Supporting information for: Combined small-angle X-ray and neutron scattering restraints in molecular dynamics simulations

Po-chia Chen,[*,†,⊥] Roman Shevchuk,[‡,⊥] Felix M. Strnad,[‡] Charlotte Lorenz,[¶,#] Lukas Karge,[§] Ralph Gilles,[§] Andreas M. Stadler,[¶,#] Janosch Hennig,[†] and Jochen S. Hub[*,‖]

† *Structural and Computational Biology Unit, EMBL Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany*

‡ *Institute for Microbiology and Genetics, Georg-August-Universität Göttingen, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany*

¶ *Jülich Centre for Neutron Science (JCNS-1) and Institute for Complex Systems ICS (ICS-1) Forschungszentrum Jülich GmbH, 52425 Jülich , Germany*

§ *Heinz Maier-Leibnitz Zentrum, Technische Universität München, Lichtenbergstr. 1, 85748 Garching, Germany*

‖ *Saarland University, Theoretical Physics and Center for Biophysics, Campus E2 6, 66123 Saarbrücken, Germany*

⊥ *These authors contributed equally to this work.*

\# *Institute of Physical Chemistry, RWTH Aachen University, Landoltweg 2, 52056 Aachen, Germany*

E-mail: pchen@embl.de; jochen.hub@uni-saarland.de

Supporting Figures

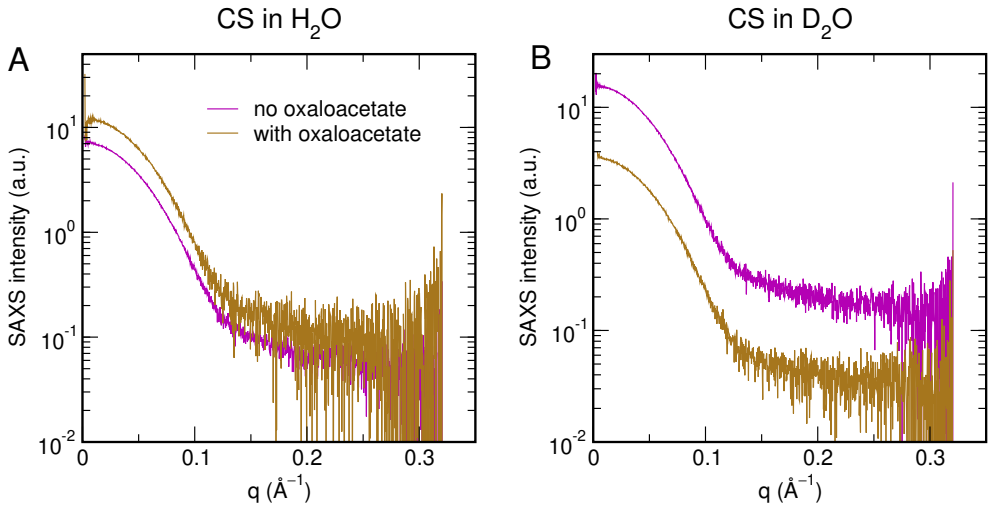

Figure S1: SAXS data of citrate synthase (CS) collected at SWING beamline at SOLEIL (Paris). (A) CS in H_2O buffer, and (B) in D_2O buffer, in the apo form (magenta) and in the oxaloacetate-bound state (tan). Differences in signal-to-noise ratios arise from the use of 20, 40, and 60 frames in the averaging and merging procedure. See Table S1 for details.

