**Supporting Information: SASDDQ4 and SASDDR4**

SAXS samples and matching buffers (**Table S1a**) were loaded into a 96 well plate and drawn into a 1.5 mm capillary for measurement at the SAXS-WAXS beamline of the Australian synchrotron (acquisition parameters in **Table S1b**.) Data reduction for data scaling, normalisation and solvent subtraction to obtain the protein scattering profile followed by analysis and interpretation used the software listed in **Table S1c**. The wild-type and mutant forms of the protein yield similar structural parameters and properties (**Tables S1d, S2** and **S3**). A small concentration dependence to the structural parameters was observed in particular for the mutant form, indicative of weak attractive forces that resulted in an increase in *Rg* of ~1-2 Å over the measured concentration range (**Tables S2** and **S3**) for the mutant form. Plots of *I*(0) versus protein concentration showed a linear dependence of the scattering intensity and the data were therefore subjected to extrapolation to infinite dilution to minimise the concentration-dependent effects (**Figure S1**). This step resulted in some amplification of errors in the lowest *q*-data.

**Figures S2** and **S3** show the standard plots of the SAXS data and for all concentrations measured and the extrapolation to infinite dilution. The plots are consistent with samples being essentially monodisperse, folded particles: the **A** panels show log-log plots that show the expected near zero slope at low *q*; **B** panels show the required linear Guinier plots; **C** panels show well-behaved *P*(*r*) profiles that approach zero around ~80 Å; and **D** panels show dimensionless Kratky plots with the characteristic bell-shaped curve, with a maximum ~1.4 and shifted to the right of *qRg* = 1.75 indicative of a mostly folded, elongated particle with a small upward trajectory for *qRg* values > 7 indicative of some flexible region(s). Molecular weight determinations from *I*(0) over the entire concertation range are within 20% of the values expected from chemical composition (**Table S2** and **S3**). The Porod volumes, *Vp*, for both forms are within 10% of the values expected from chemical composition, which for a well-folded protein is expected to be ~50% larger and is thus consistent with the presence of some flexibility as indicated in the dimensionless Kratky plots that impacts the accuracy *VP*.

