



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:37 AM GMT

PDB ID : 3FDF  
Title : Crystal structure of the serine phosphatase of RNA polymerase II CTD (SSU72 superfamily) from *Drosophila melanogaster*. Orthorhombic crystal form. Northeast Structural Genomics Consortium target FR253.  
Authors : Kuzin, A.P.; Chen, Y.; Seetharaman, J.; Forouhar, F.; Chinag, Y.; Fang, Y.; Cunningham, K.; Ma, L.-C.; Xiao, R.; Liu, J.; Baran, M.C.; Acton, T.B.; Rost, B.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)  
Deposited on : 2008-11-25  
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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

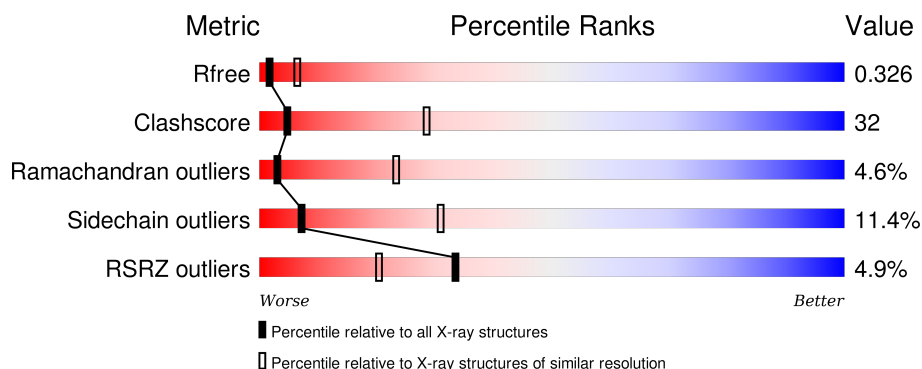
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 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	195	<div> <div>38%</div> <div>51%</div> <div>8%</div> <div>..</div> </div>
1	B	195	<div> <div>42%</div> <div>44%</div> <div>11%</div> <div>..</div> </div>
1	C	195	<div> <div>41%</div> <div>47%</div> <div>9%</div> <div>..</div> </div>
1	D	195	<div> <div>17%</div> <div>41%</div> <div>50%</div> <div>7%</div> <div>.</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 6240 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.

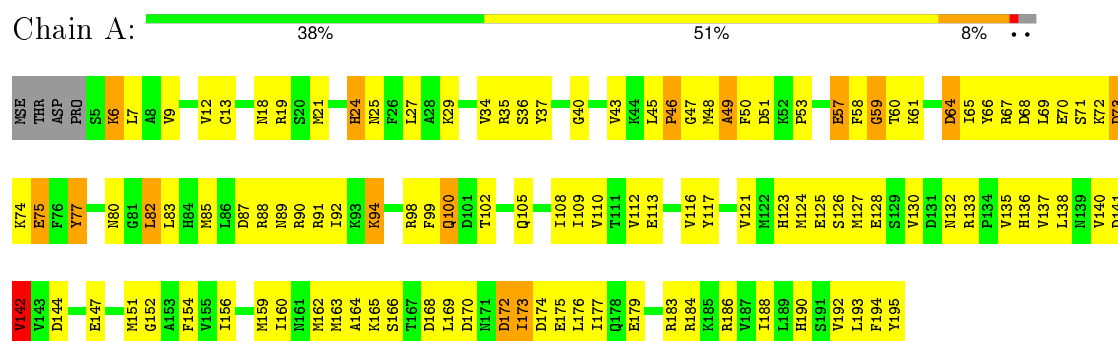
- Molecule 1 is a protein called FR253.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	191	Total	C	N	O	S	Se	0	0	0
			1560	978	269	300	2	11			
1	B	191	Total	C	N	O	S	Se	0	0	0
			1560	978	269	300	2	11			
1	C	191	Total	C	N	O	S	Se	0	0	0
			1560	978	269	300	2	11			
1	D	191	Total	C	N	O	S	Se	0	0	0
			1560	978	269	300	2	11			