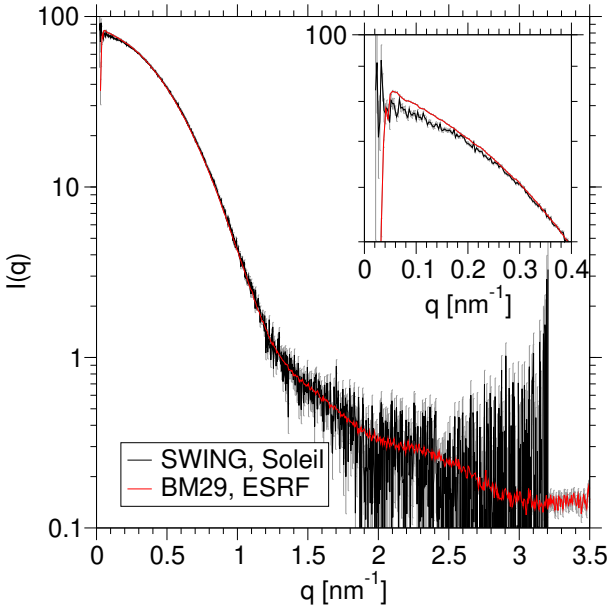

Figure S2: Comparison of apo CS measured at ESRF (red with grey errorbars) versus Soleil (black with grey errorbars), fitted to account for differences in concentration and background scattering between the two detector setups. Inset depicts a zoom in at low angles, exhibiting a small departure from quadratic decay at ESRF measurements.

