



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 12, 2023 – 11:22 am GMT

PDB ID : 3ZEF  
Title : Crystal structure of Prp8:Aar2 complex: second crystal form at 3.1 Angstrom resolution  
Authors : Galej, W.P.; Oubridge, C.; Newman, A.J.; Nagai, K.  
Deposited on : 2012-12-05  
Resolution : 3.10 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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) : 2.36

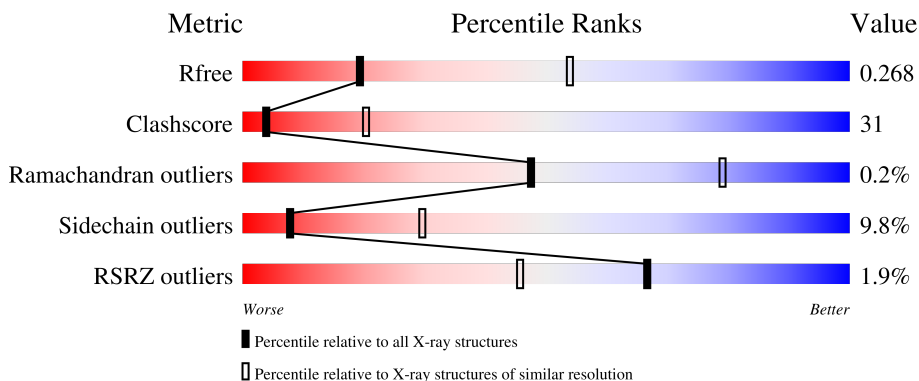
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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  $\geq 3$ , 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	355	 2% 75% 15% 8%
1	D	355	 67% 18% 11%
2	B	1531	 2% 52% 33% 6% 9%
2	E	1531	 2% 58% 29% 5% 7%

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 28285 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 A1 CISTRON-SPLICING FACTOR AAR2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	327	2684	1722	438	508	16	0	0	0
1	D	317	2618	1682	426	493	17	0	0	0

- Molecule 2 is a protein called PRE-MRNA-SPLICING FACTOR 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1398	11401	7329	1926	2109	37	0	0	0
2	E	1420	11582	7445	1958	2142	37	0	0	0

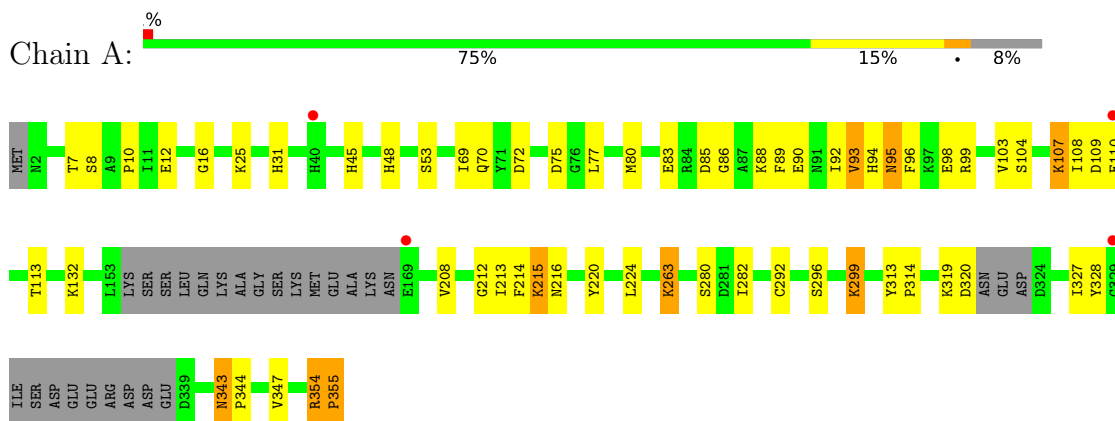
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	883	SER	-	expression tag	UNP P33334
B	884	GLY	-	expression tag	UNP P33334
B	1961	ASN	LEU	engineered mutation	UNP P33334
B	1999	LEU	ILE	engineered mutation	UNP P33334
E	883	SER	-	expression tag	UNP P33334
E	884	GLY	-	expression tag	UNP P33334
E	1961	ASN	LEU	engineered mutation	UNP P33334
E	1999	LEU	ILE	engineered mutation	UNP P33334

