In-house high energy remote SAD-phasing using the magic triangle: how to tackle the P1 low symmetry using multiple orientations on the same human IBA57 crystal to increase multiplicity

Gourdoupis Spyridon [#], Nasta Veronica [#], Ciofi-Baffoni Simone [#§], Banci Lucia [#§] and Calderone Vito [#§*]

[#] CERM - Center for Magnetic resonance, University of Florence, via Luigi Sacconi 6, 50019 Sesto Fiorentino (FI), Italy

[§] Department of Chemistry, University of Florence, via della Lastruccia 3, 50019 Sesto Fiorentino (FI), Italy

* Corresponding author

email address: calderone@cerm.unifi.it (Calderone Vito)

Keywords: SAD phasing, I3C, magic triangle, P1, IBA57, SIRAS

Abstract

This paper describes the approach used to solve in-house the structure of human IBA57 through 5-amino-2,4,6-triiodoisophthalic acid (I3C) high energy remote SAD-phasing. Multiple orientations of the same triclinic crystal have been exploited to acquire sufficient real data multiplicity for phasing. It is described how the collection of an in-house native dataset and its joint use with the I3C derivative through a SIRAS approach decreases the data multiplicity needed by almost 50%. Furthermore, it is illustrated that there is a clear data multiplicity threshold value for success and failure in phasing and how adding further data does not significantly affect substructure solution and model building.

To our knowledge, this is the only structure present in the PDB which has been solved in-house, through remote SAD, in space group P1 and using one crystal only. All raw data used, deriving from the different orientations, have been deposited to Zenodo (DOI 10.5281/zenodo.2531553) in order to enable software developers to improve methods for data processing and structure solution and for educational purposes.

Introduction

About two-thirds of recently deposited structures in the PDB have been solved using molecular replacement which requires the availability of an already solved structure of a protein homologous to the unknown one. However, experimental phasing has the advantage of being free from model bias and is required for samples that do not have any structurally related entries. Single-wavelenght anomalous dispersion (SAD) has become the most used method since the first years of this century for de novo determination of protein crystal structures and it accounts for over 70% of structures determined by experimental phasing, deposited in the PDB (Berman, 2000, Bunkoczi *et al.*, 2015).

Heavy-atom soaks normally show a low success rate, but systematic heavy-atom screening with conventional heavy-metal ions has been performed using gel electrophoresis (Boggon & Shapiro, 2000), mass spectrometry (Agniswamy *et al.*, 2008) or a database approach (Sugahara *et al.*, 2005).

The combination of phase information from a partial molecular-replacement solution and weak experimental phases has also proven to be successful in a number of cases (Tereshko *et al.*, 2008, Schuermann & Tanner, 2003, Roversi *et al.*, 2010).

Since soaking with heavy-atoms often suffers from non-specific binding, a new class of compounds has been developed that combines heavy atoms with functional groups for binding to proteins. The molecule 5-amino-2,4,6-triiodoisophthalic acid (I3C) contains three functional groups (two carboxylate groups and one amino group) that interact with proteins through hydrogen bonds. Three iodine atoms suitable for anomalous dispersion phasing are arranged in an equilateral triangle and are thus easily identifiable in the substructure (Beck *et al.*, 2008, Beck *et al.*, 2010) . Anomalous scatterers such as halide ions or halogenated fragments have been more and more exploited for experimental phasing (Dauter & Dauter, 2007, Kim *et al.*, 2013, Liu *et al.*, 2011, Ramagopal *et al.*, 2003, Salgado *et al.*, 2005, Bauman *et al.*, 2016).

Halides show a very large anomalous signal at the common wavelengths adopted at synchrotrons (often corresponding to some of their absorption edges) but they still show a sufficient anomalous

difference also at the lower energies typical for in-house diffractometers. The success of such phasing depends on many factor such as the occupancy, on the orderliness of the surface bound halide atoms and of course on the absence of radiation damage.

More in general SAD success depends on how data are acquired, merged and scaled, on their real multiplicity, on their completeness and on the anomalous effect intensity (which also depends on how far the energy at which we are acquiring data is from the absorption edge of the anomalous scatterer). A high multiplicity is easily achievable when the crystal belongs to a high symmetry space group, but lower symmetry examples have also proven successful (Lakomek *et al.*, 2009).

When the crystal belongs to a low symmetry space group or when a high symmetry one decays rapidly during data collection the multi-crystal averaging approach can be helpful: it involves merging data sets from statistically equivalent crystals (Liu *et al.*, 2012, Liu *et al.*, 2013), enhancing the anomalous signal-to-noise ratio and minimizing anomalous signal decay due to radiation damage.

A second approach involves collecting many low-dose data sets from a single crystal in multiple orientations, mitigating radiation damage while also reducing systematic errors in data collection (Debreczeni *et al.*, 2003, Brockhauser *et al.*, 2013, Weinert *et al.*, 2015, Finke *et al.*, 2016). By changing the crystal orientation, it is possible to measure the same reflections in different diffraction directions on different areas of the detector and with different sample absorption, thus minimizing systematic errors.

Although the several recent advancements in data collection strategies, difficult cases such as low-symmetry space groups, low diffraction resolution, small crystals, weak anomalous signal or some combination thereof still remain a challenge for SAD phasing.

The two strategies illustrated above allow the extraction of as much anomalous signal as possible with the highest possible accuracy from as many crystals as needed for successful phasing (Klinke *et al.*, 2015).

In order to solve the structure of human IBA57, which crystallizes in space group P1, we have adopted a methodological approach based on an in-house remote SAD-phasing using the magic triangle I3C

and on multiple orientation of the same crystal in order to acquire real multiplicity and thus minimizing the difficulties of merging datasets coming from different crystals.

IBA57 is a 325-residue human protein which localizes to the mitochondrion and is part of the iron-sulfur cluster assembly pathway (Rouault, 2015, Andreini *et al.*, 2017).

The maturation of mitochondrial iron-sulfur proteins requires a complex protein machinery. In the late steps of this machinery, a [2Fe-2S] cluster is converted into a [4Fe-4S] cluster. Human IBA57 protein acts in this step as an iron-sulfur cluster assembly component along with ISCA1 and ISCA2 (Brancaccio *et al.*, 2014, Brancaccio *et al.*, 2017, Ciofi-Baffoni *et al.*, 2018).

The structure of this protein was still unknown and the closest homologue whose structure was determined showed 25% sequence homology only (which made molecular replacement very unlikely to be successful). For this reason, experimental phasing was the most likely way to solve it. We had thus adopted an approach based on an in-house I3C SAD data collection on a single P1 crystal in multiple orientations in order to gain sufficient multiplicity.

Materials and Methods

Cloning, expression and purification

Human IBA57 has been cloned, overexpressed and purified at homogeneity for crystallization experiments as described elsewhere (Gourdoupis *et al.*, 2018).

Crystallization, data collection and structure determination

The purified IBA57 protein sample had a concentration of 0.4 mM and was dissolved in a solution containing 50 mM phosphate buffer, 150 mM NaCl and 2.5 mM TCEP. Crystallization trials on IBA57 were performed by the sitting drop vapor diffusion method at 20 °C by mixing an equal volume of the sample and of a solution containing 0.1 M MES pH 6.0 and 20 % PEG3350. Bunches of large, thin and superposed crystals (Fig. 1) started to grow overnight and grew to final size in a few days. The protruding edges of these bunches of superposed crystals were cut so as to obtain the

best possible approximation to a single crystal useful for diffraction. The chosen crystal was then soaked in the mother liquor containing 25 % ethylene glycol and 0.5 M 5-amino-2,4,6-triiodoisophthalic acid (I3C) (previously activated with LiOH) for 2-3 minutes as a longer time caused the crystal to crack.