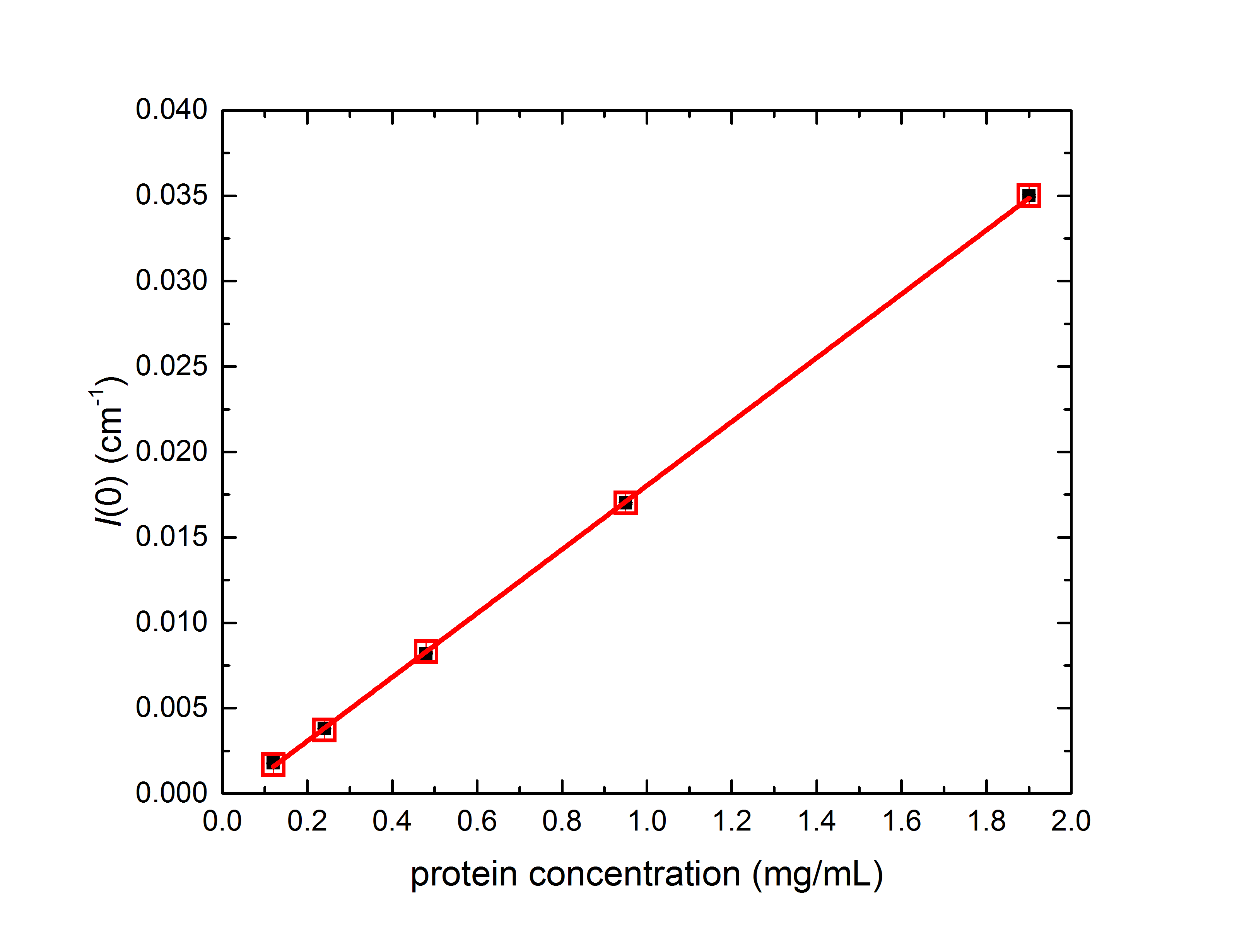
**Table S2d** gives the structural parameters derived from Guinier and *P*(*r*) analysis of the extrapolated data sets, which are the same for the two forms within experimental error and the overlaid *P*(*r*) plots (**Figure S4A**) likewise show no significant differences consistent with the mutation having no significant impact on the overall conformation of the protein in solution. *Ab initio* shape modelling using DAMMIN (**Table S2e**) yielded envelops that show very similar extended structures (**Figure S4B**). Initially 20 independent DAMMIN calculations were done using an initial spherical search space that yielded similar shapes as indicated by normalized spatial discrepancy (NSD) values. DAMSTART was used to generate a starting envelop consistent with the average of the 20 models which was used as input to additional DAMMIN refinement to obtain a final optimized envelop. Minimal *χ*2 values for the final optimized model are larger than the expected ideal value of ~1 where errors are accurately propagated Poisson counting statistics. To evaluate whether this indicated the models could be improved with respect to the data, the model fits were evaluated using CorMap. The *P*-values of 0.54 obtained using CorMap indicate excellent model fits to the data such that data are randomly distributed about the model profile (**Figure S4C** through **F**).

