wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 09:09 PM GMT

PDB ID	:	1NKQ
Title	:	Crystal structure of yeast ynq8, a fumarylacetoacetate hydrolase family protein
Authors	:	Eswaramoorthy, S.; Kumaran, D.; Daniels, B.; Studier, F.W.; Swaminathan, S.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2003-01-03
Resolution	:	2.20 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
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

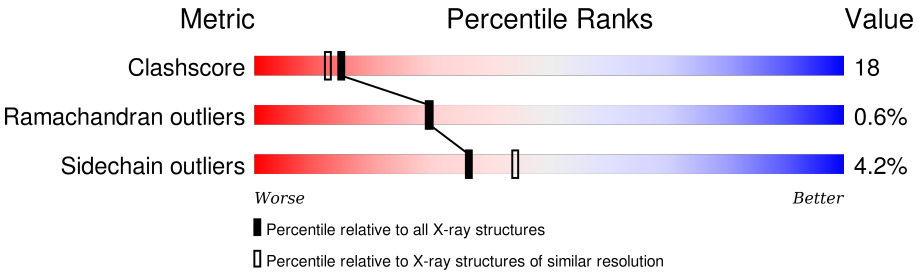
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (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))
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (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 on the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	259	68% 25% • 5%
1	B	259	69% 23% • 6%
1	C	259	62% 30% • 5%
1	D	259	64% 27% • 6%
1	E	259	64% 29% • 5%
1	F	259	60% 29% • 7%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	ACY	B	703	-	-	X	-
4	ACY	C	705	-	-	X	-
4	ACY	E	709	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12088 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 Hypothetical 28.8 kDa protein in PSD1-SKO1 intergenic region.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	247	Total	C	N	O	S	Se	0	0	0
			1911	1214	328	360	4	5			
1	B	244	Total	C	N	O	S	Se	0	0	0
			1894	1204	326	355	4	5			
1	C	245	Total	C	N	O	S	Se	0	0	0
			1894	1204	324	357	4	5			
1	D	244	Total	C	N	O	S	Se	0	0	0
			1888	1201	323	355	4	5			
1	E	246	Total	C	N	O	S	Se	0	0	0
			1908	1212	330	357	4	5			
1	F	240	Total	C	N	O	S	Se	0	0	0
			1870	1187	325	349	4	5			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P53889
A	104	MSE	MET	MODIFIED RESIDUE	UNP P53889
A	148	MSE	MET	MODIFIED RESIDUE	UNP P53889
A	189	MSE	MET	MODIFIED RESIDUE	UNP P53889
A	203	MSE	MET	MODIFIED RESIDUE	UNP P53889
A	243	MSE	MET	MODIFIED RESIDUE	UNP P53889
B	301	MSE	MET	MODIFIED RESIDUE	UNP P53889
B	404	MSE	MET	MODIFIED RESIDUE	UNP P53889
B	448	MSE	MET	MODIFIED RESIDUE	UNP P53889
B	489	MSE	MET	MODIFIED RESIDUE	UNP P53889
B	503	MSE	MET	MODIFIED RESIDUE	UNP P53889
B	543	MSE	MET	MODIFIED RESIDUE	UNP P53889
C	1	MSE	MET	MODIFIED RESIDUE	UNP P53889
C	104	MSE	MET	MODIFIED RESIDUE	UNP P53889
C	148	MSE	MET	MODIFIED RESIDUE	UNP P53889
C	189	MSE	MET	MODIFIED RESIDUE	UNP P53889
C	203	MSE	MET	MODIFIED RESIDUE	UNP P53889

Continued on next page...

