



# wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 10:43 AM GMT

PDB ID : 2HGQ  
Title : Crystal structure of the 70S *Thermus thermophilus* ribosome with translocated and rotated Shine-Dalgarno Duplex. This entry 2HGQ contains 50S ribosomal subunit. The 30S ribosomal subunit can be found in PDB entry 2HGP.  
Authors : Jenner, L.; Yusupova, G.; Rees, B.; Moras, D.; Yusupov, M.  
Deposited on : 2006-06-27  
Resolution : 5.50 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

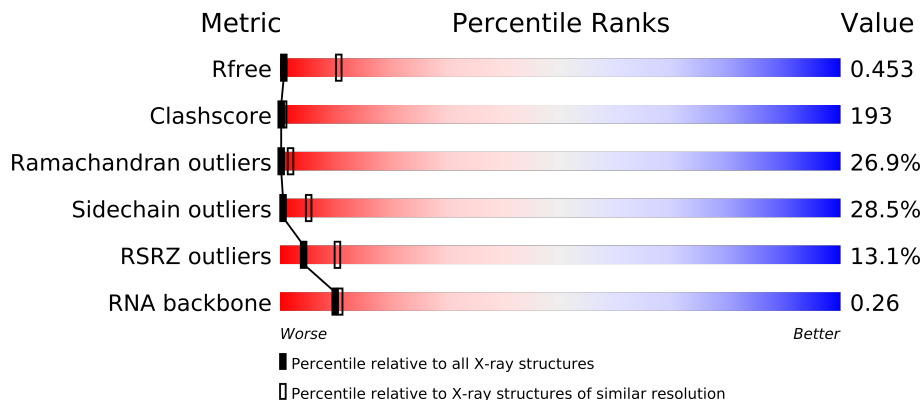
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance



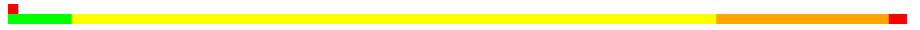
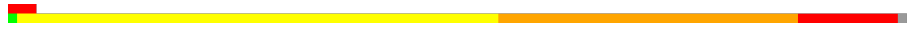

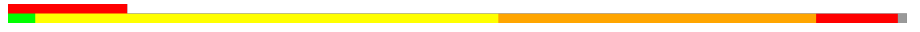
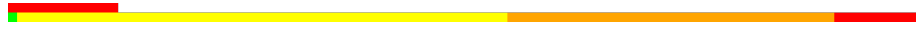





The reported resolution of this entry is 5.50 Å.

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))
$R_{free}$	66092	1083 (7.50-3.50)
Clashscore	79885	1008 (7.40-3.52)
Ramachandran outliers	78287	1275 (7.50-3.50)
Sidechain outliers	78261	1251 (7.50-3.50)
RSRZ outliers	66119	1082 (7.50-3.50)
RNA backbone	1838	1040 (7.50-2.80)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	2916	
2	B	123	
3	C	229	
4	D	276	
5	E	206	
6	F	210	
7	G	182	
8	H	180	
9	K	148	
10	L	147	
11	M	140	
12	N	122	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
13	O	150	
14	P	141	
15	Q	118	
16	R	112	
17	S	146	
18	T	118	
19	U	101	
20	V	113	
21	W	96	
22	X	110	
23	Y	206	
24	Z	85	
25	1	67	
26	2	60	
27	3	71	
28	4	60	
29	5	54	
30	6	49	
31	7	65	
32	8	37	

## 2 Entry composition

There are 32 unique types of molecules in this entry. The entry contains 94012 atoms, of which 0 are hydrogen and 0 are deuterium.

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 RNA chain called 23S ribosomal RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	2889	Total	C	N	O	P	0	0	0
			62218	27691	11629	20009	2889			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	493	G	-	INSERTION	GB 48268

- Molecule 2 is a RNA chain called 5S ribosomal RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	123	Total	C	N	O	P	0	0	0
			2641	1175	488	855	123			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	A	-	INSERTION	GB 48271
B	120	U	-	INSERTION	GB 48271

- Molecule 3 is a protein called 50S ribosomal protein L1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	228	Total	C	N	O	S	0	0	0
			1742	1102	318	319	3			

- Molecule 4 is a protein called 50S ribosomal protein L2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	272	Total	C	N	O	S	0	0	0
			2124	1339	424	358	3			

- Molecule 5 is a protein called 50S ribosomal protein L3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	206	Total	C	N	O	S	0	0	0
			1578	997	302	273	6			

- Molecule 6 is a protein called 50S ribosomal protein L4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	208	Total	C	N	O	S	0	0	0
			1625	1034	303	286	2			

- Molecule 7 is a protein called 50S ribosomal protein L5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	182	Total	C	N	O	S	0	0	0
			1482	947	269	261	5			

- Molecule 8 is a protein called 50S ribosomal protein L6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	174	Total	C	N	O	S	0	0	0
			1328	844	248	235	1			

- Molecule 9 is a protein called 50S ribosomal protein L9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	K	148	Total	C	N	O	S	0	0	0
			1155	737	205	212	1			

- Molecule 10 is a protein called 50S ribosomal protein L11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	L	138	Total	C	N	O	S	0	0	0
			1025	654	181	185	5			

- Molecule 11 is a protein called 50S ribosomal protein L13.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	M	139	Total	C	N	O	S	0	0	0
			1113	717	207	186	3			

- Molecule 12 is a protein called 50S ribosomal protein L14.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	N	122	Total	C	N	O	S	0	0	0
			932	587	171	170	4			

- Molecule 13 is a protein called 50S ribosomal protein L15.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	O	145	Total	C	N	O	S	0	0	0
			1106	688	226	190	2			

- Molecule 14 is a protein called 50S ribosomal protein L16.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	P	136	Total	C	N	O	S	0	0	0
			1080	688	204	183	5			

- Molecule 15 is a protein called 50S ribosomal protein L17.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
15	Q	117	Total	C	N	O	S	0	0	0
			960	599	202	159				

- Molecule 16 is a protein called 50S ribosomal protein L18.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
16	R	110	Total	C	N	O	S	0	0	0
			877	553	175	149				

- Molecule 17 is a protein called 50S ribosomal protein L19.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
17	S	117	Total	C	N	O	S	0	0	0
			976	614	197	164	1			

- Molecule 18 is a protein called 50S ribosomal protein L20.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
18	T	117	Total	C	N	O	S	0	0	0
			964	610	202	151	1			

- Molecule 19 is a protein called 50S ribosomal protein L21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
19	U	101	Total	C	N	O	S	0	0	0
			779	501	142	135	1			

- Molecule 20 is a protein called 50S ribosomal protein L22.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
20	V	110	Total	C	N	O	S	0	0	0
			876	552	171	151	2			

- Molecule 21 is a protein called 50S ribosomal protein L23.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
21	W	94	Total	C	N	O	S	0	0	0
			742	483	133	126				

- Molecule 22 is a protein called 50S ribosomal protein L24.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
22	X	110	Total	C	N	O	S	0	0	0
			844	539	158	141	6			

- Molecule 23 is a protein called 50S ribosomal protein L25.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
23	Y	180	Total	C	N	O	S	0	0	0
			1435	916	256	260	3			

- Molecule 24 is a protein called Ribosomal protein L27.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
24	Z	85	Total	C	N	O	S	0	0	0
			670	415	141	112	2			

- Molecule 25 is a protein called 50S ribosomal protein L29.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
25	1	67	Total	C	N	O	S	0	0	0
			567	350	116	99	2			

- Molecule 26 is a protein called 50S ribosomal protein L30.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
26	2	59	Total	C	N	O			
			469	298	90	81	0	0	0

- Molecule 27 is a protein called 50S ribosomal protein L31.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
27	3	71	Total	C	N	O	S			
			581	364	108	104	5	0	0	0

