Structure, Volume *20*

Supplemental Information

Structure of the Lectin Regulatory Domain of the Cholesterol-Dependent Cytolysin Lectinolysin Reveals the Basis for Its Lewis Antigen Specificity

Susanne C. Feil, Sara Lawrence, Terrence D. Mulhern, Jessica K. Holien, Eileen M. Hotze, Stephen Farrand, Rodney K. Tweten, and Michael W. Parker

Inventory of Supplemental Information

Supplemental Data

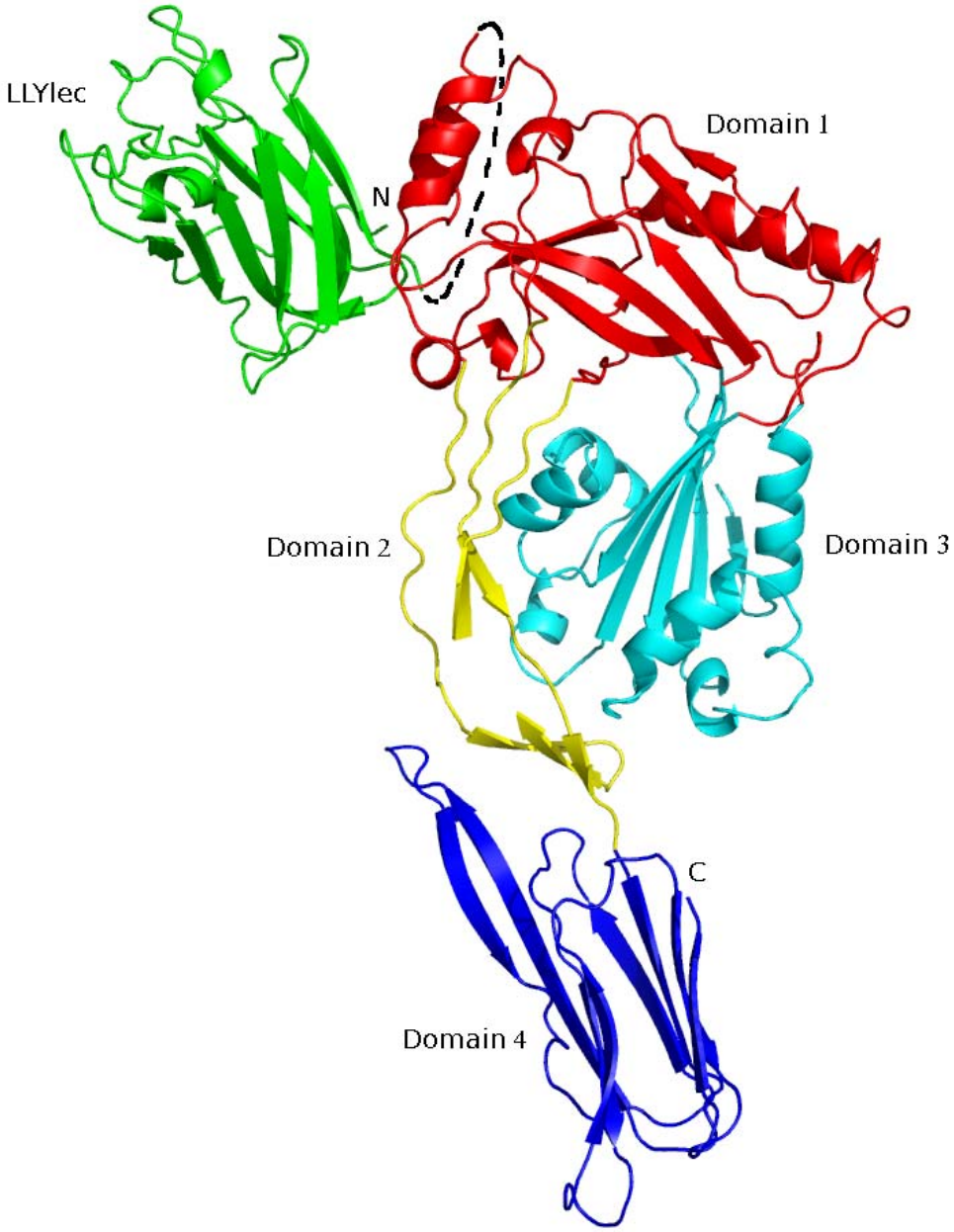
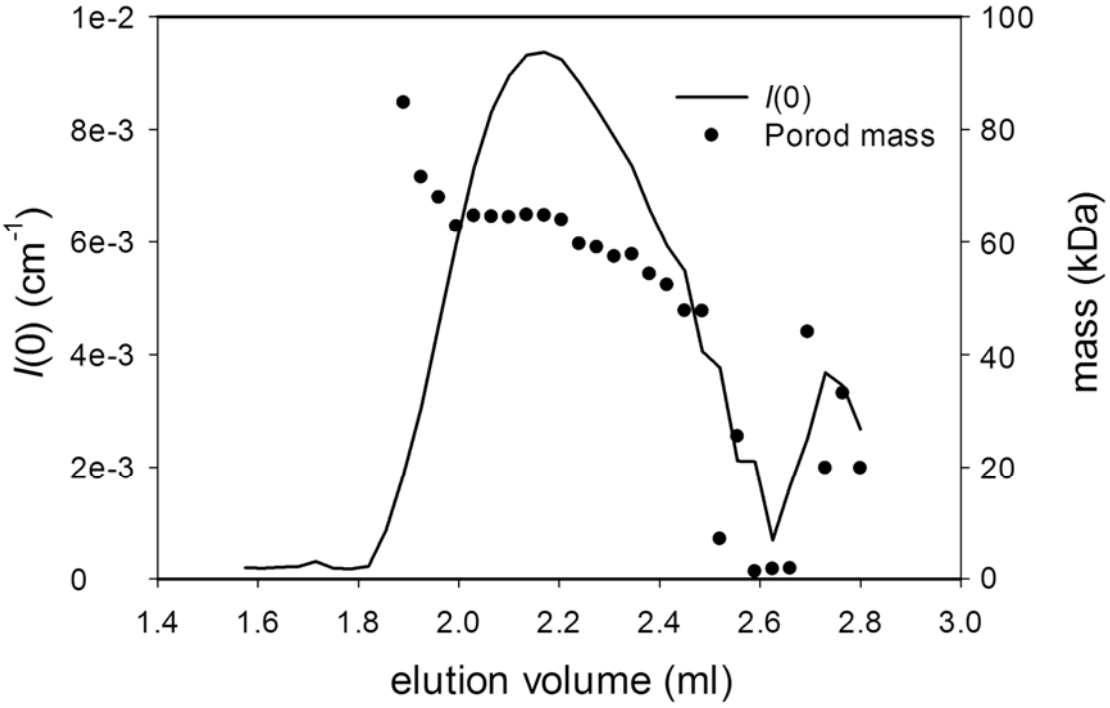