**Table S1**: SAS data acquisition, sample details, data analysis, modelling fitting and software used.

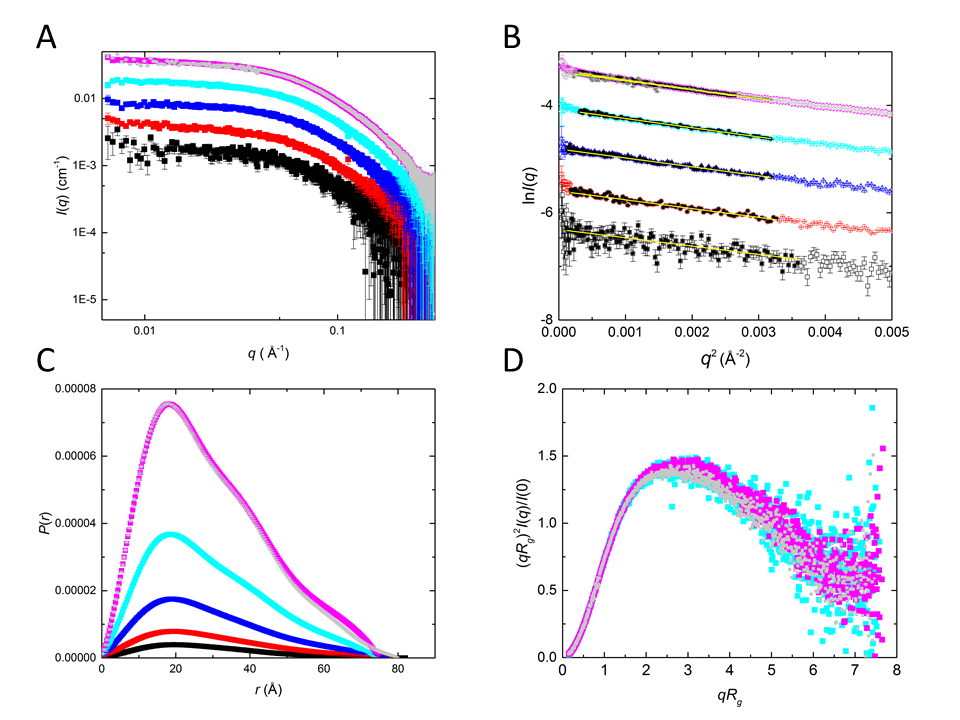
|  |  |  |  |
| --- | --- | --- | --- |
| (*a*) Sample details | | | |
|  | | LHX4-ISL1(wt) | LX4-ISL1(R282G) |
| Organism | |  |  |
| Source (Catalogue No. or reference) | |  |  |
| Sequence (Uniprot ID + uncleaved tags, bound ligands/modifications, *etc.*) | |  |  |
| Extinction coefficient ε (A280, 0.1% cm-1) | | 1.025 | 1.031 |
| Partial specific volume  (cm3 g-1) | | 0.719 | 0.718 |
| Mean solute and solvent scattering length (1010 cm-2) | | 12.561, 9.465 | 12.563, 9.465 |
| mean scattering contrast  (1010 cm-2) | | 3.096 | 3.098 |
| Molecular mass *M* (chemical composition (Da)) | | 18988 | 18627 |
| Protein concentration range (mg/mL) from A280 | | 0.12-1.9 | 0.12-1.9 |
| Solvent composition – all samples | | 20mM Tris HCl, 150 mM NaCl, 0.5 mM TCEP, 15 mM NaN3 | |
| (*b*) SAS data collection parameters | | | |
| Source, instrument and description or reference | Australian synchrotron SAXS-WAXS Beamline with Dectris Pilatus 1m detector (Kirby et al., 2013) | | |
| Wavelength (Å) | 1.0332 | | |
| Beam geometry (size, sample-to-detector distance) | 544 µm x 274 µm, 2.68 m | | |
| *q*-measurement range (Å-1) | 0.006-0.328 | | |
| Absolute scaling method | Comparison with scattering from pure H2O | | |
| Normalization | To transmitted intensity by beam stop counter | | |
| Radiation damage monitoring | Data frame by frame comparison | | |
| Exposure time, number | 1s integrations, 24-32 frames per sample | | |
| Sample configuration | Samples were drawn from 100 µL aliquots in 96 well plate into a 1.5 mm quartz capillary at a flow-rate of 7 µL/s. | | |
| Sample temperature (ºC) | 13.5 | | |
| (*c*) Software employed for SAS data reduction, analysis and interpretation | | | |
| SAS data reduction | I(q) vs q using ScatterBrain 2.82 (<http://www.synchrotron.org.au/aussyncbeamlines/saxswaxs/software-saxswaxs>), solvent subtraction and extrapolation PRIMUSqt (ATSAS 2.8.0) (Petoukhov et al., 2012) | | |
| Calculation of ε from sequence | ProtParam (Gasteiger *et al.*, 2005) | | |
| Calculation of and  values | MULCh 1.1 (06/10/16) (Whitten *et al.*, 2008) | | |
| Guinier, *P*(*r*), Porod volume (*V*P) | PRIMUSqt from ATSAS 2.8.3 (Franke et al., 2017) | | |
| Shape/bead modelling | DAMMIF (Franke & Svergun, 2009) and DAMMIN (Svergun, 1999) via ATSAS on-line <https://www.embl-hamburg.de/biosaxs/atsas-online/>. Averaging/filtering/optimisation of average of DAMMIN models used local versions of DAMAVER/DAMFILT/DAMSTART from ATSAS 2.8.3 | | |
| Molecular graphics | PyMOL v1.70.0.5 Win64 | | |
| (*d*) Structural parameters – infinite dilution, extrapolated data set from 5 concentrations (Tables S1,S2) | | | |
