



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 29, 2021 – 04:05 PM EDT

PDB ID : 1PVQ
Title : BASIS FOR A SWITCH IN SUBSTRATE SPECIFICITY: CRYSTAL STRUCTURE OF SELECTED VARIANT OF CRE SITE-SPECIFIC RECOMBINASE, LNSGG BOUND TO THE ENGINEERED RECOGNITION SITE LOXM7
Authors : Baldwin, E.P.; Martin, S.S.; Abel, J.; Gelato, K.A.; Kim, H.; Schultz, P.G.; Santoro, S.W.
Deposited on : 2003-06-28
Resolution : 2.75 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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	:	2.18
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)	:	2.18

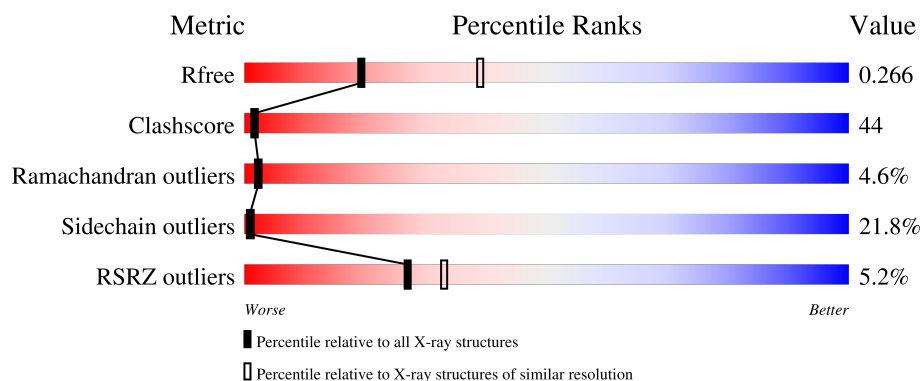
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. 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	C	34	<div> <div>9%</div> <div>12%</div> <div>47%</div> <div>38%</div> <div>.</div> </div>
2	D	34	<div> <div>3%</div> <div>24%</div> <div>59%</div> <div>18%</div> </div>
3	A	349	<div> <div>8%</div> <div>34%</div> <div>42%</div> <div>15%</div> <div>7%</div> <div>.</div> </div>
3	B	349	<div> <div>%</div> <div>40%</div> <div>42%</div> <div>9%</div> <div>9%</div> <div>.</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6633 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 DNA chain called DNA 34-MER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	34	Total	C	N	O	P	0	4	0
			737	357	129	216	35			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	7	DC	DT	engineered mutation	GB M10145
C	8	DT	DC	engineered mutation	GB M10145
C	9	DA	DG	engineered mutation	GB M10145
C	26	DT	DC	engineered mutation	GB M10145
C	27	DA	DG	engineered mutation	GB M10145
C	28	DG	DA	engineered mutation	GB M10145

- Molecule 2 is a DNA chain called DNA 34-MER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	34	Total	C	N	O	P	0	5	0
			774	375	138	224	37			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	7	DC	DT	engineered mutation	GB M10494
D	8	DT	DC	engineered mutation	GB M10494
D	9	DA	DG	engineered mutation	GB M10494
D	26	DT	DC	engineered mutation	GB M10494
D	27	DA	DG	engineered mutation	GB M10494
D	28	DG	DA	engineered mutation	GB M10494

- Molecule 3 is a protein called Recombinase cre.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	323	Total 2543	C 1579	N 485	O 464	S 15	38	0	0
3	B	319	Total 2512	C 1561	N 481	O 455	S 15	4	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	MET	-	initiating methionine	UNP P06956
A	-4	HIS	-	expression tag	UNP P06956
A	-3	HIS	-	expression tag	UNP P06956
A	-2	HIS	-	expression tag	UNP P06956
A	-1	HIS	-	expression tag	UNP P06956
A	0	HIS	-	expression tag	UNP P06956
A	1	HIS	-	expression tag	UNP P06956
A	174	LEU	ILE	engineered mutation	UNP P06956
A	258	ASN	THR	engineered mutation	UNP P06956
A	259	SER	ARG	engineered mutation	UNP P06956
A	262	GLY	GLU	engineered mutation	UNP P06956
A	266	GLY	GLU	engineered mutation	UNP P06956
B	-5	MET	-	initiating methionine	UNP P06956
B	-4	HIS	-	expression tag	UNP P06956
B	-3	HIS	-	expression tag	UNP P06956
B	-2	HIS	-	expression tag	UNP P06956
B	-1	HIS	-	expression tag	UNP P06956
B	0	HIS	-	expression tag	UNP P06956
B	1	HIS	-	expression tag	UNP P06956
B	174	LEU	ILE	engineered mutation	UNP P06956
B	258	ASN	THR	engineered mutation	UNP P06956
B	259	SER	ARG	engineered mutation	UNP P06956
B	262	GLY	GLU	engineered mutation	UNP P06956
B	266	GLY	GLU	engineered mutation	UNP P06956

