Appendix A

Crystals and X-ray data for the case described in this paper were produced and collected over a period of 2 years. Full details are included in Table A1. In the table, *Visit ID* refers to the unique code assigned by the Diamond synchrotron user office to the specific experiment at a given beamline.

The crystal used for data collection is named in the *Crystal* column. It is stored in liquid nitrogen inside a *Puck*, containing several crystals. The crystal itself can be large enough for the beam to be shone through it at several *Positions*. The *Serial Number*, thus, assigns a unique number to all sweeps collected for this structure. The remaining 5 columns describe how the crystal was prepared, including details of the presence of a soaked or co-crystallised heavy atom, and cooled down to cryo-temperatures. There are three *Base Conditions*: bc1, bc2 and bc3. They are a mixture of commercial screens, additive screens (not revealed, as they are sensitive data) and gadolinium (Gd). More specifically:

(1) bc1 = A + B + C1
(2) bc2 = A + B + C2
(3) bc3 = A + B + C1 + Gd

where,

A = commercial screen
B = commercial screen for optimization
C1 = additive screen
C2 = additive screen
Gd = gadolinium

There are also three types of *Cryogenic Conditions*:

(1) cry1 = 30% glycerol + 5 mM magnesium chloride
(2) cry2 = 30% glycerol + OH
(3) cry3 = 30% glycerol + 5 mM magnesium chloride + 1 M sodium bromide

Also, some crystals have been dehydrated by addition of salts directly in the crystallization plates [51] with one of two protocols, dh1 or dh2 (column *Dehydration*).

Table A1. Information on all datasets used for the work described in this paper. Crystals were prepared with one of three different base conditions, bc1, bc2 and bc3 (see text). They were also prepared for cooling with one of three different cryogenic conditions, cry1, cry2 and cry3 (see text). The majority of crystals included heavy atoms to attempt SAD phasing. Heavy atoms were soaked in solution for most of the crystals; in a few cases, they were co-crystallised. In order to improve resolution, one of two dehydration screenings have been attempted for many crystals. The table also lists details concerning dates of the various data collections, position of the crystals in the pucks and whether crystals were shot once or more times. The serial number, thus, corresponds to a unique and specific sweep obtained from X-ray diffraction.

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
02/05/2013	mx8031-26	777	xtal1	1 2 3 4	1 2 3 4	bc1	cry2	no	no	no
			xtal3	1 2 3 4 5	5 6 7 8 9	bc1	cry2	no	no	no
			xtal6	1 2 3	10 11 12	bc1	cry2	no	no	no
			xtal8	1 b_1	13 14	bc1	cry2	no	no	no
22/05/2013	mx8681-3	777	xtal3	real_3_2 real_3c_2 real_3d_3 real_3e_2	15 16 17 18	bc1	cry1	no	no	no
			xtal4	1	19	bc1	cry1	no	no	no
			xtal6	6_1 6a_2 6b_1 6b_2 6c_1	20 21 22 23 24	bc1	cry1	no	no	no
			xtal9	9_1 9b_1 9b_2 9b_3	25 26 27 28	bc2	cry1	no	no	no
			xtal14	b_1	29	bc2	cry1	no	no	no
		778	xtal3	3_2	30	?	?	?	?	?

Table A1. *Cont.*

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
20/10/2013	mx8681-13	542	1_13	3 5 6 7 8 9 10	31 32 33 34 35 36 37	bc1	cry1	dh1	no	no
13/02/2014	mx5005-1	136	7	7 7b	38 39	bc1	cry1	dh1	no	$KICl_6$
			8	8 8b	40 41	bc1	cry1	dh1	no	$KICl_6$
			10	10 10b	42 43	bc1	cry1	dh1	no	$KICl_6$
		138	xtal15	15_1	44	bc1	cry1	dh1	no	Tantalum
		542	xtal2	1	45	bc1	cry1	dh1	no	Hg (Thi)
			xtal4	1	46	bc1	cry1	dh1	no	Hg (Thi)
		544	xtal2	_	47	bc1	cry1	dh1	no	Pt (PIP)
		546	xtal4	4_1	48	bc1	cry1	dh1	yes	$KAu(CN)_2$
		758	xtal1	1	49	bc1	cry1	dh1	no	Hg (Ace)
			xtal9	1	50	bc1	cry1	dh1	no	Hg (Thi)
		762	xtal1	2	51	bc1	cry1	dh1	no	K_2PtCl_4
			xtal2	data	52	bc1	cry1	dh1	no	K_2PtCl_4
			xtal4	1	53	bc1	cry1	dh1	no	K_2PtCl_4
			xtal13	1	54	bc1	cry1	dh1	no	$KAu(CN)_2$
			xtal14	1	55	bc1	cry1	dh1	no	$KAu(CN)_2$
			xtal15	1	56	bc1	cry1	dh1	no	$KAu(CN)_2$
		764	xtal14	3 4 5	57 58 59	bc1	cry1	no	no	no
		765	xtal5	5_1	60	bc1	cry1	dh1	no	Hg (PMA)
		766	xtal3	3_1	61	bc1	cry1	dh1	no	K_2PtI_6