| Guinier Analysis | | LHX4-ISL1(wt) | LX4-ISL1(R282G) |  |
| *R*g(Å) | | 22.2 ± 2.7 | 22.7 ± 0.9 |  |
| *q-*range (Å-1) (*qRg* maximum) | | 0.015-0.055 (1.25) | 0.01-0.056 (1.3) |  |
| Quality-of-fit parameter | | 0.95 | 0.98 |  |
| *P*(*r*) analysis | | LHX4-ISL1(wt) | LX4-ISL1(R282G) |  |
| *R*g (Å) | | 23.03 ± 0.07 | 23.63 ± 0.08 |  |
| *d*max (Å) | | 80 | 84 |  |
| *q-*range (Å-1) | | 0.0150-0.3277 | 0.0096-0.3277 |  |
| Total quality estimate (GNOM) (>0.8 indicates a “good” solution) | | 0.82 | 0.81 |  |
| *V*P (ratio to expected vol.) | | 20300 (0.90) | 20700 (0.92) |  |
| (*e*) DAMMIN results – extrapolated data set | | | |
|  | | LHX4-ISL1(wt) | LX4-ISL1(R282G) |  |
| Multiple fits, spherical search space, radius = *dmax* | |  |  |  |
| Number of calculations (spherical search space, radius = *dmax*) | | 20 | 20 |  |
| *q-*range for fitting | | 0.0150-0.3277 | 0.0096-0.3277 |  |
| Symmetry/anisotropy assumptions | | P1 | P1 |  |
| Normalised spatial discrepancy (NSD) for 20 models (standard deviation) | | 0.68 (0.015) | 0.67 (0.16) |  |
| 2 value for fits | | 2.7 | 4.4 |  |
| Constant subtraction in optimisation | | 0.000136 | 0.000176 |  |
| Total excluded DAM volume (Å3) | | 26582 | 27060 |  |
| Model resolution (from DAMRES) (Å) | | 28 ± 2 | 28 ± 2 |  |
| Final fit, DAMSTART average as search volume | |  |  |  |
| 2, *P* value (from CorMap) | | 2.7, 0.54 | 4.4, 0.54 |  |
| Total excluded DAM volume (Å3) | | 28303 | 29072 |  |
| (*g*) Data/model deposition IDs | |  |  |  |
|  | | LHX4-ISL1(wt) | LX4-ISL1(R282G) |  |
|  | |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S2: Concentration dependence, Guinier Analysis** | | | | | | | |
| Protein conc. (mg/mL) | *Rg* (Å) | *I*(0) (cm-1) | Fidelity\* | *q*-range  (Å-1) | *qRg* max | *MW*Guin*I*(0) (Da) | ratio MW to expected |
| LHX4-ISL1(wt) |  |  |  |  |  |  |  |
| 0.12 | 21.51±1.91 | 0.0018±0.00003 | 0.64 | 0.011±0.060 | 1.29 | 1.82E+04 | 0.96 |
| 0.24 | 22.69±0.94 | 0.0038±0.00002 | 0.86 | 0.013±0.057 | 1.30 | 1.92E+04 | 1.01 |
| 0.48 | 22.77±1.16 | 0.0082±0.00003 | 0.89 | 0.008±0.064 | 1.28 | 2.07E+04 | 1.11 |
| 0.95 | 23.01±1.44 | 0.017±0.000006 | 0.93 | 0.018±0.064 | 1.30 | 2.17E+04 | 1.15 |
| 1.9 | 23.26±0.37 | 0.035±0.00007 | 0.93 | 0.018±0.051 | 1.20 | 2.23E+04 | 1.18 |
| LX4-ISL1(R282G) |  |  |  |  |  |  |  |
| 0.12 | 21.24±1.24 | 0.0017±0.00003 | 0.7 | 0.011±0.061 | 1.30 | 1.72E+04 | 0.91 |
| 0.24 | 22.4±0.82 | 0.0037±0.00003 | 0.81 | 0.009±0.058 | 1.29 | 1.87E+04 | 1.02 |
| 0.48 | 23.04±0.68 | 0.0083±0.00004 | 0.92 | 0.008±0.056 | 1.29 | 2.10E+04 | 1.14 |
| 0.95 | 23.14±0.75 | 0.017±0.00004 | 0.95 | 0.012±0.054 | 1.25 | 2.17E+04 | 1.18 |
| 1.9 | 23.43±1.4 | 0.035±0.00006 | 0.97 | 0.014±0.054 | 1.26 | 2.23E+04 | 1.20 |
| \* Fidelity values as given in primusqt | | | | | | | |