The crystal, of approximate size 200x70x10 μm, diffracted to 2.0 Å resolution with data intensity being acceptably intense (1.6 I/σI in the high resolution shell) up to 2.3 Å; it belonged to the triclinic space group P1 (a = 38.08 Å, b = 43.03 Å, c = 59.24 Å, α = 77.90°, β = 75.63°, γ = 71.66°) with one molecule in the asymmetric unit and a solvent content of about 50%. Since the crystal belonged to the lowest possible symmetry space group and since SAD experiments need quite high data multiplicity in order to be successful, nine diffraction experiments at 100 K have been performed under several orientations (Table 1) using in-house K_α radiation on one crystal only, which proved to be the best crystal among quite poorly diffracting ones.

The data were collected by the rotation method on an Oxford Diffraction Enhance PX-Ultra diffractometer using 1.0° oscillation per frame for a total of 1136 degrees fully exploiting the 4-circle goniometer, thus changing the values of phi, theta, kappa and omega angles. The crystal to detector distance was 60 mm and the beam size on the sample was 0.3x0.3 mm. The detector was an Onyx CCD with a 165 mm active area.

The nine datasets were processed using XDS (Kabsch, 2010) and scaled using XSCALE (Kabsch, 2010). The relevant statistics are shown in Table 2.

The analysis of the SAD Patterson maps, density modification and chain tracing were performed with the program SHELXC/D/E (Sheldrick, 2010) as implemented in the HKL2MAP (Pape & Schneider, 2004) interface, using the first six datasets and it provided the positions of twelve iodine atoms in the asymmetric unit with good occupancies. The preliminary phases obtained were then improved by further phase refinement and density modification techniques using a solvent content of 50% with the program autoSHARP (Vonrhein *et al.*, 2007). Likewise the model tracing was performed by looping between BUCCANEER (Cowtan, 2006) and ARP/wARP (Langer *et al.*, 2008), which, in the end,

allowed for a complete chain tracing. Structure determination will be explained in more details in the Results and Discussion section.

The refinement was then carried out using BUSTER (Bricogne G. *et al.*, 2016) making use of TLS restraints and water molecules were added during refinement. In between refinement cycles the model was subjected to manual building using COOT (Emsley & Cowtan, 2004). The stereochemical quality of the refined model was assessed using the program MOLPROBITY (Chen *et al.*, 2010). The Ramachandran plot shows a good geometry and no residues in disallowed regions. Figure 2 shows the ribbon representation of the solved structure of human IBA57 with the residue for which even the main chain electron density is very poor if not absent at all highlighted in red (53-59, 61, 88-92, 115-118, 138-147, 262, 296-300, 306-311). The relevant refinement statistics are shown in Table 2 and include the above mentioned residues.

Two more datasets have been collected after the above-mentioned SAD structure determination: a native dataset has been collected in-house (whose resolution was comparable to that of the I3C derivative) in order to test to what extent SIRAS would have made structure determination easier, and, lately, another native dataset has been collected at the synchrotron using beamline ID30B at ESRF (Grenoble, France) in order to achieve a higher resolution.

In the latter case the structure was solved using the in-house experimentally SAD-solved structure as the molecular replacement template and yielding a structure refined at 1.75 Å resolution. The refinement statistics for the synchrotron collected dataset are shown in Table 2.

At the time of structure solution coordinates and structure factors were deposited and released in the Protein Data Bank under the accession codes 5OLI (for the in-house I3C derivative) and 6ESR (for the higher resolution synchrotron structure); at the time of writing the present manuscript both entries have been re-refined in order to optimize model quality and statistics and so they have been superseded by 6QE4 and 6QE3 respectively. All what discussed hereafter will then apply to re-refined models (6QE4 and 6QE3) although named with the original entry ID (5OLI and 6ESR).

Results and Discussion

SAD vs. SIRAS

Due to the immediate in-house availability of I3C the first attempt was that of a SAD experiment: it was in fact clear from the very beginning that experimental phasing of IBA57 would be the strategy for solving its structure due to the absence of available models with sequence identity higher than 25%.

SAD was performed on iodine at the Cu_{α} wavelength which, although well above the L-I absorption edge of the element (2.389 Å), still grants about six electrons difference between f' and f''. The crystal was thus soaked for less than two minutes in a solution containing I3C and then frozen in the N_2 stream of the cryojet.

While the data collection was progressing, it became clear that the crystal belonged to space group P1 and so the collection of 360° of rotation around phi would not ensure a great data multiplicity due to the triclinic lowest possible symmetry. Multiplicity is in fact essential in anomalous dispersion phasing (especially in a SAD experiment) and allows for more accurate data measurements: the little anomalous differences must be greater than the error on the intensity estimate in order to be meaningful and to eventually led to structure solution. It is worth to underline that by 'multiplicity' we intend real multiplicity i.e. the measurement of equivalent or identical reflections with different paths through the crystal, not just repeated measurements which is obtainable either with high crystal symmetry or by the use of a 4-circle goniometer (Debreczeni *et al.*, 2003).

Since the full phi rotation would not be sufficient to achieve a reasonable real multiplicity on a triclinic crystal, other experiments were setup. Using one crystal with a good diffraction, nine datasets were collected on the same crystal exploiting the potential of the 4-circle goniometer (Table 1) leading to a total of 1136 degrees rotation.

The substructure was solved using all the nine datasets which were merged and scaled together keeping Friedel pairs separated: collectively these datasets allowed us to reach an anomalous multiplicity of about 5.6. The relevant intensities were then fed into the standard routine of

SHELXC/D/E through the HKL2MAP interface. SHELXD found a solution with 12 iodine sites arranged in four triangles (Figure 3) with CC All/Weak 24.8/11.2 and CFOM= 36.0 as figure of merit and correlation respectively, using the data up to 2.5 Å resolution with the default number of cycles; this allowed SHELXE to trace over a hundred residues as poly Ala in one of the two hands out of the 325 of the entire sequence. The chain was not continuous but the electron density map of this solution did not appear random and it was possible to find patches of density in which a main chain could fit. So this partial model output by SHELXE (without the heavy atom sites) was then fed into the standard autoSHARP routine. This routine refines phases, carries out density modification, traces with BUCCANEER first while performing density modification with PARROT, then carries out another solvent flattening on the model output by BUCCANEER and in the end performs a last cycle of model building through ARP/wARP. This approach lead to a partial model traced by BUCCANEER showing 184 residues out of the 307 expected with the longest chain of 31 residues. A visual inspection of the electron density map computed with this model showed several well defined regions (Fig. 4) and was very promising despite a not encouraging R_{free} value (0.51). This partial model was then input again into the autoSHARP routine leading to a great improvement which yielded a model of 250 residues with the longest chain of 54 and an R_{free} of 0.40. Then, again, this model was used as input for a last autoSHARP routine and this eventually lead to a complete tracing of the structure with an R_{free} of 0.38. The model was then refined and water added to the final R_{free}/R_{cryst} of 26.0/21.5.

Figure 5 shows the anomalous electron density contoured at 3.0 σ level using computed using the anomalous differences and the phased of the refined structure.

From a browsing of the PDB this is the first case, to our knowledge, of a protein structure which has been experimentally solved in-house, through SAD, using one crystal only in multiple orientations and in space group P1 which is the one with the lowest possible symmetry.