Table A1. *Cont.*

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
17/02/2014	cm4982-1	CPS-0134	12	2	62	bc1	cry1	dh1	yes	$OsCl_3$
			13	2	63	bc1	cry1	dh1	yes	K_2PtCl_4
				3	64					
				4	65					
		CPS-0140	7	7_1	66	bc1	cry1	dh1	yes	K_2PtCl_4
			11	11_4	67	bc1	cry1	dh1	yes	Pt (PIP)
			12	12_1	68	bc1	cry1	dh1	yes	AgN
			15	15_1	69	bc1	cry1	dh1	yes	I3C (m.triangle)
		CPS-0761	1	2	70	bc1	cry1	dh1	yes	$GdCl_3$
			2	2	71	bc1	cry1	dh1	yes	$GdCl_3$
				line	72					
			5	1	73	bc1	cry1	dh1	yes	$GdCl_3$
				2	74					
				5	75					
			7	1	76	bc1	cry1	dh1	yes	$GdCl_3$
			9	2	77	bc1	cry1	dh1	yes	$GdCl_3$
				3	78					
02/05/2014	cm4982-2	767	data_0767_2	2	79	bc1	cry1	dh1	no	$KPtCl_4$
				3	80					
				4	81					
				9	82					
				10	83					
				13	84					
				15	85					
			data_0767_7	1	86	bc1	cry1	dh1	no	$KPtCl_4$
				2	87					
				3	88					
				4	89					
				5	90					
				6	91					
				7	92					
				8	93					
				9	94					
				10	95					
				11	96					
				12	97					
				14	98					
				15	99					
				16	100					

Table A1. *Cont.*

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
02/05/2014	cm4982-2	767	data_0767_9	1	101	bc1	cry1	dh1	no	$KPtCl_4$
				3	102					
				4	103					
				5	104					
				6	105					
			data_0767_10	1	106	bc1	cry1	dh1	no	$KPtCl_4$
				2	107					
				3	108					
				4	109					
				5	110					
				6	111					
				7	112					
				11	113					
			data_0767_11	1	114	bc1	cry1	dh1	no	$KPtCl_4$
				2	115					
				3	116					
				4	117					
				5	118					
				7	119					
				8	120					
				9	121					
				10	122					
				11	123					
			data_0767_13	2	124	bc1	cry1	dh1	no	$KPtCl_4$
				3	125					
				6	126					
			data_0767_14	1	127	bc1	cry1	dh1	no	$KPtCl_4$
				2	128					
				3	129					
				4	130					
			data_0767_15	1	131	bc1	cry1	dh1	no	$KPtCl_4$
				2	132					
				3	133					
				4	134					
		754	1	1_1	135	bc1	cry3	dh2	no	$KPtCl_4$
				1_2	136					
				1_3	137					
				1_4	138					
				1_5	139					
				1_6	140					
				1_7	141					
				1_8	142					
				1_9	143					
				1_10	144					