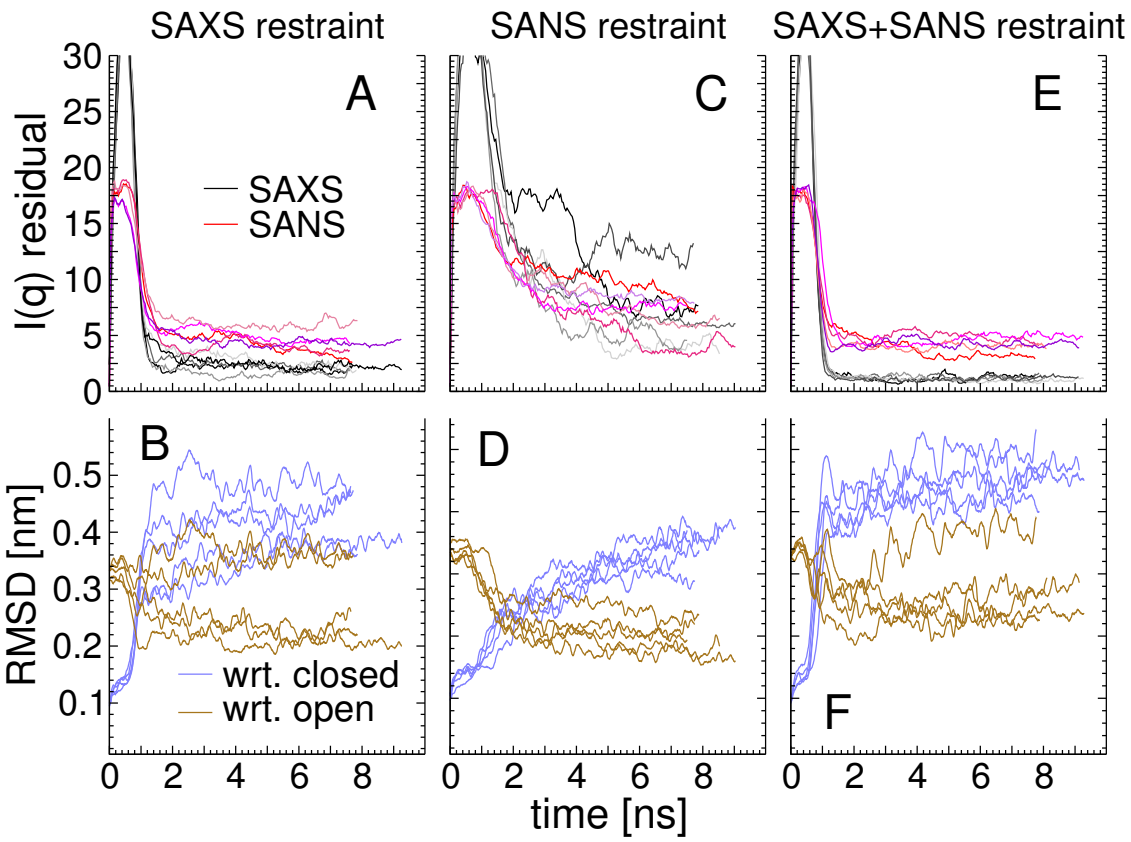

Figure S3: SAXS- and SANS-restraint simulations of citrate synthase (CS) without symmetry restraint on the protein. (A/B) Refinement against SAXS data and cross-validation against SANS, (C/D) refinement against SANS data and cross-validation against SAXS, and (E/F) simultaneous refinement against SAXS and SANS data. (A/C/E) Residuals with respect to the experimental data. (B/D/F) Backbone RMSD during refinement simulation, shown with respect to the open crystal structure (brown, PDB code 1CTS) and closed crystal structure (blue, PDB code 2CTS). Occasionally, only one of the two monomers opened up and 'overshooted' to match the experiment R_g, as evident from an increased RMSD with respect to the open crystal structure.

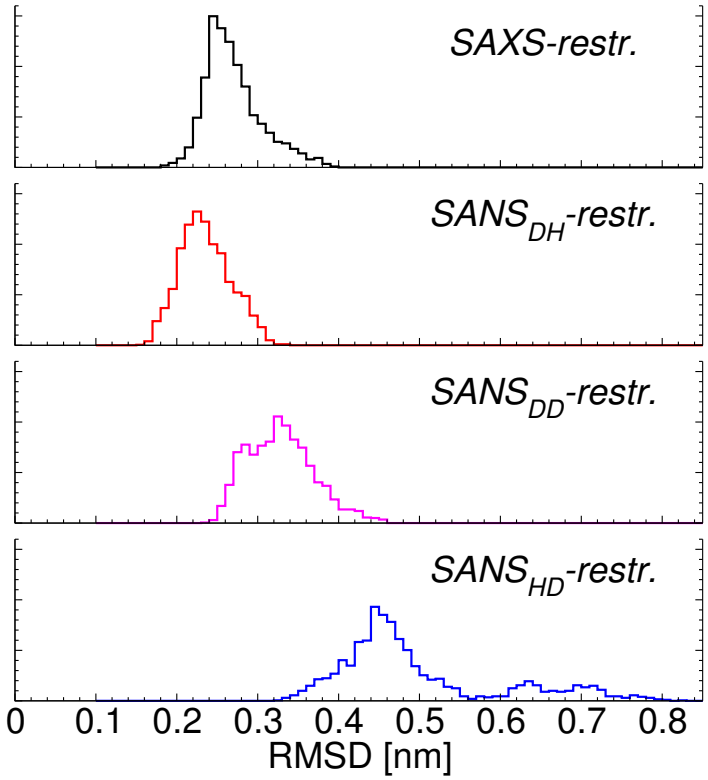

Figure S4: C_α-RMSD to PDB structure 4QQB computed from the last 2ns of SAS-restrained ensembles during the 4-way cross-validation study.