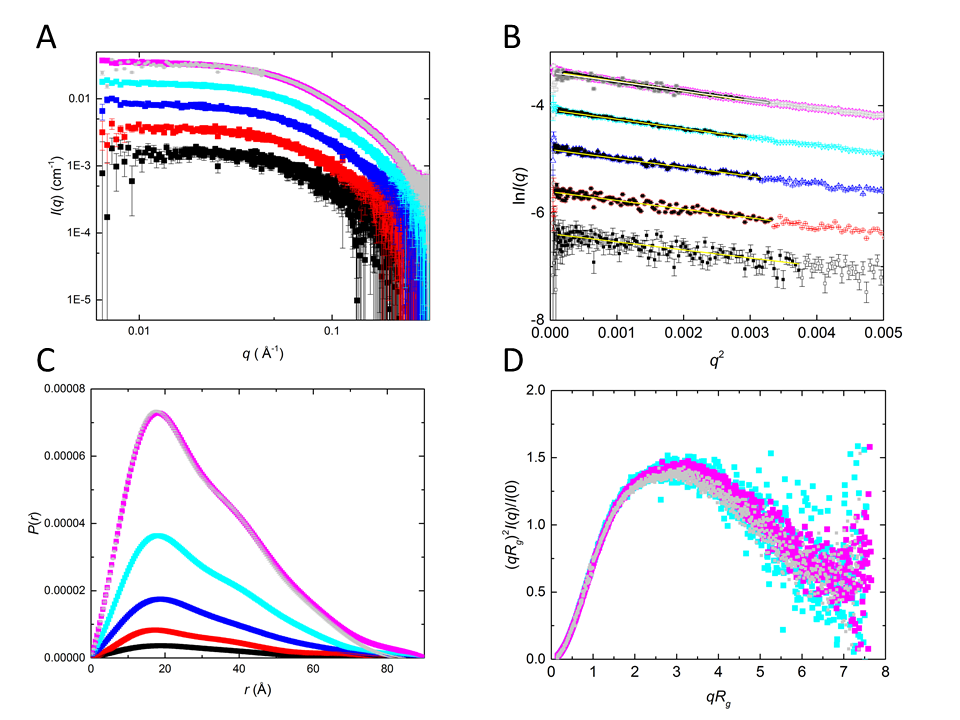
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S3: Concentration dependence, *P*(*r*) Analysis** | | | | | | | | |
| Protein conc. (mg/mL) | *Rg* (Å) | *I*(0) (cm-1) | *dmax* (Å) | Total quality estimate\* | *q*-range (Å-1) | Porod volume (Å3) | MW*P*(*r*)*I*(0) (Da) | ratio *MW* to expected |
| LHX4-ISL1(wt) |  |  |  |  |  |  |  |  |
| 0.12 | 23.09±0.68 | 0.0018±0.00004 | 82 | 0.82 | 0.011-0.328 | 22300 | 18229.4 | 0.96 |
| 0.24 | 23.35±0.15 | 0.00378±0.00002 | 78 | 0.81 | 0.014-0.328 | 21800 | 19140.87 | 1.01 |
| 0.48 | 23.37±0.11 | 0.00825±0.00003 | 77 | 0.81 | 0.012-0.328 | 21600 | 20887.85 | 1.11 |
| 0.95 | 23.18±0.07 | 0.0171±0.00004 | 74 | 0.83 | 0.018-0.328 | 21700 | 21875.28 | 1.15 |
| 1.9 | 23.44±0.04 | 0.0351±0.00005 | 75 | 0.82 | 0.018-0.328 | 21200 | 22450.94 | 1.18 |
| LX4-ISL1(R282G) |  |  |  |  |  |  |  |  |
| 0.12 | 22.86±0.92 | 0.00168±0.00004 | 88 | 0.79 | 0.0111- 0.3277 | 20200 | 17039.51 | 0.91 |
| 0.24 | 23.53±0.48 | 0.00375±0.00004 | 82 | 0.81 | 0.0135- 0.3277 | 19500 | 19017.31 | 1.02 |
| 0.48 | 24.46±0.20 | 0.00837±0.00004 | 90 | 0.73 | 0.0115- 0.3277 | 21800 | 21223.32 | 1.14 |
| 0.95 | 24.16±0.13 | 0.0172±0.00005 | 90 | 0.73 | 0.0178- 0.3277 | 20900 | 22036.06 | 1.18 |
| 1.9 | 24.53±0.07 | 0.0349±0.00006 | 90 | 0.74 | 0.0178- 0.3277 | 21200 | 22356.35 | 1.20 |
| \* Total quality estimate as provided by GNOM, values > 0.8 indicate good solutions, > 0.7 are reasonable solutions. | | | | | | | | |