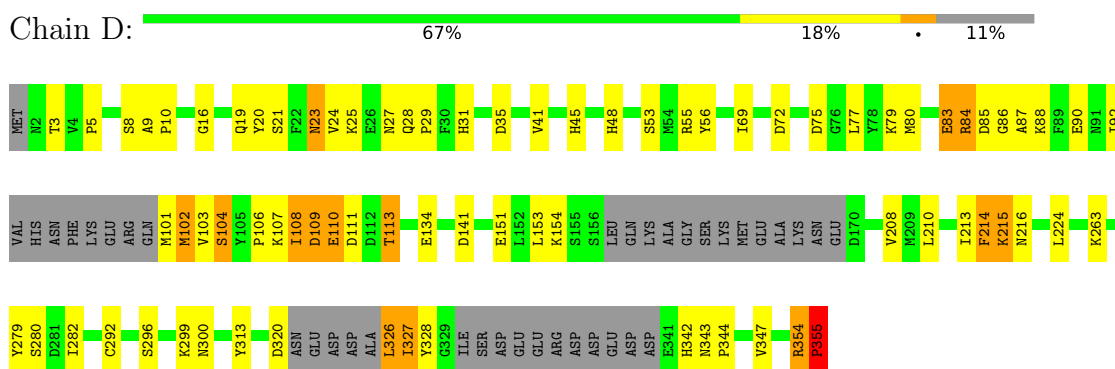
### 3 Residue-property plots [i](#)

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.

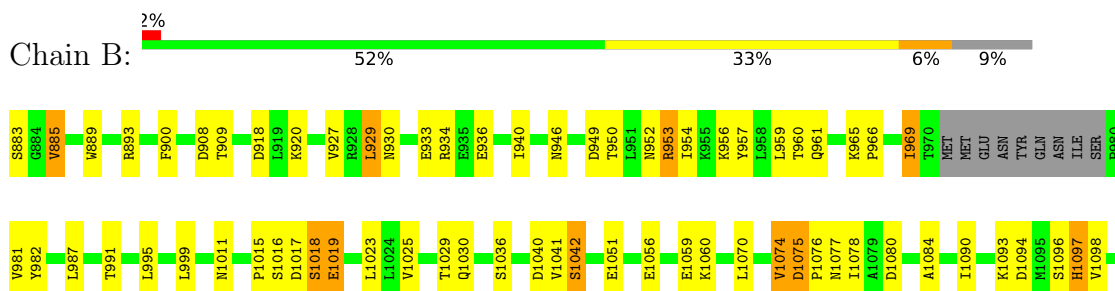
- Molecule 1: A1 CISTRON-SPLICING FACTOR AAR2