At this point it is relevant to define the smallest possible number of datasets needed to solve the structure and to identify what would be a reasonable threshold i.e. the minimum multiplicity and phasing power needed to solve the phase problem.

Thus, we tested the SAD phasing efficiency of dataset 1 alone and dataset 1 plus the other datasets which have been collected using the very same procedure applied for datasets 1 to 9. What appeared evident was that it was not possible to trace the structure with a number of datasets lower than 7 although the SHELXC/D/E statistics are more or less the same in all cases. Only starting from datasets 1-7 (multiplicity of about 4.8) it was possible to solve and trace the full structure possibly because of the increase of the phasing power along with the increasing number of datasets included. The dataset including datasets 1 to 6 seems to be the borderline for successful structure solutions (Table 3). The same situation was also found if using a random combination of datasets. We would like to point out that, even in the unsuccessful dataset ensembles, the typical phasing statistics and the R_{free}/R_{cryst} values appear very similar to those of the successful ensembles; nevertheless, chain tracing is not successful indicating a poor phase quality given by a not sufficient data multiplicity.

Furthermore, the 12 anomalous sites found in the successful ensemble including the lowest number of datasets (1-7) have been compared with the anomalous sites found in each of the remaining datasets ensembles using PHENIX.EMMA (Adams *et al.*, 2010). It turned out that even the ensembles with 1 to 4 datasets show half of the sites in common with ensemble 1-7, whereas ensemble 1-5 shows 10 sites in common and ensemble 1-6 shows 7 sites in common. On the other hand, ensemble 1-8 and 1-9 show 8 and 6 sites in common with 1-7 respectively. This would allow to infer that it is not so much the number of initial anomalous sites found which is crucial for solving the structure but it is rather the accuracy of the measured anomalous differences which in turn depends on the real multiplicity on the data.

We then obtained another crystal from the very same sample preparation of comparable quality to the one used for structure solution. We then collected a native dataset of it and analyzed to what extent SIRAS would be more efficient than SAD alone to solve the structure i.e. to what extent multiplicity could be lowered exploiting the isomorphous effect of iodine for structure solution.

The same procedure used for the first crystal (SHELXC/D/E and autoSHARP routine and subsequent loops with sequential partial models) was applied on the native dataset plus the various I3C datasets

added one by one to the native one. In this case, three I3C datasets only were enough to succeed reducing the necessary multiplicity to a value below 3 thus decreasing it by half with respect to the SAD alone (Table 4).

Also in this case a random combination of datasets was tried out and the final result did not change.

6ESR vs 5OLI

It is indeed interesting to compare the structures of 6ESR and 5OLI; as described above, three IBA57 datasets have been collected in total and all of them belong to the triclinic space group. The I3C in-house derivative 5OLI (with cell parameters a= 38.08 b= 43.03 c= 59.24, α= 77.90 β= 75.63 γ= 71.66), the in-house native (with cell parameters a= 38.18 b= 43.04 c= 58.40, α= 76.55 β= 76.02 γ= 71.28) and the synchrotron native 6ESR (with cell parameters a= 37.28 b= 43.25 c= 55.30, α= 94.88 β= 93.25 γ= 108.18). The volume of the cell of one of the two in-house dataset is 88338.5 $Å^3$ and that of the synchrotron one is 84078.5 $Å^3$ yielding a difference of about 4260 $Å^3$. It is also worth underlining that, whereas the two in-house datasets show very similar cell parameters, the synchrotron one shows significant discrepancy with respect to the other two although the space group remains the same. A plausible explanation to this discrepancy comes from the fact that the crystals for in-house diffraction come from the same protein preparation whereas the crystal used at the synchrotron comes from a different one.

Furthermore, fig. 6 shows the RMSD of their C-alpha superposition. The deviation between them is generally very low or low, but there are five regions (those with very weak or no electron density) where it becomes considerably high.

It is also interesting to compare the respective secondary structure calculated with the program STRIDE (Frishman & Argos, 1995) as illustrated in Fig. 7. As also shown in Fig. 8, it appears that, in the case of 6ESR, there is a slight loss in secondary structure elements (mainly β-strands in the N-terminus region) with respect to 5OLI and there is the appearance of a very short 3/10 helix around

residue 90. It must be pointed out anyway that those regions mostly correspond to the regions in which electron density is very weak and thus model tracing can be quite approximate.

Conclusions

This work shows the approach used for the experimental phasing of human IBA57 structure. It is the only structure to our knowledge that has been solved in-house through remote SAD on an I3C derivative using one crystal only, despite the crystal being triclinic. In order to achieve sufficient multiplicity several orientations of the same crystal were collected. The multiplicity threshold for successful phasing appeared to be around five, independent of the combination of datasets used. This value can be reduced to about half through SIRAS by exploiting the isomorphous differences with a second native dataset reaching a successful multiplicity value of less than three. All the raw data of the I3C derivative and the in-house native will be deposited to ZENODO (DOI 10.5281/zenodo.2531553) both for educational purposes and to enable other crystallographers to improve methods for data processing and structure solution and thus to benefit from these findings.

Acknowledgements

The experiments were performed on beamline ID30B at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Local Contact at the ESRF for providing assistance in using beamline ID30B.

We also thank professor Kay Diederichs for suggestions and fruitful interactions.

The authors acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre.

This research is framed within the activities of the COST Action CA15133, supported by COST (European Cooperation in Science and Technology).