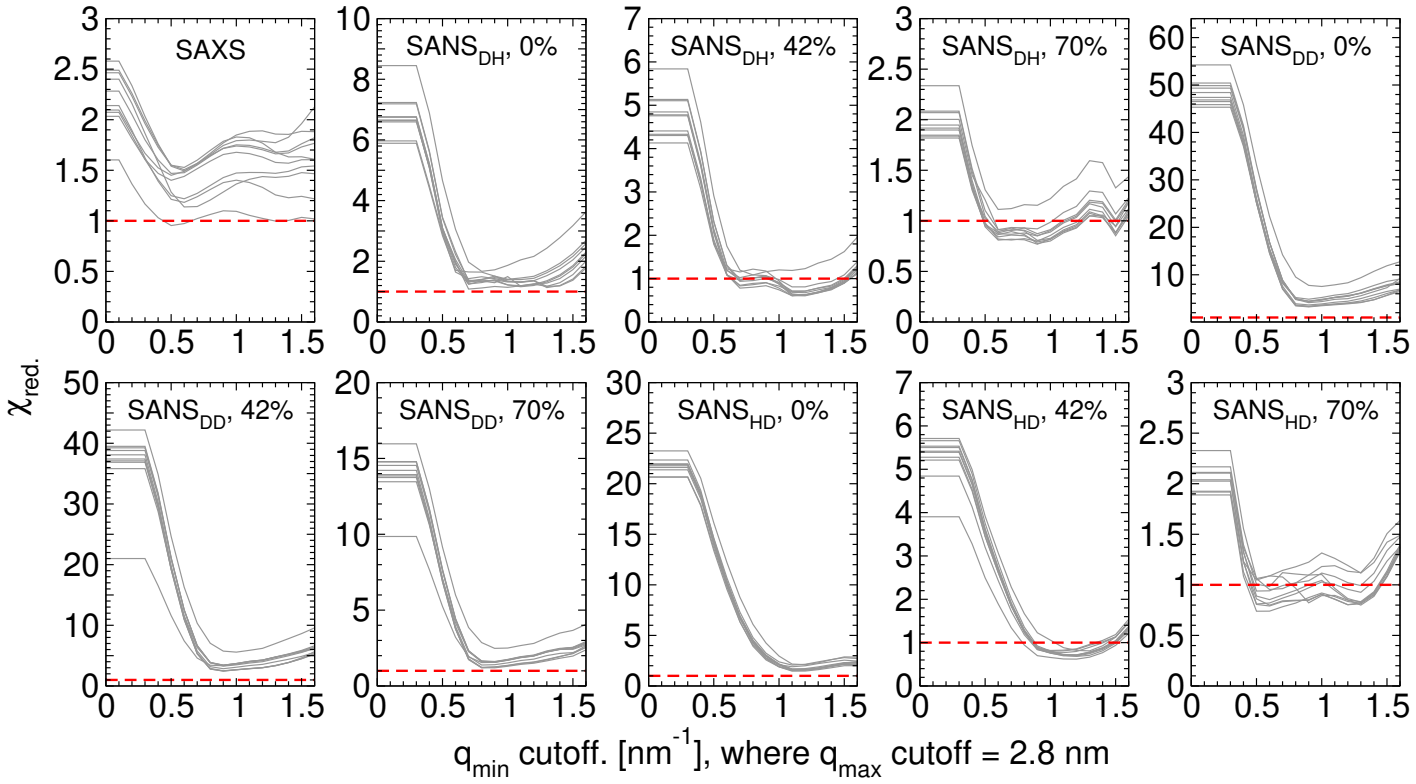

Figure S5: $\chi_{reduced}$ agreement between unrestrained MD simulations of SUM versus experiment, plotted as a function of the q-range used to fit computed curves against the experimental SAXS and SANS curves. Values of the 10 replicated are shown in solid grey lines. Ideal agreement $\chi_{red.} = 1$ plotted as dashed red. q_{max} is fixed at 2.8 nm^{-1}, while q_{min} is varied between 0 and 1.6 nm^{-1}. $\chi_{reduced}$ is defined here as $\chi * N_{Shannon}/(N_{Shannon} - 2)$, where $N_{Shannon} = (q_{max} - q_{min}) * D_{max}/\pi$ is the number of Shannon channels within the q-range being constrained, reduced by two degrees of freedom due to the fitting of scaling and constant buffer subtraction.

Supporting Tables

Table S1: Detailed reporting of citrate synthase (CS) measurements, covering free and oxaloacetate-bound data from SAXS at BM29 ESRF, SANS at SANS-1 MLZ, and SAXS at SWING SOLEIL. Bound CS was generated by adding 40 mM oxaloacetate, and 99% D_2O was achieved via dialysis in 100-fold excess volume. Scattering curves produced from multiple measurements have their individual source values reported, and separated by slashes.

