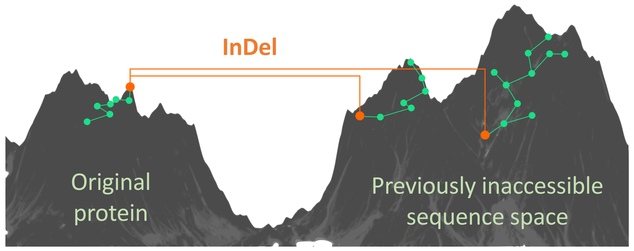
Recording the Fitness Landscapes of Small Deletions and Substitutions in GFP

Maya Petek, Stéphane Emond, Florian Hollfelder
Department of Biochemistry, University of Cambridge, UK

Insertions and deletions open a new area of fitness landscapes

The combination of an area of sequence space with a functional score defines a fitness landscape. Properties of fitness landscapes can be deduced from analysis of mutational pathways in directed evolution (below).



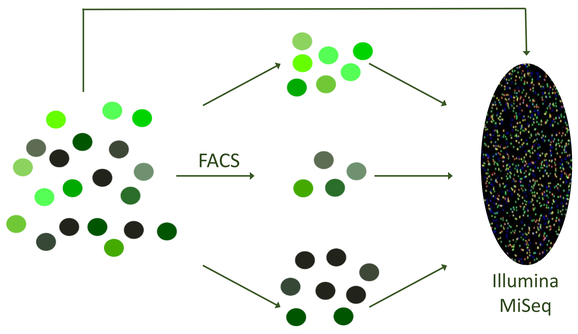
Mutagenesis methods that introduce substitutions are routinely used in protein engineering. Insertions and deletions (InDels), in addition to changing side chain chemistry, alter the protein backbone. They may open new areas of functional space, but risk disrupting existing activity and folding of the protein.

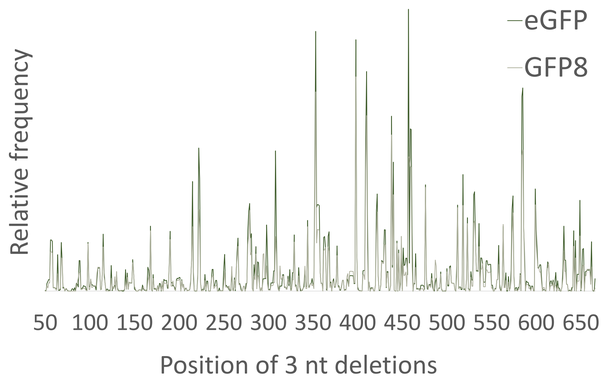
Questions:

- To what extent is folding/stability necessary to allow InDels?
- In which part of the protein are mutations tolerated?
- Are InDel libraries a good starting point for directed evolution (assume stabilized starting point)? Which readout (brightness / different colour) will lead to diverse and/or successful trajectories and functionally improved proteins?
- Epistasis with adjacent substitution: does remodelling of the InDel sequence context increase tolerance?

Experimental method

- Random transposon insertion followed by precise excision of the transposon was used to generate libraries of single mutants, by deletion 3, 6 or 9 nucleotides (nt), exchange of 3 nt or insertion of 3, 6 or 9 nt
- Mu transposon shows a weak sequence preference such that 87% possible mutation sites are sampled
- Libraries were expressed in E. coli and sorted using Fluorescence Activated Cell Sorting to separate inactive, medium and high fluorescence variants
- Starting and sorted libraries were sequenced with Illumina MiSeq to determine the composition of libraries

