wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 10:01 AM GMT

PDB ID	:	3KKA
Title	:	Co-crystal structure of the sam domains of EPHA1 AND EPHA2
Authors	:	Walker, J.R.; Yermekbayeva, L.; Butler-Cole, C.; Weigelt, J.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.M.; Bochkarev, A.; Dhe-Paganon, S.; Structural Genomics Consortium (SGC)
Deposited on	:	2009-11-05
Resolution	:	2.40 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 (1)) were used in the production of this report:

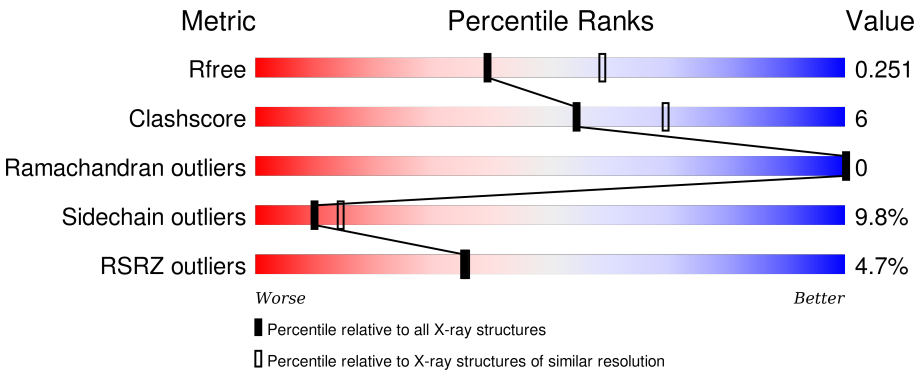
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.40 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}	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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.

Mol	Chain	Length	Quality of chain
1	A	82	% 63% 15% 22%
1	B	82	% 70% 7% 23%
2	C	86	10% 45% 20% • 31%
2	D	86	2% 63% 10% 6% 21%
2	E	86	2% 60% 15% • 23%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2562 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 EPHRIN TYPE-A RECEPTOR 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	64	514	327	87	95	5	0	1	0
1	B	63	498	319	84	90	5	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	893	MET	-	EXPRESSION TAG	UNP P21709
A	894	HIS	-	EXPRESSION TAG	UNP P21709
A	895	HIS	-	EXPRESSION TAG	UNP P21709
A	896	HIS	-	EXPRESSION TAG	UNP P21709
A	897	HIS	-	EXPRESSION TAG	UNP P21709
A	898	HIS	-	EXPRESSION TAG	UNP P21709
A	899	HIS	-	EXPRESSION TAG	UNP P21709
A	900	SER	-	EXPRESSION TAG	UNP P21709
A	901	SER	-	EXPRESSION TAG	UNP P21709
A	902	GLY	-	EXPRESSION TAG	UNP P21709
A	903	ARG	-	EXPRESSION TAG	UNP P21709
A	904	GLU	-	EXPRESSION TAG	UNP P21709
A	905	ASN	-	EXPRESSION TAG	UNP P21709
A	906	LEU	-	EXPRESSION TAG	UNP P21709
A	907	TYR	-	EXPRESSION TAG	UNP P21709
A	908	PHE	-	EXPRESSION TAG	UNP P21709
A	909	GLN	-	EXPRESSION TAG	UNP P21709
A	910	GLY	-	EXPRESSION TAG	UNP P21709
B	893	MET	-	EXPRESSION TAG	UNP P21709
B	894	HIS	-	EXPRESSION TAG	UNP P21709
B	895	HIS	-	EXPRESSION TAG	UNP P21709
B	896	HIS	-	EXPRESSION TAG	UNP P21709
B	897	HIS	-	EXPRESSION TAG	UNP P21709
B	898	HIS	-	EXPRESSION TAG	UNP P21709
B	899	HIS	-	EXPRESSION TAG	UNP P21709

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	900	SER	-	EXPRESSION TAG	UNP P21709
B	901	SER	-	EXPRESSION TAG	UNP P21709
B	902	GLY	-	EXPRESSION TAG	UNP P21709
B	903	ARG	-	EXPRESSION TAG	UNP P21709
B	904	GLU	-	EXPRESSION TAG	UNP P21709
B	905	ASN	-	EXPRESSION TAG	UNP P21709
B	906	LEU	-	EXPRESSION TAG	UNP P21709
B	907	TYR	-	EXPRESSION TAG	UNP P21709
B	908	PHE	-	EXPRESSION TAG	UNP P21709
B	909	GLN	-	EXPRESSION TAG	UNP P21709
B	910	GLY	-	EXPRESSION TAG	UNP P21709

- Molecule 2 is a protein called EPHRIN TYPE-A RECEPTOR 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	59	462	296	79	84	3	0	0	0
			Total	C	N	O	S			
2	D	68	534	342	90	99	3	0	0	0
			Total	C	N	O	S			
2	E	66	519	332	92	92	3	0	0	0