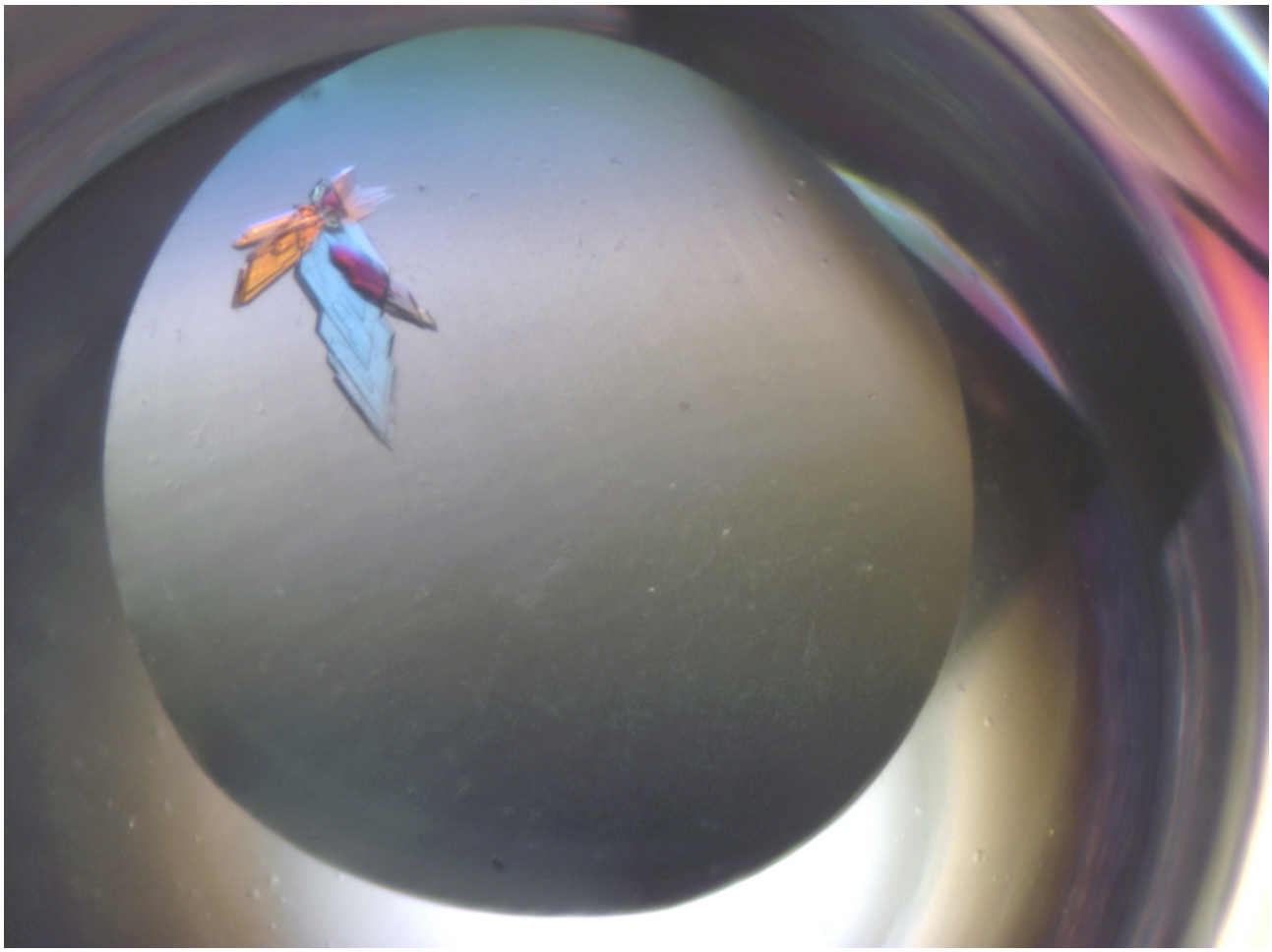

Fig. 1 Screenshot of IBA57 crystals

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8	Dataset 9
Theta (°)	0	0	0	0	0	0	0	0	0
Phi (°)	0 to 360	0	90	45	135	0	90	45	135
Kappa (°)	0	30	30	30	30	-30	-30	-30	-30
Omega (°)	0	-24 to 73	-24 to 73	-24 to 73	-24 to 73	-73 to 24	-73 to 24	-73 to 24	-73 to 24

Table 1 Description of the four angle ranges for each of the nine IBA57-I3C datasets collected

	IN-HOUSE IBA57-I3C	IBA57 SYNCHROTRON NATIVE
Diffraction source	Sealed tube	ID30B ESRF

Wavelength (Å)	1.54	0.97
Temperature (K)	100	100
Detector	Onyx CCD	Pilatus 6M
Crystal-to-detector distance (mm)	60	319.25
Oscillation range (°)	1	0.1
Total rotation range (°)	1136	360
Exposure time/image (s)	120	0.08
Space group	P1	P1
a, b, c (Å)	38.08 43.03 59.24	37.28 43.25 55.30
α, β, γ (°)	77.90 75.63 71.66	94.88 93.25 108.18
Mosaicity (°)	0.3	0.2
Resolution range (Å)	50.0 - 2.3	50.0 – 1.7
Total reflections	181082 (17654)	126801 (15749)
Unique reflections	15237 (1521)	31785 (4346)
Completeness (%)	99.9 (99.9)	96.7 (95.6)
Anomalous reflections	14701 (1521)	-
Anomalous multiplicity	5.7 (5.5)	-
Anomalous completeness	97.5 (96.5)	-
Anomalous CC1/2	0.99 (0.72)	-
CC1/2	-	0.99 (0.81)
I/σ(I)	10.7 (1.6)	15.7 (2.6)
$R_{p.i.m.}$	0.08 (0.72)	0.03
$R_{p.i.m.}$Ov	0.06 (0.51)	-
R_{merge} †	0.19 (0.91)	0.07 (0.65)
Wilson plot B factor (Å^2)	40.0	35.7
N° of sites	12	-
f'/f'' (refined)	-0.5 / 6.7	-
FOM (SHARP)	0.25	-
FOM (PARROT)	0.65	-
R_{cryst} / R_{free}‡ (%)	20.8/22.8	17.4/19.7
Protein atoms	2346	2343
Water molecules	88	129
Ligand atoms	63	-
RMSD bond lengths (Å)	0.010	0.009
RMSD bond angles (º)	1.090	0.995

Table 2 Data collection and refinement statistics of the in-house collected I3C derivative and of the synchrotron collected high resolution native dataset

Values in parenthesis refer to the high resolution shell (2.42-2.34 Å for 5OLI and 1.85-1.75 for 6ESR)

†$R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the mean intensity of the ith observation of symmetry-related reflections hkl.

‡$R_{cryst} = \sum_{hkl} \left| |F_{\text{obs}}| - |F_{\text{calc}}| \right| / \sum_{hkl} |F_{\text{obs}}|$, where F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with a randomly selected 7% of the reflections).

SAD (IN-HOUSE IBA57-I3C)									
IBA57-I3C dataset(s)	1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9
Anomalous multiplicity (AIMLESS)	2.0	2.5	2.9	3.4	3.8	4.4	4.8	5.2	5.6
Anomalous CC1/2	0.98 (0.46)	0.98 (0.55)	0.98 (0.56)	0.98 (0.55)	0.99 (0.56)	0.99 (0.56)	0.99 (0.62)	0.99 (0.65)	0.99 (0.70)
SHELXD CC All/Weak CFOM	24.6/11.2 35.8	25.4/10.6 36.0	25.4/10.5 35.9	25.6/12.2 37.8	25.9/11.6 37.5	24.8/11.2 36.0	25.2/10.0 35.2	26.2/10.2 36.4	25.3/11.7 37.0
SHELXE n° of residues n° of chains CC partial structure vs native data (%)	65 8 7.1	87 10 7.5	89 12 6.1	72 9 5.0	98 13 6.9	75 10 6.7	89 11 6.5	54 7 6.3	78 10 7.3
AutoSHARP FOM	FAILED	FAILED	0.20	0.21	0.19	0.24	0.24	0.19	0.24
PP anom	FAILED	FAILED	0.49	0.51	0.55	0.53	0.54	0.58	0.64
Buccaneer	FAILED	FAILED	156	142	190	154	184	136	165
Longest chain	FAILED	FAILED	17	25	29	17	31	38	30
Sequenced	FAILED	FAILED	0	0	0	0	31	0	0
R_{free}/R_{cryst}	FAILED	FAILED	0.52/0.52	0.52/0.50	0.52/0.50	0.50/0.50	0.51/0.49	0.50/0.50	0.51/0.51
ARP/wARP	FAILED	FAILED	114	113	123	111	132	103	111
Longest chain	FAILED	FAILED	9	13	15	19	20	10	18
Sequenced	FAILED	FAILED	0	0	0	0	0	9	0
R_{free}/R_{cryst}	FAILED	FAILED	0.43/0.30	0.45/0.33	0.48/0.33	0.46/0.30	0.46/0.34	0.44/0.29	0.44/0.34

Table 3 Statistics for each ensemble of datasets of the substructure solution, phase refinement and chain tracing for SAD. The red characters show the failure in chain tracing with the relevant multiplicity value; the green show the successful values. For the anomalous CC1/2 line the values in parenthesis refer to the highest resolution shell (2.4-2.3 Å)