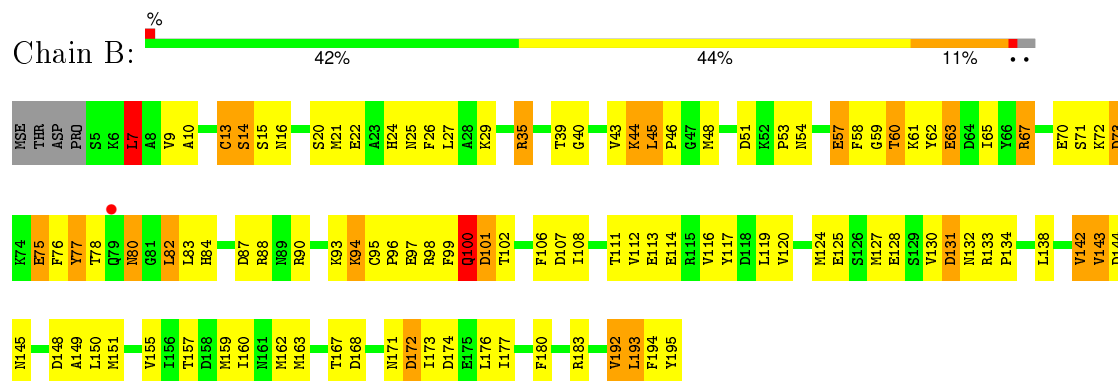
### 3 Residue-property plots

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.

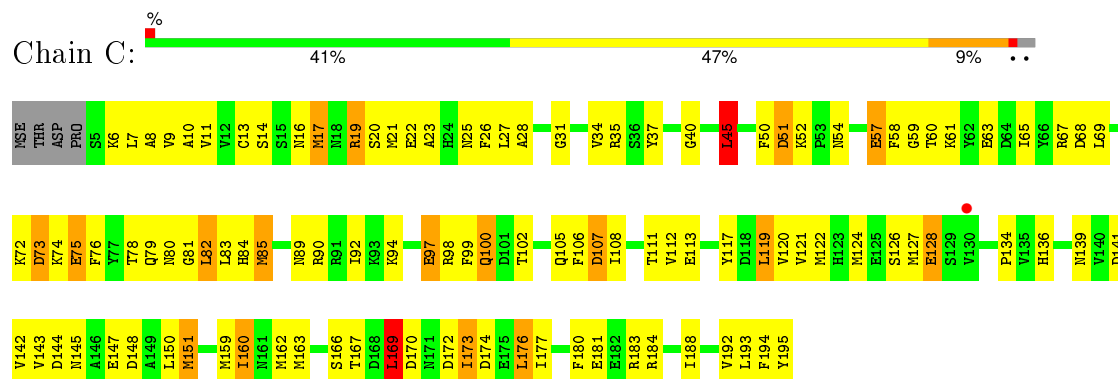
#### • Molecule 1: FR253



#### • Molecule 1: FR253



#### • Molecule 1: FR253



Chain D:

Metric	Category
V142	Red
V143	Red
V144	Red
V145	Red
V146	Red
E147	Red
D148	Red
A149	Red
L150	Red
M151	Red
G152	Red
V155	Red
V156	Red
L157	Red
D158	Red
M159	Red
M161	Red
M162	Red
M163	Red
A164	Red
K165	Red
S166	Red
L167	Red
D168	Red
L169	Red
D170	Red
M171	Red
D172	Red
L173	Red
D174	Red
E175	Red
L176	Red
L177	Red
Q178	Red
E179	Red
D180	Red
E181	Red
E182	Red
R183	Red
R184	Red
R185	Red
R186	Red
L187	Red
L188	Red
L189	Red
H190	Red
S191	Red
V192	Red
L193	Red
F194	Red
V195	Red
Y66	Yellow
L69	Yellow
E70	Yellow
S71	Yellow
K72	Yellow
D73	Yellow
K74	Yellow
T78	Yellow
Q79	Yellow
M80	Yellow
G81	Yellow
L82	Yellow
L83	Yellow
H84	Yellow
M85	Yellow
L86	Yellow
D87	Yellow
R88	Yellow
M89	Yellow
M90	Yellow
R90	Yellow
R91	Yellow
I92	Yellow
K93	Yellow
K94	Yellow
C95	Yellow
P96	Yellow
E97	Yellow
R98	Yellow
F99	Yellow
Q100	Yellow
D101	Yellow
T102	Yellow
K103	Yellow
F106	Yellow
D107	Yellow
I108	Yellow
T111	Yellow
V112	Yellow
E113	Yellow
E114	Yellow
R115	Yellow
V116	Yellow
Y117	Yellow
V120	Yellow
M124	Yellow
S129	Yellow
V130	Yellow
D131	Yellow
P134	Yellow
V135	Yellow
H136	Yellow
R42	Yellow
V43	Yellow
K44	Yellow
L45	Yellow
P46	Yellow
G47	Yellow
M48	Yellow
A49	Yellow
F50	Yellow
D51	Yellow
P53	Yellow
N54	Yellow
V55	Yellow
E56	Yellow
E57	Yellow
F58	Yellow
G59	Yellow
T60	Yellow
K61	Yellow
I65	Yellow

