Using AlphaFold2 to identify promising *de novo* protein candidates

Background: Over the past decade, evidence has accumulated that new protein coding genes can emerge *de novo* from previously non-coding DNA. Most studies have focused on large scale computational predictions of *de novo* protein coding genes across a wide range of organisms. In contrast, experimental data concerning the folding and function of *de novo* proteins is scarce. This might be due to difficulties in handling *de novo* proteins in vitro, as most are predicted to be short and disordered. Recent leaps in *in silico* structure prediction can be employed to identify highly folded and highly disordered *de novo* proteins that will be used for experimental analysis. Further experimental exploration of *de novo* proteins will both, shed new light on molecular evolution, and enable the development of new techniques in protein engineering for biotechnological applications.

Objectives: During this project we will use pLDDT scores of b factor positions of the AlphaFold2 predicted structures of *de novo* proteins to find candidates that are (i) highly structured or (ii) highly disordered. Molecular dynamics simulations (Gromacs) will be conducted to explore structural flexibility. Candidate´s *de novo* status will be confirmed manually using established synteny-based approaches. Candidate sequences will be ordered, amplified and cloned into plasmids for expression. Optimal conditions for protein expression will be identified using SDS PAGE and Western Blot. Proteins will further be purified using His-tag and properties analysed in the first place using Thermal Shift Assay (TSA), twin arginine translocation assay (tat assay) and Circular dichroism (CD)

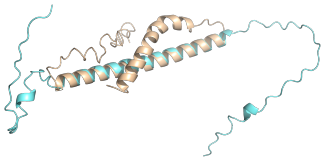

Figure 1: AlphaFold2 prediction of *de novo* protein Goddard (cyan) aligned with its partial structural determination (Lange et al. 2021)

Requirements:

- Experience in bash scripting, good command of unix command line
- Interest in evolution at the level of individual proteins and in protein structure & folding
- Interest in lab work on DNA and protein level and basic knowledge of PCR, DNA-cloning, protein expression & purification

Methods:

- Structural bioinformatics
- Biochemical characterization of proteins via SDS-PAGE, Western Blot, TSA, tat, CD

Supervision:
Lars Eicholt, Room 20, l.eicholt@wwu.de,
Molecular Evolution and Bioinformatics Group (http://bornberglab.org/).

Selected Literature:
1 Lars A. Eicholt, Margaux Aubel, Katrin Berk, Erich Bornberg-Bauer, Andreas Lange
Heterologous expression of naturally evolved putative de novo proteins with chaperones
Protein Science, **2022**
2 Erich Bornberg-Bauer, Klara Hlouchova, and Andreas Lange
Structure and function of naturally evolved de novo proteins
COSB, **2021**
3 Andreas Lange, Prajal H Patel, Brennen Heames, Adam M Damry, Thorsten Saenger, Colin J Jackson, Geoffrey D Findlay, Erich Bornberg-Bauer
Structural and functional characterization of a putative de novo gene in Drosophila
Nat. Comms, **2021**