

GapRepairer - repair protein structures and their topology.

Supplementary Data

Aleksandra I. Jarmolinska, Michal Kadlof, Pawel Dabrowski-Tumanski and Joanna I. Sulkowska

Contents

- 1 Protein structure with doubtful knot fingerprint. 1**
- 2 Protein structure with doubtful knot type stemming from large gap. 3**

1 Protein structure with doubtful knot fingerprint.

The structure of thermostabilized EAAT1 mutant with PDB code 5LLM has an unusual knot fingerprint $K + 3_1 + 3_1 + 3_1$. It is unique among all proteins gathered in KnotProt database, therefore it is possible, that the fingerprint is an artifact of the gap. Indeed, the straight interval connecting the termini of one gap is located in between other pieces of the chain (Fig 1 bottom panel). Modeling the missing fragments with GapRepairer yields a structure with regular slipknot $S + 3_1$ motif (Fig 1 top panel).

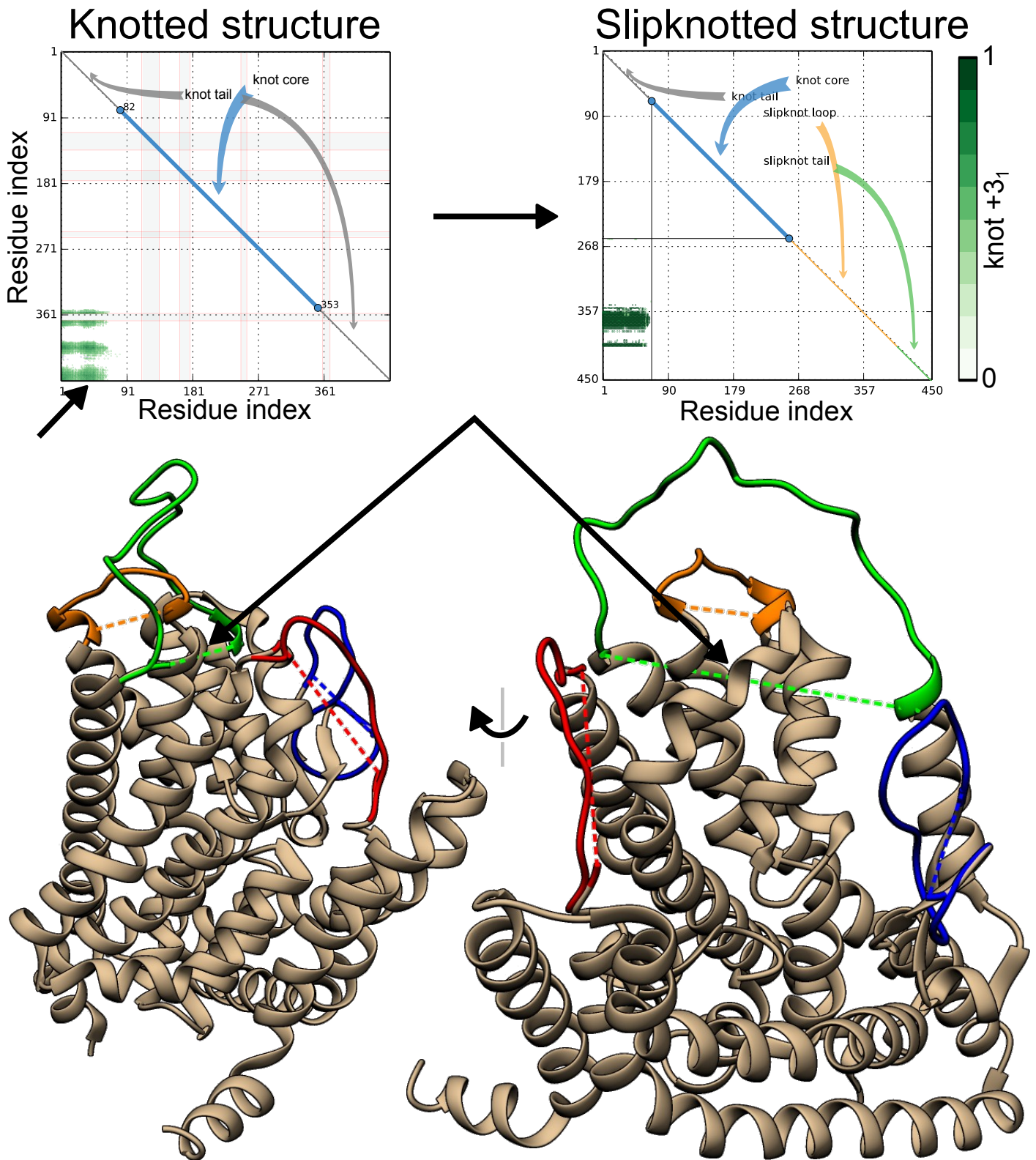


Figure 1: **Comparison of EAAT1 mutant crystal structure (PDB code 5LLM) with its model with filled gaps.** Top row - knot fingerprint matrices of the original structure (left) and the best model (right). The arrow indicates the dubious 3_1 knotted part. Bottom row - the original structure viewed from different angles. The dashed lines denote the gaps and straight intervals connecting the gap termini. The color chain pieces are the corresponding gap filling in the best model. Note that the green dashed line is surrounded by other parts of the chain (indicated by the arrows).

