PRe-ART

Predictive Reagent Antibody Replacement Technology

Anupama Chembath[1], Mohammed Ashraf[1], Ben P.G. Wagstaffe[1], Erich Michel[2], Yvonne Stark[2], Noelia Ferruz-Capapey[3], Josef Kynast[3], Birte Hoecker[3], Anna V Hine[1] and Andreas Plückthun[2]

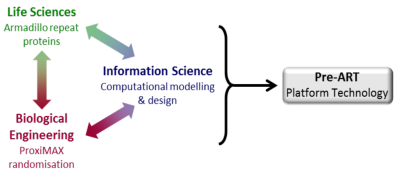
[1]School of Life & Health Sciences, Aston University, B4 7ET , Birmingham, UK.

[2]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

[3] Biochemistry, University of Bayreuth, Bldg. NW III, Universitaetsstrasse 30, 95440, Bayreuth, Germany

is led by the University of Zurich, with partner institutions of Aston University (UK) and the University of Bayreuth (Germany). We gratefully acknowledge H2020 FET-OPEN for our funding.

Abstract: Reagent antibodies are used extensively in research, typically to bind or capture protein targets. Unlike well-characterised therapeutic antibodies, about half of the commercially available *reagent* antibodies have previously been shown to not function correctly either in terms of their specificity or else in recognising their target at all. PRe-ART (Predictive Reagent Antibody Replacement Technology) aims to replace conventional reagent antibodies with sequence-defined, designed armadillo repeat proteins (dArmRPs), capable of conserved and specific binding of extended peptides. PRe-ART uses a feedback loop of experimental synthesis and evaluation with computational modelling to iteratively develop the properties of novel dArmRPs. Ultimately, PRe-ART aims to create an encyclopaedia of pre-designed and experimentally pre-selected modules that may be combined to generate unique binding molecules that function as antibody replacements. Progress to date in terms of library synthesis, screening and modelling will be presented.



What are dArmRPs? Armadillo repeat proteins are alpha solenoid proteins, possessing subunits of 42 amino acids each consisting of 3 alpha helices (H1-H3) (Figure 1). Each repeat is capable of recognising an individual dipeptide, stacking to form a continued interface for target interactions. dArmRPs are optimised proteins based on their natural counterparts.

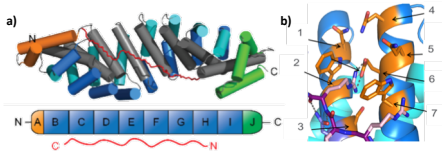


Figure 1: a) Schematic of dArmRP showing capping domains (orange/green), internal repeats with individual helices coloured (blue, turquoise, grey) with an extended peptide (red) bound antiparallel to dArmRP. b) Close up of binding pocket (two adjacent H3 helices) identifying residues targeted for mutagenesis. Images: UZH.

PRe-ART Objectives: By engineering the specificity of the dipeptide subunits of the dArmRPs, PRe-ART aims to create a set of sequence defined, pre-screened modules which may be combined on demand to form novel dArmRP peptide-binding proteins, with the desired affinity and specificity, but without any need for further screening (Figure 2).

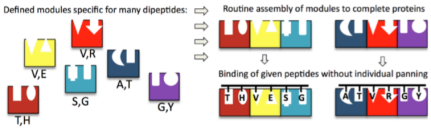


Figure 2: Schematic of PRe-ART's objective to create specific dipeptide binding subunits which are joined to form sequence defined, specific dArmRPs with a conserved binding mode. Images: UZH.

MAX Randomisation: MAX randomisation works via a process of selectional hybridisation followed by ligation and asymmetric amplification (**Figure 3**). Because MAX codons are non-degenerate, MAX randomisation can add all 20 chosen codons (or any required subset) at each saturated position.

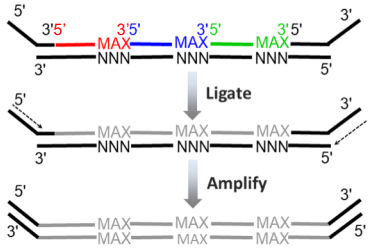


Figure 3: Schematic representation of the MAX randomisation process. A single template oligonucleotide is synthesized to be fully-degenerate at the designated, saturated codons. Meanwhile, a set of up to 20 small selection oligonucleotides are synthesised individually, for each saturated position. Each selection oligonucleotide consists of a short (typically in the order of 6bp) addressing region that is fully-complementary to the template and one MAX codon. The selection oligonucleotides are mixed as required and alongside two terminal oligonucleotides, are hybridised with the template and ligated together. The ligated strand is then selectively amplified with primers complementary to the terminal oligonucleotides, to generate a randomisation cassette (Hughes et al., 2003). Image: Aston.

Library Synthesis: an initial "general" peptide-binding library was constructed using MAX randomisation, to contain 7 saturated codons as illustrated in Figure 1b. Randomisation of these key binding residues was performed in a non-degenerative fashion to encode from 4 to 18 amino acids at the pre-selected locations.

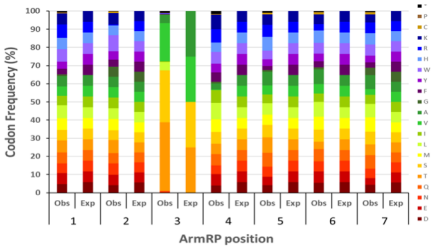


Figure 4: Illumina sequencing data showing observed and expected codon frequencies for the initial DNA library containing 7 saturated positions. Data: Aston.

