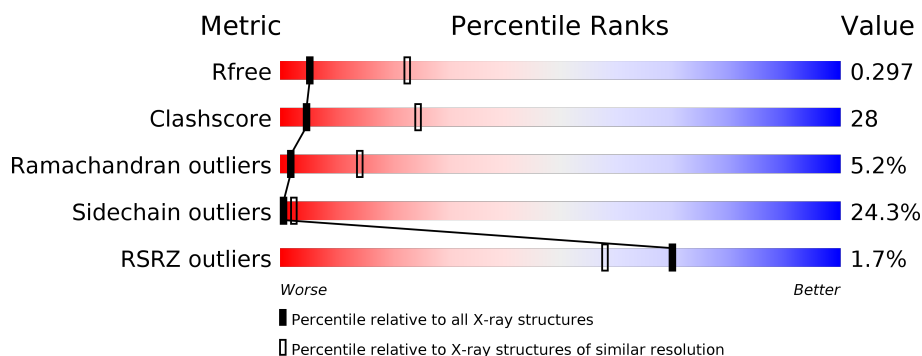
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.21 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	1335 (3.24-3.20)
Clashscore	141614	1460 (3.24-3.20)
Ramachandran outliers	138981	1437 (3.24-3.20)
Sidechain outliers	138945	1436 (3.24-3.20)
RSRZ outliers	127900	1291 (3.24-3.20)

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  $\leq 5\%$ . 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	585	<div> <div>2%</div> <div> <div>40%</div> <div>36%</div> <div>16%</div> <div>7%</div> </div> </div>
1	B	585	<div> <div>%</div> <div> <div>40%</div> <div>37%</div> <div>15%</div> <div>7%</div> </div> </div>
1	C	585	<div> <div>%</div> <div> <div>39%</div> <div>37%</div> <div>17%</div> <div>7%</div> </div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 13293 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.

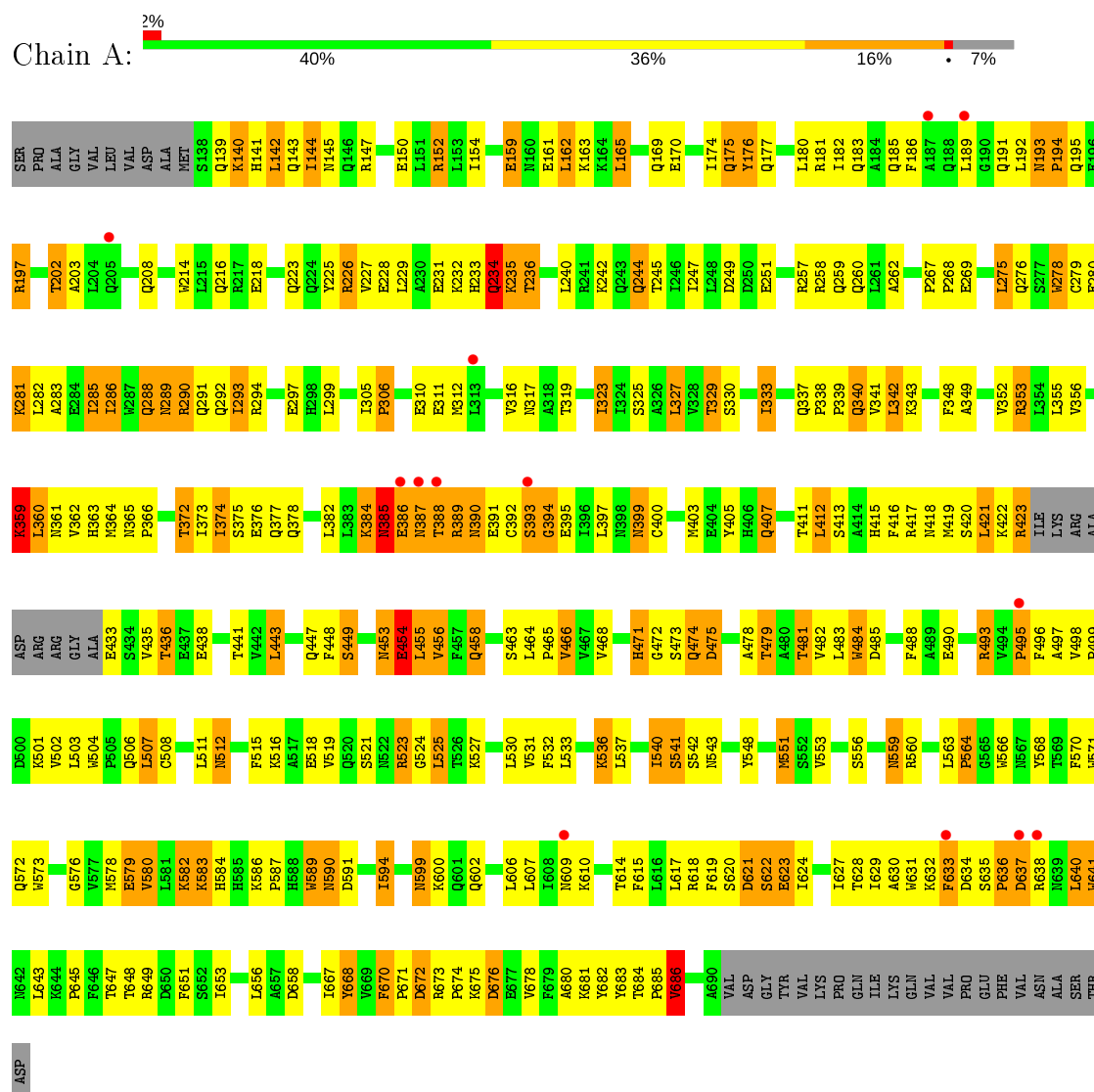
- Molecule 1 is a protein called Signal transducer and activator of transcription 5A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	544	Total	C	N	O	S	0	0	0
			4431	2818	778	821	14			
1	B	544	Total	C	N	O	S	0	0	0
			4431	2818	778	821	14			
1	C	544	Total	C	N	O	S	0	0	0
			4431	2818	778	821	14			