SIRAS (IBA57 IN-HOUSE IBA57-I3C + IN-HOUSE NATIVE IBA57)									
IBA57-I3C dataset (s)	1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9
Anomalous multiplicity (AIMLESS)	2.0	2.5	2.9	3.4	3.8	4.4	4.8	5.2	5.6
Anomalous CC1/2	0.98 (0.46)	0.98 (0.55)	0.98 (0.56)	0.98 (0.55)	0.99 (0.56)	0.99 (0.56)	0.99 (0.62)	0.99 (0.65)	0.99 (0.70)
SHELXD CC All/Weak CFOM	22.1/12.6 34.7	20.9/11.4 32.2	21.4/11.9 33.3	21.2/11.4 32.6	22.2/12.3 34.5	22.2/18.8 35.0	23.0/13.4 36.4	23.6/14.2 37.7	22.7/13.2 35.9
SHELXE n° of residues n° of chains CC partial structure vs native data (%)	110 15 8.8	96 14 7.9	103 13 10.3	86 10 10.2	94 12 7.8	101 12 7.50	92 9 9.0	95 13 14.6	85 11 13.1
AutoSHARP FOM	0.20	0.24	0.21	0.21	0.20	0.18	0.19	0.22	0.22
PP anom	0.62	0.65	0.60	0.61	0.61	0.60	0.58	0.65	0.66
PP iso	1.05	1.15	1.0	1.01	1.00	0.92	0.86	1.00	0.99
Buccaneer	61	127	277	241	153	190	172	285	180
Longest chain	14	24	81	22	26	33	18	64	24
Sequenced	0	14	200	0	17	30	2	255	0
R_{free}/R_{cryst}	0.53/0.52	0.52/0.53	0.47/0.47	0.51/0.51	0.54/0.53	0.51/0.50	0.49/0.48	0.45/0.45	0.52/0.52
ARP/wARP	64	80	179	134	110	145	111	213	94
Longest chain	11	9	27	13	10	27	11	48	9
Sequenced	0	0	44	0	0	0	0	118	7
R_{free}/R_{cryst}	0.49/0.33	0.50/0.33	0.44/0.30	0.46/0.31	0.49/0.35	0.47/0.33	0.51/0.35	0.38/0.30	0.49/0.35

Table 4 Statistics for each ensemble of datasets of the substructure solution, phase refinement and chain tracing for SIRAS. The red characters show the failure in chain tracing with the relevant multiplicity value; the green show the successful values. For the anomalous CC1/2 line the values in parenthesis refer to the highest resolution shell (2.4-2.3 Å)

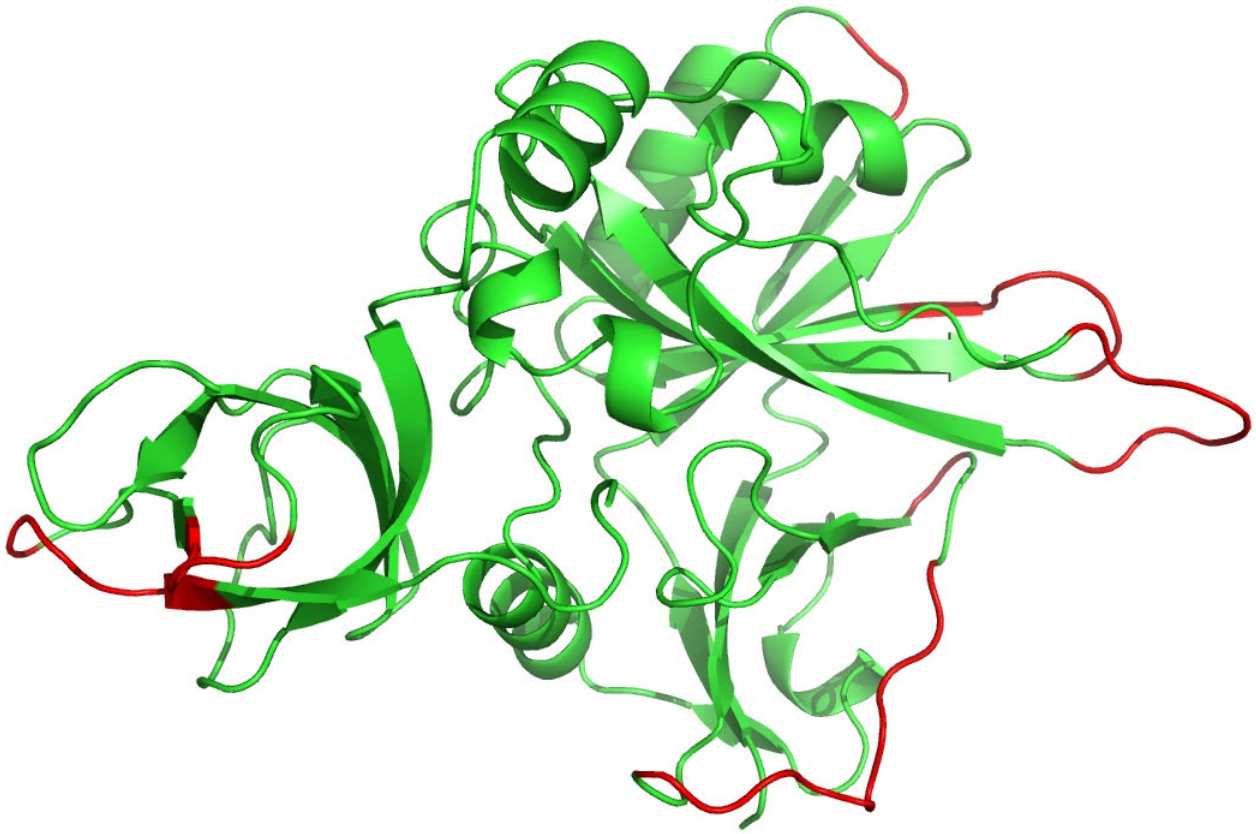

Fig. 2 Secondary structure ribbon representation of the structure of human IBA57. Highlighted in red the residues for which very poor or no electron density at all is observed.

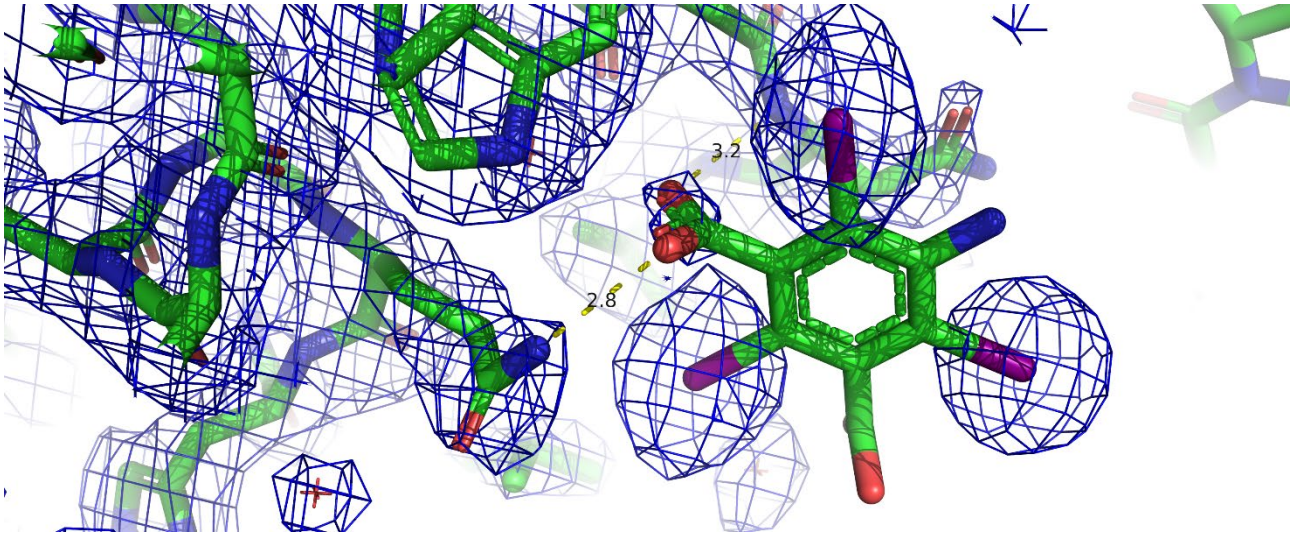

Fig. 3 Detail of the electron density of one of the four molecules of I3C bound through hydrogen bonds to two residues on the surface of IBA57