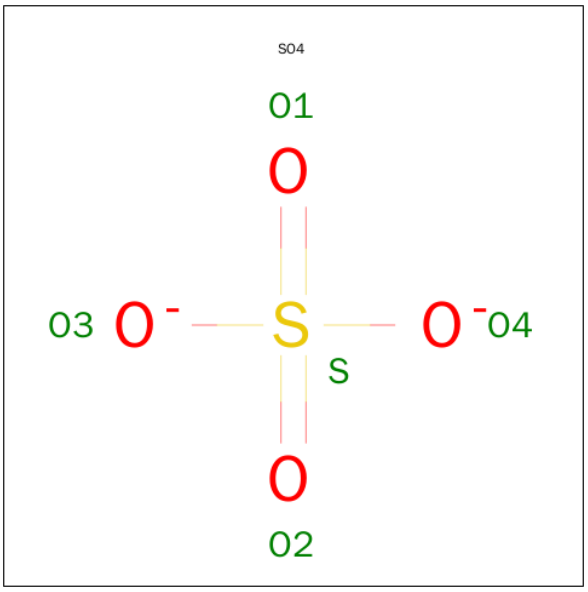
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	243	MSE	MET	MODIFIED RESIDUE	UNP P53889
D	301	MSE	MET	MODIFIED RESIDUE	UNP P53889
D	404	MSE	MET	MODIFIED RESIDUE	UNP P53889
D	448	MSE	MET	MODIFIED RESIDUE	UNP P53889
D	489	MSE	MET	MODIFIED RESIDUE	UNP P53889
D	503	MSE	MET	MODIFIED RESIDUE	UNP P53889
D	543	MSE	MET	MODIFIED RESIDUE	UNP P53889
E	1	MSE	MET	MODIFIED RESIDUE	UNP P53889
E	104	MSE	MET	MODIFIED RESIDUE	UNP P53889
E	148	MSE	MET	MODIFIED RESIDUE	UNP P53889
E	189	MSE	MET	MODIFIED RESIDUE	UNP P53889
E	203	MSE	MET	MODIFIED RESIDUE	UNP P53889
E	243	MSE	MET	MODIFIED RESIDUE	UNP P53889
F	301	MSE	MET	MODIFIED RESIDUE	UNP P53889
F	404	MSE	MET	MODIFIED RESIDUE	UNP P53889
F	448	MSE	MET	MODIFIED RESIDUE	UNP P53889
F	489	MSE	MET	MODIFIED RESIDUE	UNP P53889
F	503	MSE	MET	MODIFIED RESIDUE	UNP P53889
F	543	MSE	MET	MODIFIED RESIDUE	UNP P53889

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

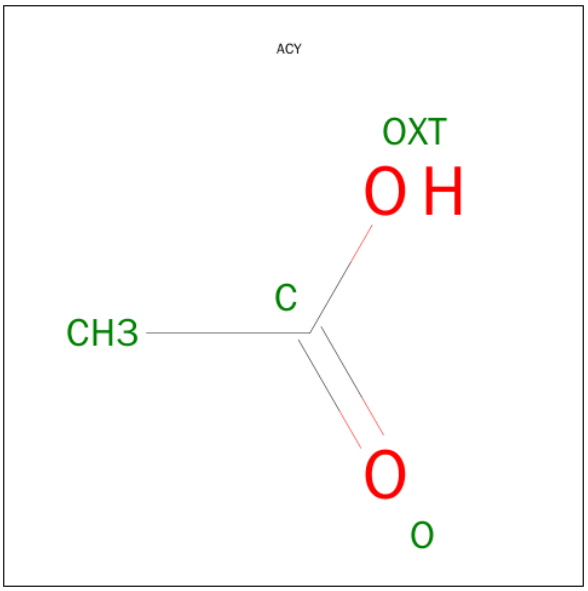
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	1	Total Ca 1 1	0	0
2	E	1	Total Ca 1 1	0	0
2	B	1	Total Ca 1 1	0	0
2	C	1	Total Ca 1 1	0	0
2	A	1	Total Ca 1 1	0	0
2	F	1	Total Ca 1 1	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	C	1	Total O S 5 4 1	0	0
3	E	1	Total O S 5 4 1	0	0

- Molecule 4 is ACETIC ACID (three-letter code: ACY) (formula: $C_2H_4O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	C	1	Total C O 4 2 2	0	0
4	C	1	Total C O 4 2 2	0	0
4	D	1	Total C O 4 2 2	0	0
4	D	1	Total C O 4 2 2	0	0
4	E	1	Total C O 4 2 2	0	0
4	E	1	Total C O 4 2 2	0	0
4	F	1	Total C O 4 2 2	0	0
4	F	1	Total C O 4 2 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	112	Total O 112 112	0	0
5	B	106	Total O 106 106	0	0
5	C	99	Total O 99 99	0	0
5	D	122	Total O 122 122	0	0
5	E	107	Total O 107 107	0	0
5	F	88	Total O 88 88	0	0

