WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 07:53 am BST

PDB ID	:	3AUP
Title	:	Crystal structure of Basic 7S globulin from soybean
Authors	:	Yoshizawa, T.; Shimizu, T.; Taichi, M.; Nishiuchi, Y.; Yamabe, M.; Shichijo, N.; Unzai, S.; Hirano, H.; Sato, M.; Hashimoto, H.
Deposited on	:	2011-02-14
Resolution	:	1.91 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

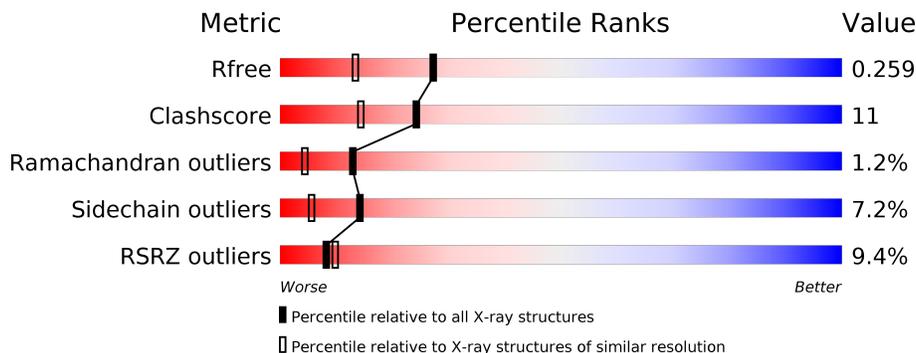|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 1.91 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	Percentile Ranks	Value
Rfree		0.259
Clashscore		11
Ramachandran outliers		1.2%
Sidechain outliers		7.2%
RSRZ outliers		9.4%

Worse — *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	7937 (1.94-1.90)
Clashscore	141614	8644 (1.94-1.90)
Ramachandran outliers	138981	8530 (1.94-1.90)
Sidechain outliers	138945	8530 (1.94-1.90)
RSRZ outliers	127900	7793 (1.94-1.90)

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 >=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	403	3% — 75% / 17% • 6%
1	B	403	9% — 74% / 16% • 8%
1	C	403	12% — 66% / 22% • 8%
1	D	403	10% — 68% / 19% • 11%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 11918 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 Basic 7S globulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	377	Total 2888	C 1824	N 512	O 531	S 21	133	3	0
1	B	372	Total 2842	C 1796	N 503	O 522	S 21	126	2	0
1	C	372	Total 2830	C 1781	N 503	O 525	S 21	70	1	0
1	D	360	Total 2737	C 1728	N 482	O 506	S 21	96	1	0

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	180	Total 180	O 180	0	0
2	B	163	Total 163	O 163	0	0
2	C	170	Total 170	O 170	0	0
2	D	108	Total 108	O 108	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 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.
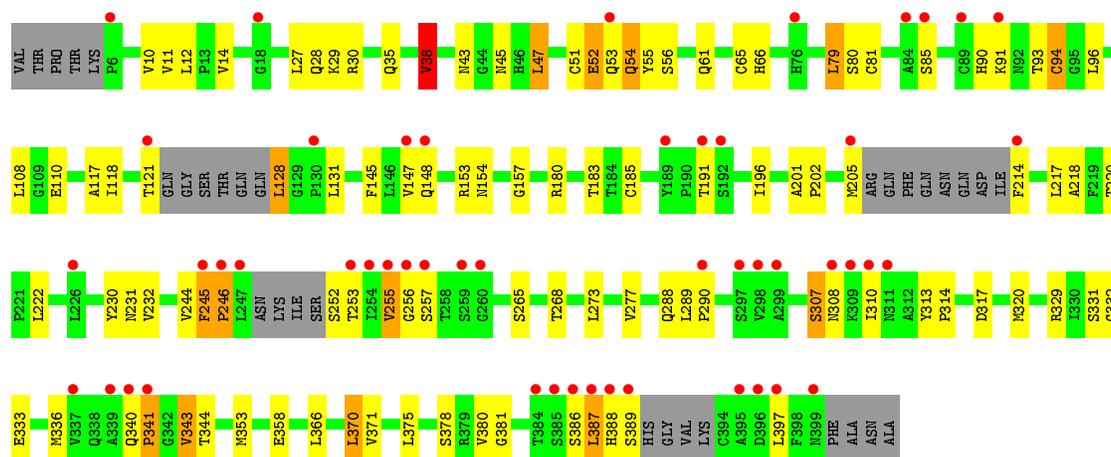
- Molecule 1: Basic 7S globulin