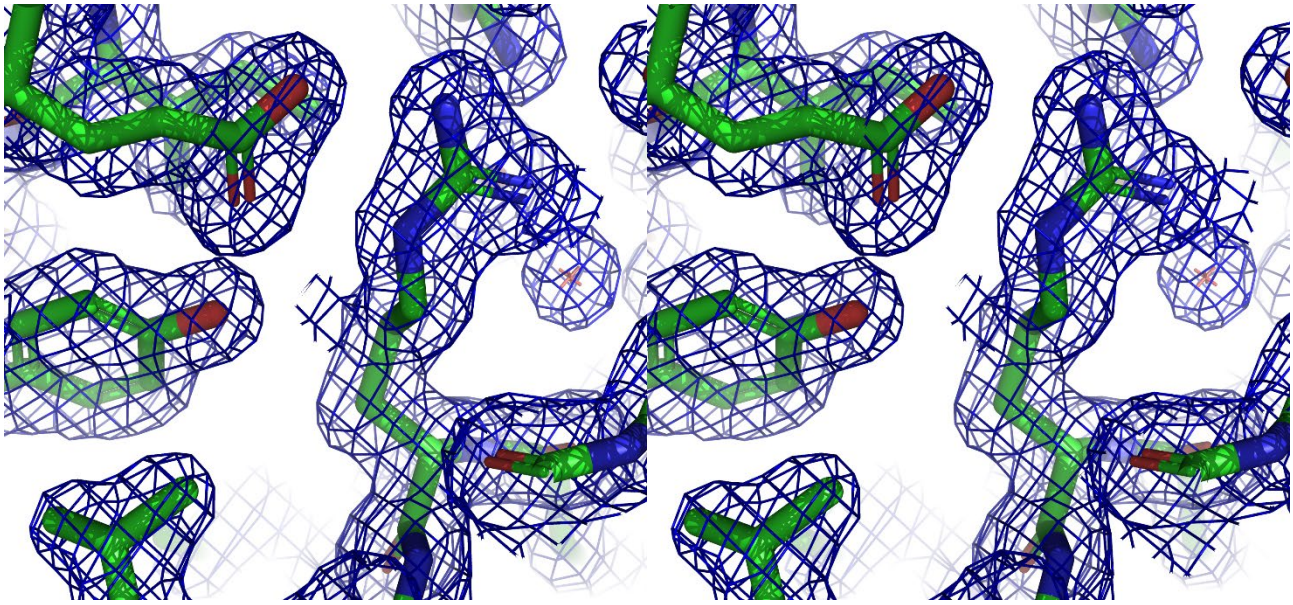

Fig. 4 Stereo representation of the quality of the electron density of 6ESR

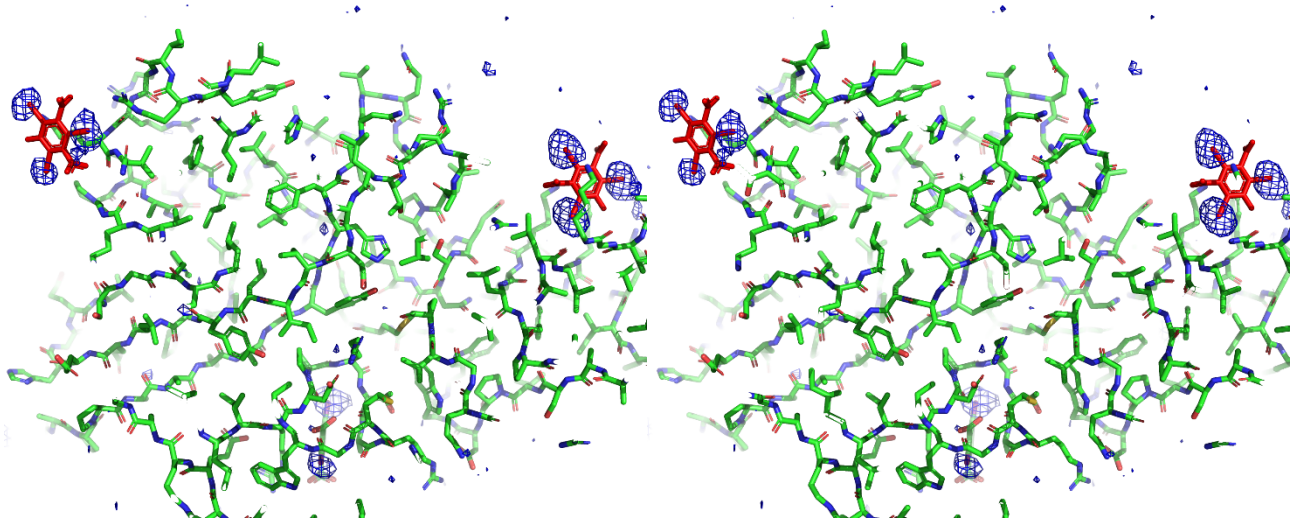

Fig. 5 Stereo view of the anomalous electron density contoured at 3.0 σ level. Two I3C molecules and the respective iodine atoms positions are clearly visible.