- Molecule 28 is a protein called 50S ribosomal protein L32.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
28	4	57	Total	C	N	O	S			
			445	279	87	74	5	0	0	0

- Molecule 29 is a protein called 50S ribosomal protein L33.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
29	5	49	Total	C	N	O	S			
			426	265	87	70	4	0	0	0

- Molecule 30 is a protein called 50S ribosomal protein L34.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
30	6	49	Total	C	N	O	S			
			430	263	108	57	2	0	0	0

- Molecule 31 is a protein called 50S ribosomal protein L35.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
31	7	64	Total	C	N	O	S			
			515	331	102	79	3	0	0	0

- Molecule 32 is a protein called 50S ribosomal protein L36.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
32	8	37	Total	C	N	O	S			
			307	188	68	47	4	0	0	0

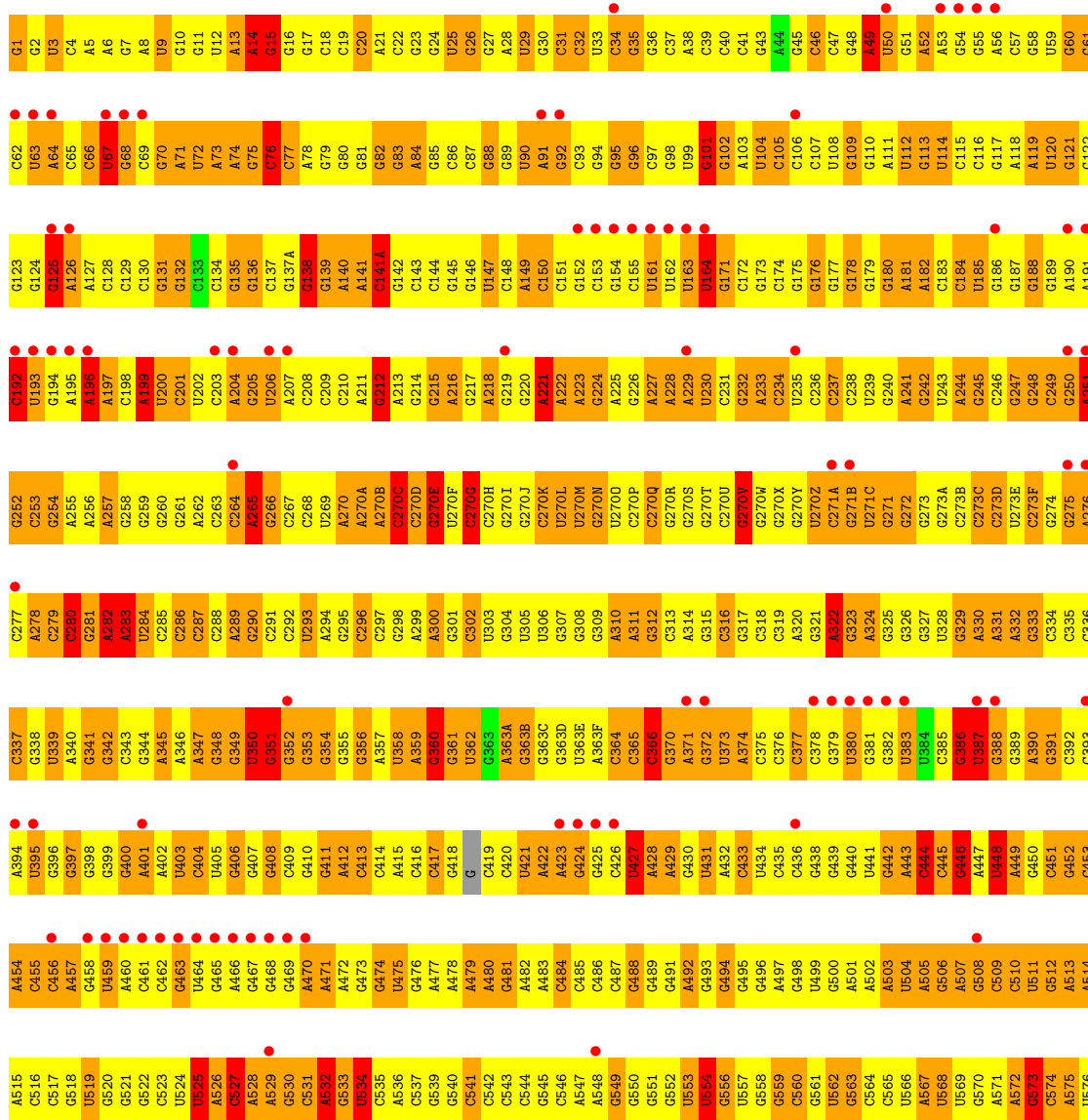


### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

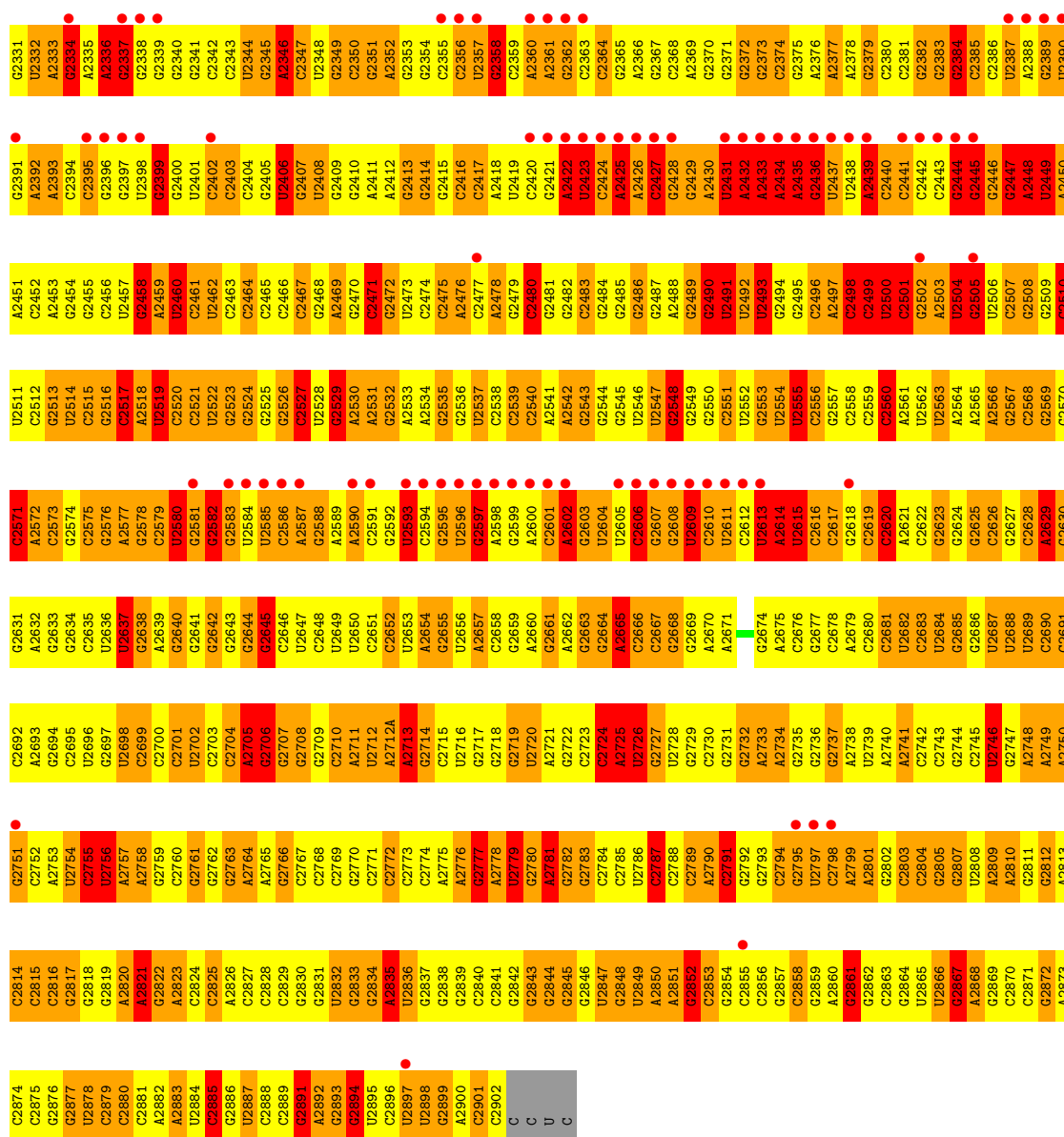
#### • Molecule 1: 23S ribosomal RNA

Chain A:



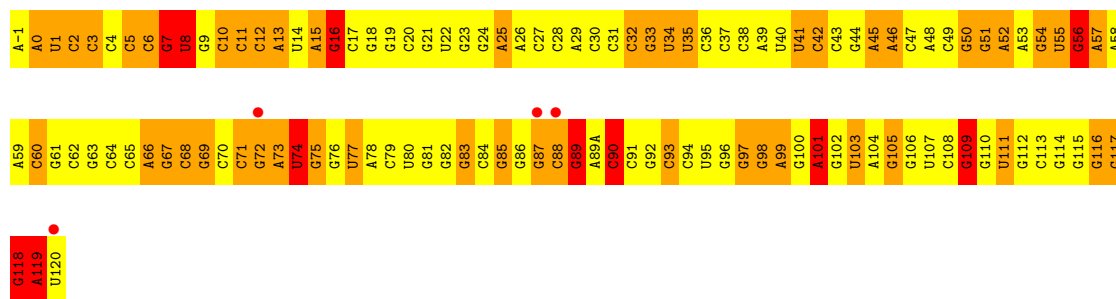


G2271	U2272	A2273	A2274	G2275	G2276	G2277	A2278	G2279	G2280	G2281	G2282	G2283	A2284	G2285	A2286	A2287	A2288	G2289	G2290	U2291	G2292	G2293	G2294	G2295	G2296	A2297	G2298	G2299	G2300	G2301	G2302	G2303	G2304	G2305	G2306	G2307	G2308	A2309	A2310	A2311	U2312	G2313	G2314	G2315	G2316	G2317	G2318	G2319	A2320	G2321	G2322	G2323	G2324	G2325	G2326	A2327	A2328	G2329	G2330																	
G2141	G2142	A2143	U2144	G2145	G2146	G2147	A2148	G2149	G2150	G2151	G2152	G2153	G2154	A2155	G2156	G2157	A2158	G2159	G2160	G2161	G2162	G2163	G2164	G2165	G2166	G2167	G2168	G2169	G2170	A2171	G2172	U2173	G2174	G2175	G2176	G2177	G2178	G2179	U2180	G2181	G2182	G2183	G2184	A2185	G2186	G2187	G2188	G2189	G2190	G2191	G2192	G2193	G2194	A2195	G2196	A2197	A2198	G2199	G2200																	
G2080	G2081	A2082	G2083	G2084	G2085	G2086	G2087	G2088	G2089	G2090	G2091	U2092	G2093	G2094	G2095	U2096	G2097	G2098	G2099	G2100	G2101	G2102	G2103	G2104	G2105	G2106	G2107	G2108	G2109	G2110	G2111	G2112	U2113	G2114	G2115	G2116	G2117	G2118	A2119	G2120	G2121	G2122	G2123	G2124	G2125	A2126	G2127	G2128	G2129	G2130	G2131	G2132	G2133	G2134	A2135	G2136	G2137	G2138	G2139	G2140																
A2020	G2021	U2022	G2023	G2024	G2025	G2026	G2027	G2028	G2029	A2030	A2031	G2032	A2033	G2034	G2035	G2036	G2037	G2038	G2039	G2040	U2041	G2042	G2043	G2044	G2045	G2046	U2047	G2048	G2049	G2050	A2051	G2052	G2053	G2054	G2055	G2056	G2057	A2058	G2059	A2060	G2061	A2062	G2063	G2064	G2065	G2066	G2067	U2068	G2069	G2070	A2071	G2072	G2073	U2074	G2075	A2076	A2077	G2078	U2079																	
G1959	G1960	A1961	A1962	A1963	G1964	G1965	G1966	G1967	A1968	G1969	A1970	A1971	G1972	G1973	G1974	U1975	U1976	A1977	A1978	A1979	A1980	A1981	G1982	G1983	G1984	G1985	A1986	G1987	G1988	G1989	G1990	U1991	G1992	G1993	U1994	U1995	G1996	G1997	G1998	A1999	G1999	G2000	A2001	G2002	G2003	G2004	A2005	G2006	G2007	G2008	G2009	G2010	U2011	G2012	A2013	A2014	A2015	U2016	U2017	G2018	G2019															
G1899	A1900	A1901	G1902	G1903	G1904	G1905	G1906	G1907	G1908	G1909	G1910	U1911	G1912	A1913	A1914	U1915	A1916	A1917	A1918	A1919	G1920	G1921	G1922	A1923	G1924	G1925	U1926	U1927	A1928	G1929	A1930	U1931	G1932	G1933	G1934	G1935	A1936	A1937	A1938	U1939	U1940	G1941	G1942	G1943	G1944	G1945	U1946	G1947	G1948	G1949	G1950	U1951	G1952	A1953	G1954	U1955	U1956	G1957	G1958																	
G1830	G1831	G1832	G1833	U1834	G1835	G1836	G1837	G1838	G1839	G1840	U1841	G1842	G1843	G1844	G1845	G1846	A1847	A1848	G1849	G1850	A1851	G1852	A1853	A1854	G1855	G1856	G1857	G1858	A1859	G1860	G1861	A1862	G1863	U1864	G1865	G1866	G1867	G1868	G1869	G1870	A1871	A1872	G1873	G1874	G1875	G1876	A1877	G1878	A1879	G1880	G1881	G1882	G1883	G1884	A1885	G1886	G1887	G1888	G1889	G1890	G1891	G1892	G1893	G1894	G1895	G1896	G1897	G1898								
G1770	G1771	G1772	U1773	G1774	U1775	G1776	U1777	U1778	G1779	A1780	U1781	G1782	U1783	A1784	U1785	U1786	A1787	G1788	A1789	G1790	A1791	G1792	G1793	U1794	G1795	U1796	G1797	U1798	G1799	G1800	A1801	A1802	A1803	G1804	U1805	G1806	G1807	U1808	A1809	A1810	G1811	G1812	G1813	G1814	A1815	G1816	G1817	U1818	A1819	G1820	A1821	G1822	G1823	G1824	A1825	G1826	G1827	G1828	A1829	G1830																
G1694	G1695	G1696	G1697	A1698	G1699	G1700	A1701	G1702	G1703	G1704	G1705	U1706	G1707	G1708	U1709	G1710	G1711	G1712	A1713	G1714	G1715	A1716	G1717	G1718	G1719	G1720	G1721	G1722	G1723	G1724	G1725	G1726	U1727	G1728	A1729	G1730	G1731	A1732	G1733	G1734	G1735	G1736	G1737	G1738	A1739	G1740	G1741	G1742	G1743	G1744	G1745	G1746	G1747	G1748	A1749	G1750	G1751	G1752	G1753	G1754	A1755	G1756	U1757	G1758	A1759	G1760	G1761	G1762	G1763	G1764	G1765	U1766	G1767	G1768	U1769	G1770
A1634	G1635	G1636	A1637	G1638	U1639	A1640	G1641	G1642	G1643	G1644	G1645	G1646	G1647	G1648	G1649	G1650	G1651	A1652	G1653	A1654	G1655	A1656	G1657	G1658	G1659	G1660	G1661	G1662	G1663	A1664	A1665	G1666	G1667	A1668	A1669	A1670	G1671	G1672	U1673	G1674	G1675	A1676	A1677	G1678	U1679	G1680	G1681	G1682	G1683	G1684	G1685	G1686	G1687	U1688	A1689	A1690	G1691	G1692	G1693	A1694	A1695	A1696	G1697	G1698	U1699	A1700										
C1574	C1575	U1576	U1577	C1578	A1579	C1580	C1581	C1582	C1583	C1584	C1585	A1586	A1587	C1588	C1589	U1590	G1591	C1592	C1593	G1594	C1595	A1596	U1597	C1598	C1599	C1600	U1601	U1602	A1603	C1604	C1605	G1606	C1607	A1608	A1609	A1610	C1611	C1612	C1613	C1614	C1615	C1616	A1617	C1618	C1619	G1620	U1621	G1622	G1623	G1624	C1625	C1626	G1627	G1628	U1629	C1630	A1631	G1632	C1633	C1634	C1635	C1636	C1637	C1638	C1639	C1640										
C1515	U1516	C1517	C1518	C1519	U1520	C1521	C1522	C1523	C1524	C1525	C1526	C1527	C1528	A1529	C1530	C1531	C1532	C1533	C1534	U1535	A1536	C1537	C1538	C1539	C1540	U1541	G1542	A1543	C1544	C1545	C1546	C1547	C1548	C1549	C1550	C1551	C1552	A1553	C1554	C1555	C1556	C1557	C1558	G1559	G1560	C1561	C1562	C1563	C1564	C1565	A1566	A1567	G1568	A1569	A1570	A1571	C1572	C1573	C1574	C1575	C1576	C1577	C1578	C1579	C1580											
U1454	G1455	G1456	A1457	C1458	G1459	A1460	A1461	C1462	C1463	C1464	C1465	G1466	C1467	A1468	A1469	G1470	C1471	A1472	G1473	C1474	U1475	C1476	C1477	A1478	G1479	A1480	U1481	G1482	G1483	G1484	G1485	A1486	G1487	G1488	U1489	A1490	G1491	G1492	C1493	A1494	A1495	A1496	U1497	C1498	C1499	G1499	C1500	C1501	C1502	U1503	C1504	C1505	C1506	A1507	A1508	C1509	A1510	A1511	G1512	C1513	C1514	C1515	C1516	C1517	C1518	C1519	C1520									