- Molecule 1: A1 CISTRON-SPLICING FACTOR AAR2



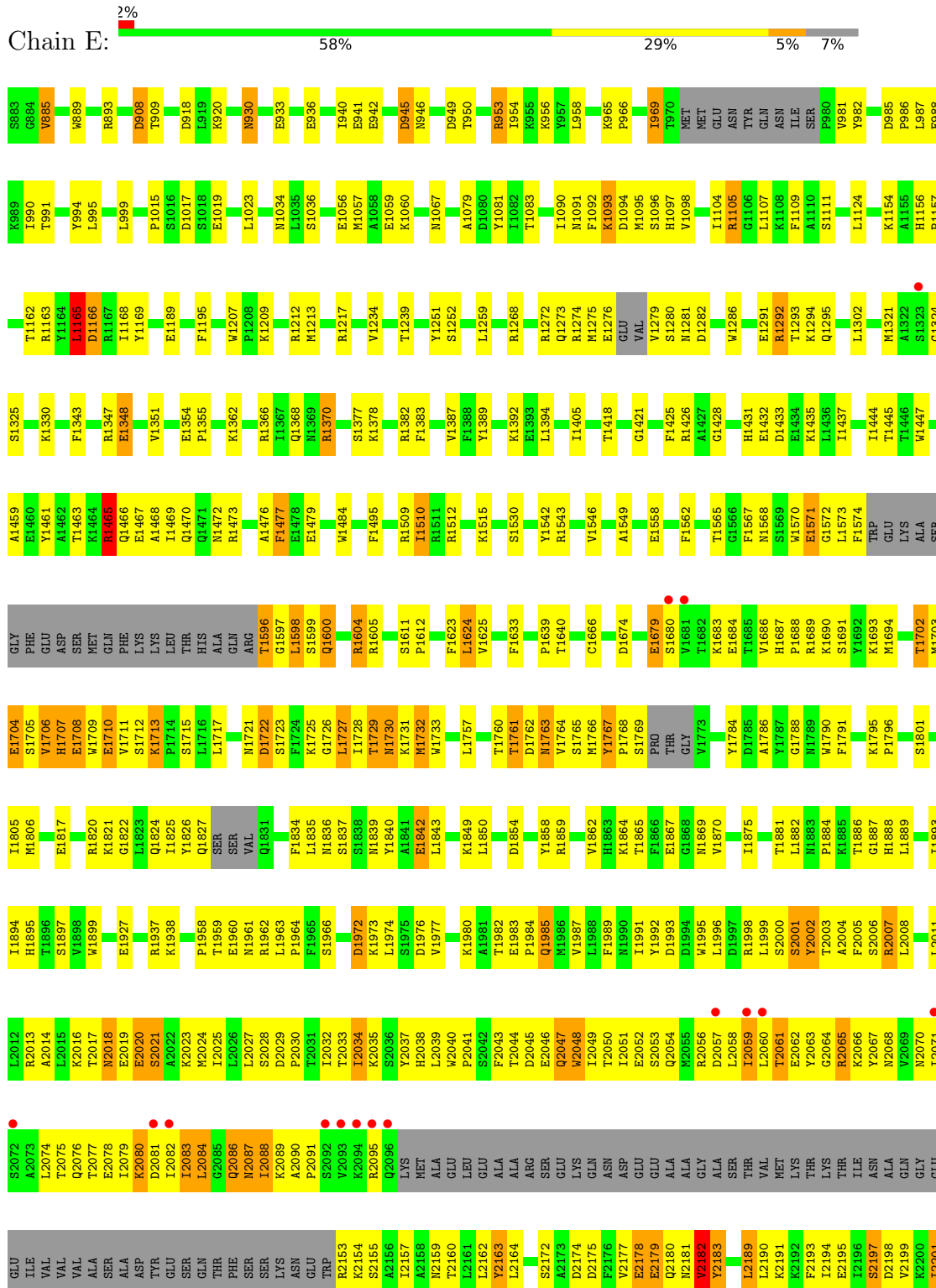
- Molecule 2: PRE-MRNA-SPLICING FACTOR 8