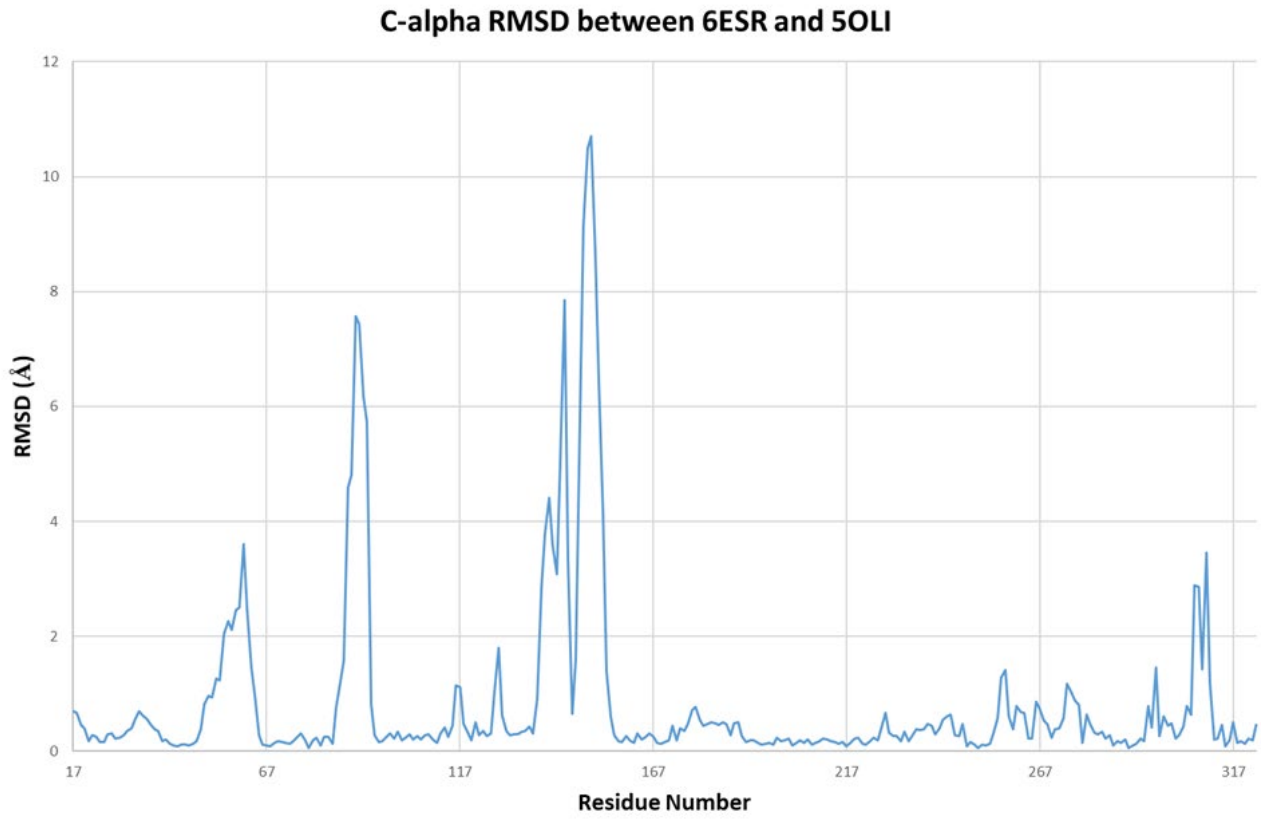

Fig 6. C-alpha RMSD between 6ESR and 5OLI showing clustered regions of large deviation between them.