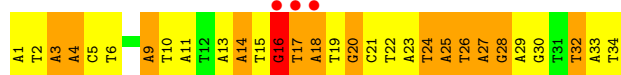
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	7	Total 7	O 7	0	0
4	D	8	Total 8	O 8	0	0
4	A	22	Total 22	O 22	0	0
4	B	30	Total 30	O 30	0	0

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

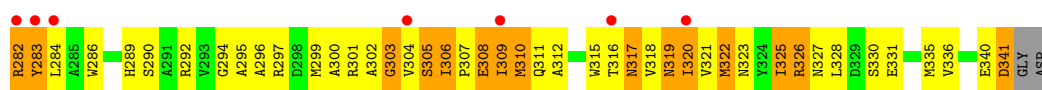
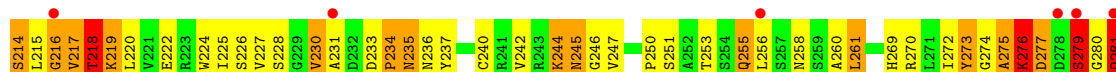
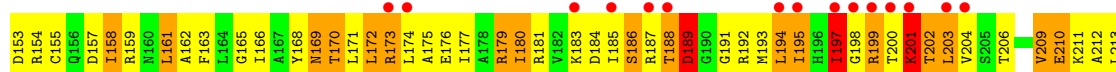
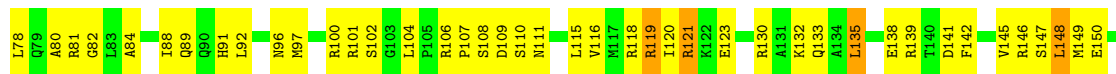
• Molecule 1: DNA 34-MER



• Molecule 2: DNA 34-MER



• Molecule 3: Recombinase cre



• Molecule 3: Recombinase cre



A295	A296	R297	D298	M299		S305	I306	P307	E308		A312	G313	G314	W315	T316	N317	V318	N319	I320	V321	N322	N323	Y324	I325	R326	N327	LEU	ASP	SER	GLU	T332	G333	A334	N335	V336	R337	L338	L339	E340	D341	GLY	ASP																		
V209	E210	K211	A212	L213	S214		V217	T218	K219	L220	V221	E222	R223	W224		V230	A231	D232	D233	P234	N235	N236	Y237	L238		R241	V242	R243	K244	N245	G246	V247		S254	Q255		N258	S259	A260	L261		T268		L271		K276	D277	D278	S279	G280	Q281	R282	Y283	L284		H289		R292		
E138	R139	T140		D143	Q144	Y145	R146	S147	L148	M149	E150		R154	C155		I158	R159	Y160	L161		M169	T170	L171	L172	R173	L174	A175	E176	I177	A178	R179	I180	R181	V182	K183	D184	I185	S186	R187	T188	D189		R192	M193	L194	I195	H196	I197	G198	R199	T200	K201	T202	L203	V204	S205	T206	A207	G208	
F64	P65	A66		P68		R72		Y77	L78	Q79	A80	R81	G82	L83	A84	V85	K86	T87	I88		Q89	Q90	H91	L92	G93	Q94	L95	N96	M97	L98	H99		S102		R106	P107	S108	D109	S110		V116	M117	R118	R119	I120	R121	K122	E123	N124	V125	D126	A127	G128	E129	R130	A131	K132	Q133	A134	
MET	HIS	HIS	HIS	HIS	HIS	HIS	SER	ASN	ASN	LEU	LEU	THR	HIS	GLN	ASN	LEU	PRO	ALA	LEU	PRO	PRO	VAL	ASP	ALA	T19	S20	D21	E22	V23	R24		L27	M28	D29	M30	F31	R32		Q35	A36	F37	S38	E39	H40		K43	M44	L45		V48	C49	R50	S51		A54		M60	R61	K62	W63