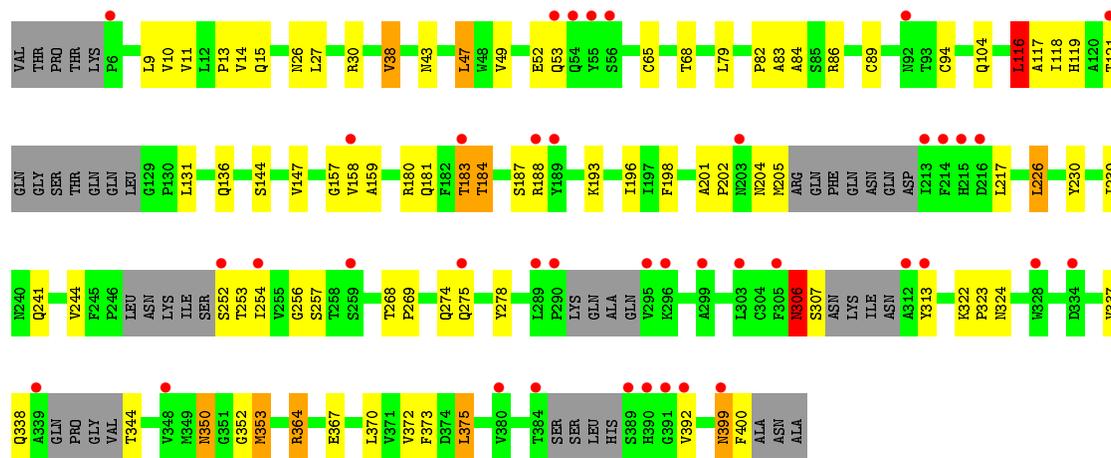Chain A:

- Molecule 1: Basic 7S globulin

Chain B:

- Molecule 1: Basic 7S globulin

Chain C:
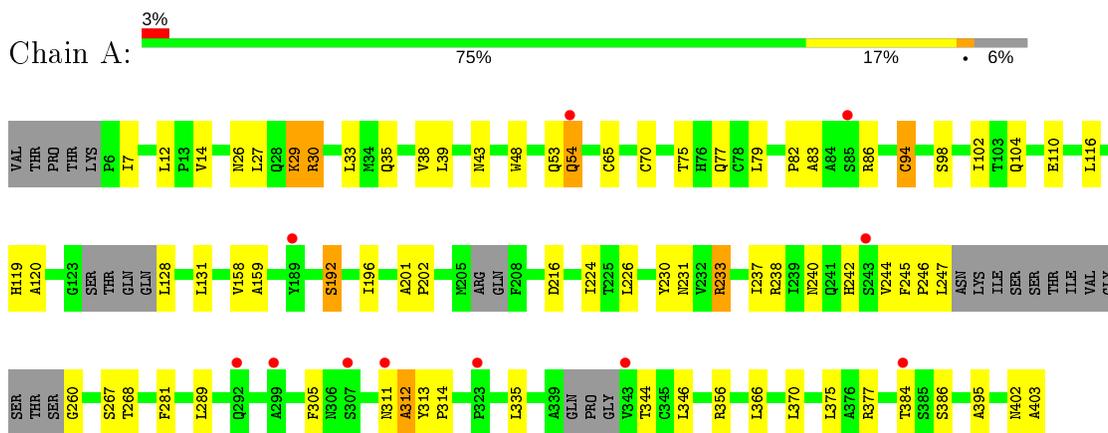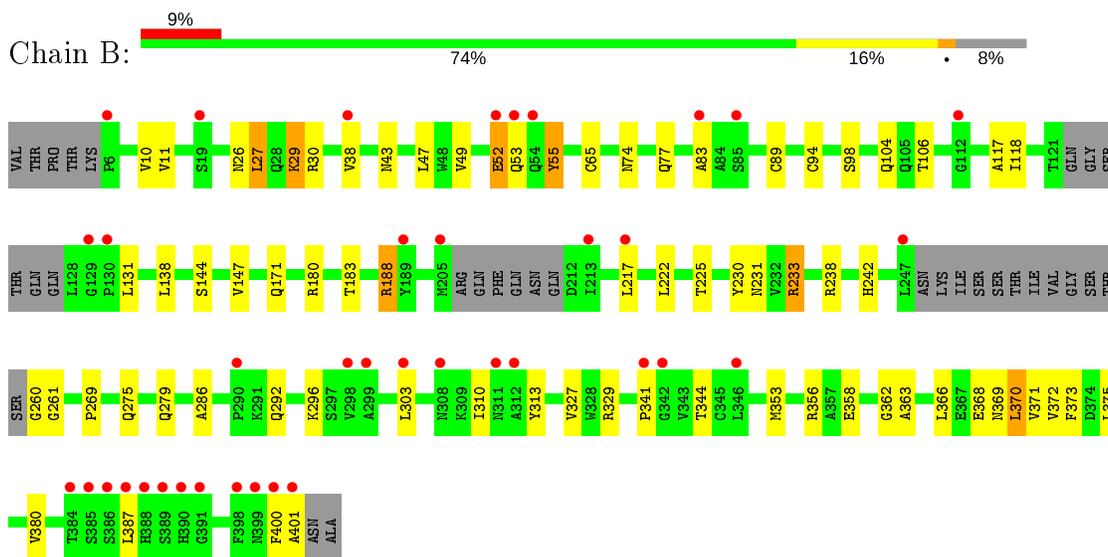
• Molecule 1: Basic 7S globulin