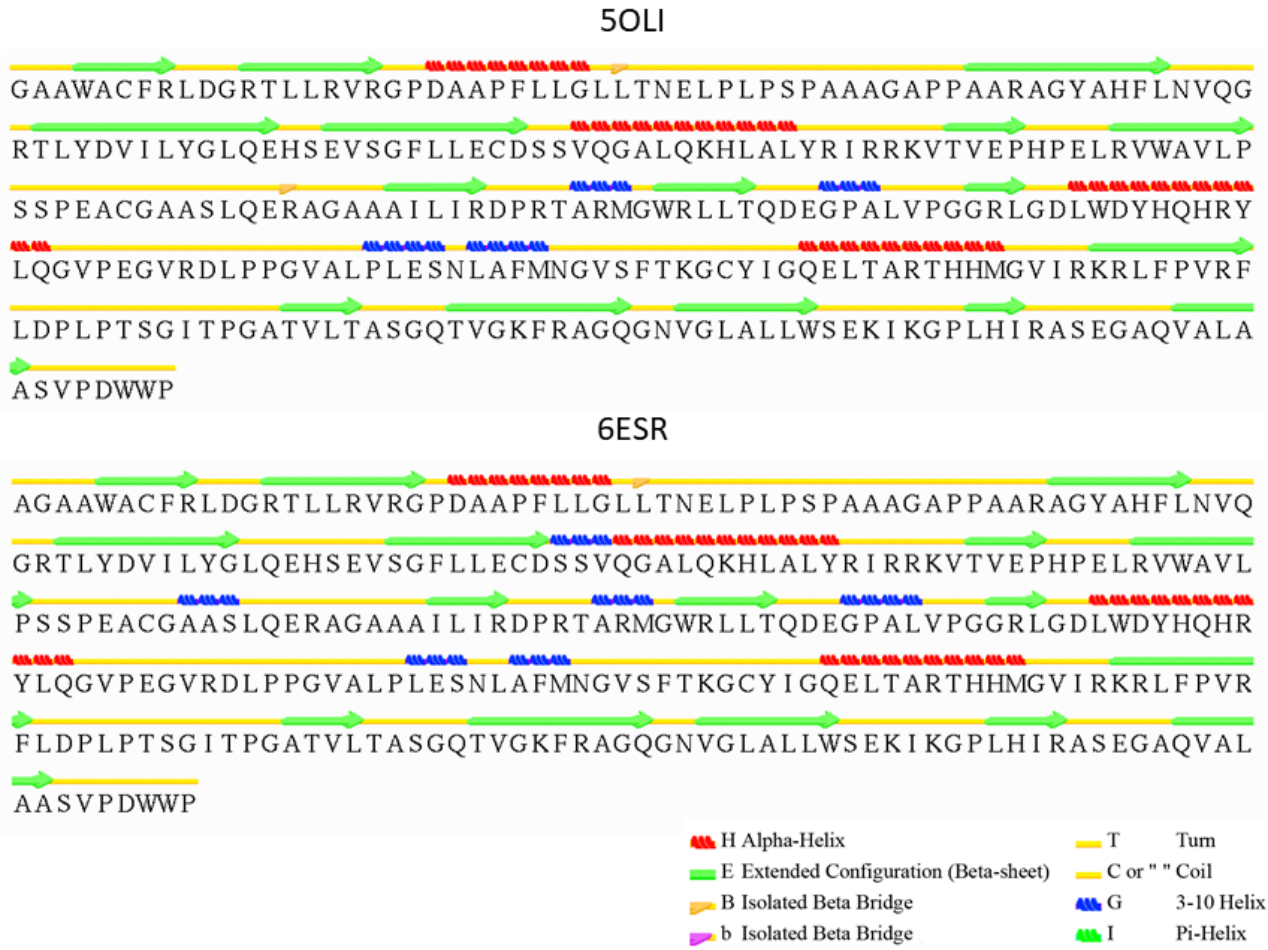

Fig. 7 Secondary structure comparison between 5OLI and 6ESR

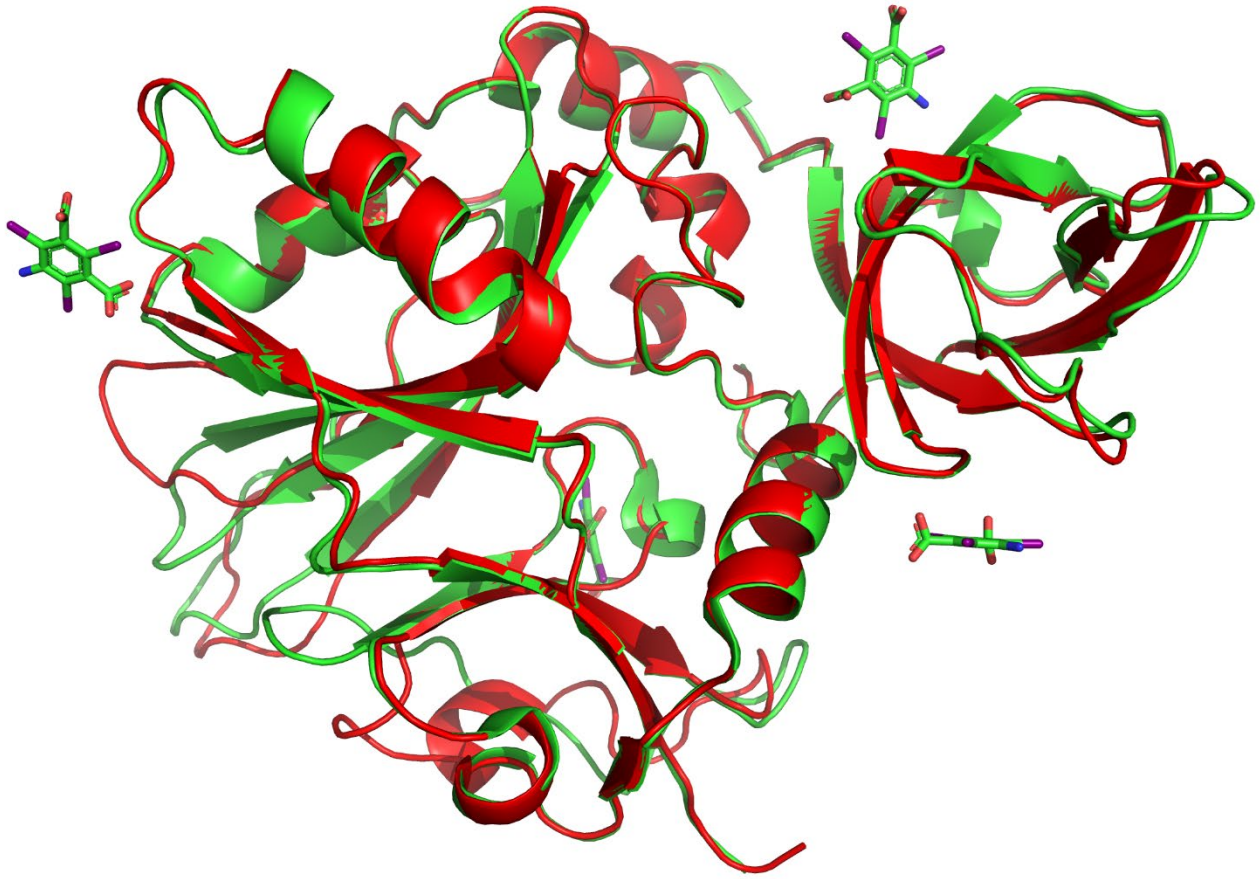

Fig. 8 Superposition between 6ESR (red) and 5OLI (green) secondary structures.

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.

Agniswamy, J., Joyce, M. G., Hammer, C. H. & Sun, P. D. (2008). *Acta Crystallogr D Biol Crystallogr* **64**, 354-367.

Andreini, C., Rosato, A. & Banci, L. (2017). *PLoS One* **12**, e0171279.

Bauman, J. D., Harrison, J. J. & Arnold, E. (2016). *IUCrJ* **3**, 51-60.

Beck, T., Gruene, T. & Sheldrick, G. M. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 374-380.

Beck, T., Krasauskas, A., Gruene, T. & Sheldrick, G. M. (2008). *Acta Crystallogr D Biol Crystallogr* **64**, 1179-1182.

Berman, H. M. (2000). *Nucleic Acids Research* **28**, 235-242.

Boggon, T. J. & Shapiro, L. (2000). *Structure* **8**, R143-R149.

Brancaccio, D., Gallo, A., Mikolajczyk, M., Zovo, K., Palumaa, P., Novellino, E., Piccioli, M., Ciofi-Baffoni, S. & Banci, L. (2014). *J Am Chem Soc* **136**, 16240-16250.

Brancaccio, D., Gallo, A., Piccioli, M., Novellino, E., Ciofi-Baffoni, S. & Banci, L. (2017). *J Am Chem Soc* **139**, 719-730.

Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P., Sharff A., Smart O.S., Vonrhein C. & T.O., W. (2016). *Cambridge, United Kingdom: Global Phasing Ltd.*

Brockhauser, S., Ravelli, R. B. & McCarthy, A. A. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 1241-1251.

Bunkoczi, G., McCoy, A. J., Echols, N., Grosse-Kunstleve, R. W., Adams, P. D., Holton, J. M., Read, R. J. & Terwilliger, T. C. (2015). *Nat Methods* **12**, 127-130.

Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallogr. D Biol. Crystallogr* **66**, 12-21.

Ciofi-Baffoni, S., Nasta, V. & Banci, L. (2018). *Metallomics* **10**, 49-72.

Cowtan, K. (2006). *Acta Cryst. D* **62**, 1002-1011.

Dauter, M. & Dauter, Z. (2007). *Methods Mol Biol* **364**, 149-158.

Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Crystallographica Section D Biological Crystallography* **59**, 688-696.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr. D. Biol. Crystallogr* **60**, 2126-2132.

Finke, A. D., Panepucci, E., Vonrhein, C., Wang, M., Bricogne, G. & Olieric, V. (2016). *Methods Mol Biol* **1320**, 175-191.

Frishman, D. & Argos, P. (1995). *Proteins* **23**, 566-579.

Gourdoupis, S., Nasta, V., Calderone, V., Ciofi-Baffoni, S. & Banci, L. (2018). *J Am Chem Soc* **140**, 14401-14412.

Kabsch, W. (2010). *Acta Crystallogr. D. Biol. Crystallogr* **66**, 125-132.

Kim, M. K., Lee, S., An, Y. J., Jeong, C. S., Ji, C. J., Lee, J. W. & Cha, S. S. (2013). *Mol Cells* **36**, 74-81.

Klinke, S., Foos, N., Rinaldi, J. J., Paris, G., Goldbaum, F. A., Legrand, P., Guimaraes, B. G. & Thompson, A. (2015). *Acta Crystallogr D Biol Crystallogr* **71**, 1433-1443.

Lakomek, K., Dickmanns, A., Mueller, U., Kollmann, K., Deuschl, F., Berndt, A., Lubke, T. & Ficner, R. (2009). *Acta Crystallogr D Biol Crystallogr* **65**, 220-228.

Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protocols* **3**, 1171-1179.

Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancia, F. & Hendrickson, W. A. (2012). *Science* **336**, 1033-1037.

Liu, Q., Liu, Q. & Hendrickson, W. A. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 1314-1332.

Liu, X., Zhang, H., Wang, X. J., Li, L. F. & Su, X. D. (2011). *PLoS One* **6**, e24227.

Pape, T. & Schneider, T. R. (2004). *Journal of Applied Crystallography* **37**, 843-844.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Crystallographica Section D Biological Crystallography* **59**, 868-875.

Rouault, T. A. (2015). *Nat Rev Mol Cell Biol* **16**, 45-55.

Roversi, P., Johnson, S. & Lea, S. M. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 420-425.

Salgado, P. S., Walsh, M. A., Laurila, M. R., Stuart, D. I. & Grimes, J. M. (2005). *Acta Crystallogr D Biol Crystallogr* **61**, 108-111.

Schuermann, J. P. & Tanner, J. J. (2003). *Acta Crystallographica Section D Biological Crystallography* **59**, 1731-1736.

Sheldrick, G. M. (2010). *Acta Crystallogr. D. Biol. Crystallogr* **66**, 479-485.

Sugahara, M., Asada, Y., Ayama, H., Ukawa, H., Taka, H. & Kunishima, N. (2005). *Acta Crystallogr D Biol Crystallogr* **61**, 1302-1305.

Tereshko, V., Uysal, S., Koide, A., Margalef, K., Koide, S. & Kossiakoff, A. A. (2008). *Protein Sci* **17**, 1175-1187.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol* **364**, 215-230.

Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Rose, J., Ebihara, A., Kuramitsu, S., Li, D., Howe, N., Schnapp, G., Pautsch, A., Bargsten, K., Prota, A. E., Surana, P., Kottur, J., Nair, D. T., Basilico, F., Cecatiello, V., Pasqualato, S., Boland, A., Weichenrieder, O., Wang, B. C., Steinmetz, M. O., Caffrey, M. & Wang, M. (2015). *Nat Methods* **12**, 131-133.