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.60 Å   157.28 Å   61.51 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	19.98 – 3.20 29.78 – 3.19	Depositor EDS
% Data completeness (in resolution range)	91.7 (19.98-3.20) 97.6 (29.78-3.19)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.98 (at 3.18 Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.252   ,   0.313 0.272   ,   0.326	Depositor DCC
$R_{free}$ test set	734 reflections (4.87%)	DCC
Wilson B-factor (Å <sup>2</sup> )	81.5	Xtriage
Anisotropy	0.368	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 41.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 28476 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	6240	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	104.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.78% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

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

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	0.55	0/1574	0.78	0/2096
1	B	0.52	0/1574	0.79	1/2096 (0.0%)
1	C	0.48	0/1574	0.72	1/2096 (0.0%)
1	D	0.37	0/1574	0.65	0/2096
All	All	0.49	0/6296	0.74	2/8384 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	45	LEU	CA-CB-CG	6.26	129.71	115.30
1	B	51	ASP	N-CA-C	-5.53	96.08	111.00

There are no chirality outliers.

There are no planarity outliers.

### 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	1560	0	1534	106	0
1	B	1560	0	1534	93	0
1	C	1560	0	1534	98	1
1	D	1560	0	1534	100	1
All	All	6240	0	6136	392	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 32.

The worst 5 of 392 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:67:ARG:HG3	1:B:67:ARG:HH11	1.33	0.92
1:D:124:MSE:HE2	1:D:135:VAL:HG11	1.53	0.88
1:B:25:ASN:HD21	1:B:29:LYS:HE2	1.37	0.88
1:A:163:MSE:HB3	1:A:169:LEU:HD21	1.58	0.86
1:C:7:LEU:H	1:C:7:LEU:HD23	1.43	0.83

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:C:167:THR:CG2	1:D:184:ARG:NH1[3_445]	2.19	0.01

## 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	189/195 (97%)	150 (79%)	26 (14%)	13 (7%)	1	10
1	B	189/195 (97%)	149 (79%)	33 (18%)	7 (4%)	4	29
1	C	189/195 (97%)	159 (84%)	25 (13%)	5 (3%)	7	40
1	D	189/195 (97%)	142 (75%)	37 (20%)	10 (5%)	2	19
All	All	756/780 (97%)	600 (79%)	121 (16%)	35 (5%)	3	23

5 of 35 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	46	PRO
1	B	73	ASP

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	C	173	ILE
1	D	32	PHE
1	D	73	ASP

### 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	175/167 (105%)	158 (90%)	17 (10%)	10	39
1	B	175/167 (105%)	148 (85%)	27 (15%)	3	16
1	C	175/167 (105%)	151 (86%)	24 (14%)	4	21
1	D	175/167 (105%)	163 (93%)	12 (7%)	19	59
All	All	700/668 (105%)	620 (89%)	80 (11%)	7	31

5 of 80 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	128	GLU
1	C	6	LYS
1	D	66	TYR
1	B	142	VAL
1	B	172	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	171	ASN
1	C	24	HIS
1	D	100	GLN
1	C	16	ASN
1	C	18	ASN

### 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	180/195 (92%)	-0.14	0 100 100	42, 78, 120, 136	0
1	B	180/195 (92%)	-0.10	1 (0%) 90 84	44, 79, 118, 130	0
1	C	180/195 (92%)	-0.12	1 (0%) 90 84	46, 91, 115, 138	0
1	D	180/195 (92%)	1.19	33 (18%) 2 1	100, 158, 198, 201	0
All	All	720/780 (92%)	0.21	35 (4%) 33 20	42, 94, 188, 201	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	7	LEU	8.5
1	D	131	ASP	7.3
1	D	8	ALA	7.1
1	D	130	VAL	6.2
1	D	129	SER	6.2

### 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.