ParaMAX:

Non-degenerative saturation mutagenesis applied to an alpha helical repeat protein

Ben P.G. Wagstaffe[1], Anupama Chembath[1], Mohammed Ashraf[1], Erich Michel[2], Yvonne Stark[2], Andreas Plückthun[2] and Anna V Hine[1]
[1]School of Life & Health Sciences, Aston University, Birmingham, B4 7ET, UK.
[2]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Abstract: Alpha helical repeat proteins possess alpha solenoid motifs, which comprise repeating alpha helical modules arrayed in a highly structured and uniformed manner. They, thus create a desirable interface for high affinity and specific protein/ligand interactions due to the modularity's predictability and reproducibility in target binding. Engineering of such repeat proteins is most efficient using a high-throughput, library-based approach. Saturation mutagenesis allows synthesis of massively diverse gene libraries via codon randomisation. The resulting libraries are then expressed and screened via either cloning (e.g.phage display) or purely *in vitro* display technologies. Saturation mutagenesis is itself most efficient when minimising use of the genetic code and we have previously presented non-degenerate methodologies to saturate either contiguous or non-contiguous (ProxiMAX and MAX randomisations respectively). However, ProxiMAX, which is predominantly used to engineer antibodies, is best-suited to an automated format, whilst MAX randomisation is primarily for engineering active sites. Alpha helical repeat proteins require a combination of both contiguous and non-contiguous codons to be saturated. We therefore present ParaMAX randomisation, an extension of MAX randomisation, but with the added benefit of being able to saturate contiguous codons. ParaMAX randomisation is currently in the experimental stages of saturating up to 8 contiguous codons. Results to date and accompanying quality control data will be presented.

MAX randomisation

What are MAX codons? MAX codons are a **non-degenerate** mixture of **selected** codons (exactly one codon per amino acid). MAX uses a mixture of codons that encode all 20 amino acids, or any lesser set of amino acids selected. The DNA sequence of the codons is unimportant and can be selected for maximal expression in any organism.

2nd Base

1st Base	A	C	G	T	3rd Base
A	Lys	Thr	Arg	Ile	A
	Asn	**Thr**	**Ser**	Ile	C
	Lys	Thr	Arg	Met	G
	Asn	Thr	Ser	**Ile**	T
C	Gln	Pro	Arg	Leu	A
	His	Pro	**Arg**	Leu	C
	Gln	**Pro**	Arg	**Leu**	G
	His	Pro	Arg	Leu	T
G	**Glu**	Ala	Gly	Val	A
	Asp	Ala	**Gly**	Val	C
	Glu	**Ala**	Gly	**Val**	G
	Asp	Ala	Gly	Val	T
T	*	Ser	*	Leu	A
	Tyr	Ser	Cys	Phe	C
	*	Ser	**Trp**	Leu	G
	Tyr	Ser	Cys	**Phe**	T

Figure 1: Exemplar MAX codons. Codons chosen to encode all amino **acids except Cys and Met, selected according to** *E. coli* codon preference, for MAXimal expression. Note that the NNN would be required as a conventional degenerate codon to include all in this selection.

MAX methodology: MAX randomisation works via a process of selectional hybridisation followed by ligation and asymmetric amplification (**Figure 2**). Because MAX codons are non-degenerate, MAX randomisation can add all 20 chosen codons (or any required subset) at each saturated position.

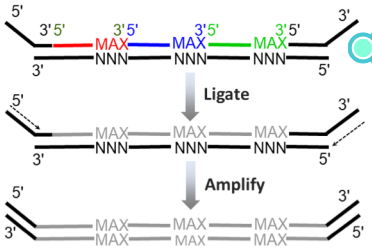

Figure 2: Schematic representation of the MAX randomisation process. A single template oligonucleotide is synthesized to be fully-degenerate at the designated, saturated codons. Meanwhile, a set of up to 20 small selection oligonucleotides are synthesised individually, for each saturated position. Each 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).

MAX randomisation saturates separated codons, such as those encoding residues on the surface of α-helices. PRe-ART, employs MAX randomisation to saturate peptide-binding residues within Armadillo Repeat Proteins (ArmRPs). ArmRPs are naturally-occurring eukaryotic proteins characterised by a right-handed superhelix formed by 4–12 stacked tandem armadillo repeat motifs. Each repeat consists of approximately 42 amino acids folded into three α-helices in a triangular assembly.

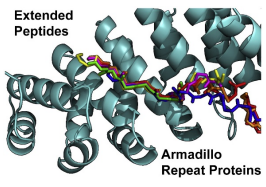

Figure 3: Peptides in extended conformation bound by a designed Armadillo Repeat Protein (Varadamsetty et al., 2012).

ParaMAX randomisation

ParaMAX randomisation saturates contiguous codons, using adapted MAX randomisation methodology. ParaMAX randomisation has also been used to saturate peptide-binding residues within ArmRPs.

ParaMAX methodology: ParaMAX randomisation employs selectional hybridisation and rounds of ligations, asymmetric amplifications and restrictions, to generate regions of randomised contiguous codons within ArmRPs (**Figure 4**). Being an adaptation of MAX randomisation means the perks of non-degeneracy still apply allowing all 20 chosen codons (or any required subset) to be used to saturate a position.

