

The image displays a central 3D ribbon model of a protein structure, possibly a viral capsid, rendered in orange and yellow. This central structure is surrounded by several smaller, blue ribbon diagrams of protein fragments, each labeled with a specific amino acid sequence. The background is a light blue gradient with a subtle pattern of white lines.

The labels on the protein fragments include:

- Top left: **QWINKA**, **YAPCO**, **CXCCTCCV**, **YKRCSDI**
- Top center: **LSST**, **MYATI**, **ALN**, **ALSC**
- Top right: **APLNRICE**, **TFRPGCPPL**, **SKXYTIPSY**
- Bottom left: **VLACI**, **YWA**, **YPSAI**, **YVW**, **YVDS**
- Bottom center: **GEDIR**, **LEIGDI**, **GEDIR**, **LEIGDI**, **GEDIR**, **LEIGDI**
- Bottom right: **LAEL**, **LAAKI**, **LAES**, **QAPN**, **LES**
- Bottom center (rotated): **VGINV**, **YKCHTPI**, **YKCHTPI**, **YKCHTPI**, **YKCHTPI**, **YKCHTPI**



ROYAL SOCIETY
OF CHEMISTRY

Diwakar Shukla *et al.*
Substrate prediction for RiPP biosynthetic enzymes via
masked language modeling and transfer learning