Full wwPDB X-ray Structure Validation Report (i)

Jun 22, 2024 – 09:02 PM EDT

PDB ID	:	4YH1
Title	:	Structure of Human Scp1 bound to cis-proline peptidomimetic CTD phospho-Ser5 peptide
Authors	:	Mayfield, J.E.; Zhang, Y.
Deposited on	:	2015-02-26
Resolution	:	2.20 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

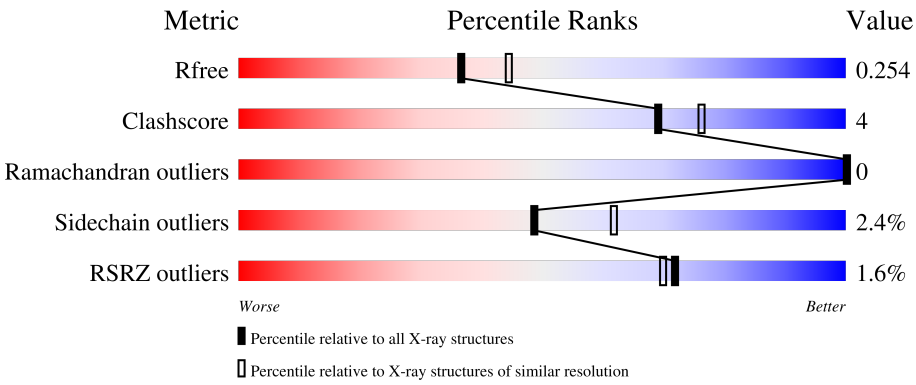
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	189	2%, 83%, 11%, • 5%
1	B	189	87%, 6%, •• 5%
2	C	12	25%, 8%, 67%
2	D	12	25%, 17%, 8%, 8%, 67%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3123 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 Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	179	1454	935	248	266	5	0	0	0
1	B	179	1454	935	248	266	5	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	73	GLY	-	expression tag	UNP Q9GZU7
A	74	SER	-	expression tag	UNP Q9GZU7
A	75	HIS	-	expression tag	UNP Q9GZU7
A	76	GLY	-	expression tag	UNP Q9GZU7
A	96	ASN	ASP	engineered mutation	UNP Q9GZU7
A	256	GLN	-	expression tag	UNP Q9GZU7
A	257	PRO	-	expression tag	UNP Q9GZU7
A	258	ARG	-	expression tag	UNP Q9GZU7
A	259	PRO	-	expression tag	UNP Q9GZU7
A	260	GLY	-	expression tag	UNP Q9GZU7
A	261	SER	-	expression tag	UNP Q9GZU7
B	73	GLY	-	expression tag	UNP Q9GZU7
B	74	SER	-	expression tag	UNP Q9GZU7
B	75	HIS	-	expression tag	UNP Q9GZU7
B	76	GLY	-	expression tag	UNP Q9GZU7
B	96	ASN	ASP	engineered mutation	UNP Q9GZU7
B	256	GLN	-	expression tag	UNP Q9GZU7
B	257	PRO	-	expression tag	UNP Q9GZU7
B	258	ARG	-	expression tag	UNP Q9GZU7
B	259	PRO	-	expression tag	UNP Q9GZU7
B	260	GLY	-	expression tag	UNP Q9GZU7
B	261	SER	-	expression tag	UNP Q9GZU7

- Molecule 2 is a protein called A small phosphatase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	4	36	21	4	10	1	0	0	0
2	D	4	36	21	4	10	1	0	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
3	A	1	1	1	0	0
3	B	1	1	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	68	68	68	0	0
4	B	69	69	69	0	0
4	C	4	4	4	0	0

3 Residue-property plots (i)

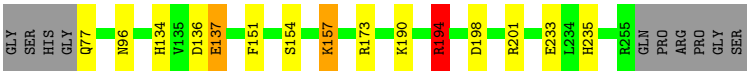
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 1

Chain A:

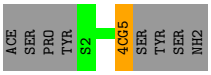
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 1

Chain B:



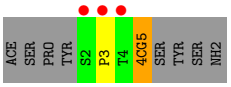
• Molecule 2: A small phosphatase 1

Chain C:



• Molecule 2: A small phosphatase 1

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	125.11Å 78.80Å 62.85Å 90.00° 112.54° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 48.89 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.0 (50.00-2.20) 97.0 (48.89-2.20)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.07 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.193 , 0.246 0.199 , 0.254	Depositor DCC
R_{free} test set	1469 reflections (5.28%)	wwPDB-VP
Wilson B-factor (Å^2)	27.6	Xtriage
Anisotropy	0.456	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 39.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3123	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.13% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: 4CG, MG

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	0.84	2/1490 (0.1%)	0.94	3/2026 (0.1%)
1	B	1.02	3/1490 (0.2%)	1.11	10/2026 (0.5%)
2	C	0.93	0/20	0.71	0/27
2	D	0.96	0/20	0.95	0/27
All	All	0.94	5/3020 (0.2%)	1.02	13/4106 (0.3%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	137	GLU	CD-OE1	16.58	1.43	1.25
1	B	137	GLU	CG-CD	8.97	1.65	1.51
1	A	137	GLU	CD-OE1	7.76	1.34	1.25
1	B	233	GLU	CD-OE2	5.54	1.31	1.25
1	A	137	GLU	CG-CD	5.12	1.59	1.51

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	137	GLU	CG-CD-OE1	13.61	145.51	118.30
1	B	137	GLU	CG-CD-OE2	-12.98	92.33	118.30
1	B	173	ARG	NE-CZ-NH1	8.37	124.48	120.30
1	B	194	ARG	NE-CZ-NH1	8.01	124.30	120.30
1	B	173	ARG	NE-CZ-NH2	-7.14	116.73	120.30
1	A	160	ASP	CB-CG-OD1	6.98	124.59	118.30
1	B	198	ASP	CB-CG-OD2	-5.94	112.95	118.30
1	B	233	GLU	OE1-CD-OE2	5.89	130.37	123.30
1	B	136	ASP	CB-CG-OD1	5.55	123.29	118.30
1	B	198	ASP	CB-CG-OD1	5.53	123.27	118.30
1	A	173	ARG	NE-CZ-NH1	5.16	122.88	120.30
1	B	194	ARG	CG-CD-NE	5.09	122.48	111.80

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	200	ARG	NE-CZ-NH2	-5.04	117.78	120.30

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	1454	0	1427	12	0
1	B	1454	0	1427	9	0
2	C	36	0	29	2	0
2	D	36	0	29	3	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	68	0	0	0	1
4	B	69	0	0	2	1
4	C	4	0	0	0	0
All	All	3123	0	2912	22	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 4.

All (22) 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:134:HIS:ND1	1:B:137:GLU:OE1	2.04	0.91
1:B:194:ARG:HD2	4:B:407:HOH:O	1.76	0.84
1:B:137:GLU:OE1	1:B:235:HIS:ND1	2.16	0.73
1:A:198:ASP:OD1	1:A:200:ARG:HD3	1.92	0.68
1:A:96:ASN:HD21	2:C:5:4CG:P05	2.22	0.63
1:B:77:GLN:HA	1:B:77:GLN:OE1	2.00	0.62
1:A:200:ARG:O	1:A:250:VAL:HG22	2.04	0.57
1:A:137:GLU:OE1	1:A:235:HIS:ND1	2.35	0.55
2:C:5:4CG:O	2:C:5:4CG:H10	2.12	0.49
1:A:255:ARG:HB2	1:A:255:ARG:NH1	2.28	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:107:LYS:HZ3	1:B:201:ARG:HH22	1.63	0.47
2:D:5:4CG:H10	2:D:5:4CG:O	2.14	0.46
1:A:132:ARG:CZ	1:A:225:TRP:HB2	2.45	0.45
1:A:209:PRO:HA	1:A:212:TYR:CZ	2.53	0.44
1:A:98:ASP:O	1:A:99:GLU:HB2	2.18	0.44
1:A:93:VAL:HG23	1:A:146:PHE:CG	2.55	0.41
1:B:137:GLU:CD	1:B:235:HIS:HD1	2.19	0.41
1:A:160:ASP:HB3	1:A:161:PRO:HD3	2.03	0.41
1:A:198:ASP:OD1	1:A:200:ARG:CD	2.64	0.41
1:B:96:ASN:OD1	2:D:5:4CG:P05	2.78	0.41
1:B:157:LYS:NZ	4:B:457:HOH:O	2.54	0.40
1:B:154:SER:HB2	2:D:3:PRO:O	2.21	0.40