Figure S1

(A)


1e-2
8e-3
6e-3
4e-3
2e-3
0
I(0) (cm-1)
100
80
60
40
20
0
mass (kDa)
I(0)
Porod mass
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
elution volume (ml)

(B)

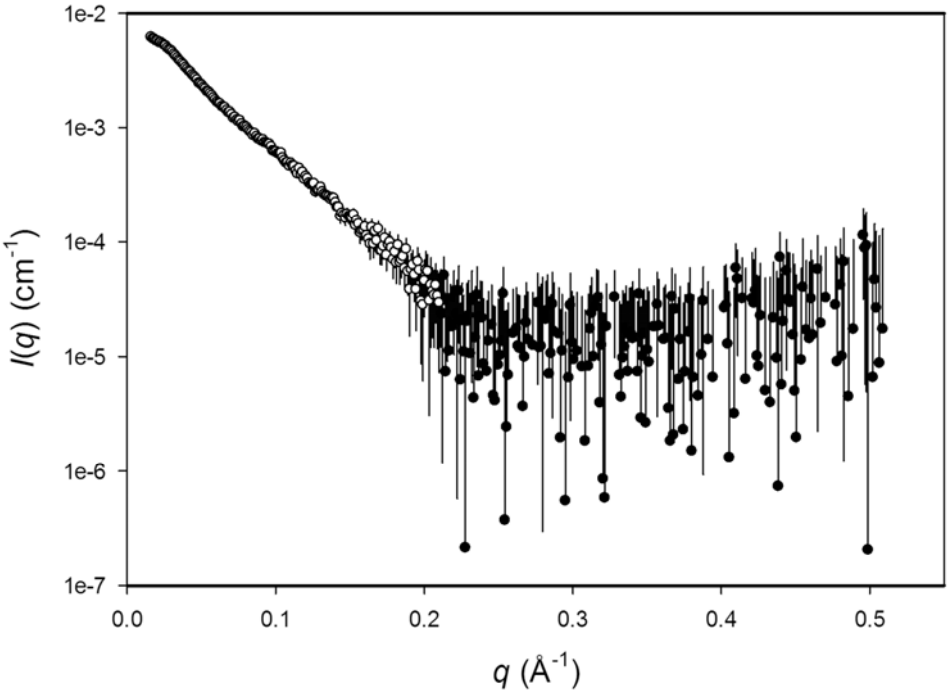

1e-2
1e-3
1e-4
1e-5
1e-6
1e-7
I(q) (cm-1)
0.0
0.1
0.2
0.3
0.4
0.5
q (Å-1)

Figure S2

(C)

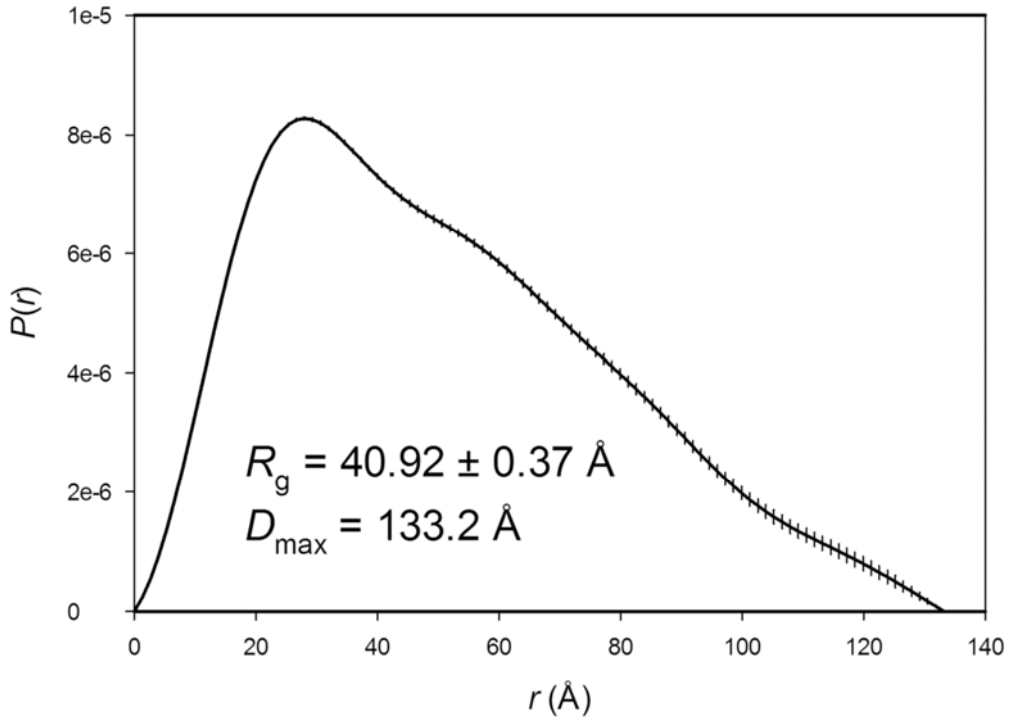

1e-5
8e-6
6e-6
4e-6
2e-6
0
P(r)
R_g = 40.92 ± 0.37 Å
D_{max} = 133.2 Å
0
20
40
60
80
100
120
140
r (Å)

Figure S2

Figure Legends

Figure S1, related to Figure 1. Cartoon diagram of a 3D model of LLY^{lec} showing how the N-terminal LLY^{lec} domain is positioned relative to the rest of the toxin molecule. The rest of the toxin was modeled based on the related CDC crystal structure of ALO (Bourdeau *et al*., 2009). The positioning of LLY^{lec} with respect to the rest of the toxin is based on the SAXS data presented in the text. The black dashed line signifies the (unmodeled) linker between the C-terminus of LLY^{lec} and the N-terminus of the CDC domain 1.

