Transposon-based Approaches for Generating Novel Molecular Diversity during

Directed Evolution

D. Dafydd Jones[¶], James A.J. Arpino[¶], Amy J. Baldwin[¶] & Matthew C. Edmundson[¶]

[¶] School of Biosciences, Cardiff University, Cardiff, UK.

Corresponding author: Dafydd Jones (jonesdd@cardiff.ac.uk)

Running Title: Transposon-based random mutagenesis.

i. Summary

This chapter introduces a set of transposon-based methods that were developed to sample trinucleotide deletion, trinucleotide replacement and domain insertion. Each approach has a common initial step that utilises an engineered version of the Mu transposon called MuDel. The inherent low sequence specificity of MuDel results in its random insertion into target DNA during *in vitro* transposition. Removal of the transposon using a type IIS restriction endonuclease generates blunt end random breaks at a frequency of one per target gene and the concomitant loss of 3 bp. Self-ligation or insertion of another DNA cassette results in the sampling of trinucleotide deletion or trinucleotide substitution/domain insertion, respectively.

ii. Key words: Directed evolution, molecular diversity, Mu transposon, trinucleotide deletion, trinucleotide exchange, domain insertion, expanded genetic code.

1. Introduction

Generating molecular diversity lies at the heart of directed evolution (*1-4*). With an ever-expanding directed evolution toolbox, different mutational approaches are required. These will access sequence space beyond that sampled by simple gene point mutations that ultimately restrict sequence, structural and functional diversity (*1, 5, 6*). Sampling randomly introduced point mutations severely limits the amino acid range that can be tested at a particular residue position as statistically only one base per codon will ever be mutated (*4*). Furthermore, many traditional directed evolution approaches ignore mutations that alter the protein backbone. Such mutations, termed indels (insertion/deletion), are commonly observed amongst natural protein homologues (*7, 8*) and range from single amino acid deletions to the insertion of extended protein segments or domains. Mutation of the protein backbone changes the structure of a protein in a manner distinct to that of side-chain substitutions, sampling new conformational and hence functional space (*9*). Until recently, generating indel mutations for directed evolution libraries was a major challenge for the field. Rational site-directed approaches can be used but generally require detailed information regarding structure and function to guide mutagenesis. Even armed with such information, the local and global structural consequences of rational indel mutagenesis can be difficult to predict.

To address these challenges, a set of transposon-based approaches were developed to sample trinucleotide deletion (*10, 11*), trinucleotide replacement (*12, 13*) and domain insertion (*14-16*) (**Fig. 1**). Each of these approaches relies on a common first step that ultimately defines genetic diversity sampled across a target gene and thus protein backbone: transposition with an engineered version of the Mu transposon, termed MuDel (*10*).

The Mu transposon, or more accurately mini-Mu (*17*), is a DNA construct that can be inserted randomly into a target DNA sequence with high efficiency and accuracy but with low target site preference. Transposition is performed *in vitro* using the commercially available MuA transposase. To implement the various mutagenesis strategies mini-Mu is modified

symmetrically close to both termini to introduce the recognition sequence for MlyI, a type IIS restriction endonuclease that cuts outside its recognition sequence. Placement of the MlyI sites within MuDel, in combination with the transposon insertion mechanism, results in the removal of 3 bp at random positions within a target gene upon MlyI digestion (**Fig. 2**). Self-ligation of the resulting DNA generates a library of proteins with single amino acids deleted at random positions.

Deletion of trinucleotide sequences opens up the possibility of their replacement with a new trinucleotide sequence and is the basis for trinucleotide exchange (TriNEx) (*(13)*; **Fig. 2**). This approach uses the expanded amino acid range of site-directed saturation mutagenesis in combination with the broader sampling capacity of whole-gene mutagenesis. To implement TriNEx an additional DNA cassette termed SubSeq is used, which donates a new trinucleotide sequence back to the target gene. MlyI is used to achieve trinucleotide exchange by means of the mechanism outlined in **Fig. 2**. The donating sequence in SubSeq is determined by the user and can be randomized *(13)* or defined *(12)*. The latter has been used to perform non-natural amino acid whole gene directed evolution.

Domain insertion *(18-20)* is a powerful approach to generate new protein scaffolds in which the functions of normally disparate proteins can be coupled generating new components for use in, for example, biosensing, synthetic biology and bionanotechnology. Domain insertion is achieved using the MuDel approach by replacing SubSeq with a DNA cassette encoding a protein segment of choice *(14-16)* (**Fig. 1**). As well as sampling a variety of insertion positions across a target gene, the nature of the DNA cassette itself can form the basis of a mini-library (**Fig. 3**). The ends of the DNA cassette can be modified to: (1) overcome reading frame artefacts due to cross codon deletion events during MuDel excision so increasing the diversity of insertion positions sampled; (2) vary the nature of the linking sequences between the host protein and the introduced protein segment/domain.

2. Materials

2.1. Transposition and mutagenesis components

1. pNOM *(10)* (*see* **Note 1**).

2. pNOM-XP3 *(15)* (*see* **Notes 1-3**).