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)
4:A:416:HOH:O	4:B:417:HOH:O[1_556]	2.13	0.07

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	177/189 (94%)	167 (94%)	10 (6%)	0	100	100
1	B	177/189 (94%)	171 (97%)	6 (3%)	0	100	100
2	C	2/12 (17%)	2 (100%)	0	0	100	100
2	D	2/12 (17%)	2 (100%)	0	0	100	100
All	All	358/402 (89%)	342 (96%)	16 (4%)	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	162/169 (96%)	158 (98%)	4 (2%)	47	60
1	B	162/169 (96%)	158 (98%)	4 (2%)	47	60
2	C	3/9 (33%)	3 (100%)	0	100	100
2	D	3/9 (33%)	3 (100%)	0	100	100
All	All	330/356 (93%)	322 (98%)	8 (2%)	49	62

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	151	PHE
1	A	208	SER
1	A	230	SER
1	A	242	GLU
1	B	151	PHE
1	B	157	LYS
1	B	190	LYS
1	B	194	ARG

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

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

2 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	4CG	C	5	3,2	14,16,17	0.87	0	13,22,24	2.74	3 (23%)
2	4CG	D	5	3,2	14,16,17	1.25	2 (14%)	13,22,24	2.62	4 (30%)

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	4CG	C	5	3,2	-	3/9/22/24	0/1/1/1
2	4CG	D	5	3,2	-	3/9/22/24	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	5	4CG	CA-C09	2.94	1.54	1.50
2	D	5	4CG	C09-C10	2.03	1.36	1.32

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	5	4CG	O04-C03-CA	8.13	116.05	108.14
2	D	5	4CG	O04-C03-CA	7.58	115.52	108.14
2	D	5	4CG	C12-C13-C14	4.09	110.53	103.92
2	C	5	4CG	C12-C13-C14	3.88	110.20	103.92
2	C	5	4CG	O08-P05-O07	2.40	116.80	107.64
2	D	5	4CG	CA-C09-C10	2.37	128.77	126.29
2	D	5	4CG	P05-O04-C03	2.08	124.02	118.30

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	5	4CG	O04-C03-CA-C09
2	C	5	4CG	O04-C03-CA-N
2	D	5	4CG	C03-O04-P05-O07

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	D	5	4CG	C03-O04-P05-O08
2	D	5	4CG	C03-O04-P05-O06
2	C	5	4CG	C03-O04-P05-O08

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	5	4CG	2	0
2	D	5	4CG	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 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, 95th 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^2)	Q<0.9
1	A	179/189 (94%)	-0.30	3 (1%) 70 68	19, 35, 53, 70	0
1	B	179/189 (94%)	-0.54	0 100 100	20, 27, 40, 68	0
2	C	3/12 (25%)	0.43	0 100 100	58, 58, 58, 62	0
2	D	3/12 (25%)	3.03	3 (100%) 0 0	72, 72, 76, 77	0
All	All	364/402 (90%)	-0.39	6 (1%) 72 70	19, 30, 53, 77	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	2	SER	4.7
1	A	246	ARG	3.5
1	A	253	VAL	2.9
1	A	243	GLN	2.3
2	D	3	PRO	2.3
2	D	4	THR	2.1

6.2 Non-standard residues in protein, DNA, RNA chains (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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å^2)	Q<0.9
2	4CG	C	5	16/17	0.93	0.13	50,68,80,84	0
2	4CG	D	5	16/17	0.93	0.20	39,72,89,89	0

6.3 Carbohydrates (i)

There are no monosaccharides 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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å^2)	Q<0.9
3	MG	A	301	1/1	0.92	0.15	52,52,52,52	0
3	MG	B	301	1/1	0.97	0.08	54,54,54,54	0

6.5 Other polymers (i)

There are no such residues in this entry.