IUCrJ

Volume 3 (2016)

Supporting information for article:

Structure of a heterogeneous, glycosylated, lipid-bound, *in vivo*-grown protein crystal at atomic resolution from the viviparous cockroach *Diploptera punctata*

Sanchari Banerjee, Nathan P. Coussens, François-Xavier Gallat, Nitish Sathyanarayanan, Jandhyam Srikanth, Koichiro J. Yagi, James S. S. Gray, Stephen S. Tobe, Barbara Stay, Leonard M. G. Chavas and Subramanian Ramaswamy

Table S1 Root mean square deviation values for the superimposition of Lili-Mip on various lipocalins.

	Cα	**Overall RMS Deviation**
Siderocalin	15.4 Å	16.1 Å
Human tear lipocalin	3.9 Å	4.5 Å
Boar salivary lipocalin	14.4 Å	8.0 Å
Prostaglandin D synthase	6.1 Å	6.5 Å
Bacterial lipocalin	4.5 Å	4.6 Å

Table S2 Approximate averaged distance between the residues forming the Lili-Mip hydrophobic cavity and linoleic acid (LIN) or oleic acid (OLE).

Lili-Mip residue	LIN	OLE
Leu 22	4.0 Å	3.9 Å
Leu 29	5.1 Å	5.0 Å
Gln 32	3.6 Å	5.8 Å
Val 33	4.1 Å	4.5 Å
Ile 36	4.1 Å	4.4 Å
Val 51	3.9 Å	4.5 Å
Asp 53	4.1 Å	3.7 Å
His 63	3.9 Å	3.9 Å
Phe 76	3.7 Å	3.5 Å
Met 78	4.8 Å	4.5 Å
Thr 81	4.1 Å	4.1 Å
Glu 83	2.7 Å	3.4 Å

Tyr 84	4.4 Å	3.8 Å
Tyr 88	4.0 Å	3.9 Å
Phe 100	3.6 Å	3.0 Å
Leu 113	4.0 Å	3.6 Å

Table S3 Energetic values for mammalian milks as compared to Lili-Mip crystals.

Species	**kcal (per 100 g)**	**kJ (per 100 g)**
Lili-Mip	232	973
Cow	66	275
Goat	60	253
Sheep	95	396
Water Buffalo	110	463
Human	72	303

Table S4 Energetic constants of Lili-Mip.

Calories per gram of protein	4.0
Calories per gram of fat	9.0
Calories per gram of carbohydrates	4.0
Kcal to kJ conversion factor	4.19

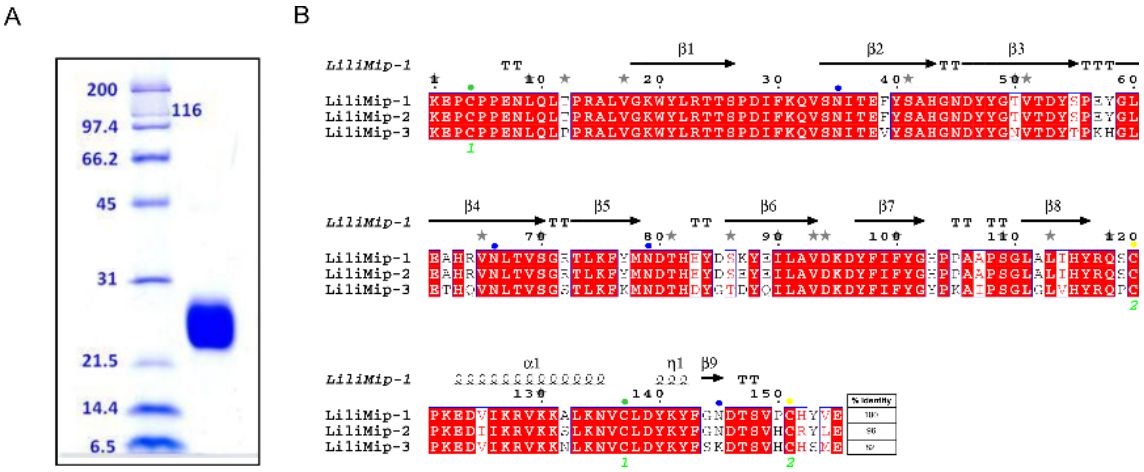

Figure S1 *In vivo* Lili-Mip crystals. (A) An SDS-PAGE gel of Lili-Mip protein after solubilizing multiple crystals. The streak shows a mixture of molecular masses between 22 kDa - 30 kDa, as indicated by a broad-range molecular ladder. (B) Primary sequence alignment of the Lili-Mips 1, 2 and 3, the three abundant sequences observed in the crystals. The sequence identities between two sequences are indicated to the right of the alignment. The red background-shadings indicate identical amino acids. Secondary structure elements for Lili-Mip are displayed above the alignments. Blue spheres signify glycosylated residues and green and yellow spheres show the cysteines forming 1st and 2nd disulfide bridges, respectively. The sequence alignments were carried out using the program *Clustal Omega* (Sievers *et al.*, 2011) and the graphical output was generated using ESPript version 3.0 (Robert & Gouet, 2014).