	ESRF SAXS CS_{free}	ESRF SAXS CS_{bound}	MLZ SANS CS_{free}	MLZ SANS CS_{bound}	SOLEIL CS_{free} (H_2O)	SOLEIL CS_{bound} (H_2O)	SOLEIL CS_{free} (D_2O)	SOLEIL CS_{bound} (D_2O)
(a) Sample Details								
Source organism	*Sus scrofa* (Pig)							
Expression organism	*Sus scrofa* (Pig)							
Source	Sigma-Aldrich							
Description	Obtained from commercial supplier, *c.f.* Uniprot ID: P00889							
Extinction coefficient ϵ_{280nm} ($M^{-1}\,cm^{-1}$)	156 120							
Molecular mass M from chemical composition (kDa)	98							
loading concentrations (mg ml^{-1})	2.5	2.6 / 5.3 / 10.4 / 51.2	2.5 / 4.9 / 9.7 / 47.4	2.5 / 5.0	1.25 / 2.5 / 5.0	1.25 / 2.5 / 5.0	1.25 / 2.5 / 5.0	
injection volume (μl)	75	–	–	75				
Concentration (μM)	25.5	26 / 54 / 106 / 522	26 / 50 / 99 / 483	25.5 / 51	12.8 / 25.5 / 51	12.8 / 25.5 / 51	12.8 / 25.5 / 51	
Solvent composition and source	50 mM TRIS, 50 mM NaCl, pH 7.6	99% D_2O		50 mM TRIS, 50 mM NaCl, pH 7.6		99% D_2O		
(b) SAS data collection parameters								
Source and instrument	Grenoble ESRF BM29 with Dectris Pilatus 1M	SANS-1 MLZ. See Mühlbauer *et al.* (2016).	SOLEIL SWING with Eiger 4M					
Wavelength (Å)	0.992	4.5 / 6	1.03					
Sample-detector distance (m)	2.867	2 / 8 / 20	3.37					
q-measurement range (nm^{-1})	0.0282–4.5250	0.1092–4.209	0.0217–3.2041					0.0325–3.2041
Radiation damage monitoring	frame-by-frame comparison							
Exposure time (s)	2.0	–	0.1					
Frames	10	–	20 / 20	20 / 20 / 60	20 / 20 / 20	40 / 40 / 20		

	ESRF SAXS CS_{free}	ESRF SAXS CS_{bound}	MLZ SANS CS_{free}	MLZ SANS CS_{bound}	SOLEIL CS_{free} (H_2O)	SOLEIL CS_{bound} (H_2O)	SOLEIL CS_{free} (D_2O)	SOLEIL CS_{bound} (D_2O)
Sample configuration	continuous flow capillary		steady state		steady state			
Data composition	*n.a.*		Extrapolate to infinite dilution		Merge of multiple concentrations by averaging			

(c) Software employed for SAS data reduction, analysis and interpretation

SAXS data processing	ATSAS 2.8
Calculation of ϵ from sequence	ExPASy ProtParam server
Basic analyses	Primus with AutoRg and DATGNOM

(d) Structural parameters

AutoRg Guinier analysis. **: error values are over-exaggerated due to AutoRg protocols including residual aggregation.*								
$I(0)$ (arb.)	79.49±0.04	80.19±0.03	0.12±0.00	0.11±0.00	7.20±0.00	12.26±0.02	15.64±0.01	3.57±0.00
R_g (Å)*	30.0±0.1	29.2±0.0	27.8±0.6	26.1±1.0	29.33±3.33	28.45±0.74	29.34±1.90	28.83±0.18
q-range ($Å^{-1}$)	0.1975–0.4185	0.1758–0.4449	0.1508–0.4626	0.1821–0.4937	0.0758–0.4412	0.0785–0.4547	0.0623–0.0441	0.0568–0.4492
Coefficient of correl. R^2	0.89	0.92	0.96	0.95	0.96	0.87	0.96	0.93
DATGNOM $P(r)$ Analysis								
$I(0)$ (arb.)	78.93±0.02	79.84±0.02	0.1186	0.1126	7.183±0.007	12.21±0.02	15.59±0.01	3.554±0.003
R_g (Å)	29.68	29.00	28.0±0.6	26.8	29.21±0.06	28.10±0.04	29.14±0.03	28.51±0.04
D_{max} (Å)	93.6	91.3	109.4	94.3	105.6	88.2	93.8	90.1
GNOM total est.	0.6299	0.6385	0.7196	0.6830	0.9088	0.8313	0.9139	0.9714