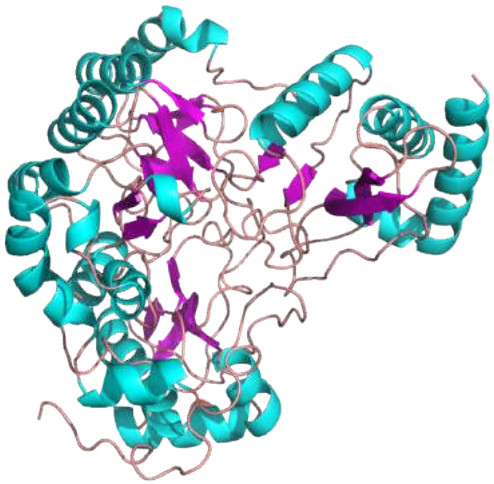
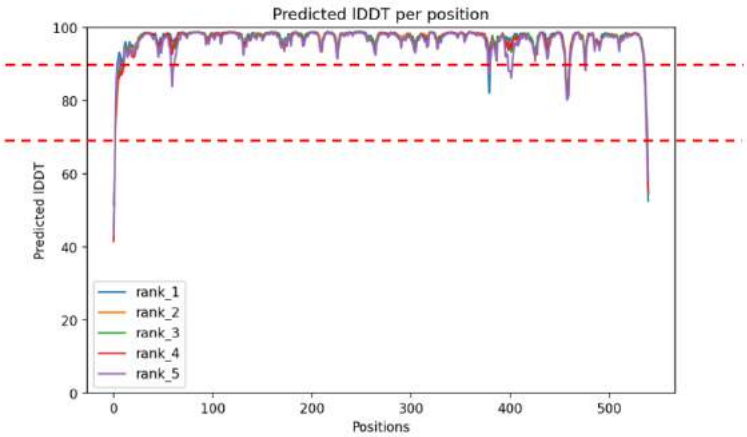
Modelling 3D protein structures utilising AlphaFold technology



pIDDT (predicted IDDT matching score)

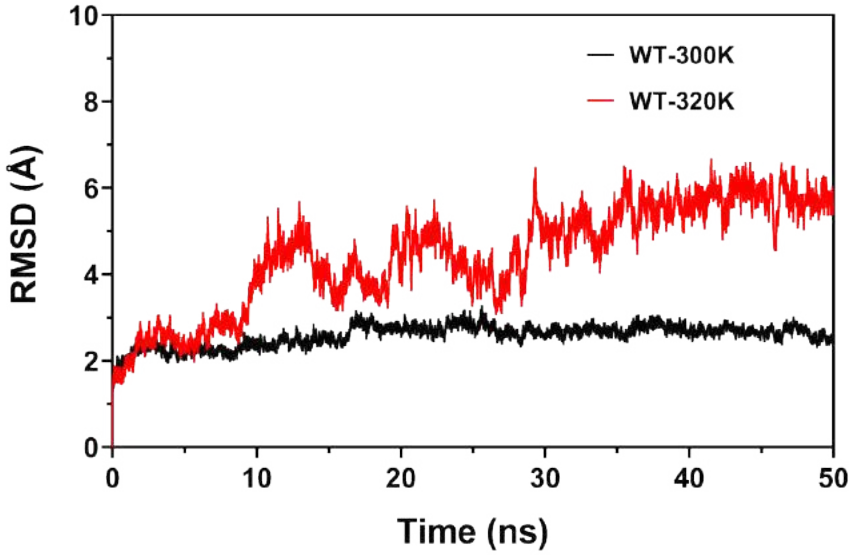
The anticipated structure comprises atomic coordinates and confidence estimates per residue, ranging from 0 to 100. The credibility grows as the score increases. This measure of confidence, referred to as pIDDT, corresponds to the per-residue score predicted by the model on the lDDT-C metric. The idDT metric is a pre-existing tool utilised in protein structure prediction. Residues with a pIDDT score of ≥ 90 demonstrate an exceptionally high level of confidence in the model, whilst those with scores between 90 and 70 are classified as having high confidence. Residues with scores between 70 and 50 on the pLDDT metric have low confidence levels, whereas those with scores less than 50 correspond to extremely low confidence levels.