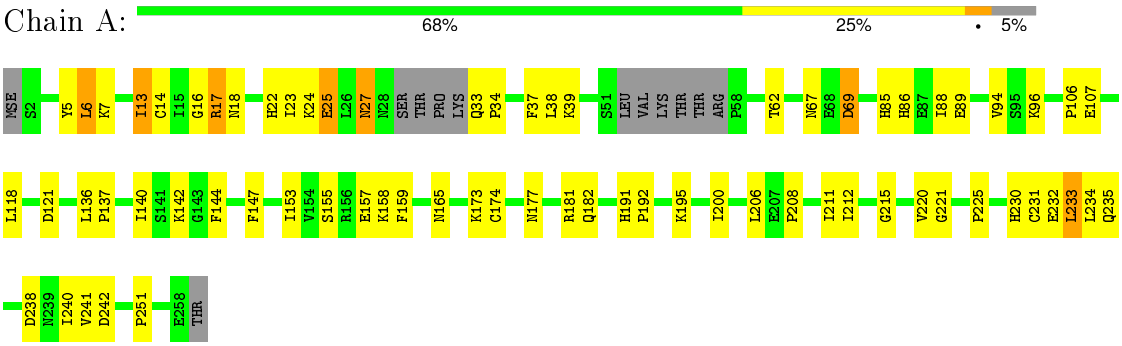
3 Residue-property plots (i)

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 (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.

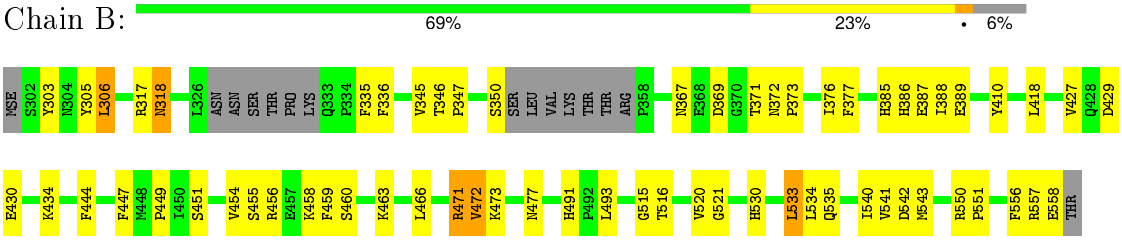
- Molecule 1: Hypothetical 28.8 kDa protein in PSD1-SKO1 intergenic region

Chain A:



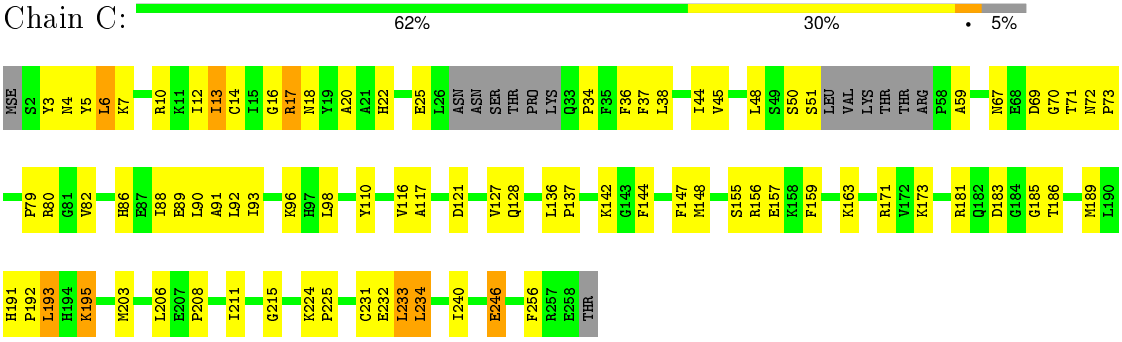
- Molecule 1: Hypothetical 28.8 kDa protein in PSD1-SKO1 intergenic region