2 Protein structure with doubtful knot type stemming from large gap.

The crystal structure of transport protein from *arabidopsis thaliana* (PDB code 5L25) has 6 gaps in total, one comprising 94 residues (green gap in Fig 2). Moreover, the straight interval filling of the second largest gap (red in Fig 2) results is surrounded by other chain parts and results in unnatural knot type 5_1 (red spots in the knot fingerprint matrix in Fig 2).

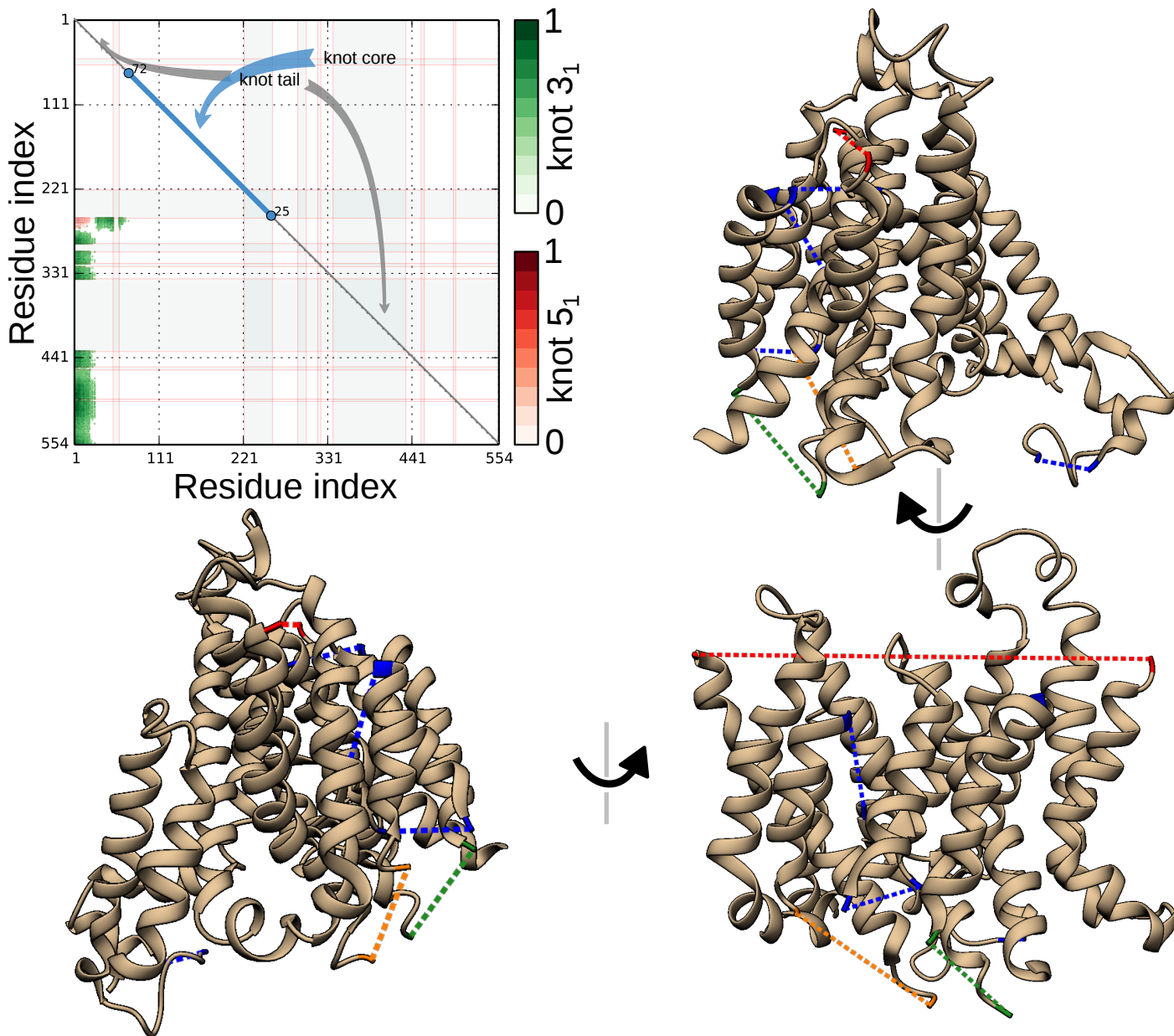


Figure 2: **Knot fingerprint matrix and the (gapped) structure of transport protein with PDB code 5L25.** The knot fingerprint (top left panel) comprises unnatural 5_1 knotted motif (red spots in the matrix). It is the result of the arrangement of straight interval filling of the 36-residue long gap (indicated by red dashed line). Note also, that the termini of the longest gap (94 residues, green dashed line) are spatially very close each other.