There are 51 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	885	MET	-	EXPRESSION TAG	UNP P29317
C	886	HIS	-	EXPRESSION TAG	UNP P29317
C	887	HIS	-	EXPRESSION TAG	UNP P29317
C	888	HIS	-	EXPRESSION TAG	UNP P29317
C	889	HIS	-	EXPRESSION TAG	UNP P29317
C	890	HIS	-	EXPRESSION TAG	UNP P29317
C	891	HIS	-	EXPRESSION TAG	UNP P29317
C	892	SER	-	EXPRESSION TAG	UNP P29317
C	893	SER	-	EXPRESSION TAG	UNP P29317
C	894	GLY	-	EXPRESSION TAG	UNP P29317
C	895	ARG	-	EXPRESSION TAG	UNP P29317
C	896	GLU	-	EXPRESSION TAG	UNP P29317
C	897	ASN	-	EXPRESSION TAG	UNP P29317
C	898	LEU	-	EXPRESSION TAG	UNP P29317
C	899	TYR	-	EXPRESSION TAG	UNP P29317
C	900	PHE	-	EXPRESSION TAG	UNP P29317
C	901	GLN	-	EXPRESSION TAG	UNP P29317

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	886	MET	-	EXPRESSION TAG	UNP P29317
D	887	HIS	-	EXPRESSION TAG	UNP P29317
D	888	HIS	-	EXPRESSION TAG	UNP P29317
D	889	HIS	-	EXPRESSION TAG	UNP P29317
D	890	HIS	-	EXPRESSION TAG	UNP P29317
D	891	HIS	-	EXPRESSION TAG	UNP P29317
D	892	HIS	-	EXPRESSION TAG	UNP P29317
D	893	SER	-	EXPRESSION TAG	UNP P29317
D	894	SER	-	EXPRESSION TAG	UNP P29317
D	895	GLY	-	EXPRESSION TAG	UNP P29317
D	896	ARG	-	EXPRESSION TAG	UNP P29317
D	897	GLU	-	EXPRESSION TAG	UNP P29317
D	898	ASN	-	EXPRESSION TAG	UNP P29317
D	899	LEU	-	EXPRESSION TAG	UNP P29317
D	900	TYR	-	EXPRESSION TAG	UNP P29317
D	901	PHE	-	EXPRESSION TAG	UNP P29317
D	902	GLN	-	EXPRESSION TAG	UNP P29317
E	886	MET	-	EXPRESSION TAG	UNP P29317
E	887	HIS	-	EXPRESSION TAG	UNP P29317
E	888	HIS	-	EXPRESSION TAG	UNP P29317
E	889	HIS	-	EXPRESSION TAG	UNP P29317
E	890	HIS	-	EXPRESSION TAG	UNP P29317
E	891	HIS	-	EXPRESSION TAG	UNP P29317
E	892	HIS	-	EXPRESSION TAG	UNP P29317
E	893	SER	-	EXPRESSION TAG	UNP P29317
E	894	SER	-	EXPRESSION TAG	UNP P29317
E	895	GLY	-	EXPRESSION TAG	UNP P29317
E	896	ARG	-	EXPRESSION TAG	UNP P29317
E	897	GLU	-	EXPRESSION TAG	UNP P29317
E	898	ASN	-	EXPRESSION TAG	UNP P29317
E	899	LEU	-	EXPRESSION TAG	UNP P29317
E	900	TYR	-	EXPRESSION TAG	UNP P29317
E	901	PHE	-	EXPRESSION TAG	UNP P29317
E	902	GLN	-	EXPRESSION TAG	UNP P29317

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
3	B	1	1	1	0	0

- Molecule 4 is water.

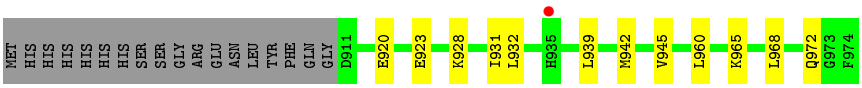
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	13	13	13	0	0
			Total	O		
4	B	8	8	8	0	0
			Total	O		
4	C	1	1	1	0	0
			Total	O		
4	D	8	8	8	0	0
			Total	O		
4	E	4	4	4	0	0

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.

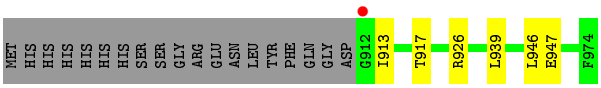
- Molecule 1: EPHRIN TYPE-A RECEPTOR 1

Chain A:



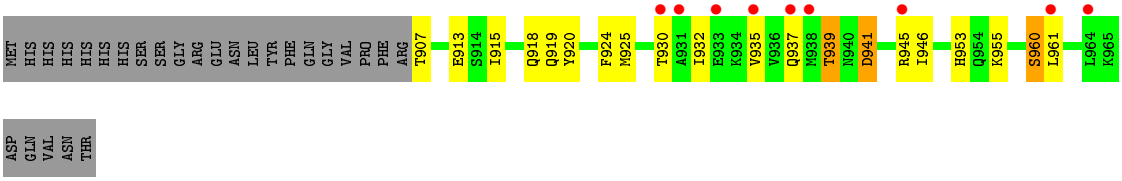
- Molecule 1: EPHRIN TYPE-A RECEPTOR 1

Chain B:



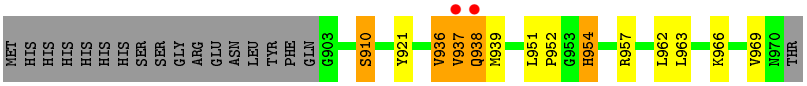
- Molecule 2: EPHRIN TYPE-A RECEPTOR 2

Chain C:



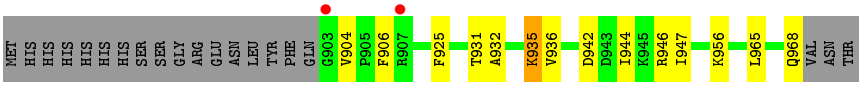
- Molecule 2: EPHRIN TYPE-A RECEPTOR 2

Chain D:



- Molecule 2: EPHRIN TYPE-A RECEPTOR 2

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	57.55Å 56.07Å 107.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.44 – 2.40 30.44 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.9 (30.44-2.40) 99.9 (30.44-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	2.34 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.200 , 0.235 0.226 , 0.251	Depositor DCC
R_{free} test set	705 reflections (5.24%)	DCC
Wilson B-factor (Å^2)	44.5	Xtriage
Anisotropy	0.103	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 56.4	EDS
Estimated twinning fraction	0.023 for k,h,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 14167 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2562	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.85% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CL

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	0.73	0/524	0.70	0/706
1	B	0.75	0/508	0.72	0/684
2	C	0.54	0/470	0.60	0/634
2	D	0.68	0/544	0.70	0/735
2	E	0.59	0/529	0.64	0/714
All	All	0.66	0/2575	0.67	0/3473

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 (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	514	0	507	7	0
1	B	498	0	498	3	0
2	C	462	0	462	11	0
2	D	534	0	527	7	0
2	E	519	0	515	7	0
3	B	1	0	0	0	0
4	A	13	0	0	0	0
4	B	8	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	C	1	0	0	0	0
4	D	8	0	0	0	0
4	E	4	0	0	1	0
All	All	2562	0	2509	32	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 6.

The worst 5 of 32 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:E:944:ILE:O	2:E:947:ILE:HG22	1.80	0.81
2:E:925:PHE:HA	2:E:947:ILE:HD11	1.71	0.72
1:A:923:GLU:HG3	1:A:928:LYS:HD2	1.72	0.72
2:E:942:ASP:HB3	2:E:946:ARG:NH2	2.08	0.69
2:C:915:ILE:HG21	2:C:960:SER:OG	1.96	0.66

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	63/82 (77%)	62 (98%)	1 (2%)	0	100	100
1	B	61/82 (74%)	59 (97%)	2 (3%)	0	100	100
2	C	57/86 (66%)	56 (98%)	1 (2%)	0	100	100
2	D	66/86 (77%)	65 (98%)	1 (2%)	0	100	100
2	E	64/86 (74%)	63 (98%)	1 (2%)	0	100	100
All	All	311/422 (74%)	305 (98%)	6 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	56/73 (77%)	54 (96%)	2 (4%)	42	63
1	B	54/73 (74%)	52 (96%)	2 (4%)	41	62
2	C	48/76 (63%)	40 (83%)	8 (17%)	3	3
2	D	56/76 (74%)	47 (84%)	9 (16%)	3	3
2	E	53/76 (70%)	48 (91%)	5 (9%)	11	16
All	All	267/374 (71%)	241 (90%)	26 (10%)	10	15

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	961	LEU
2	D	937	VAL
2	E	965	LEU
2	D	910	SER
2	D	936	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	953	HIS
2	D	938	GLN
2	D	954	HIS

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 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	64/82 (78%)	0.04	1 (1%)	74	74	28, 42, 59, 76	0
1	B	63/82 (76%)	0.05	1 (1%)	74	74	30, 44, 65, 75	0
2	C	59/86 (68%)	0.82	9 (15%)	3	3	44, 77, 100, 114	0
2	D	68/86 (79%)	0.07	2 (2%)	55	54	34, 51, 66, 68	0
2	E	66/86 (76%)	0.31	2 (3%)	54	53	38, 62, 83, 90	0
All	All	320/422 (75%)	0.25	15 (4%)	35	36	28, 52, 85, 114	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	912	GLY	4.0
2	C	931	ALA	3.9
2	C	945	ARG	3.9
2	C	933	GLU	3.2
2	C	938	MET	3.0

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 carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron

density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
3	CL	B	63	1/1	0.57	0.38	-	89,89,89,89	0

6.5 Other polymers (i)

There are no such residues in this entry.