The figure demonstrates that the majority of the target protein AF prediction models exhibit high model confidence. However, there are a few specific positions (such as around position ~70 compared to around positions ~380, 450) and the ends of the chain where the confidence is not as high.



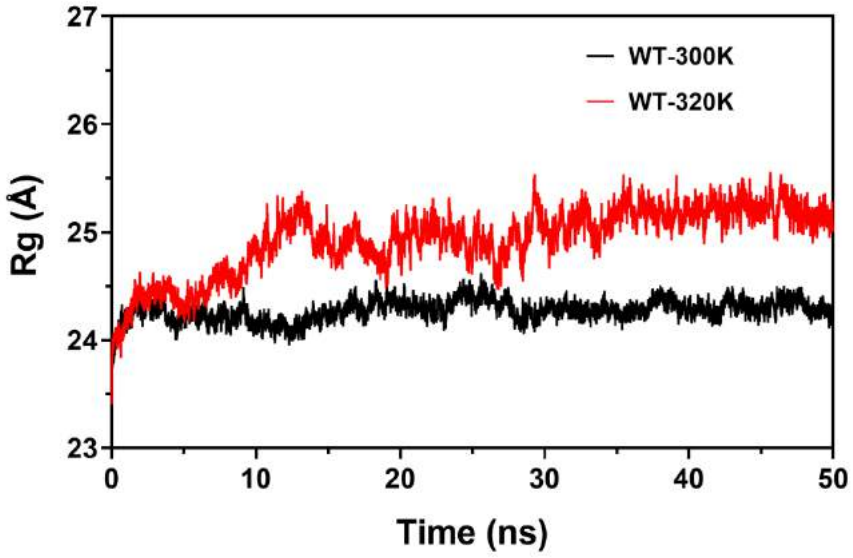
This value is calculated as the root mean square deviation. The RMSD value represents the degree of deviation of data from the mean, equivalent to the standard deviation, in statistics. The RMSD value is a widely used parameter in the analysis, modelling and alignment of protein structures, as well as in molecular dynamics simulations. It is employed to quantify deviations of atoms from their relative positions. The larger the RMSD value, the greater the spatial extent of

an atom's motion, and the lower its spatial resistance. Technical abbreviations will always be explained upon first use. The protein underwent simulations at 320K and 300K for 50 nanoseconds. The root mean square deviation (RMSD) was stable after 10 nanoseconds at 300K. Nonetheless, the considerable fluctuation range observed at 320K indicates greater protein movement than at 300K. Consequently, the protein's stability is relatively lower at 320K.



RG (Radius of Gyration)：

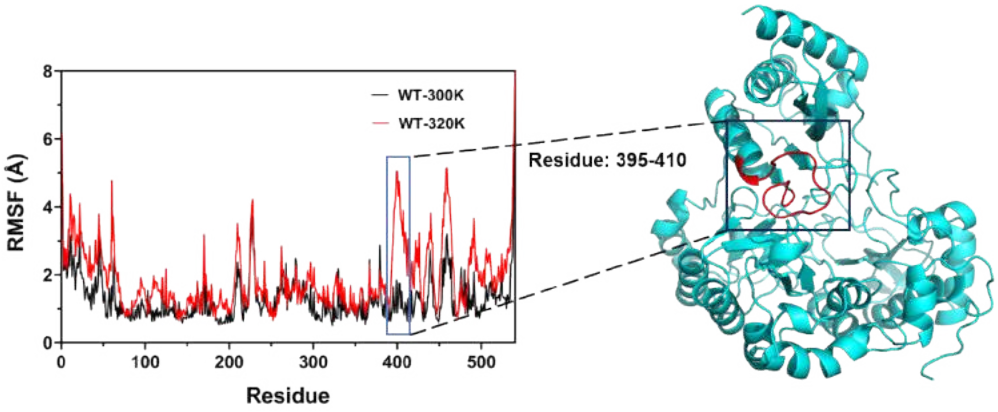
RG is a parameter that assesses the degree of compactness in molecules and proteins, typically measured in angstroms (Å). It represents the average distance between the mass distribution of all atoms in a molecule and a fixed center. RG proves useful in monitoring global conformational changes, especially when folding of proteins or aggregation occurs. A smaller RG usually indicates a more compact structure.



