The case of 1lkr held at the PDB and its variable amino acid occupancies; re refinement of 4ow9 to correct this

John R Helliwell and Simon W M Tanley

School of Chemistry, University of Manchester, Manchester, M13 9PL, UK

John.helliwell@manchester.ac.uk

Abstract

There is a problem with 1lkr held in the PDB. Quite a number of amino acids, in both of the 1lkr molecules in the monoclinic asymmetric unit, are affected by chemically erroneous large variations of individual atomic occupancies along the length of many of its amino acid side chains. This wholly unexpected effect propagated into our 4ow9. We have now made a re refinement of 4ow9 to correct this and deposited this at the PDB (new PDB code 5M1Y) to supercede 4ow9. This does not affect our article (Tanley and Helliwell *Acta Cryst F* 70, 1127-1131 (2014)) contents or conclusions and this Zenodo preprint is therefore an Addendum to our original article and explains why 4ow9 needs to be superceded at the PDB. We also note that 1lkr will need these amino acid occupancy errors correcting by the PDB.

Keywords

1lkr; chemically erroneous atomic occupancies; 4ow9

Introduction

There is a problem with 1lkr held in the PDB. Quite a number of amino acids, in both of the 1lkr molecules in the monoclinic asymmetric unit, are affected by chemically erroneous large variations of individual atomic occupancies along the length of many of its amino acid side chains (See figure 1a and c). This effect propagated into our 4ow9 model (see Figure 1b and d) i.e. was not detected as a fault by the molecular replacement software we used or our model refinement software or by the PDB validation report check of our deposition or by our co-editor or by our referees of our article submission. We noticed this problem in a 'by eye' scan of the PDB file which only by chance did we undertake. We have now made a re refinement of 4ow9 to correct this and deposited this at the PDB (new PDB code 5M1Y) to supercede 4ow9. The PDB code 1lkr was deposited in the PDB as part of a study by Steinrauf 1998.

Methods

The erroneous amino acid atomic occupancies were reset to 1.0 by a simple, but tedious, hand editing of the 4ow9 PDB file. To quantify the level of tediousness we note that the hand edits took an hour i.e. not such a big effort in fact. Phenix_Refine (Afonine et al 2012) was used for the model re-refinement. COOT (Emsley and Cowtan 2004) was used to inspect the molecular model and the

electron density maps. Minor adjustments were made to the model as a result of the helpful Phenix_Refine diagnostics of the model quality.

Results

A corrected refined model to supercede 4ow9 has been made. The details of the model refinement are given in Table 1.

Conclusions

Our new PDB code 5M1Y supercedes 4ow9. This does not affect our article (Tanley and Helliwell *Acta Cryst F70*, 1127-1131 (2014)) contents or conclusions and this Zenodo preprint is therefore an Addendum to our original article and explains why 4ow9 needs to be superceded at the PDB. We also note that 1lkr will need these amino acid occupancy errors correcting by the PDB.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012) *Acta Cryst.* D68, 352–367.

Emsley, P. and Cowtan, K. (2004). *Acta Cryst.* D60, 2126–2132.

Steinrauf, L.K. (1998) *Acta Cryst* D54, 767-779.

Tanley, S. W. M. and Helliwell, J. R. *Acta Cryst.* (2014). F70, 1127-1131.

Figure 1 The nature of the problem with 1lkr held in the PDB illustrated with one amino acid residue from each molecule. Many amino acids, in both of these 1lkr molecules in the monoclinic asymmetric unit, were affected by this physically erroneous error of a large variation of individual atomic occupancies along the length of the amino acid side chains. This propagated into our 4ow9 model. (a) an example, Arg61, from 1lkr molecule 1 and (b) the same example, Arg 61, as (a) from 4ow9 molecule 1 and (c) from 1lkr molecule 2, Lys B1 (d) the same example, LysB1, as (c) from 4ow9 molecule 2.

ATOM	469	N	ARG	A	61	3.010	22.216	15.585	1.00	3.84		N
ANISOU	469	N	ARG	A	61	3660	2717	2349	93	-28	555	N
ATOM	470	CA	ARG	A	61	2.970	20.743	15.570	1.00	4.46		C
ANISOU	470	CA	ARG	A	61	3754	2657	2544	41	66	553	C
ATOM	471	C	ARG	A	61	1.647	20.354	16.222	1.00	5.74		C
ANISOU	471	C	ARG	A	61	3906	2703	2816	-197	242	386	C
ATOM	472	O	ARG	A	61	1.120	19.410	15.709	1.00	6.25		O
ANISOU	472	O	ARG	A	61	4396	2236	2979	18	497	261	O
ATOM	473	CB	ARG	A	61	4.049	20.111	16.529	1.00	7.05		C
ANISOU	473	CB	ARG	A	61	4092	2862	2946	247	-96	676	C
ATOM	474	CG	ARG	A	61	3.946	18.421	16.398	0.91	8.79		C
ANISOU	474	CG	ARG	A	61	4426	2888	3225	342	-100	836	C
ATOM	475	CD	ARG	A	61	5.765	18.048	16.582	0.58	0.96		C
ANISOU	475	CD	ARG	A	61	4501	3381	3451	394	-70	1115	C
ATOM	476	NE	ARG	A	61	6.212	17.894	15.042	0.84	2.97		N
ANISOU	476	NE	ARG	A	61	4657	3872	3541	399	83	1111	N
ATOM	477	CZ	ARG	A	61	7.462	18.444	14.928	0.65	3.39		C
ANISOU	477	CZ	ARG	A	61	4680	3959	3585	547	79	1134	C
ATOM	478	NH1	ARG	A	61	7.848	18.880	16.049	0.48	3.86		N
ANISOU	478	NH1	ARG	A	61	4935	3616	3842	725	-573	1344	N
ATOM	479	NH2	ARG	A	61	7.990	18.416	13.868	0.85	6.42		N
ANISOU	479	NH2	ARG	A	61	5031	4469	3832	1016	454	1486	N