Chain B:



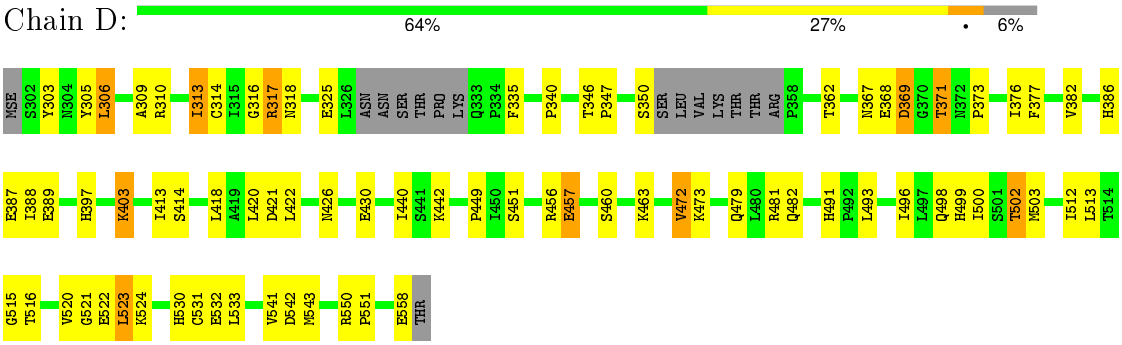
- Molecule 1: Hypothetical 28.8 kDa protein in PSD1-SKO1 intergenic region

Chain C:



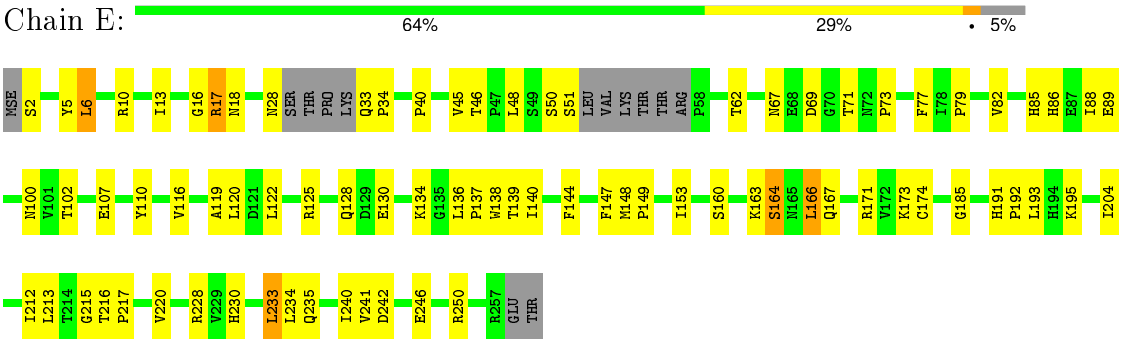
- Molecule 1: Hypothetical 28.8 kDa protein in PSD1-SKO1 intergenic region

Chain D:



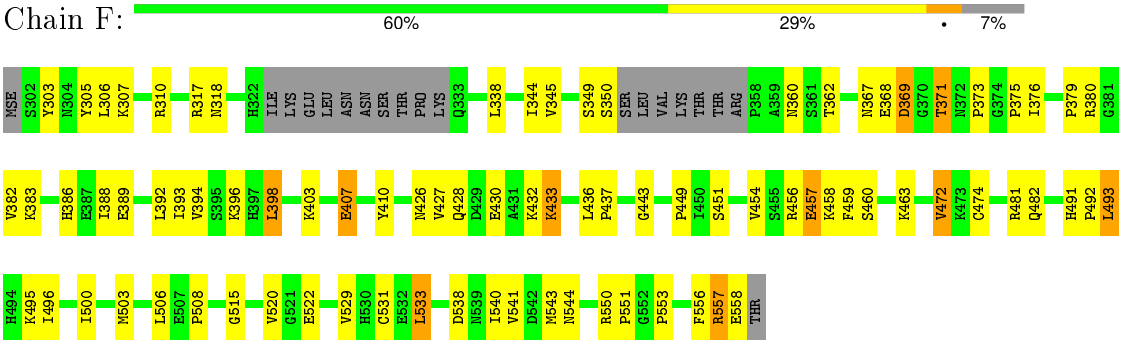
• Molecule 1: Hypothetical 28.8 kDa protein in PSD1-SKO1 intergenic region