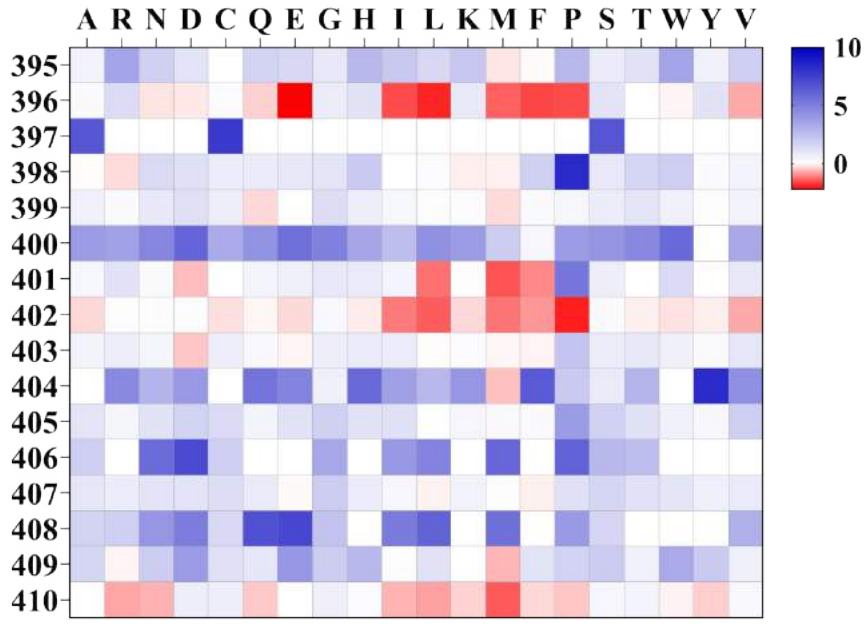
RMSF (Root Mean Square Fluctuation)：

Root mean square fluctuation (RMSF) is employed to determine the average degree of fluctuation of each atom in a molecule during a simulation. It accurately represents the flexibility and vibration of atoms in the molecule. Utilising RMSF, it is possible to identify and locate molecular regions where instability or flexibility occurs during simulation. This method has practical value in detecting active sites, flexible regions and conformational changes in proteins. RMSF values are used to express atomic coordinate fluctuations in the amino acids that are characteristic of the protein. Larger RMSF values indicate increased flexibility. The first time a technical term is used, its abbreviation is explained. The text is written in simple sentences with a

logical sequence of information and causal connections between statements. Biased and emotional language is avoided, and high-level, standard language and consistent technical terminology is used. Consistent formatting, style, and citation is maintained throughout. The language is formal and grammatically correct. A virtual mutation was conducted on the flexible region, identified by its red colour in the protein marker. The virtual mutation indicates that lower energy results in higher stability, which was subsequently verified through experimentation.



Foldx virtual mutation (more red means lower energy, good mutation site)



Primers designed with reference to virtual mutation

突变	引物(正向)
396E	CATTGGTCACGCGGTTAGCTGCGAAGGTATTGAATATTGCAACCTGGC
396I	GCCAGGTTGCAATATTCAATACCGATGCAGCTAACCGCGTGACCAATG
396L	CATTGGTCACGCGGTTAGCTGCCTCGGTATTGAATATTGCAACCTGGC
396M	CATTGGTCACGCGGTTAGCTGCATGGGTATTGAATATTGCAACCTGGC
396F	CATTGGTCACGCGGTTAGCTGCTTTGGTATTGAATATTGCAACCTGGC

396P	CATTGGTCACGCGGTTAGCTGCCCAGGTATTGAATATTGCAACCTGGC
401L	TGCACCGGTATTGAATATCTCAACCTGGCGCTGGTCGA
401M	TGCACCGGTATTGAATATATGAACCTGGCGCTGGTCGA
402I	GCACCGGTATTGAATATTGCATTCTGGCGCTGGTCGAAACCA
402L	GCACCGGTATTGAATATTGCCTCCTGGCGCTGGTCGAAACCA
402M	GCACCGGTATTGAATATTGCATGCTGGCGCTGGTCGAAACCA
402P	GCACCGGTATTGAATATTGCCCCCTGGCGCTGGTCGAAACCA
410M	GGCGCTGGTCGAAACCAAAATGCGTCTGCGCAAAATCGCGG