Figure S2, related to Figure 6. SAXS of LLY. (A) The elution of LLY from the Superdex S200 size exclusion column into the X-ray beam was tracked by monitoring the forward scattering intensity ($I(0)$) as a function of elution volume (solid line). The averaged scattering of the buffer eluting before the protein peak was used for background subtraction. The mass of the eluting species (circles) was estimated from Porod analysis the scattering profiles. This indicated that monomeric LLY eluted at 2.0-2.2 ml (circles). SDS-PAGE did not indicate the presence of smaller species in the LLY sample and the apparent drop in the mass for late eluting material is due to buffer subtraction mismatch between the injected sample buffer and the column buffer. (B) Averaged SAXS data for monomeric LLY. The mean intensities as a function of the magnitude of the scattering vector ($I(q)$ vs. q) are shown as circles and the error bars indicate ± 1 standard deviation. The final data set used for shape modeling is indicated with the open circles (N=176, q-range 0.016-0.210 $Å^{-1}$). (C) LLY pair distance vector distribution function. The radius of gyration (R_g) and maximum dimension (D_{max}) values derived from the $P(r)$ analysis are indicated.

Supplemental Experimental Procedures

Generation and purification of recombinant of the LLY^{lec} domain for crystallography

A mutant of LLY^{lec} (Q190C) was used in this study. The mutant was originally prepared for labeling with fluorescent dye (maleimidine derivative of Alexa-488 via the cysteine sulfhydryl LLY^{Q190C}) and the mutation was shown not to influence activity of the toxin. DNA of this mutant, corresponding to residues 38 to 190 (GenBank accession number AB051299.1), was amplified by PCR with an amino terminal 6 x His tag and a TEV protease cleavage site for His-tag removal. The fusion protein-coding region was cloned into pMSg67 and the sequence was verified after cloning.

The fusion protein was expressed in *E. coli* BL21 (DE3) in Turbo Broth medium (Athena Enzyme Systems) at 37°C. The cell pellet was resuspended in ice cold Buffer A (500 mM NaCl, 25 mM Tris pH 7.5, 20 mM imidazole) with EDTA-free protease inhibitor tablets (Roche) (as per manufacturers instructions). Cells were lysed by passage through an Aventi C5 Cell Crusher 103mPa. HexaHis-LLY^{lec} was purified on a 5 ml HisTrap ^{TM} HP column (GE), equilibrated in Buffer A. The column was washed with 50 ml Buffer A until the optical density at 280 nm was less than 600 mAU. HexaHis-LLY^{lec} was eluted from the column with a continuous gradient from 20 mM imidazole (Buffer A) to 400 mM imidazole (Buffer B: 500 mM NaCl, 25 mM Tris pH 7.5, 400 mM imidazole). Fractions containing LLY^{lec} were pooled and dialyzed into 5 litres 150 mM NaCl, 25 mM Tris pH 7.5 at 4°C for 16 hours. TEV protease was added 1:100 w/w to the dialyzed protein and the His-tag cleaved at 4°C for 72 hours. The His-tag and TEV protease was removed by a second purification step on a 5 ml HisTrap ^{TM} HP column. Fractions containing cleaved LLY^{lec} protein were pooled and dialyzed into 10 mM NaCl, 25 mM Tris pH 7.5. Purified LLY^{lec} was concentrated to 10 mg/ml and stored at -80°C.

Generation and purification of recombinant of the LLY^{lec} domain for SAXS

Hexahis _TEV_LLY^{T168C}, a mutant of LLY, was expressed and purified for SAXS analysis of full length LLY, to locate the position of the lectin domain (LLY^{lec}) , relative to the CDC domains 1 to 4 (LLY^{CDC}). LLY^{T168C} was originally prepared for labeling with fluorescent dye (maleimidine derivative of Alexa-488 via the cysteine sulfhydryl LLY^{T168C}) and the mutation was shown not to influence activity of the toxin (Farrand *et al*., 2008). "The coding region for cloned *LLY* gene, corresponding to residues 38 to 665 (GenBank accession number AB051299.1) with the T168C mutation was subcloned into *pMCSg7* vector with an amino terminal 6 x His tag and a TEV protease cleavage site for His-tag removal. (*pMCSg7* is a ligation-independent cloning vector – for details see http://plasmid.med.harvard.edu/PLASMID/GetVectorDetail.do?vectorid=367).

LLY^{T168C} coding region was amplified by PCR using 5' oligonucleotide sequence: TACTTCCAATCCAATGAGCAAGGGAATCGTCCAGTT and 3'oligonucleotide sequence: TTATCCACTTCCAATGTTACTCATTCACAATTTTTTCAT and cloned into PMCSg7. The sequence was verified after cloning.