ATOM	469	N	ARG	A	61	-10.741	8.823	14.627	1.00	16.97		N
ANISOU	469	N	ARG	A	61	2376	2310	1761	-150	-161	210	N
ATOM	470	CA	ARG	A	61	-10.770	10.300	14.612	1.00	19.39		C
ANISOU	470	CA	ARG	A	61	2784	2560	2024	-185	-178	200	C
ATOM	471	C	ARG	A	61	-12.150	10.866	14.356	1.00	19.03		C
ANISOU	471	C	ARG	A	61	2816	2476	1938	-121	-150	196	C
ATOM	472	O	ARG	A	61	-12.566	11.837	14.992	1.00	19.76		O
ANISOU	472	O	ARG	A	61	3010	2519	1976	-106	-167	187	O
ATOM	473	CB	ARG	A	61	-9.781	10.925	13.639	1.00	22.17		C
ANISOU	473	CB	ARG	A	61	3123	2904	2394	-268	-175	202	C
ATOM	474	CG	ARG	A	61	-9.905	12.449	13.687	0.91	24.48		C
ANISOU	474	CG	ARG	A	61	3541	3126	2632	-302	-191	192	C
ATOM	475	CD	ARG	A	61	-8.806	13.200	12.998	0.58	26.89		C
ANISOU	475	CD	ARG	A	61	3850	3414	2951	-404	-192	193	C
ATOM	476	NE	ARG	A	61	-7.716	13.495	13.927	0.84	29.72		N
ANISOU	476	NE	ARG	A	61	4196	3777	3318	-479	-250	185	N
ATOM	477	CZ	ARG	A	61	-6.484	12.995	13.809	0.65	31.04		C
ANISOU	477	CZ	ARG	A	61	4251	3997	3545	-544	-261	190	C
ATOM	478	NH1	ARG	A	61	-6.177	12.173	12.806	0.48	30.47		N
ANISOU	478	NH1	ARG	A	61	4077	3975	3523	-539	-211	204	N
ATOM	479	NH2	ARG	A	61	-5.548	13.324	14.688	0.85	32.22		N
ANISOU	479	NH2	ARG	A	61	4390	4150	3702	-611	-326	179	N

ATOM	1003	N	LYS	B	1	18.882	41.841	43.702	1.00	8.49		N
ANISOU	1003	N	LYS	B	1	4694	3203	2531	-264	196	94	N
ATOM	1004	CA	LYS	B	1	18.522	43.123	44.483	1.00	6.03		C
ANISOU	1004	CA	LYS	B	1	3989	2773	2768	-63	197	-282	C
ATOM	1005	C	LYS	B	1	19.014	42.719	45.884	1.00	4.62		C
ANISOU	1005	C	LYS	B	1	3872	2575	2566	-111	409	-341	C
ATOM	1006	O	LYS	B	1	18.824	41.599	46.384	1.00	4.39		O
ANISOU	1006	O	LYS	B	1	4130	2462	2335	-363	369	-473	O
ATOM	1007	CB	LYS	B	1	16.950	43.202	44.464	1.00	7.65		C
ANISOU	1007	CB	LYS	B	1	3953	3150	3020	215	279	-304	C
ATOM	1008	CG	LYS	B	1	16.500	43.954	45.451	1.00	8.67		C
ANISOU	1008	CG	LYS	B	1	3869	3451	3176	155	296	-177	C
ATOM	1009	CD	LYS	B	1	15.018	44.433	45.572	1.00	0.16		C
ANISOU	1009	CD	LYS	B	1	3903	3815	3323	401	218	24	C
ATOM	1010	CE	LYS	B	1	14.518	45.985	46.365	0.76	0.89		C
ANISOU	1010	CE	LYS	B	1	3846	4224	3239	180	249	211	C
ATOM	1011	NZ	LYS	B	1	13.298	45.897	46.772	0.80	5.91		N
ANISOU	1011	NZ	LYS	B	1	3671	5020	4454	-218	147	314	N