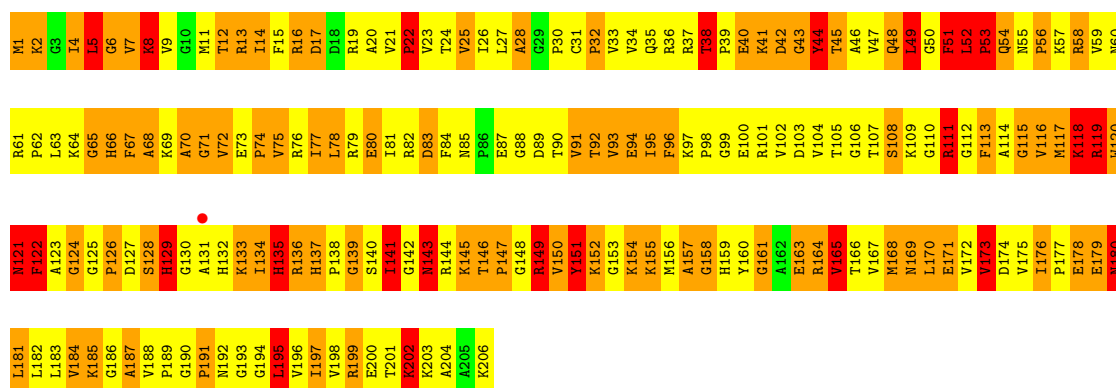
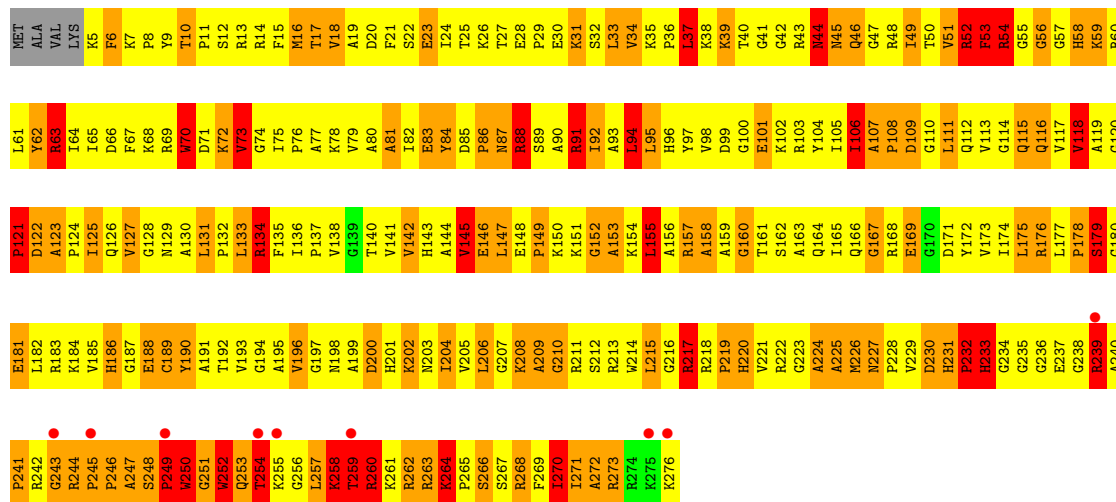
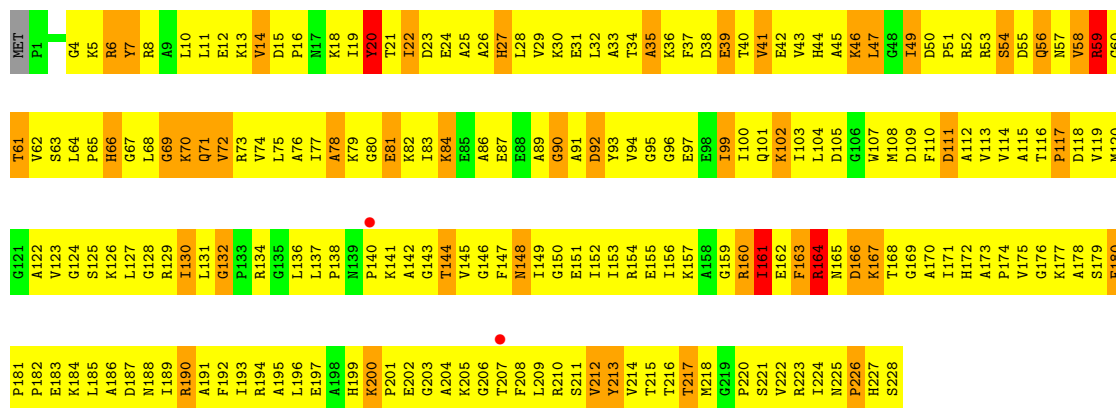
• Molecule 2: 5S ribosomal RNA

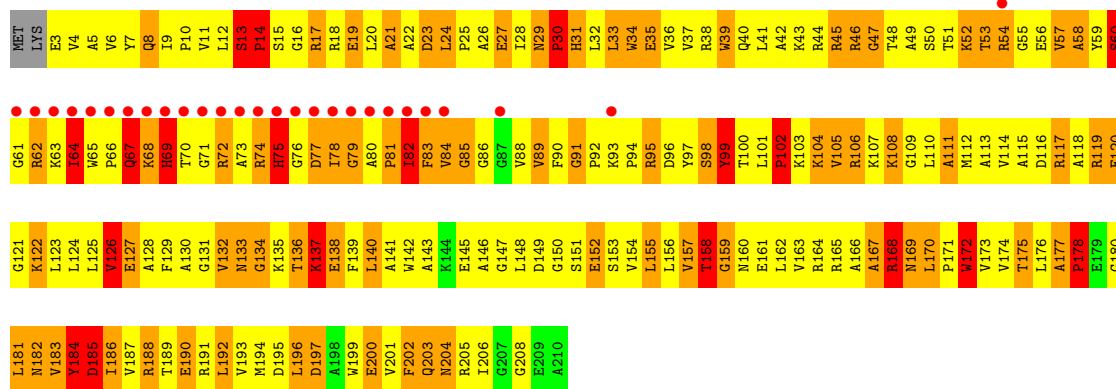
Chain B:



• Molecule 3: 50S ribosomal protein L1

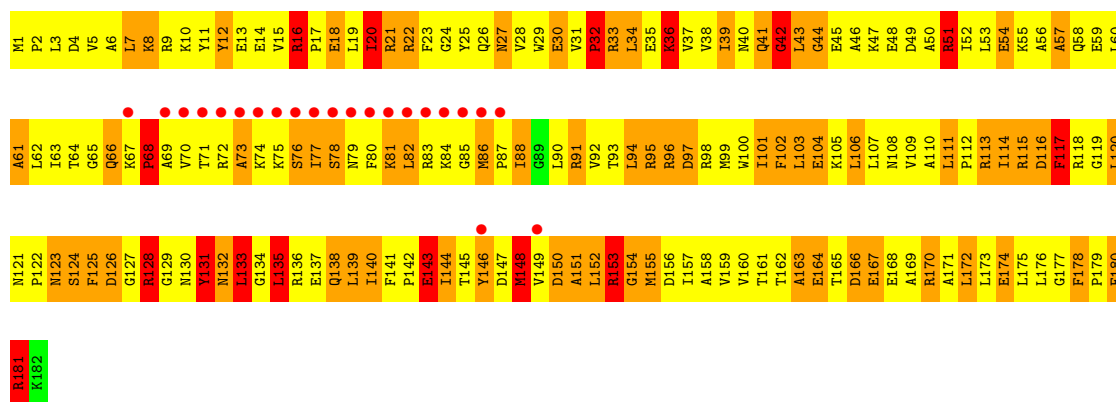
Chain C:





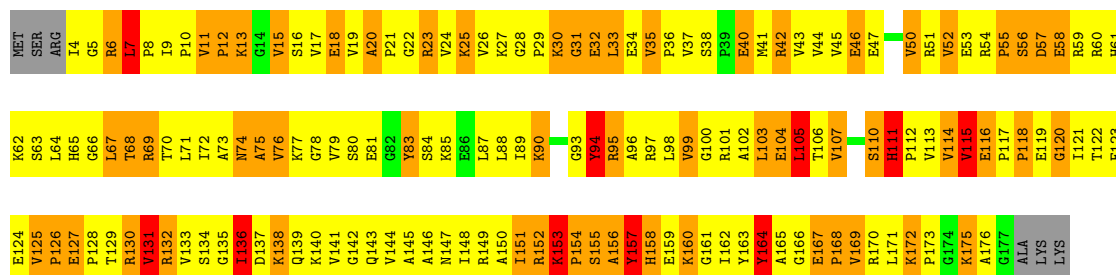
• Molecule 7: 50S ribosomal protein L5

Chain G:



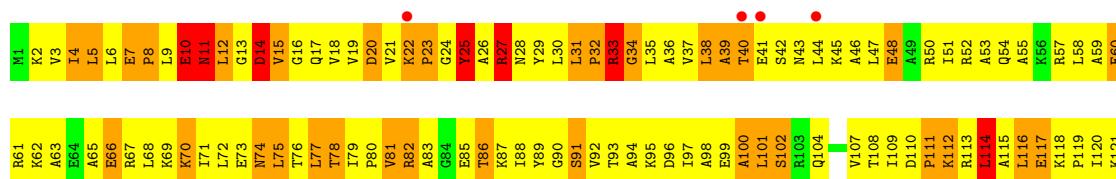
• Molecule 8: 50S ribosomal protein L6

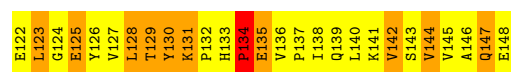
Chain H:



• Molecule 9: 50S ribosomal protein L9

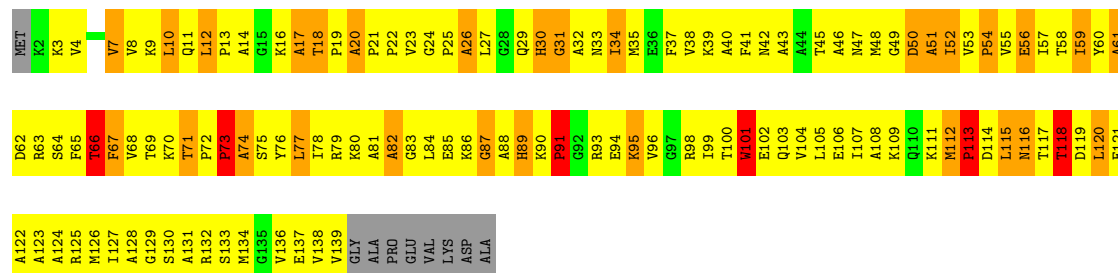
Chain K:





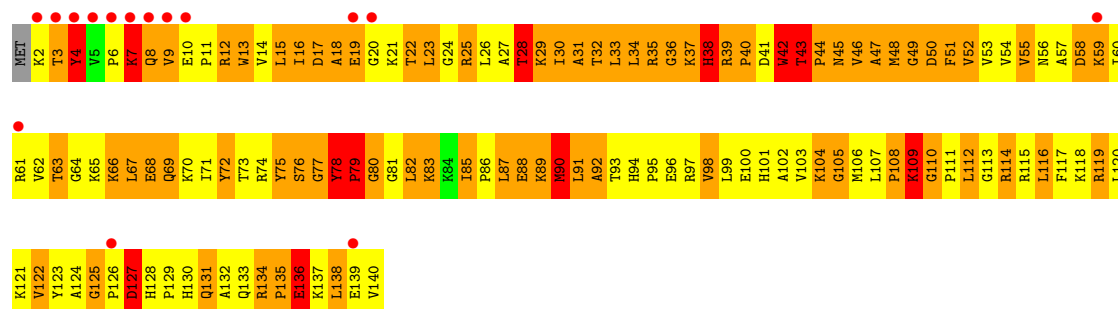
• Molecule 10: 50S ribosomal protein L11

Chain L:



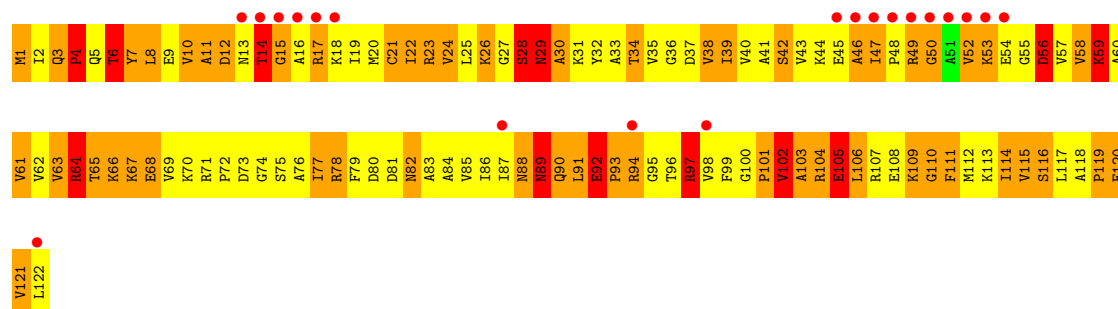
• Molecule 11: 50S ribosomal protein L13

Chain M:



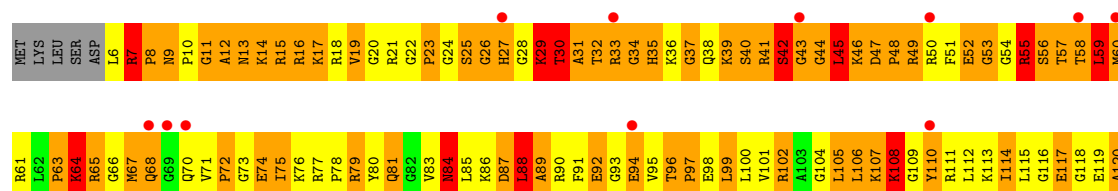
• Molecule 12: 50S ribosomal protein L14

Chain N:



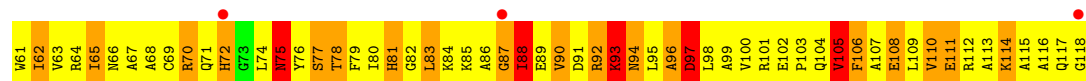
• Molecule 13: 50S ribosomal protein L15

Chain O:









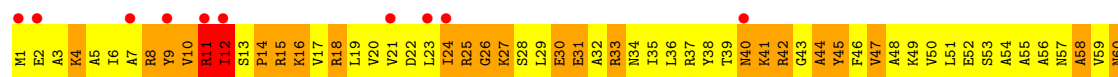
• Molecule 19: 50S ribosomal protein L21

Chain U:



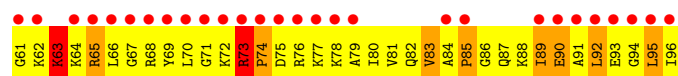
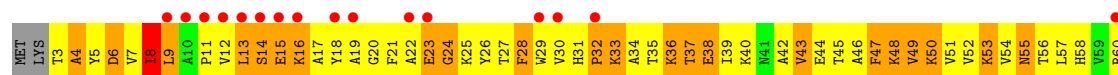
• Molecule 20: 50S ribosomal protein L22

Chain V:



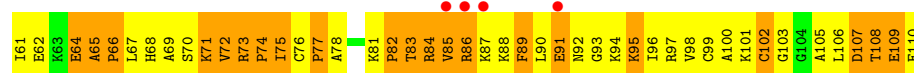
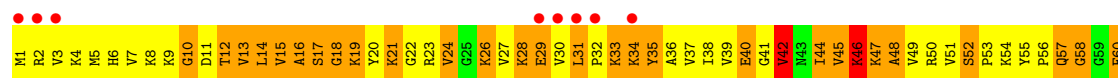
• Molecule 21: 50S ribosomal protein L23

Chain W:



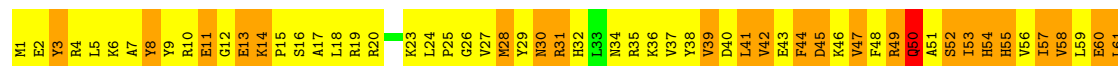
• Molecule 22: 50S ribosomal protein L24

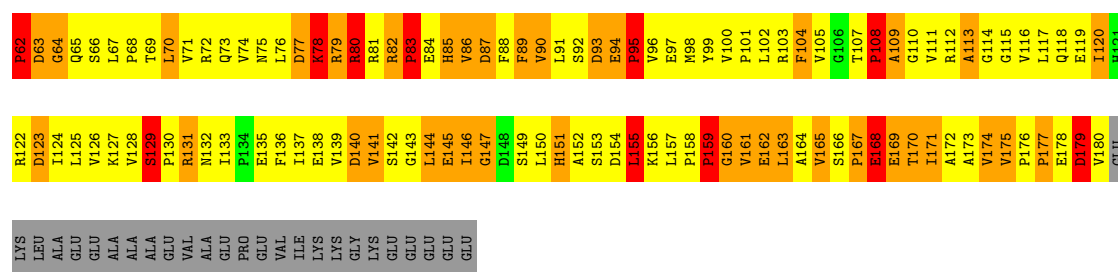
Chain X:



• Molecule 23: 50S ribosomal protein L25

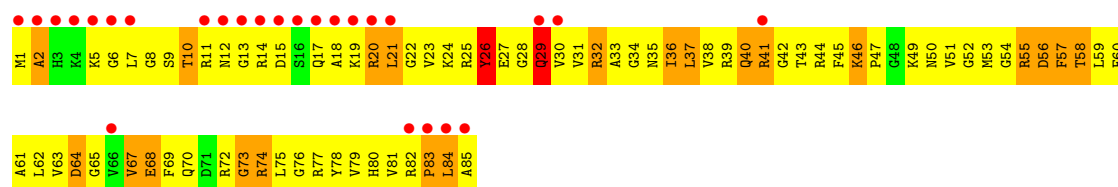
Chain Y:





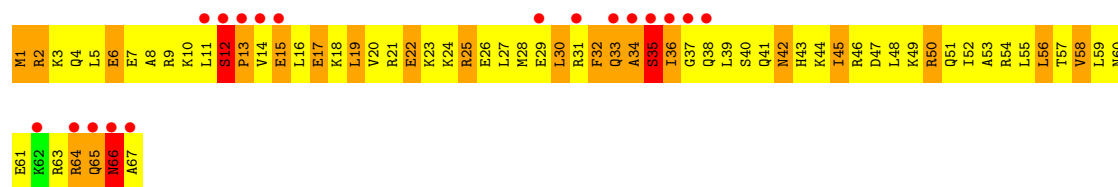
• Molecule 24: Ribosomal protein L27

Chain Z:



• Molecule 25: 50S ribosomal protein L29

Chain 1:



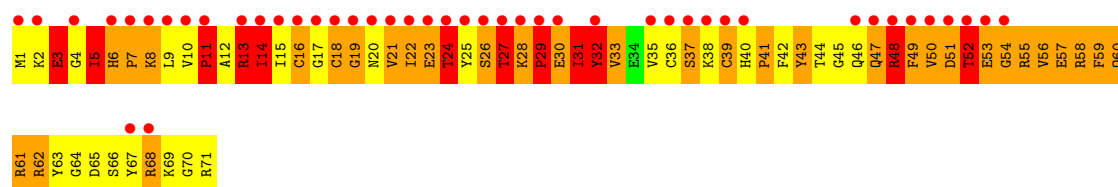
• Molecule 26: 50S ribosomal protein L30

Chain 2:



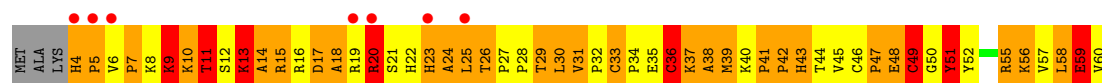
• Molecule 27: 50S ribosomal protein L31

Chain 3:



• Molecule 28: 50S ribosomal protein L32

Chain 4:



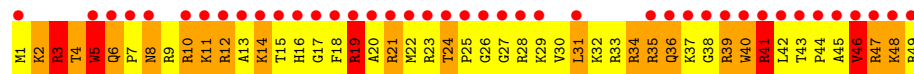
• Molecule 29: 50S ribosomal protein L33

Chain 5:



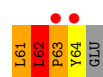
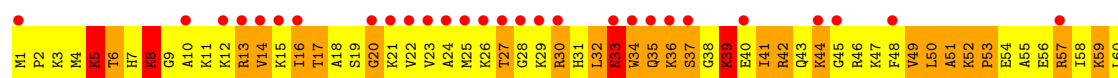
- Molecule 30: 50S ribosomal protein L34

Chain 6:



- Molecule 31: 50S ribosomal protein L35

Chain 7:



- Molecule 32: 50S ribosomal protein L36

Chain 8:



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	508.54Å 508.54Å 806.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	300.00 – 5.50 430.13 – 4.98	Depositor EDS
% Data completeness (in resolution range)	96.7 (300.00-5.50) 99.1 (430.13-4.98)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.00 (at 4.88Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.243 , 0.326 0.430 , 0.453	Depositor DCC
$R_{free}$ test set	7981 reflections (4.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	108.9	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.06 , 12.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.53$ , $\langle L^2 \rangle = 0.37$	Xtriage
Outliers	1 of 226745 reflections (0.000%)	Xtriage
$F_o, F_c$ correlation	0.74	EDS
Total number of atoms	94012	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	222.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.63% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