4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	107.41Å 121.35Å 180.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	5.00 – 2.75 90.31 – 2.75	Depositor EDS
% Data completeness (in resolution range)	97.0 (5.00-2.75) 95.6 (90.31-2.75)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.70 (at 2.73Å)	Xtriage
Refinement program	TNT	Depositor
R, R_{free}	0.224 , 0.281 0.228 , 0.266	Depositor DCC
R_{free} test set	1435 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å ²)	54.3	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 88.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	6633	wwPDB-VP
Average B, all atoms (Å ²)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.19% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

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	C	0.67	1/826 (0.1%)	2.47	28/1274 (2.2%)
2	D	0.62	0/868	1.36	8/1337 (0.6%)
3	A	0.37	0/2584	0.77	7/3484 (0.2%)
3	B	0.45	0/2552	0.72	1/3439 (0.0%)
All	All	0.48	1/6830 (0.0%)	1.21	44/9534 (0.5%)

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	C	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	18[B]	DA	O3'-P	8.00	1.70	1.61

The worst 5 of 44 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	18[B]	DA	O3'-P-O5'	36.17	172.72	104.00
1	C	16[A]	DG	OP2-P-O3'	-25.79	48.46	105.20
1	C	16[B]	DG	OP2-P-O3'	-25.79	48.46	105.20
1	C	16[A]	DG	O3'-P-O5'	23.03	147.76	104.00
1	C	16[B]	DG	O3'-P-O5'	23.03	147.76	104.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	16[A]	DG	Sidechain

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	C	737	0	414	51	0
2	D	774	0	424	76	0
3	A	2543	0	2563	281	0
3	B	2512	0	2536	192	0
4	A	22	0	0	4	0
4	B	30	0	0	3	0
4	C	7	0	0	0	0
4	D	8	0	0	0	0
All	All	6633	0	5937	544	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 44.

The worst 5 of 544 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)
3:A:171:LEU:HD13	3:A:312:ALA:HB1	1.20	1.11
3:A:325:ILE:HG23	3:A:331:GLU:HG3	1.22	1.11
3:A:201:LYS:HE3	3:A:201:LYS:HA	1.33	1.08
3:A:213:LEU:HB3	3:A:217:VAL:HG23	1.33	1.06
2:D:33:DA:H2''	2:D:34:DT:H5''	1.36	1.05

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	
3	A	321/349 (92%)	264 (82%)	34 (11%)	23 (7%)	1	1
3	B	315/349 (90%)	284 (90%)	25 (8%)	6 (2%)	8	14
All	All	636/698 (91%)	548 (86%)	59 (9%)	29 (5%)	2	2

5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	188	THR
3	A	189	ASP
3	A	217	VAL
3	A	218	THR
3	A	234	PRO

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	
3	A	268/291 (92%)	203 (76%)	65 (24%)	0	1
3	B	264/291 (91%)	213 (81%)	51 (19%)	1	2
All	All	532/582 (91%)	416 (78%)	116 (22%)	1	1

5 of 116 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	320	ILE
3	B	325	ILE
3	B	39	GLU
3	B	322	MET
3	B	258	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	26	ASN
3	B	255	GLN
3	B	35	GLN
3	B	327	ASN
3	B	90	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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	15:DT	O3'	16[B]:DG	P	1.38

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	C	34/34 (100%)	0.31	3 (8%) 10 11	29, 52, 85, 94	2 (5%)
2	D	34/34 (100%)	0.04	1 (2%) 51 61	31, 46, 70, 90	0
3	A	322/349 (92%)	0.57	28 (8%) 10 12	23, 67, 98, 99	13 (4%)
3	B	319/349 (91%)	0.09	5 (1%) 72 79	20, 43, 90, 99	1 (0%)
All	All	709/766 (92%)	0.32	37 (5%) 27 33	20, 53, 96, 99	16 (2%)

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	282	ARG	8.7
3	A	283	TYR	6.4
3	A	281	GLN	6.2
1	C	17[B]	DT	5.2
3	B	327	ASN	4.7

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.