ATOM	1003	N	LYS	B	1	4.601	-10.738	-11.943	1.00	18.64	N	
ANISOU	1003	N	LYS	B	1	3267	1885	1930	119	-127	-23	N
ATOM	1004	CA	LYS	B	1	4.347	-11.970	-12.727	1.00	18.53	C	
ANISOU	1004	CA	LYS	B	1	3280	1862	1897	88	-79	-18	C
ATOM	1005	C	LYS	B	1	4.882	-11.812	-14.175	1.00	17.12	C	
ANISOU	1005	C	LYS	B	1	2959	1746	1798	85	-94	-20	C
ATOM	1006	O	LYS	B	1	4.608	-10.786	-14.819	1.00	15.06	O	
ANISOU	1006	O	LYS	B	1	2577	1547	1598	54	-82	-17	O
ATOM	1007	CB	LYS	B	1	2.840	-12.240	-12.814	1.00	19.68	C	
ANISOU	1007	CB	LYS	B	1	3450	2012	2015	-1	29	-10	C
ATOM	1008	CG	LYS	B	1	2.490	-13.315	-13.818	1.00	20.79	C	
ANISOU	1008	CG	LYS	B	1	3587	2159	2154	-46	79	-10	C
ATOM	1009	CD	LYS	B	1	1.064	-13.807	-13.678	1.00	21.01	C	
ANISOU	1009	CD	LYS	B	1	3662	2172	2146	-133	185	-15	C
ATOM	1010	CE	LYS	B	1	0.796	-14.763	-14.836	0.76	21.33	C	
ANISOU	1010	CE	LYS	B	1	3677	2228	2198	-178	223	-20	C
ATOM	1011	NZ	LYS	B	1	-0.648	-15.012	-15.015	0.80	21.73	N	
ANISOU	1011	NZ	LYS	B	1	3715	2293	2248	-274	321	-37	N

Table 1 (a) 4ow9 re-refinement; this Table was prepared using Phenix_Refine's utility (Afonine et al 2012).

4ow9_revisit_occs_set_to_1 Cycle 11

New PDB code 5M1Y

Wavelength (Å)	1.542
Resolution range (Å)	31.27 - 1.904 (1.972 - 1.904)
Space group	P 1 21 1
Unit cell	27.25 62.54 58.98 90 90.92 90
Total reflections	110261
Unique reflections	11097 (329)
Multiplicity	6.9(1.6)
Completeness (%) ^{*1}	71.17 (21.27)
Mean I/sigma(I)	16.73 (1.96)
Wilson B-factor	10.87
R-merge	0.130 (0.229)
Reflections used for R-free	547
R-work	0.2159 (0.2809)
R-free	0.2621 (0.1938)
Number of non-hydrogen atoms	2095
macromolecules	2002

ligands	57
water	36
Protein residues	258
RMS(bonds)	0.003
RMS(angles)	0.72
Ramachandran favored (%)	97
Ramachandran allowed (%)	3
Ramachandran outliers (%)	0
Clashscore	2.51
Average B-factor	14.70
macromolecules	14.60
ligands	20.40
solvent	14.20

Footnotes

*1 the resolution where the data completeness crosses 50% is 2.1 Å.

Responses to the PDB Validation report from the authors:

The 16 amino acids flagged in the RSRZ outlier table, comprising 13 from molecule A and 3 from molecule B, were investigated by omit maps. The difference density for these wasn't very clear and basically there was no further guidance as to an improved positioning from their current position.

Remaining difference map peaks (blue 2Fo-Fc 1.2 rms; green Fo-Fc 3 sigma; orange anomalous map 3.0 sigma):-

Peak 1:-

Previously we had a DMSO here (as shown) but leads to a clash and the density isn't completely certain in terms of shape to be a DMSO. There is a weak anomalous peak so maybe a low occupancy iodine but then the density shape isn't correct. So, since we are not completely chemically certain we leave no assignment to this in the model.

Peak 2:-

Previously we had a low occupancy iodine atom here (as shown) but leads to a clash and there is no anomalous peak. So, since we are not completely chemically certain we leave no assignment to this in the model.

Peak 3:-

Previously we had a partially occupied chlorine atom as shown in the COOT caption. This seems chemically uncertain and we remove this.

Peak 4:-

Previously we had a partially occupied iodine as the COOT caption indicates. This does have a weak anomalous peak as shown but Phenix_refine's excellent diagnostics on such things highlights this as a clash (the near neighbour distance of 2.6Å confirms this). So, we have removed it.

Peak 5:-

This is in a chemically non-sensical position and so we ignore it. It also may quite well indicate the start of the noise level in the Fo-Fc map (COOT's default cut off is 5 sigma).

Peak 6:-

This is in a chemically non-sensical position and so we ignore it. It also may quite well, as peak 5 also, indicates the start of the noise level in the Fo-Fc map (COOT's default cut off is 5 sigma).

Peak 7:-

This is in a chemically non-sensical position and so we ignore it. It also may quite well, as peaks 5 and 6 also, indicates the start of the noise level in the Fo-Fc map (COOT's default cut off is 5 sigma).

Since the last three peaks (5,6 and 7) seem to be noise we now discontinue our responses to the rest of the listed Fo-Fc map peaks.