The artificial entanglement introduced by straight interval filling of the second largest gap (red in Fig 2) is removed by modeling of this gap. However, carefree modeling results in the improper structure of the fragment filling the largest, 94-residue long gap (green in Fig 2) and leads to even more complicated knot fingerprint (Fig 3). Comparison between two results of modeling shows, how easy it is to artificially entangle protein chain if it has a large gap (Fig 3).

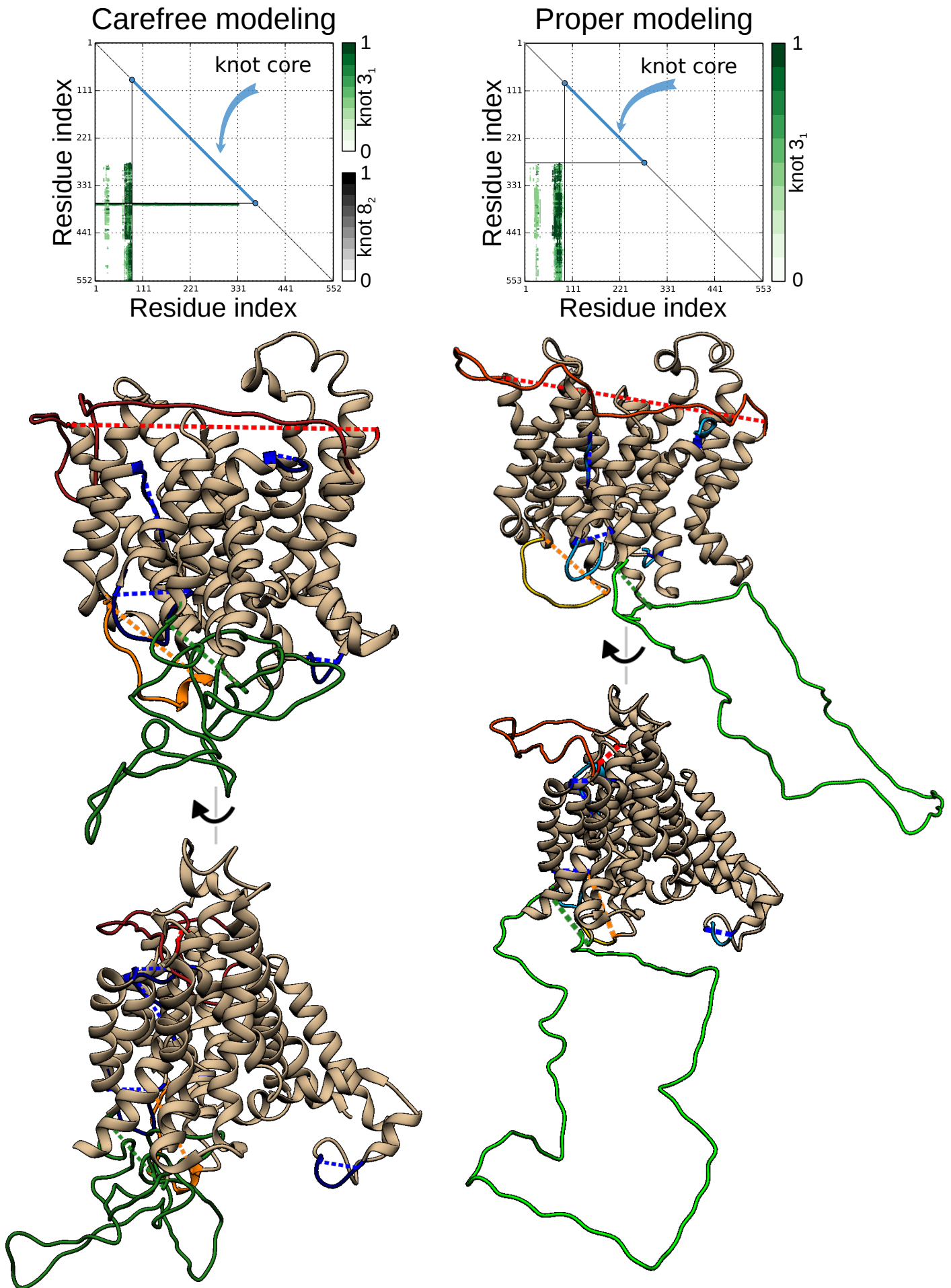


Figure 3: **Comparison of two gap fillings of transport protein with PDB code 5L25.** In the carefree modeling (left panel), the filling of 94-residue long gap (green) results in complex knot fingerprint, comprising unnatural 8_2 knot. For comparison, proper modeling of this loop (right panel) results in regular $S3_13_1$ motif.