Table A1. *Cont.*

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
02/05/2014	cm4982-2	754	4	4_1 4_2 4_3 4_4 4_5 4_6 4_7 4_8 4_9 4_10 4_11	145 146 147 148 149 150 151 152 153 154 155	bc1	cry3	dh2	no	$KPtCl_4$
			5	5_1 5_2 5_3	156 157 158	bc1	cry3	dh2	no	$KPtCl_4$
		758	02	2_2 2_3	159 160	bc3	cry1	dh1	no	Os
			03	3_1	161	bc3	cry1	dh1	no	Os
			04	4_2 4_4 4_5	162 163 164	bc3	cry1	dh1	no	Os
			05	5_1 5_2 5_3 5_4 5_5	165 166 167 168 169	bc3	cry1	dh1	no	Os
			06	6_1 6_2 6_3 6_4 6_5	170 171 172 173 174	bc3	cry1	dh1	no	Os
			08	8_1	175	bc3	cry1	dh1	no	Os
			10	10_1 10_2	176 177	bc3	cry1	dh1	no	Os
			11	11_1 11_2	178 179	bc3	cry1	dh1	no	Os
			13	13_1 13_2	180 181	bc3	cry1	dh1	no	Os
			15	15_2 15_3 15_4 15_5 15_6	182 183 184 185 186	bc3	cry1	dh1	no	Os

Table A1. *Cont.*

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
02/05/2014	cm4982-2	765	1	1_1	187	bc3	cry1	dh1	no	$KPtCl_4$
				1_3	188					
				1_4	189					
				1_5	190					
				1_6	191					
				1_7	192					
				1_8	193					
				1_9	194					
				1_10	195					
				1_11	196					
				1_12	197					
			2	2_3	198	bc3	cry1	dh1	no	$KPtCl_4$
				2_4	199					
				2_5	200					
				2_6	201					
				2_7	202					
				2_8	203					
				2_9	204					
				2_10	205					
				2_11	206					
				2_12	207					
				2_13	208					
				2_14	209					
			5	5_4	210	bc1	cry1	dh1	no	Os
				5_6	211					
				5_7	212					
				5_8	213					
				5_9	214					
30/06/2014	cm4982-3	758	0758_3	3_1	215	bc1	cry1	dh2	no	$KPtCl_4$
				3_2	216					
				3_3	217					
				3_4	218					
			0758_4	4_1	219	bc3	cry1	dh2	no	Os

Table A1. *Cont.*

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
30/06/2014	cm4982-3	758	0758_5	5_1	220	bc3	cry1	dh2	no	Os
				5_2	221					
				5_3	222					
				5_4	223					
				5_5	224					
				5_6	225					
				5_7	226					
				5_8	227					
				5_9	228					
				5_10	229					
				5_11	230					
				5_12	231					
				5_13	232					
				5_14	233					
			0758_6	6_1	234	bc3	cry1	dh2	no	Os
				6_2	235					
				6_4	236					
				6_5	237					
				6_6	238					
			0758_8	8_1	239	bc3	cry1	dh2	no	Os
				8_2	240					
				8_3	241					
			0758_9	9_1	242	bc3	cry1	dh2	no	Os
				9_2	243					
				9_3	244					
				9_4	245					
				9_5	246					
				9_6	247					
				9_7	248					
			0758_10	10_1	249	bc3	cry1	dh2	no	Os
				10_2	250					
				10_3	251					
				10_4	252					
				10_5	253					
				10_6	254					
			0758_11	11_1	255	bc3	cry1	dh2	no	Os
				11_2	256					
				11_3	257					
				11_4	258					
				11_5	259					
				11_6	260					

Table A1. *Cont.*

Date	Visit ID	Puck	Crystal	Position	Serial N.	Base Condition	Cryogenic Condition	Dehydration	Co-Crystallized	Heavy Atom
30/06/2014	cm4982-3	758	0758_12	12_1	261	bc1	cry1	dh2	no	Os
				12_2	262					
				12_3	263					
				12_4	264					
				12_5	265					
				12_6	266					
				12_7	267					
			0758_13	line	268	bc1	cry1	dh2	no	Os
			0758_14	14_1	269	bc1	cry1	dh2	no	Os
				14_2	270					
			0758_15	15_1	271	bc1	cry1	dh2	no	Os

Appendix B

A table different from Table A1, but related to it, is Table 1, included in Section 4.3. Table 1 is a representation of a reshaped dataframe, an object present in the *R programming language* [46]. In this appendix, it will be explained how the reshaped dataframe is obtained. The starting point is the manual construction of a dataframe associated with Table A1. Several solutions can be envisaged to avoid this time-consuming task, all of them making use of database algorithms. These will be implemented shortly in *BLEND*, but for the work described in this article, preparation of the initial dataframe and the subsequent formation of the reshaped dataframe were carried out manually. A few lines of the code for the initial dataframe are shown in Figure A1.