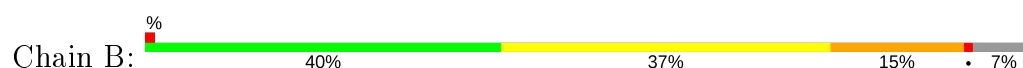
### 3 Residue-property plots

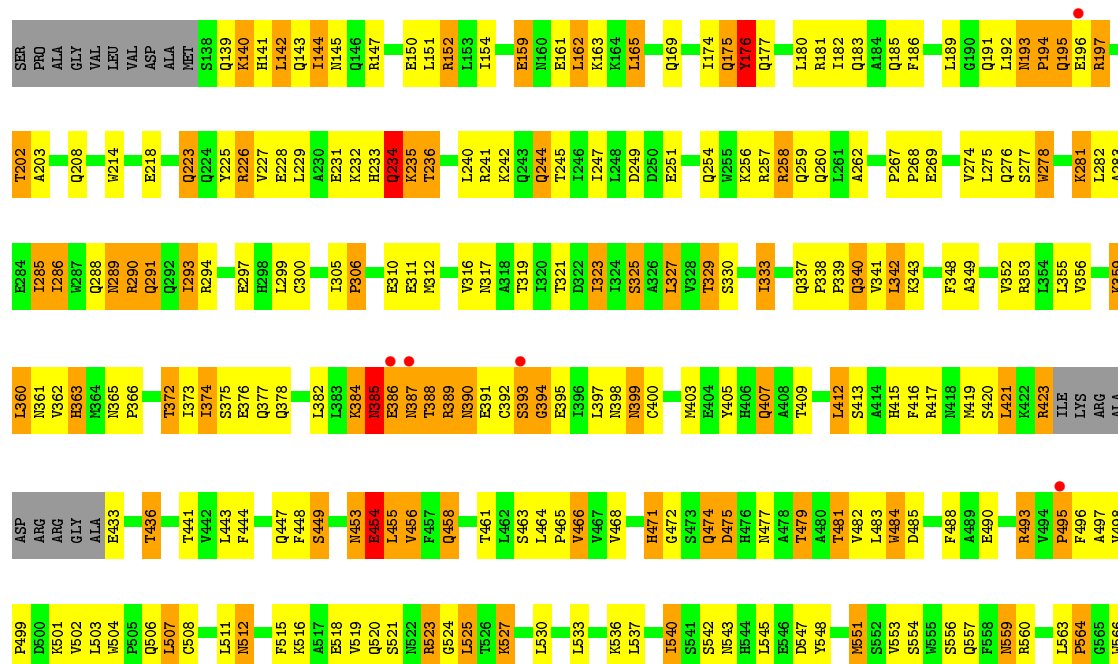
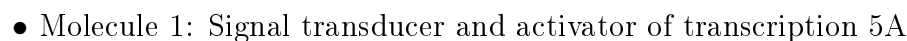
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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.