Chain E:



• Molecule 1: Hypothetical 28.8 kDa protein in PSD1-SKO1 intergenic region

Chain F:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	57.70Å 85.99Å 316.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20	Depositor
% Data completeness (in resolution range)	(Not available) (50.00-2.20)	Depositor
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.216 , 0.262	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	12088	wwPDB-VP
Average B, all atoms (Å^2)	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: ACY, CA, SO4

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	0.35	0/1946	0.65	0/2625
1	B	0.35	0/1929	0.65	0/2602
1	C	0.34	0/1929	0.64	0/2603
1	D	0.36	0/1923	0.66	0/2595
1	E	0.35	0/1943	0.64	0/2620
1	F	0.41	0/1905	0.67	0/2568
All	All	0.36	0/11575	0.65	0/15613

There are no bond length outliers.

There are no bond angle outliers.

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	1911	0	1905	68	0
1	B	1894	0	1892	52	0
1	C	1894	0	1886	79	0
1	D	1888	0	1881	64	0
1	E	1908	0	1910	80	0
1	F	1870	0	1873	90	0
2	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	10	0	0	1	0
3	B	15	0	0	0	0
3	C	5	0	0	0	0
3	E	5	0	0	1	0
4	A	8	0	6	1	0
4	B	8	0	6	2	0
4	C	8	0	6	3	0
4	D	8	0	6	1	0
4	E	8	0	6	3	0
4	F	8	0	6	1	0
5	A	112	0	0	4	0
5	B	106	0	0	2	0
5	C	99	0	0	5	0
5	D	122	0	0	3	0
5	E	107	0	0	4	0
5	F	88	0	0	6	0
All	All	12088	0	11383	407	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 18.

The worst 5 of 407 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:E:148:MSE:HE3	1:E:213:LEU:HD11	1.38	1.01
1:F:367:ASN:HD22	1:F:371:THR:HG22	1.23	0.99
1:C:224:LYS:HD3	1:C:225:PRO:HD2	1.45	0.98
1:F:318:ASN:HD21	1:F:491:HIS:H	1.12	0.96
1:E:67:ASN:HD22	1:E:71:THR:HG23	1.31	0.94

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	241/259 (93%)	229 (95%)	10 (4%)	2 (1%)	24	22
1	B	238/259 (92%)	223 (94%)	14 (6%)	1 (0%)	39	42
1	C	239/259 (92%)	229 (96%)	8 (3%)	2 (1%)	24	22
1	D	238/259 (92%)	230 (97%)	6 (2%)	2 (1%)	24	22
1	E	240/259 (93%)	229 (95%)	9 (4%)	2 (1%)	24	22
1	F	234/259 (90%)	222 (95%)	12 (5%)	0	100	100
All	All	1430/1554 (92%)	1362 (95%)	59 (4%)	9 (1%)	30	29

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	59	ALA
1	B	551	PRO
1	A	17	ARG
1	C	17	ARG
1	A	165	ASN

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	214/224 (96%)	208 (97%)	6 (3%)	51	63
1	B	212/224 (95%)	203 (96%)	9 (4%)	36	44
1	C	212/224 (95%)	205 (97%)	7 (3%)	45	56

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	D	211/224 (94%)	198 (94%)	13 (6%)	23 25
1	E	214/224 (96%)	210 (98%)	4 (2%)	65 77
1	F	210/224 (94%)	195 (93%)	15 (7%)	18 19
All	All	1273/1344 (95%)	1219 (96%)	54 (4%)	36 44

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	325	GLU
1	D	472	VAL
1	F	493	LEU
1	D	368	GLU
1	D	371	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	182	GLN
1	D	367	ASN
1	F	386	HIS
1	C	230	HIS
1	D	318	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)

Of 25 ligands modelled in this entry, 6 are monoatomic - leaving 19 for Mogul analysis.