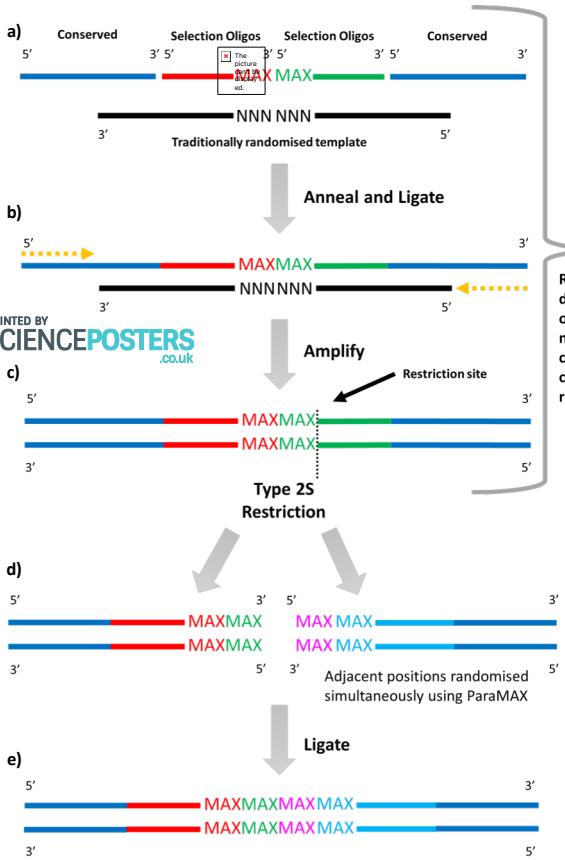

Figure 4: Schematic representation of ParaMAX randomisation. a) Single template oligonucleotide is synthesised to be fully-degenerate at the designated saturated codons. Meanwhile, a set of up to 20 small selection oligonucleotides are synthesised individually, for each saturated position. Each selection 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. b) The ligated strand is then selectively amplified with primers complementary to the terminal oligonucleotides, to generate a randomisation cassette containing two positions of randomisation. c) Individual randomisation cassettes undergo Type IIs restriction, removing a terminal sequences and exposing two MAX codons, producing two cleaved randomisation cassettes ready for d) blunt end ligation. e) Formation of a randomisation cassette containing four randomised contiguous codons. Two adjacent cassettes each containing four randomised positions can then undergo restriction and ligation to form a cassette of 8 randomised positions (not shown).

Cassette Construction

ParaMAX randomisation was used to randomise 8 contiguous codons using amino acid subsets containing up to 19 different amino acids, by creating fragments of the cassette (couplets) containing two positons of randomisation and joining them together to form larger randomisation cassettes.

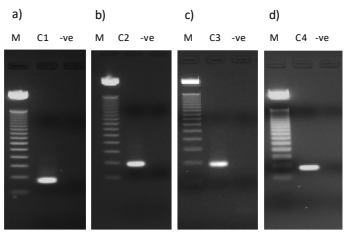

Figure 5: Synthesis of MAX couplets. MAX couplets are two adjacent contiguous randomised positions generated via MAX randomisation and PCR. a) Construction of couplet 1 by PCR (84bp). Lanes: M-50bp marker, couplet 1 (C1), -ve control b) Construction of couplet 2 by PCR (92bp). Lanes: (see a)). c) Construction of couplet 3 by PCR (92bp). Lanes: (see a)) d) Construction of couplet 4 by PCR (84bp). Lanes: (see a)).

Repeated depending on the number of contiguous codons to be randomised

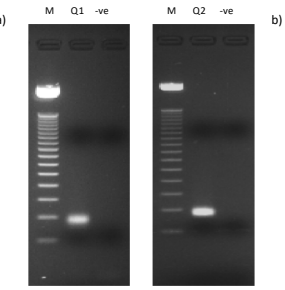

Figure 6: Synthesis of MAX quads. MAX quads are 4 randomised contiguous positions formed from the digestion and ligation of two adjacent MAX couplets. a) Construction of MAX quad 1 (Q1) (90bp). Lanes: M- 50bp marker, Q1, -ve control. b) Construction of MAX quad 2 (Q2) (90bp). Lanes: (see a)).

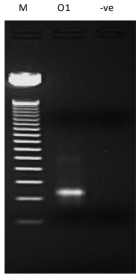

Figure 7: Synthesis of MAX octo. MAX octo (O1) is a cassette containing 8 randomised contiguous positions formed from the digestion and ligation of two adjacent MAX quads. Lanes: M-50bp marker, O1 (120bp), -ve control.

References:
- Ashraf, M, Frigotto, L, Smith, ME, Patel, S, Hughes, MD, Poole, AJ, Hebaishi, HRM, Ullman, CG & Hine, AV (2013) *Biochem. Soc. Trans.* **41**, 1189–1194.
- Hughes, MD, Nagel, DA, Santos, AF, Sutherland, AJ & Hine, AV (2003) *J. Mol. Biol.* **331**, 967-972.
- Varadamsetty, G, Tremmel, D, Hansen, S, Parmeggiani, F & Plückthun, A. (2012). *J. Mol Biol* **424**, 68-87.