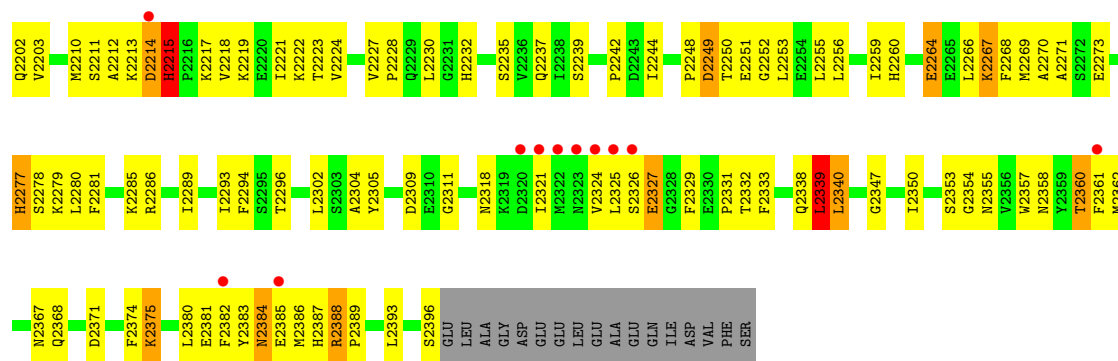


M2322	M2323	V2324	L2325	F2329	E2330	P2331	T2332	S2333	S2334	T2335	H2336	A2337	Q2338	L2339	L2340	L2341	S2342	D2343	R2344	T2345	T2346	G2347	N2348	F2349	I2350	I2351	P2352	S2353	G2354	N2355	V2356	F2357	N2358	F2361	M2362	N2367	Q2368	E2369	G2370	N2373	F2374	K2375	Y2376	P2379	L2380	E2381	F2382	G2383	E2384	E2385	M2386	H2387	ARG	PRO		
L2196	S2197	D2198	V2199	K2200	L2201	Q2202	V2203	F2206	L2207	Y2208	G2209	M2210	D2214	H2215	P2216	K2217	V2218	K2219	E2220	L2221	K2222	T2223	V2224	V2227	P2228	Q2229	L2230	G2231	H2232	V2233	S2234	S2235	S2236	Q2237	I2238	S2239	N2240	L2241	P2242	D2243	I2244	G2245	D2246	L2247	P2248	D2249	T2250	G2252	W2314	G2315	E2316	E2317	L2256	G2257	W2258	L2259
H2260	T2261	Q2262	T2263	E2264	L2266	K2267	F2268	M2269	A2270	E2273	V2274	A2275	T2276	H2277	S2278	K2279	L2280	F2281	A2282	D2283	K2284	K2285	R2286	D2287	C2288	L2289	D2290	L2291	S2292	L2293	F2294	S2295	T2296	P2297	V2300	S2301	L2302	S2303	A2304	Y2305	N2306	L2307	D2309	E2310	G2311	Y2312	Q2313	W2314	G2315	E2316	E2317	M2318	K2319	D2320	L2321	
E1842	L1843	F1844	M1845	M1846	L1848	K1849	L1850	F1851	V1852	D1853	D1854	T1855	M1856	F1857	L1858	R1859	K1864	T1865	F1866	E1867	H1868	P1869	R1869	K1870	S1871	M1876	L1877	F1880	T1881	M1882	M1883	P1884	K1885	T1886	G1887	H1888	L1889	F1890	L1891	Q1892	L1893	I1894	H1895	T1896	S1897	V1898	W1899	ALA	GLY	PHO	GLN	L1905	S1906	Q1907	L1908	A1909
V1917	V1921	V1921	K1926	E1927	E1928	Q1929	P1930	K1931	Q1932	L1941	L1944	E1945	M1948	R1957	P1958	T1959	E1960	N1961	R1962	L1963	M1969	D1972	K1973	L1974	S1975	D1976	V1977	V1978	M1979	K1980	E1983	P1984	Q1985	M1986	V1987	L1988	F1989	Q1990	S1991	I1991	Y1992	D1993	D1994	W1995	L1996	L1997	D1998	E2000	S2001	Y2002						
T2003	K2004	S2006	L2008	T2009	L2010	L2011	K2016	T2017	N2018	E2019	E2020	K2023	N2024	L2025	L2026	L2027	S2028	D2029	P2030	T2031	L2032	T2033	L2034	K2035	S2036	Y2037	H2038	L2039	W2040	P2041	S2042	F2043	T2044	D2045	E2046	Q2047	W2048	L2049	T2050	L2051	E2052	S2053	Q2054	M2055	R2056	D2057	L2058	L2059	L2060	T2061	E2062	Y2063	G2064	Y2067		
N2068	V2069	N2070	L2071	S2072	Q2076	T2079	K2080	Q2086	N2087	L2088	K2089	A2090	P2091	SER	VAL	LYS	ARG	GLN	LYS	MET	ALA	GLU	LEU	GLU	ALA	ALA	ALA	ALA	ARG	ARG	GLU	LYS	GLN	ASN	ASP	GLP	GLU	GLU	ALA	ALA	GLY	ALA	THR	THR	VAL	MET	LYS	THR	N2188	L2189	L2190	K2191	K2192	F2193	E2195	
GLU	GLU	ILE	VAL	VAL	VAL	ALA	SER	ALA	ASP	TYR	GLU	SER	GLN	THR	PHE	SER	SER	LYS	N2150	E2151	W2152	R2153	K2154	S2155	A2156	L2157	A2158	N2159	T2160	L2161	L2162	Y2163	L2164	R2165	L2166	K2167	D2174	F2176	V2177	E2178	E2179	Q2180	N2181	V2182	Y2183	V2184	L2185	P2186	K2187	N2188	L2189	L2190	K2191	K2192	F2193	E2195
L2122	M2123	Q1218	W1228	R1233	S1237	L1238	T1239	Y1251	S1252	K1253	L1259	F1260	Y1261	S1262	N1263	N1260	S1149	K1150	E1151	V1152	E1153	K1154	Q1273	M1275	L1276	GLU	VAL	SER	M1281	D1282	W1286	D1287	L1288	W1286	L1167	R1167	Y1169	D1177	E1180	E1181	D1154	E1185	Y1186	L1187	A1188	E1189	M1190	P1191	D1192	F1195	S1325	W1207				
I1140	F1343	E1348	E1354	P1355	Q1368	S1377	F1383	V1387	N1388	Y1389	L1394	I1405	L1406	T1430	H1431	E1432	D1433	E1434	K1435	L1436	I1437	I1405	L1406	H1431	E1432	D1433	E1434	K1435	L1436	I1437	I1444	W1447	D1453	T1463	K1464	R1465	Q1466	E1467	Q1471	M1472	R1473	R1474	L1475	A1476	F1477	E1478	E1479	R1486	F1495	A1503						
R1506	R1509	I1510	R1512	K1515	R1521	Y1542	V1546	E1558	K1563	M1568	M1569	M1570	E1571	L1573	F1574	M1575	LYS	ALA	SER	GLY	PHE	PHE	LYS	LEU	THR	HIS	ALA	Q1594	R1595	T1596	G1597	L1598	L1599	Q1599	Q1600	K1604	R1605	S1611	P1612	A1684																
L1624	V1625	F1633	P1639	T1640	I1661	C1666	L1673	D1674	Q1677	L1678	E1679	K1683	E1684	T1685	V1686	H1687	P1688	R1689	K1690	S1691	Y1692	K1693	M1703	E1704	S1705	V1706	H1707	E1708	W1709	E1710	K1713	L1715	L1717	H1718	E1719	T1720	M1721	D1722	L1727	L1728	T1729	K1730	M1731	M1732	W1733	R1739	D1742									
D1747	R1750	Y1759	D1762	M1763	V1764	P1768	T1771	G1772	Y1784	D1785	A1786	F1791	K1795	P1796	L1797	M1800	S1801	M1802	R1803	T1804	I1805	M1806	K1807	A1808	M1809	T1886	E1817	K1821	G1822	L1823	Q1824	L1825	T1825	GLN	SER	GLY	SER	VAL	GLN	GLU	PHO	PHE	GLN	LYS	ASN	SER	ASN	Y1840	A1841							