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	0.98	108/69685 (0.2%)	1.10	361/108786 (0.3%)
2	B	0.83	4/2954 (0.1%)	0.99	7/4606 (0.2%)
3	C	0.54	0/1775	0.86	0/2393
4	D	0.72	2/2174 (0.1%)	1.19	12/2927 (0.4%)
5	E	0.75	0/1611	1.16	13/2171 (0.6%)
6	F	0.64	0/1660	1.03	4/2247 (0.2%)
7	G	0.62	0/1507	1.06	4/2027 (0.2%)
8	H	0.59	0/1354	0.99	4/1831 (0.2%)
9	K	0.60	0/1170	1.02	4/1581 (0.3%)
10	L	0.61	0/1044	0.94	1/1415 (0.1%)
11	M	0.77	1/1140 (0.1%)	1.16	8/1537 (0.5%)
12	N	0.92	1/942 (0.1%)	1.31	10/1268 (0.8%)
13	O	0.71	0/1123	1.12	5/1493 (0.3%)
14	P	0.72	0/1100	1.19	8/1470 (0.5%)
15	Q	0.70	0/974	1.06	2/1302 (0.2%)
16	R	0.72	0/887	1.06	4/1180 (0.3%)
17	S	0.85	0/990	1.31	9/1325 (0.7%)
18	T	0.76	0/982	1.08	0/1306
19	U	0.87	1/790 (0.1%)	1.28	9/1057 (0.9%)
20	V	0.66	0/886	1.04	1/1189 (0.1%)
21	W	0.57	0/756	0.93	0/1015
22	X	0.54	0/857	1.04	2/1142 (0.2%)
23	Y	0.66	0/1467	1.11	7/1992 (0.4%)
24	Z	0.65	0/679	1.04	1/902 (0.1%)
25	1	0.59	0/569	0.88	0/751
26	2	0.59	0/474	1.09	2/635 (0.3%)
27	3	0.84	1/594 (0.2%)	1.31	8/795 (1.0%)
28	4	0.71	0/459	1.16	3/621 (0.5%)
29	5	0.85	1/433 (0.2%)	1.36	5/576 (0.9%)
30	6	0.73	0/438	1.01	0/575
31	7	0.60	0/523	1.14	5/690 (0.7%)
32	8	0.59	0/310	1.08	1/407 (0.2%)
All	All	0.90	119/102307 (0.1%)	1.09	500/153212 (0.3%)

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	463
2	B	0	18
3	C	0	1
5	E	0	1
11	M	0	1
14	P	0	1
17	S	0	1
19	U	0	1
27	3	0	1
28	4	0	1
All	All	0	489

The worst 5 of 119 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1203	G	O3'-P	-34.70	1.19	1.61
1	A	2500	U	C4-O4	18.52	1.38	1.23
1	A	2500	U	O3'-P	-15.17	1.43	1.61
1	A	607	U	N3-C4	-14.53	1.25	1.38
1	A	2448	A	O3'-P	-14.42	1.43	1.61

The worst 5 of 500 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1203	G	P-O3'-C3'	27.70	152.94	119.70
1	A	2448	A	C5'-C4'-O4'	-27.31	76.33	109.10
1	A	1204	A	N9-C1'-C2'	14.59	132.97	114.00
1	A	607	U	C5-C6-N1	-12.65	116.38	122.70
1	A	2384	G	N9-C1'-C2'	12.21	129.87	114.00

There are no chirality outliers.

5 of 489 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	14	A	Sidechain
1	A	25	U	Sidechain
1	A	3	U	Sidechain
1	A	31	C	Sidechain

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	9	U	Sidechain

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	62218	0	31356	15157	0
2	B	2641	0	1337	605	1
3	C	1742	0	1796	722	0
4	D	2124	0	2207	1489	0
5	E	1578	0	1647	1084	0
6	F	1625	0	1666	822	0
7	G	1482	0	1546	857	0
8	H	1328	0	1408	697	0
9	K	1155	0	1244	502	0
10	L	1025	0	1074	441	0
11	M	1113	0	1183	782	0
12	N	932	0	994	709	0
13	O	1106	0	1183	795	0
14	P	1080	0	1127	749	0
15	Q	960	0	1021	603	0
16	R	877	0	938	517	0
17	S	976	0	1033	631	0
18	T	964	0	1022	742	0
19	U	779	0	852	574	0
20	V	876	0	941	459	0
21	W	742	0	800	360	0
22	X	844	0	930	421	0
23	Y	1435	0	1463	717	0
24	Z	670	0	700	350	0
25	1	567	0	621	312	0
26	2	469	0	518	320	0
27	3	581	0	577	392	0
28	4	445	0	459	277	0
29	5	426	0	452	279	0
30	6	430	0	480	273	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
31	7	515	0	587	396	0
32	8	307	0	335	150	0
All	All	94012	0	63497	30326	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 193.

The worst 5 of 30326 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:2447:G:C8	1:A:2500:U:H3'	1.23	1.61
1:A:1202:C:C5	1:A:1203:G:H1'	1.39	1.55
1:A:1203:G:O6	1:A:1240:U:C2	1.66	1.48
1:A:2459:A:N3	1:A:2460:U:H1'	1.38	1.38
1:A:2596:U:H2'	4:D:242:ARG:CZ	1.55	1.35

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	Distance(Å)	Clash(Å)
2:B:0:A:OP1	2:B:0:A:OP1[15_545]	2.18	0.02

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	
3	C	226/229 (99%)	155 (69%)	41 (18%)	30 (13%)	0	12
4	D	270/276 (98%)	125 (46%)	59 (22%)	86 (32%)	0	1
5	E	204/206 (99%)	117 (57%)	31 (15%)	56 (28%)	0	1
6	F	206/210 (98%)	109 (53%)	46 (22%)	51 (25%)	0	2
7	G	180/182 (99%)	79 (44%)	47 (26%)	54 (30%)	0	1
8	H	172/180 (96%)	80 (46%)	46 (27%)	46 (27%)	0	2
9	K	146/148 (99%)	91 (62%)	30 (20%)	25 (17%)	0	6

*Continued on next page...*



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
10	L	136/147 (92%)	69 (51%)	38 (28%)	29 (21%)	0	4
11	M	137/140 (98%)	64 (47%)	28 (20%)	45 (33%)	0	0
12	N	120/122 (98%)	59 (49%)	23 (19%)	38 (32%)	0	1
13	O	143/150 (95%)	57 (40%)	36 (25%)	50 (35%)	0	0
14	P	134/141 (95%)	49 (37%)	33 (25%)	52 (39%)	0	0
15	Q	115/118 (98%)	57 (50%)	39 (34%)	19 (16%)	0	7
16	R	108/112 (96%)	48 (44%)	29 (27%)	31 (29%)	0	1
17	S	115/146 (79%)	52 (45%)	26 (23%)	37 (32%)	0	1
18	T	115/118 (98%)	35 (30%)	50 (44%)	30 (26%)	0	2
19	U	99/101 (98%)	52 (52%)	19 (19%)	28 (28%)	0	1
20	V	108/113 (96%)	63 (58%)	24 (22%)	21 (19%)	0	5
21	W	92/96 (96%)	57 (62%)	16 (17%)	19 (21%)	0	4
22	X	108/110 (98%)	43 (40%)	32 (30%)	33 (31%)	0	1
23	Y	178/206 (86%)	95 (53%)	44 (25%)	39 (22%)	0	3
24	Z	83/85 (98%)	52 (63%)	21 (25%)	10 (12%)	1	14
25	1	65/67 (97%)	36 (55%)	20 (31%)	9 (14%)	0	11
26	2	57/60 (95%)	34 (60%)	8 (14%)	15 (26%)	0	2
27	3	69/71 (97%)	23 (33%)	16 (23%)	30 (44%)	0	0
28	4	55/60 (92%)	14 (26%)	19 (34%)	22 (40%)	0	0
29	5	47/54 (87%)	14 (30%)	7 (15%)	26 (55%)	0	0
30	6	47/49 (96%)	20 (43%)	12 (26%)	15 (32%)	0	1
31	7	62/65 (95%)	23 (37%)	18 (29%)	21 (34%)	0	0
32	8	35/37 (95%)	20 (57%)	5 (14%)	10 (29%)	0	1
All	All	3632/3799 (96%)	1792 (49%)	863 (24%)	977 (27%)	0	2