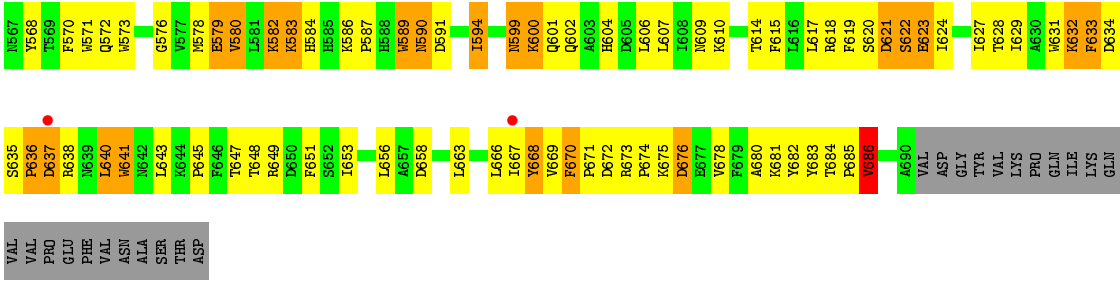
- Molecule 1: Signal transducer and activator of transcription 5A



- Molecule 1: Signal transducer and activator of transcription 5A







## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	143.02Å 235.43Å 111.42Å 90.00° 108.76° 90.00°	Depositor
Resolution (Å)	78.56 – 3.21 78.56 – 3.21	Depositor EDS
% Data completeness (in resolution range)	98.6 (78.56-3.21) 98.6 (78.56-3.21)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.74 (at 3.19Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.266 , 0.299 0.264 , 0.297	Depositor DCC
$R_{free}$ test set	5010 reflections (8.82%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	53.2	Xtriage
Anisotropy	0.216	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 45.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.84	EDS
Total number of atoms	13293	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 16.43% 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

### 5.1 Standard geometry

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	0.66	1/4527 (0.0%)	0.80	1/6134 (0.0%)
1	B	0.65	1/4527 (0.0%)	0.80	2/6134 (0.0%)
1	C	0.65	0/4527	0.80	2/6134 (0.0%)
All	All	0.66	2/13581 (0.0%)	0.80	5/18402 (0.0%)

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	5
1	B	0	6
1	C	0	5
All	All	0	16

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	279	CYS	CB-SG	-7.35	1.69	1.82
1	B	579	GLU	CG-CD	5.12	1.59	1.51

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	472	GLY	N-CA-C	-6.00	98.10	113.10
1	C	472	GLY	N-CA-C	-5.58	99.14	113.10
1	B	176	TYR	CA-CB-CG	5.38	123.63	113.40
1	B	472	GLY	N-CA-C	-5.25	99.98	113.10
1	C	176	TYR	CA-CB-CG	5.15	123.18	113.40

There are no chirality outliers.



5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	197	ARG	Peptide
1	A	360	LEU	Peptide
1	A	387	ASN	Peptide
1	A	471	HIS	Peptide
1	A	623	GLU	Peptide

## 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	4431	0	4415	258	0
1	B	4431	0	4415	253	0
1	C	4431	0	4415	247	0
All	All	13293	0	13245	744	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 28.

The worst 5 of 744 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:668:TYR:HB3	1:A:674:PRO:HA	1.21	1.14
1:C:668:TYR:HB3	1:C:674:PRO:HA	1.19	1.13
1:A:226:ARG:HH12	1:A:305:ILE:HG12	1.04	1.13
1:B:668:TYR:HB3	1:B:674:PRO:HA	1.21	1.13
1:C:226:ARG:HH12	1:C:305:ILE:HG12	0.96	1.12

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	540/585 (92%)	434 (80%)	78 (14%)	28 (5%)	2	14
1	B	540/585 (92%)	430 (80%)	84 (16%)	26 (5%)	2	16
1	C	540/585 (92%)	440 (82%)	69 (13%)	31 (6%)	1	13
All	All	1620/1755 (92%)	1304 (80%)	231 (14%)	85 (5%)	2	14

5 of 85 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	193	ASN
1	A	194	PRO
1	A	359	LYS
1	A	385	ASN
1	A	389	ARG

### 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	490/523 (94%)	370 (76%)	120 (24%)	0	2
1	B	490/523 (94%)	373 (76%)	117 (24%)	0	3
1	C	490/523 (94%)	370 (76%)	120 (24%)	0	2
All	All	1470/1569 (94%)	1113 (76%)	357 (24%)	0	2

5 of 357 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	340	GLN
1	B	495	PRO
1	C	543	ASN
1	B	353	ARG
1	B	412	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 98 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	292	GLN
1	B	415	HIS
1	C	486	ASN
1	B	298	HIS
1	B	363	HIS

### 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](#)

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 [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	544/585 (92%)	0.39	13 (2%) 59 45	31, 53, 77, 94	0
1	B	544/585 (92%)	0.34	8 (1%) 73 62	31, 53, 77, 94	0
1	C	544/585 (92%)	0.33	7 (1%) 77 66	31, 53, 77, 94	0
All	All	1632/1755 (92%)	0.35	28 (1%) 70 58	31, 53, 77, 94	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	196	GLU	3.9
1	A	393	SER	3.7
1	C	386	GLU	3.0
1	A	388	THR	2.8
1	A	187	ALA	2.8

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.