VAL  
HIS  
PHE  
LEU  
GLN  
PHE  
SER  
GLU  
LEU  
LEU  
ALA  
GLY  
ASP  
GLU  
GLU  
LEU  
LEU  
ALA  
GLU  
GLN  
ILE  
ASP  
VAL  
PHE  
SER

● Molecule 2: PRE-MRNA-SPLICING FACTOR 8





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	125.42Å 177.84Å 216.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	137.45 – 3.10 92.50 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (137.45-3.10) 99.8 (92.50-3.00)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.69 (at 3.01Å)	Xtrriage
Refinement program	REFMAC 5.8.0016	Depositor
R, $R_{free}$	0.220 , 0.267 0.221 , 0.268	Depositor DCC
$R_{free}$ test set	4874 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	79.1	Xtrriage
Anisotropy	0.449	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 67.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	28285	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	96.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 20.50 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.5812e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 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	0/2754	0.72	2/3725 (0.1%)
1	D	0.56	0/2685	0.75	4/3626 (0.1%)
2	B	0.59	1/11671 (0.0%)	0.79	5/15827 (0.0%)
2	E	0.58	3/11856 (0.0%)	0.77	8/16074 (0.0%)
All	All	0.58	4/28966 (0.0%)	0.77	19/39252 (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	2
1	D	0	1
2	B	0	6
2	E	0	5
All	All	0	14