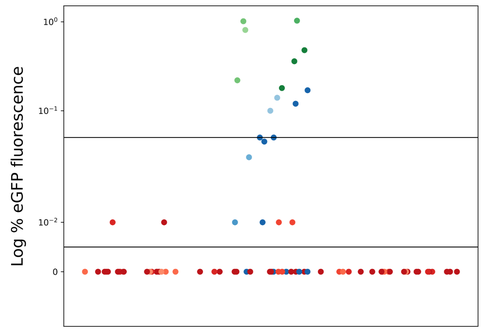




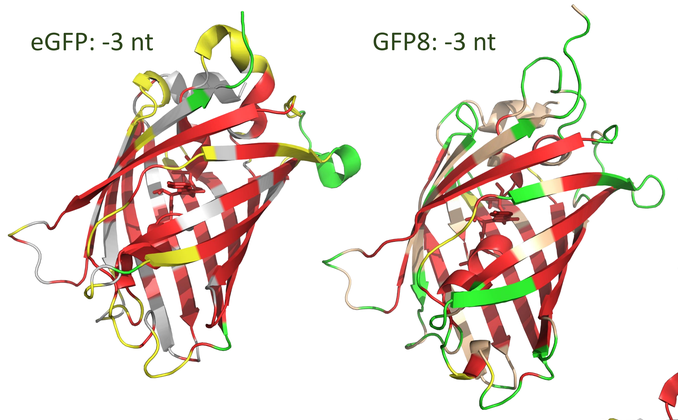
References:
Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., and Tans, S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. Nature *445*, 383–386.
Emond, S., Kay, E., Heames, B., Devenish, S., Petek, M., Tokuriki, N., Hollfelder, F. TRIAD: a transposition-based approach for gene mutagenesis by random short in-frame insertions and deletions. (in preparation)
Jones, D.D. (2005). Triplet nucleotide removal at random positions in a target gene: The tolerance of TEM-1 β-lactamase to an amino acid deletion. Nucleic Acids Res. *33*, 1–8.

Position of fluorescent, intermediate and dark deletions in eGFP and GFP8

Right: Distribution of log % of fluorescence for individually characterised and Sanger sequenced variants. The predicted activity of variants (green = high, blue = medium, red = inactive) fits well with measured green fluorescence.

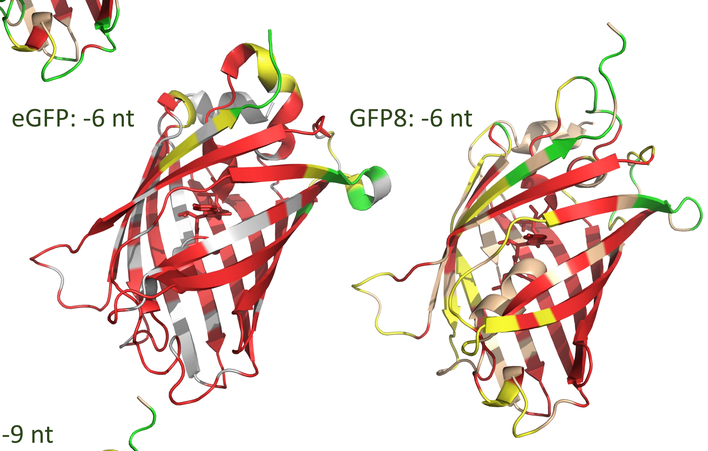


Below: Heat map of predicted fluorescence of deletion variants.

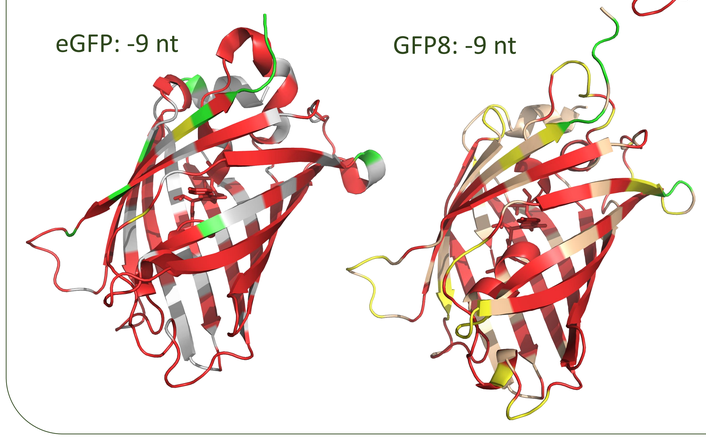


High
Medium
Inactive

Single amino acid deletions (-3 nt) are more likely to be highly fluorescent in GFP8 and intermediate in less stable eGFP. All loops show some tolerance of 3 nt deletions.



Two amino acid deletions (-6 nt) are more deleterious to eGFP, but tend to show medium fluorescence intensity GFP8, including in position where single amino acids deletions are not tolerated. Here, they rearrange the backbone but preserve the orientation of β-strands.



Some three amino acid deletions retain reduced fluorescence in the stable variant, but a few are still tolerated in eGFP, primarily in the final β strand.