	Date	VisitID	Puck	Crystal	Position	BaseCondition	CryogenicCondition	Dehydration	CoCrystallization	HeavyAtom
1	02/05/2013	mx8031-26	777	xtal1	1	bc1	cry2	no	no	no
20	22/05/2013	mx8681-3	777	xtal6	6_1	bc1	cry1	no	no	no
50	13/02/2014	mx5005-1	758	xtal9	1	bc1	cry1	dh1	no	Hg(Thi)
130	02/05/2014	cm4982-2	767	data_0767_14	4	bc1	cry1	dh1	no	K2PtCl4
200	02/05/2014	cm4982-2	765	2	2_5	bc3	cry1	dh1	no	K2PtCl4
201	02/05/2014	cm4982-2	765	2	2_6	bc3	cry1	dh1	no	K2PtCl4
202	02/05/2014	cm4982-2	765	2	2_7	bc3	cry1	dh1	no	K2PtCl4
203	02/05/2014	cm4982-2	765	2	2_8	bc3	cry1	dh1	no	K2PtCl4
204	02/05/2014	cm4982-2	765	2	2_9	bc3	cry1	dh1	no	K2PtCl4
205	02/05/2014	cm4982-2	765	2	2_10	bc3	cry1	dh1	no	K2PtCl4

Figure A1. Initial R dataframe, corresponding to Table A1. Just a few lines of the dataframe are shown in this figure.

The dataframe is a simple matrix in which each row corresponds to a single dataset. As multiple datasets can be associated with a same *Date*, *VisitID*, *Puck*, etc., then values for these columns are, often, repeated. Next, a dataframe including all possible combinations from the unique conditions in the initial dataframe, is created. Let us call this dataframe *theoretical conditions dataframe*. It turns out that the base conditions (BC) comprise 3 unique values (bc1, bc2, bc3), the cryogenic conditions (CC) also comprise 3 unique values (cry1, cry2, cry3), the dehydration protocol includes 3 unique values (no = no dehydration, dh1, dh2), the co-crystallisation flag (CO) includes two values (yes, no), and the heavy atom types (HA) are 15 (no = no heavy atom, KlCl6, Tantalum, Hg(Thi), Pt(PIP), KAu(CN)2, Hg(Ace), K2PtCl4, Hg(PMA), K2PtI6, OsCl3, AgN, IC3(m_triangle), GdCl3, Os). The possible combinations from all values listed above are $3 \times 3 \times 3 \times 2 \times 15 = 810$. This means that the theoretical conditions dataframe has 810 rows. Not all possible combinations will be present in the data collected for this work, because the total number of datasets is 271. For this reason, the initial dataframe entries are matched against the theoretical conditions dataframe; the result of this comparison is the new dataframe, simply called *conditions dataframe*, shown in Table 1.

Appendix C

With molecular replacement, models are oriented and placed at specific locations of the unit cell. Two solutions from molecular replacement runs do not necessarily overlap, even if they correspond to the same correct structure. The reason for this is that the asymmetric units selected by the molecular replacement program could be different. Furthermore, the absolute location of the oriented molecule depends on where the unit cell origin has been placed. The origin can be selected arbitrarily to be compatible with the specific symmetry. Thus, to verify whether two molecules overlap, all symmetry equivalents of the molecules and all allowed origin shifts must be tried. Within the CCP4 group of programs, this task is carried out by the program *CSYMMATCH* [52]. The input consists of the two files containing the atomic coordinates of the two models to be compared; one is considered the moving model, the other the reference model. The output consists of a PDB file corresponding to the moving model, transformed to the closest possible location to the reference model still compatible with symmetry and allowed unit cell origin. To compute the RMSD between all atoms of the reference structure and all atoms of the moved structure, we have used the CCP4 program *COMPAR*. This is an old program with no related documentation on the official CCP4 website. Details on how to run

this program have been learned via the CCP4 Bulletin Board [53]. The value for the two structures discussed in this paper is RMSD = 0.773 Å.