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	$\#\|Z\| > 2$
4	ACY	A	700	-	1,3,3	2.66	1 (100%)	0,3,3	0.00	-
4	ACY	A	701	2	1,3,3	2.52	1 (100%)	0,3,3	0.00	-
3	SO4	A	801	-	4,4,4	0.21	0	6,6,6	0.16	0
3	SO4	A	803	-	4,4,4	0.19	0	6,6,6	0.17	0
4	ACY	B	702	-	1,3,3	2.20	1 (100%)	0,3,3	0.00	-
4	ACY	B	703	2	1,3,3	2.72	1 (100%)	0,3,3	0.00	-
3	SO4	B	800	-	4,4,4	0.18	0	6,6,6	0.08	0
3	SO4	B	802	-	4,4,4	0.21	0	6,6,6	0.15	0
3	SO4	B	805	-	4,4,4	0.23	0	6,6,6	0.15	0
4	ACY	C	704	-	1,3,3	2.89	1 (100%)	0,3,3	0.00	-
4	ACY	C	705	2	1,3,3	2.45	1 (100%)	0,3,3	0.00	-
3	SO4	C	806	-	4,4,4	0.21	0	6,6,6	0.12	0
4	ACY	D	706	-	1,3,3	2.67	1 (100%)	0,3,3	0.00	-
4	ACY	D	707	2	1,3,3	2.80	1 (100%)	0,3,3	0.00	-
4	ACY	E	708	-	1,3,3	3.02	1 (100%)	0,3,3	0.00	-
4	ACY	E	709	2	1,3,3	2.52	1 (100%)	0,3,3	0.00	-
3	SO4	E	807	-	4,4,4	0.13	0	6,6,6	0.09	0
4	ACY	F	710	-	1,3,3	2.55	1 (100%)	0,3,3	0.00	-
4	ACY	F	711	2	1,3,3	2.66	1 (100%)	0,3,3	0.00	-

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
4	ACY	A	700	-	-	0/0/0/0	0/0/0/0
4	ACY	A	701	2	-	0/0/0/0	0/0/0/0
3	SO4	A	801	-	-	0/0/0/0	0/0/0/0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SO4	A	803	-	-	0/0/0/0	0/0/0/0
4	ACY	B	702	-	-	0/0/0/0	0/0/0/0
4	ACY	B	703	2	-	0/0/0/0	0/0/0/0
3	SO4	B	800	-	-	0/0/0/0	0/0/0/0
3	SO4	B	802	-	-	0/0/0/0	0/0/0/0
3	SO4	B	805	-	-	0/0/0/0	0/0/0/0
4	ACY	C	704	-	-	0/0/0/0	0/0/0/0
4	ACY	C	705	2	-	0/0/0/0	0/0/0/0
3	SO4	C	806	-	-	0/0/0/0	0/0/0/0
4	ACY	D	706	-	-	0/0/0/0	0/0/0/0
4	ACY	D	707	2	-	0/0/0/0	0/0/0/0
4	ACY	E	708	-	-	0/0/0/0	0/0/0/0
4	ACY	E	709	2	-	0/0/0/0	0/0/0/0
3	SO4	E	807	-	-	0/0/0/0	0/0/0/0
4	ACY	F	710	-	-	0/0/0/0	0/0/0/0
4	ACY	F	711	2	-	0/0/0/0	0/0/0/0

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	702	ACY	CH3-C	2.20	1.51	1.48
4	C	705	ACY	CH3-C	2.45	1.52	1.48
4	A	701	ACY	CH3-C	2.52	1.52	1.48
4	E	709	ACY	CH3-C	2.52	1.52	1.48
4	F	710	ACY	CH3-C	2.55	1.52	1.48

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

10 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	701	ACY	1	0
3	A	801	SO4	1	0
4	B	703	ACY	2	0
4	C	704	ACY	1	0
4	C	705	ACY	2	0
4	D	707	ACY	1	0
4	E	708	ACY	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	E	709	ACY	2	0
3	E	807	SO4	1	0
4	F	711	ACY	1	0

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.