****

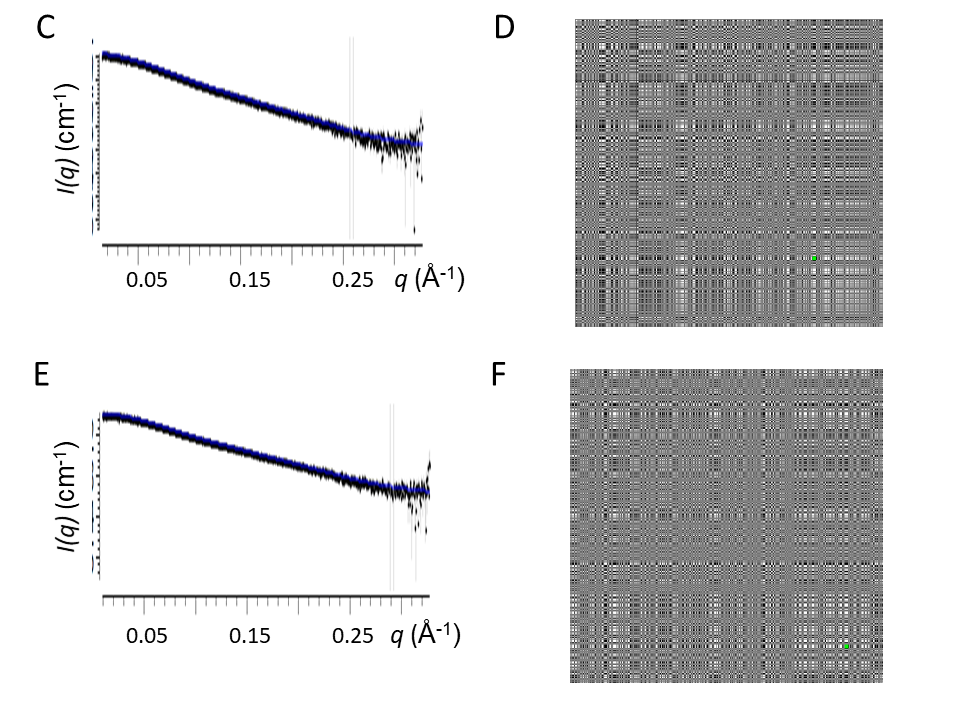
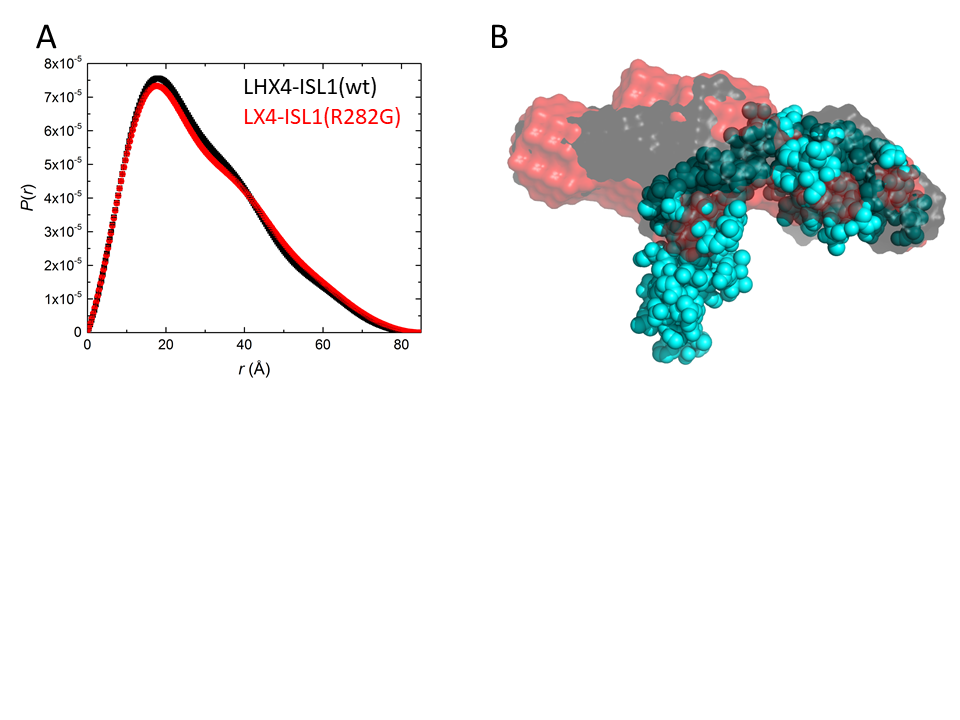
**Figure S1:** Concentration dependence of scattering intensity for LHX4-ISL1(wt) (black filled symbols) and LHX4-ISL1(R282G) (red open symbols) showing linear fit (red line). The Pearson coefficient for the two linear fit to both data sets are 0.999.