5 of 977 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	35	ALA
3	C	39	GLU
3	C	54	SER
3	C	61	THR
3	C	66	HIS

### 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	
3	C	180/181 (99%)	151 (84%)	29 (16%)	3	26
4	D	215/218 (99%)	153 (71%)	62 (29%)	0	5
5	E	166/166 (100%)	102 (61%)	64 (39%)	0	1
6	F	164/166 (99%)	105 (64%)	59 (36%)	0	2
7	G	156/156 (100%)	112 (72%)	44 (28%)	0	5
8	H	143/148 (97%)	109 (76%)	34 (24%)	1	9
9	K	124/124 (100%)	92 (74%)	32 (26%)	1	7
10	L	105/111 (95%)	89 (85%)	16 (15%)	4	28
11	M	118/119 (99%)	79 (67%)	39 (33%)	0	3
12	N	100/100 (100%)	69 (69%)	31 (31%)	0	4
13	O	111/116 (96%)	71 (64%)	40 (36%)	0	2
14	P	106/111 (96%)	65 (61%)	41 (39%)	0	1
15	Q	100/101 (99%)	71 (71%)	29 (29%)	0	5
16	R	87/88 (99%)	63 (72%)	24 (28%)	0	6
17	S	105/127 (83%)	68 (65%)	37 (35%)	0	2
18	T	93/94 (99%)	64 (69%)	29 (31%)	0	4
19	U	82/82 (100%)	57 (70%)	25 (30%)	0	5
20	V	90/92 (98%)	64 (71%)	26 (29%)	0	5
21	W	76/78 (97%)	57 (75%)	19 (25%)	1	8
22	X	91/91 (100%)	72 (79%)	19 (21%)	1	12
23	Y	159/179 (89%)	120 (76%)	39 (24%)	1	8
24	Z	67/67 (100%)	51 (76%)	16 (24%)	1	9
25	1	62/62 (100%)	44 (71%)	18 (29%)	0	5
26	2	51/52 (98%)	36 (71%)	15 (29%)	0	5
27	3	63/63 (100%)	45 (71%)	18 (29%)	0	5
28	4	50/52 (96%)	31 (62%)	19 (38%)	0	1
29	5	48/52 (92%)	32 (67%)	16 (33%)	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
30	6	42/42 (100%)	29 (69%)	13 (31%)	0	4
31	7	54/55 (98%)	44 (82%)	10 (18%)	2	17
32	8	34/34 (100%)	29 (85%)	5 (15%)	4	30
All	All	3042/3127 (97%)	2174 (72%)	868 (28%)	0	5

5 of 868 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
12	N	38	VAL
14	P	127	ILE
27	3	60	GLN
12	N	89	ASN
13	O	117	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 101 such sidechains are listed below:

Mol	Chain	Res	Type
13	O	81	GLN
18	T	44	ASN
30	6	16	HIS
14	P	12	GLN
15	Q	31	HIS

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	2888/2916 (99%)	1204 (41%)	253 (8%)
2	B	122/123 (99%)	46 (37%)	3 (2%)
All	All	3010/3039 (99%)	1250 (41%)	256 (8%)

5 of 1250 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	13	A
1	A	14	A
1	A	15	G
1	A	20	C
1	A	26	G

5 of 256 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	A	1267	U
1	A	1626	G
1	A	2690	C
1	A	1302	A
1	A	1477	A

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

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

#### 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

#### 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

#### 5.7 Other polymers ⓘ

There are no such residues in this entry.

#### 5.8 Polymer linkage issues

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, 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	2889/2916 (99%)	0.49	438 (15%) 3 8	129, 215, 297, 342	0
2	B	123/123 (100%)	0.01	4 (3%) 44 40	199, 259, 293, 316	0
3	C	228/229 (99%)	-0.17	2 (0%) 81 70	250, 281, 300, 313	0
4	D	272/276 (98%)	-0.11	9 (3%) 44 40	130, 199, 221, 259	0
5	E	206/206 (100%)	-0.46	1 (0%) 88 80	148, 186, 232, 269	0
6	F	208/210 (99%)	0.31	27 (12%) 4 10	164, 238, 263, 276	0
7	G	182/182 (100%)	0.44	22 (12%) 5 12	212, 246, 268, 276	0
8	H	174/180 (96%)	-0.39	0 100 100	196, 241, 260, 271	0
9	K	148/148 (100%)	-0.33	4 (2%) 52 44	202, 227, 249, 257	0
10	L	138/147 (93%)	-0.10	0 100 100	271, 298, 312, 327	0
11	M	139/140 (99%)	0.46	15 (10%) 6 13	167, 198, 224, 235	0
12	N	122/122 (100%)	0.58	20 (16%) 2 7	148, 177, 209, 221	0
13	O	145/150 (96%)	0.24	14 (9%) 8 15	166, 247, 276, 315	0
14	P	136/141 (96%)	-0.01	9 (6%) 18 24	166, 208, 234, 241	0
15	Q	117/118 (99%)	-0.06	3 (2%) 53 45	168, 192, 219, 226	0
16	R	110/112 (98%)	0.65	16 (14%) 3 8	186, 235, 263, 289	0
17	S	117/146 (80%)	-0.15	2 (1%) 67 56	164, 194, 227, 242	0
18	T	117/118 (99%)	0.31	6 (5%) 27 29	183, 202, 220, 232	0
19	U	101/101 (100%)	0.47	6 (5%) 22 26	168, 225, 246, 257	0
20	V	110/113 (97%)	1.88	31 (28%) 1 4	180, 209, 237, 248	0
21	W	94/96 (97%)	2.10	44 (46%) 1 3	212, 232, 280, 288	0
22	X	110/110 (100%)	0.12	12 (10%) 6 13	221, 255, 287, 305	0
23	Y	180/206 (87%)	-0.15	0 100 100	208, 246, 268, 279	0
24	Z	85/85 (100%)	1.81	26 (30%) 1 4	211, 229, 242, 271	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
25	1	67/67 (100%)	1.19	18 (26%) 1 4	218, 243, 263, 268	0
26	2	59/60 (98%)	-0.00	3 (5%) 27 29	204, 224, 249, 259	0
27	3	71/71 (100%)	3.24	45 (63%) 0 3	199, 223, 238, 246	0
28	4	57/60 (95%)	0.26	7 (12%) 5 11	164, 209, 264, 289	0
29	5	49/54 (90%)	1.59	20 (40%) 1 3	206, 232, 242, 252	0
30	6	49/49 (100%)	3.12	41 (83%) 0 2	162, 209, 221, 235	0
31	7	64/65 (98%)	2.05	30 (46%) 1 3	164, 195, 225, 242	0
32	8	37/37 (100%)	-0.02	0 100 100	205, 215, 224, 226	0
All	All	6704/6838 (98%)	0.42	875 (13%) 4 10	129, 222, 294, 342	0

The worst 5 of 875 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
24	Z	3	HIS	14.0
7	G	81	LYS	13.5
1	A	2609	U	12.5
1	A	1537	C	12.1
20	V	92	ARG	11.7

## 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 ⓘ

There are no such residues in this entry.