Preliminary Screening Results: The initial, "general" library targeted one of the two binding pockets, within one repeat, in an otherwise-conserved dArmRP. In the "wild type" dArmRP, that pocket binds an arginine residue. Screening of this initial library against peptides containing different residues in the targeted position confirmed that altering the identity of residues 1-7 within the dArmRP can indeed lead to creation of pockets with altered specificity and high specificity, as exemplified in Figure 5.

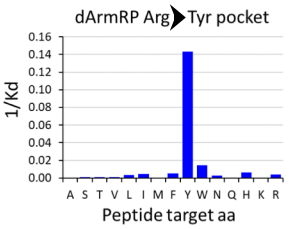


Figure 5: Example of peptide binding by an ArmRP selected from the initial library, where target specificity has been altered from Arg to Tyr at position X within the peptide NH_2-KAKXKAKRKLSF-COOH (affinity = nM^{-1}). Data: UZH.

Further Library Synthesis: MAX randomisation can be used to create both large, diverse and small, focussed libraries and has been used to synthesise an ArmRP library with theoretical diversity of 4.6 x 10^{14} unique sequences, to address two neighbouring Arginine pockets concurrently (Figure 6a). At the opposite end of the scale, multiple sets of "super-libraries" have been synthesised to contain all 140 possible variants of identified dArmRPs (Figure 6b), to ensure full exploration of 100% of the theoretical sequence space, as exemplified in Figure 5.

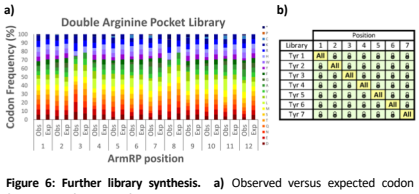


Figure 6: Further library synthesis. a) Observed versus expected codon frequencies for each of the 12 codons targeted by MAX randomisation in a "double arginine" pocket library of theoretical diversity 4.6 x 10^{14} unique sequences. **b)** Exemplar "Super library" of 140 unique sequences, created using MAX randomisation by combining 7 sets of mixtures as illustrated, where 🔒 represents fixed (locked) codon and "All" represents 20 codons to encode all 20 naturally-occurring amino acids. Images: Aston.

Experimental & Computational Validation of Preliminary Screening Results: When designing novel proteins, both affinity and specificity are of critical importance. The novel dArmRPs (Figure 5) were therefore tested against the intended target peptide and many other variant peptides, each altered at the targeted residue. Meanwhile, the determined affinities were used to train a scoring function for Atligator (ATlas-based LIGAnd binding site predictTOR). The newly developed tool Atligator (by UBT) creates distance-based interaction networks extracted from existing protein-peptide complexes. These networks are used to create new binding pockets and predict their specificity towards a certain peptide (Figure 7).

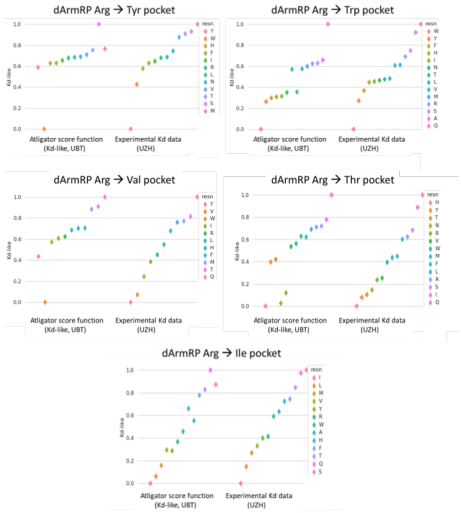


Figure 7: Atligator predictive scores compared with experimental Kd data. Each graph shows data for one altered Arg binding pocket within dArmRPs as identified in Figure 5, with each data point representing one dArmRP-peptide complex in which position X of the peptide NH_2-KAKXKAKRKLSF-COOH has been mutated to other amino acids, as indicated by "resn". Data: UBT and UZH.

Iterative Synthesis, Screening and Modelling: As illustrated in Figure 7, there is remarkable correlation between the experimentally-determined Kds and the Atligator-based predictions of peptide specificity. Atligator predictions resemble the curve shape for all binders simultaneously, with a correlation score (based on Kendall's τ) of 0.97 (ranging from -1 to 1) and ~98.5 % of Kd data pairs showing the same ranking (top-weighted).

We are now exploring whether Atligator predictions can be reversed to predict groups of likely residues in the dArmRP, tailored to bind specific residues within a target peptide. Since MAX randomisation permits any required combination of amino acids to be encoded exclusively at a given codon, we are set to refine the iterative cycle of precision library synthesis, screening and computational prediction that forms the core of the PRe-ART concept.

References:

- Varadamsetty, G, Tremmel, D, Hansen, S, Parmeggiani, F & Plückthun, A. (2012). *J. Mol Biol* **424**, 68-87.
- Hughes, MD, Nagel, DA, Santos, AF, Sutherland, AJ & Hine, AV (2003) *J. Mol. Biol.* **331**, 967-972.