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	1708	GLU	CD-OE1	8.75	1.35	1.25
2	E	1571	GLU	CD-OE1	7.11	1.33	1.25
2	B	1164	TYR	CE2-CZ	-6.19	1.30	1.38
2	E	1708	GLU	CD-OE2	6.10	1.32	1.25

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	2339	LEU	CA-CB-CG	7.05	131.52	115.30
2	E	1105	ARG	NE-CZ-NH1	6.47	123.54	120.30
1	D	355	PRO	CA-N-CD	-6.36	102.59	111.50
1	A	354	ARG	C-N-CD	6.34	141.71	128.40

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	1465	ARG	NE-CZ-NH1	6.09	123.35	120.30

There are no chirality outliers.

5 of 14 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	107	LYS	Peptide
1	A	93	VAL	Peptide
2	B	1164	TYR	Peptide
2	B	1433	ASP	Peptide
2	B	1704	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	2684	0	2530	54	0
1	D	2618	0	2494	81	0
2	B	11401	0	11345	879	1
2	E	11582	0	11528	721	1
All	All	28285	0	27897	1719	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 31.

The worst 5 of 1719 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)
2:B:2043:PHE:CE2	2:B:2051:ILE:HD11	1.31	1.63
2:B:2259:ILE:HD13	2:B:2291:ILE:CB	1.30	1.60
2:B:2183:TYR:CE1	2:B:2219:LYS:HD3	1.45	1.49
2:B:2183:TYR:HE1	2:B:2219:LYS:CD	1.26	1.47
2:B:2236:VAL:HG22	2:B:2238:ILE:CD1	1.44	1.47

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)
2:B:1685:THR:OG1	2:E:2309:ASP:OD2[4_545]	1.96	0.24

## 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	319/355 (90%)	310 (97%)	9 (3%)	0	100	100
1	D	307/355 (86%)	298 (97%)	8 (3%)	1 (0%)	41	73
2	B	1384/1531 (90%)	1345 (97%)	38 (3%)	1 (0%)	51	83
2	E	1406/1531 (92%)	1360 (97%)	42 (3%)	4 (0%)	41	73
All	All	3416/3772 (91%)	3313 (97%)	97 (3%)	6 (0%)	47	79

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	E	2088	ILE
2	E	2286	ARG
2	B	2088	ILE
1	D	29	PRO
2	E	1726	GLY

### 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	293/326 (90%)	284 (97%)	9 (3%)	40	70

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	289/326 (89%)	269 (93%)	20 (7%)	15	45
2	B	1256/1373 (92%)	1108 (88%)	148 (12%)	5	21
2	E	1278/1373 (93%)	1151 (90%)	127 (10%)	8	29
All	All	3116/3398 (92%)	2812 (90%)	304 (10%)	8	29

5 of 304 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	1729	THR
2	E	2277	HIS
2	E	1769	SER
2	E	2048	TRP
2	E	2384	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	1097	HIS
2	E	1655	GLN
2	E	2384	ASN
2	E	1140	ASN
2	E	1470	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 monosaccharides 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	327/355 (92%)	-0.32	4 (1%) 79 61	40, 75, 133, 168	1 (0%)
1	D	317/355 (89%)	-0.37	0 100 100	38, 73, 131, 170	0
2	B	1398/1531 (91%)	-0.10	35 (2%) 57 34	38, 103, 176, 253	0
2	E	1420/1531 (92%)	-0.15	26 (1%) 68 47	37, 90, 160, 245	0
All	All	3462/3772 (91%)	-0.17	65 (1%) 66 46	37, 92, 165, 253	1 (0%)

The worst 5 of 65 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	2096	GLN	15.9
2	E	2093	VAL	6.8
2	E	2323	ASN	6.5
2	B	2385	GLU	5.8
2	E	2321	ILE	5.4

### 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 monosaccharides 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.