The fusion protein was expressed in *E. coli* BL21 (DE3) pREP4 in Turbo Broth medium (Athena Enzyme Systems) at 37°C. The cell pellet was resuspended at 21°C in lysis buffer (25 mM Tris pH 8.4, 500 mM NaCl, 10 % (v/v) glycerol, 0.1% Triton X-100, 100 µM PMSF, 10 µg/ml lysozyme and 10 µg/ml DNAse 1. Cells were lysed by incubating at 21°C shaking for 1 hour, then sonicated 6 x 10 seconds. Hexahis_TEV_LLY^{T168C} was purified on a 5 ml HisTrap ^{TM} HP column (GE), equilibrated in Buffer A. The column was washed with 50 ml Buffer A (500 mM NaCl, 25 mM Tris pH 8.4, 20 mM imidazole). Hexahis_TEV_LLY^{T168C} was eluted from the column with Buffer B (500 mM NaCl, 25 mM Tris pH 8.4, 400 mM imidazole). Fractions containing Hexahis_TEV_LLY^{T168C} were pooled and dialyzed into 20 mM MES pH 5.5, 500 mM NaCl, 10% (v/v) glycerol, 5 mM EDTA, 10 mM DTT at 21°C for 16 hours. Dialysed Hexahis_TEV_LLY^{T168C} was further purified by size exclusion chromatography on a Superdex

200 25/68 column in 20 mM MES pH 5.5, 500 mM NaCl, 10% glycerol, 5 mM EDTA, 10 mM DTT. Fractions containing monomeric hexahis_TEV_LLY^{T168C} protein (as verified by SDS PAGE – data not shown) were pooled and concentrated to 6 mg/ml and stored at -80°C.

Small angle X-ray scattering of intact LLY

SAXS data were recorded on the SAXS/WAXS beamline at the Australian Synchrotron. Preliminary SAXS experiments using a static sample indicated the presence of high molecular weight aggregates confounding the analysis. Thus, synchrotron SAXS in-line with gel filtration chromatography was used to fractionate the sample and obtain scattering from the isolated monomer. These data were collected using in-line gel filtration chromatography, as described by Gunn *et al*. (2011). Injections of 50 µl of Hexahis_TEV_LLY^{T168C} at 6 mg/ml were made onto a 2.4 ml Superdex 200 3.2/30 column (GE) equilibrated in 20 mM MES pH 5.5, 500 mM NaCl, 10% glycerol, 5 mM EDTA, 10 mM DTT flowing at 0.2 ml/min. For each run 600 detector images were collected as 2 s exposures every 2.1 s. The images were analyzed as averages of blocks of five sequential exposures using the SAXS15ID software (Australian Synchrotron). Each image series was converted to 120 individual $I(q)$ vs. q profiles, where $I(q)$ is the scattered X-ray intensity as a function of the momentum transfer vector $q = (4\pi\sin\theta)/\lambda$, where the scattering angle is 2θ and the X-ray wavelength is λ (1.0332 Å). The q range over which intensities were collected was 0.016-0.524 Å^{-1}.

The chromatographic profile of the LLY sample, generated by monitoring the forward scatter ($I(0)$) as a function of elution volume is shown in Fig. S2A. Estimation of the mass of the scattering particles by Porod analysis indicated that aggregated material eluted at the front of the peak and that the monomer (Porod mass ~65 kDa *c.f.* theoretic MW of 72 kDa) eluted from 2.0-2.2 ml. These SAXS data were averaged to yield the final data set from which shapes were modeled (Fig S2B). The volume and mass of the scattering particles were estimated using AUTOPOROD and the radius of gyration (R_g) was estimated

by Guinier analysis using AUTORG. Pair distance vector distribution functions $P(r)$ were calculated using AUTOGNOM, which also yielded the maximum dimension (D_{max}) of the scattering particle. The final data set selected by AUTOGNOM and subsequently used for all model building was 176 data points over the q-range 0.016-0.210 $Å^{-1}$. (Fig S2B). The maximum dimension (D_{max}) and radius of gyration (R_g) of the LLY monomer from $P(r)$ analysis of these data were 133 Å and 40.9 ± 0.4 Å, respectively (Fig. S2C).