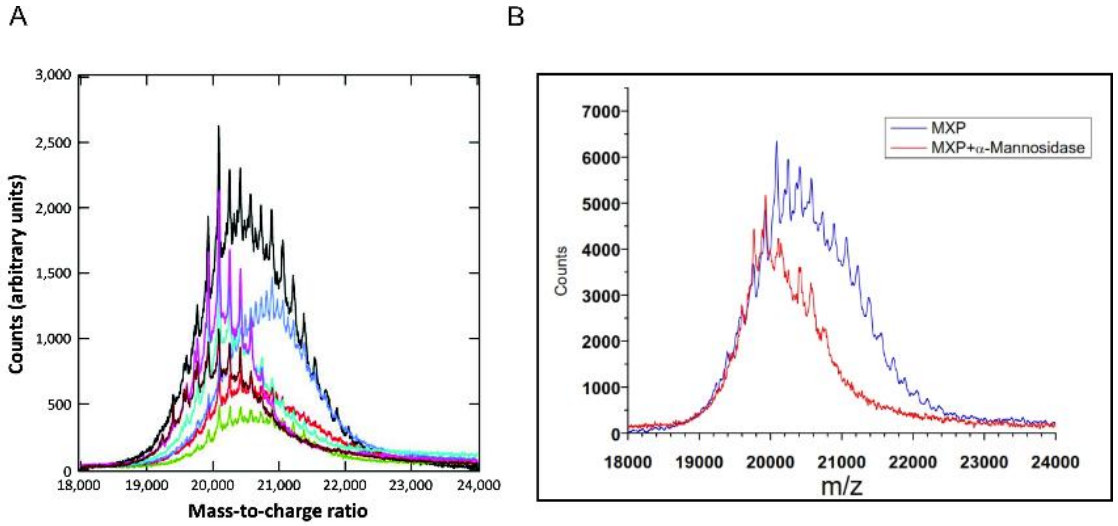

Figure S2 Mass analysis of Lili-Mip. (A) The spectrum covers the mass-to-charge range from 18,000 to 24,000 Da. Prior to mass analysis;s solubilized Lili-Mip was separated by free-flow electrophoresis (FFE). The broad envelope (4,000 Da) of the non-fractioned protein solution (black curve) and the different masses observed in the FFE fractions (colored curves) confirm the presence of multiple isoforms in the crystals. (B) Overlay of MS spectra of the Lili-Mip protein before (blue) and after (red) digestion with α-Mannosidase confirm high mannose content.

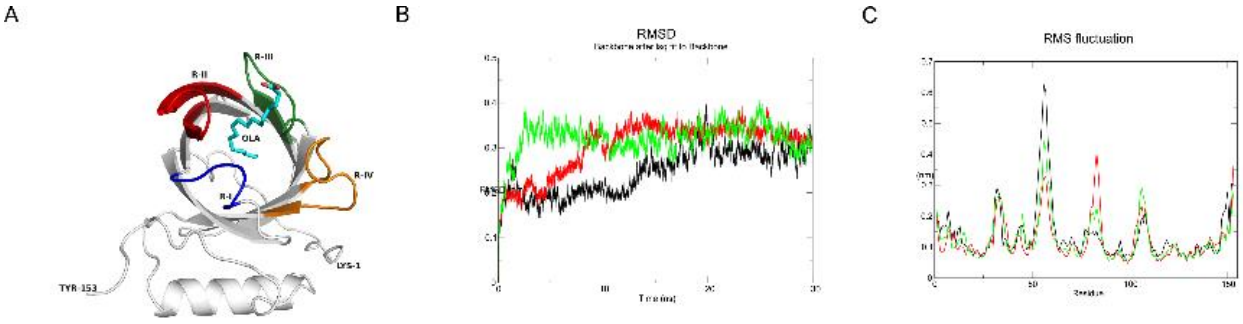

Figure S3 Analysis of trajectories from 30 ns molecular dynamic simulation studies of deglycosylated native/ligand unbound (DglyNat, black), oleic acid bound (DglyOla, red), and linoleic acid bound (DglyEic, green) Lili-Mip structures. **(A)** Ligand binding site showing the four regions I-IV. Region I, corresponding to residues 30-35 is colored in blue. Region II, corresponding to residues 50-65, is colored in red. Region III, corresponding to residues 75-85, is colored in green. Region IV, corresponding to residues 102-112, is colored in orange. **(B)** Comparison among root mean square deviation (RMSD) values for backbones from the three simulations. **(C)** Comparison among root mean square fluctuation (RMSF) values for the Cα of the residues in the three simulations. Maximum fluctuations are seen in the four regions marked as I, II, II and IV.

Movie S1. Extraction of *in vivo* grown crystals. Crystals of Lili-Mip are extracted from the midgut of *D. punctata* embryos by dissection. Crystals pouring out of the midgut in large quantities are stored in sterile water.

Movie S2. Movements observed in regions I to IV during the 30 ns of MD simulation with native protein.

Movie S3. Movements observed in regions I to IV during the 30 ns of MD simulation with the oleic acid bound structure.

Movie S4. Movements observed in regions I to IV during the 30 ns of MD simulation with the linoleic acid-bound structure.