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	135.24Å 161.17Å 84.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 1.91 19.93 – 1.91	Depositor EDS
% Data completeness (in resolution range)	95.1 (20.00-1.91) 95.2 (19.93-1.91)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.66 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.211 , 0.255 0.214 , 0.259	Depositor DCC
R_{free} test set	6912 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	31.3	Xtriage
Anisotropy	0.024	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 49.7	EDS
L-test for twinning[2]	$<|L|> = 0.48$, $<L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	11918	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.34% 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)

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	0.80	2/2961 (0.1%)	0.81	2/4028 (0.0%)
1	B	0.73	0/2913	0.80	1/3965 (0.0%)
1	C	0.71	1/2895 (0.0%)	0.83	3/3939 (0.1%)
1	D	0.67	1/2799 (0.0%)	0.78	1/3805 (0.0%)
All	All	0.73	4/11568 (0.0%)	0.80	7/15737 (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	1
1	B	0	1
1	C	0	1
All	All	0	3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	38	VAL	CB-CG1	-7.35	1.37	1.52
1	A	94	CYS	CB-SG	-6.72	1.70	1.82
1	A	110	GLU	CB-CG	-6.55	1.39	1.52
1	D	94	CYS	CB-SG	-6.34	1.71	1.82

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	53	GLN	N-CA-CB	-5.86	100.05	110.60
1	C	245	PHE	C-N-CD	-5.60	108.27	120.60
1	D	116	LEU	CA-CB-CG	5.58	128.13	115.30
1	A	94	CYS	CA-CB-SG	-5.50	104.10	114.00
1	C	47	LEU	CB-CG-CD2	5.07	119.61	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	53	GLN	Peptide
1	B	52	GLU	Peptide
1	C	245	PHE	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	2888	0	2855	59	0
1	B	2842	0	2814	56	0
1	C	2830	0	2797	65	0
1	D	2737	0	2695	82	0
2	A	180	0	0	5	0
2	B	163	0	0	6	0
2	C	170	0	0	5	0
2	D	108	0	0	9	0
All	All	11918	0	11161	236	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 11.

The worst 5 of 236 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:261:GLY:HA2	2:B:434:HOH:O	1.62	0.99
1:B:183:THR:HG22	1:B:372:VAL:HG22	1.47	0.97
1:D:201:ALA:HB1	1:D:205:MET:HE1	1.47	0.94
1:D:184:THR:HG23	2:D:432:HOH:O	1.73	0.87
1:D:399:ASN:C	1:D:399:ASN:HD22	1.77	0.87

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	370/403 (92%)	355 (96%)	12 (3%)	3 (1%)	19	9
1	B	366/403 (91%)	354 (97%)	10 (3%)	2 (0%)	29	18
1	C	363/403 (90%)	340 (94%)	15 (4%)	8 (2%)	6	1
1	D	345/403 (86%)	325 (94%)	16 (5%)	4 (1%)	13	4
All	All	1444/1612 (90%)	1374 (95%)	53 (4%)	17 (1%)	13	4

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	312	ALA
1	C	52	GLU
1	C	246	PRO
1	C	387	LEU
1	B	55	TYR

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	325/346 (94%)	311 (96%)	14 (4%)	29	18
1	B	320/346 (92%)	302 (94%)	18 (6%)	21	10
1	C	320/346 (92%)	289 (90%)	31 (10%)	8	2
1	D	309/346 (89%)	280 (91%)	29 (9%)	8	3
All	All	1274/1384 (92%)	1182 (93%)	92 (7%)	14	5

5 of 92 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	79	LEU
1	C	231	ASN
1	D	324	ASN
1	C	85	SER
1	C	147	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 45 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	26	ASN
1	C	90	HIS
1	D	324	ASN
1	C	54	GLN
1	C	104	GLN

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	377/403 (93%)	0.16	11 (2%)	51	55	25, 40, 71, 83	38 (10%)
1	B	372/403 (92%)	0.43	38 (10%)	6	8	26, 43, 76, 89	33 (8%)
1	C	372/403 (92%)	0.60	50 (13%)	3	3	25, 44, 79, 108	17 (4%)
1	D	360/403 (89%)	0.60	40 (11%)	5	6	31, 52, 84, 104	25 (6%)
All	All	1481/1612 (91%)	0.45	139 (9%)	8	10	25, 44, 77, 108	113 (7%)

The worst 5 of 139 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	387	LEU	8.7
1	C	254	ILE	7.1
1	C	341	PRO	6.8
1	C	385	SER	6.0
1	D	392	VAL	6.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.