****

**Figure S2.** **Scattering data for LX4-ISL1(wt) A** log*I*(*q*) v log*q* plots for the concentrations series showing the expected near zero slope at low-*q* expected for monodisperse scattering particles of similar size (colour code is black, red, blue, cyan, magenta for 012, 0.24, 0.48, 0.95, 1,2 mg/mL data respectively). The extrapolated data set is superimposed on the highest concentration profile for comparison (grey) **B.** Guinier plots for the data in panel A, same colour code but using open symbols and data included in the Guinier fit indicated by black filled symbols. Extrapolated data set is again superimposed on the most concentrated data set, shown in grey with filled grey symbols indicating the Guinier fit region. Yellow solid lines show the Gunier fits. **C.** P(r) versus r for the data shown in panel A. **D.** Dimensionless Kratky plots for the data shown in panel A. Panels C and D use the same symbols and colour code as in panel A.

****

**Figure S3 Scattering data for LX4-ISL1(R282G) A.** log*I*(*q*) v log*q* plots for the concentrations series showing the expected near zero slope at low-*q* expected for monodisperse scattering particles of similar size (colour code is black, red, blue, cyan, magenta for 012, 0.24, 0.48, 0.95, 1,2 mg/mL data respectively). The extrapolated data set is superimposed on the highest concentration profile for comparison (gray) **B.** Guinier plots for the data in panel A, same colour code but using open symbols and data included in the Guinier fit indicated by black filled symbols. Extrapolated data set is again superimposed on the most concentrated data set, shown in gray with filled grey symbols indicating the Guinier fit region. Yellow solid lines show the Gunier fits. **C.** *P*(*r*) versus *r* for the data shown in panel A. **D.** Dimensionless Kratky plots for the data shown in panel A. Panels C and D use the same symbols and color code as in panel A.

****

**Figure S4. Model comparisons for LX4-ISL1(wt) and LX4-ISL1(R282G) A.** *P*(*r*) versus r for the wild type (black) and mutant (red) LHX4-ISL1. **B.** Surface representations of DAMMIN envelops calculated using the *P*(*r*) solutions shown in panel A (wild type, black, and mutant red) with a space filling model for the B chain of **pdb:3MMK**. **C./D.** and **E./F.** are the plots from CorMap showing the fit of the data to the models highlighting the longest. C/D are for the wild type data and model, and E/F the mutant data and model. The vertical lines in panels C and E show the longest stretch of data points that lie on one side of the model profile (10 points out of 825 in each case) while the 2D correlation maps in D and E are generated by plotting data points as black or white depending on which side of the model profile they lie. The relatively uniform overall grey results from points being randomly distributed about the model profile.