Ab initio shape reconstructions were performed using DAMMIF (Franke & Svergun, 2009). Averaged filtered shape envelopes were generated from ensembles of DAMMIF envelopes using DAMAVER (Volkov & Svergun, 2003). Theoretical scattering profiles were generated from model coordinates and compared to experimental data using CRYSOL (Svergun *et al.,* 1995) and the statistical analysis of goodness of fit ($P_\chi(\chi^2,\nu)$) and the relative improvement between fits ($P_F(F,\nu_1,\nu_2)$) were performed as described by Mills *et al*., (2009).

Rigid body refinement was carried out using MASSHA (Konarev *et al*., 2001) and BUNCH (Petoukhov *et al.,* 2005) to identify the likely position of the LLY^{lec} domain with respect to four different LLY^{CDC} core models (ALO (pdb 3CQF); PFO (pdb 1M3J and 1PFO) and ILY (pdb 1S3R). The best fitting MASSHA refined model was obtained using ALO (Bourdeau *et al*., 2009), which was further refined using BUNCH. Ten BUNCH models were generated and compared. While the fit of the theoretical scattering profile of the ALO-based LLY^{CDC} alone to the experimental data was poor ($\chi^2_\nu = 4.95$; $P_\chi(\chi^2,\nu) \sim 0$), inclusion of the structure of the LLY^{lec} domain and simple rigid body refinement using MASSHA (Konarev *et al*., 2001) positioned the domain roughly in the same position as the computational docking and improved the fit markedly ($\chi^2_\nu = 0.588$; $P_\chi(\chi^2,\nu) = 0.999$). Additional simulated annealing refinement of the position of the LLY^{lec} domain, constrained by the interdomain linker and with inclusion of the N- and C-terminal tails, was performed using BUNCH (Petoukhov *et al.,* 2005). Ten BUNCH models were

generated and the best fitting BUNCH model was an excellent fit to the data (χ^2_ν = 0.431; $P_\chi(\chi^2,\nu)$ = 1.000), which is a statistically significant improvement over both the initial rigid body refinement (F = 1.36; P_F = 0.021) and the computationally docked model (F = 1.33; P_F = 0.032). All of the other BUNCH models had fits that were statistically indistinguishable from the best BUNCH model (F < 1.06; P_F > 0.35). The BUNCH models were optimally superimposed and an averaged and volume filtered shape envelope generated (Kozin & Svergun, 2001; Volkov & Svergun, 2003). The cross-correlation between the shape of BUNCH-derived envelope and the *ab initio* envelope was excellent, with a normalized spacial discrepancy value of 0.63. The reference BUNCH model identified in the fitting procedure (i.e. closest in shape to the average shape) is shown superimposed on the *ab initio* shape envelope (Fig. 8b).

To assess whether other physically reasonable LLY^{lec} positions were possible, the ensemble optimization method (EOM) was employed (Bernado *et al.,* 2007). A pool of 15000 full-length models was generated with randomized non-overlapping LLY^{lec} and LLY^{CDC} orientations constrained only by the interdomain linker. The D_{max} values in the pool ranged from 117-216 Å. The pool was interrogated using GAGOE to find the best fitting single model and ensembles composed of n models (n = 2, 5, 10, 20). To identify individual models, GAGOE was used iteratively with the "number of curves per ensemble" set to one. Following each run the scattering curve from the model identified was removed from the pool. All ensembles selected had narrow R_g distributions centered close to the experimental values and there was no improvement in fit for ensembles of models (n = 2-20) over the best fitting single model, strongly suggesting that the LLY^{lec} domain adopts a single orientation. The fit of the best single EOM model (χ^2_ν = 0.415) was very close to that of the best BUNCH model (χ^2_ν = 0.431). The pool was further interrogated to find the range of individual EOM models with fits statistically indistinguishable from the best model (~0.5 % of the pool had P_F < 0.05 *c.f.* the best EOM model). While the exact orientation of the LLY^{lec} domain varied, all of these models had the centre of mass of the LLY^{lec} domain positioned similarly and the C-terminus of the LLY^{lec} proximal to the N-terminus of the LLY^{CDC}. The best fitting 20 models were

optimally superimposed and an averaged shape envelope generated (Kozin & Svergun, 2001; Volkov & Svergun, 2003). The reference EOM model fits well into the *ab initio* shape envelope (Fig. 8c). The simulated annealing (BUNCH) and random search (EOM) methods identified similar LLY^{lec} positions that were consistent with the *ab initio* shape envelope.

Supplemental Reference

Kozin, M., and Svergun, D.I. (2001). Automated matching of high- and low resolution structural models. J. Appl. Crystallogr. *34*, 33–41.