3. Transposons. The key sequence details of the MuDel transposon are shown in **Fig. 2** (*see* **Notes 4 - 5**).

4. MuA transposase from Finnzymes (now ThermoScientific).

5. MuA transposase reaction buffer: 25 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 110 mM NaCl, 0.05% Triton and 10% (v/v) glycerol. This is supplied as 5x reaction buffer.

6. SubSeq DNA cassette. The SubSeq DNA cassette is required for TriNEx; the essential sequence features are shown in **Fig. 2** (*see* **Note 6**).

2.2. General materials

1. Deionised or MilliQ™ (MQ) water were used throughout. Water was sterilised by autoclave. Ampicillin, kanamycin or chloramphenicol (in ethanol) were made as a 100, 25 or 34 mg/ml stock solution respectively and filter sterilized with a 0.22 μm filter unit with the exception of chloramphenicol. The stock solutions were used to supplement bacterial growth media by dilution to an appropriate working concentration.

2. *E. coli* cell strains (electro- or chemically competent) with high transformation efficiencies (>10^9 cfu per μg pUC18 DNA) are required. DH5α and NovaBlue (Merck KGaA) strains are used for library construction, and various versions of BL21 (DE3) including BL21-Gold (DE3) (Stratagene) and Tuner™ (DE3) (Novagen) are used for protein expression.

3. Luria Bertani Broth (LB) medium: Per litre, mix 5 g yeast extract, 10 g tryptone and 10 g NaCl. Autoclave at 121°C for 25 min and let cool to ambient temperature.

4. LB-agar medium: Per litre, mix 5 g yeast extract, 10 g tryptone, 10 g NaCl and 15 g agar. Autoclave at 121°C for 25 min and let cool to 50°C in a warm water bath. Add appropriate

antibiotic selection marker and gently shake to homogenize the solution, then pour plates and

let agar solidify at room temperature. Plates can be stored at 4˚C for up to 4 weeks

5. Molecular biology grade agarose: gels were prepared by boiling 0.6-2.0 % (w/v) in 1X

TAE buffer (40 mM Tris-acetate, 1 mM EDTA) buffer followed by cooling and addition of

0.5 μg/mL ethidium bromide.

6. Restriction endonucleases (with reaction buffers) e.g. MlyI, NdeI and XhoI.

7. 100x BSA Stock solution: 10 mg/ml in water.

8. Quick T4 DNA ligase system (NE Biolabs)

9. T4 polynucleotide kinase

10. 10x T4 polynucleotide kinase reaction buffer (70 mM Tris-HCl, 10 mM MgCl2, 5 mM

DTT pH 7.6.

11. 10x ATP solution: 10mM ATP in water.

12. APex™ heat-labile alkaline phosphatase (1 unit/μg DNA) (Epicentre® Biotechnologies)

13. Oligonucleotide primers: 100 μM stock solutions.

14. GoTaq® DNA polymerase (Promega) system.

15. Phusion® High-Fidelity DNA polymerase (NE Biolabs) system.

16. PCR, gel extraction, MinElute, plasmid Miniprep and plasmid Midi-prep purification kits

(QIAGEN), used according to the manufacture's guidelines. DNA concentrations and

260/280 nm ratios after purification were estimated using a NanoDrop® ND-1000 UV-Vis

spectrophotometer.

3. Methods

3.1 MuDel insertion library construction

Prior to transposition, clone the target gene within the multiple cloning site of pNOM or a

pNOM-based vector (*see* **Notes 7 & 8**).

1. Transposition with MuA transposase (ThermoScientific). Add target plasmid DNA (100 ng equivalent to pNOM) and MuDel (20-100 ng) to MuA transposase reaction buffer containing 0.22 µg of MuA transposase in a final volume of 20 µl. Leave reactions at 30°C for 3 h and then heat inactivate at 75°C for 10 min.

2. Transform chemically competent *E. coli* cells with the equivalent of 5-20 ng of plasmid DNA from the deactivated transposition reaction (*see* **Note 9**).

3. Plate a small proportion of transformed cells (~1%) on LB agar plates supplemented with 20 µg/ml chloramphenicol to assess transposition efficiency and calculate the total number of target plasmid molecules containing MuDel. After overnight incubation at 37°C, extrapolate the number of colonies back to the number present in the total transformation mixture (*see* **Notes 10 & 11**).

4. Use the remaining portion of transformed cells to inoculate 1 L LB broth supplemented with 20 µg/ml chloramphenicol and incubate in a shaker at 37°C for 12-16 h. Purify the pooled MuDel insertion plasmid DNA library from liquid culture using Qiagen Plasmid Midi Kit and store at -20°C (*see* **Note 12**).

5. Isolate MuDel inserted within the target gene, and remove any insertions that have occurred within the plasmid, by digesting the pooled plasmid DNA with the original restriction enzymes used to clone the target gene into pNOM or its derivatives (e.g. NdeI and XhoI). Separate the resulting fragments by agarose gel electrophoresis. Restriction digestion will produce 4 DNA fragments: (A) plasmid backbone with MuDel; (B) plasmid backbone without MuDel; (C) target gene with MuDel; (D) target gene without MuDel.

6. Isolate the two bands corresponding to fragments B and C and purify the DNA from the gel matrix using a QIAquick gel extraction kit (QIAGEN)

7. Ligate the two purified fragments using the Quick Ligation™ Kit (NE Biolabs) and transform into *E. coli*. As the two DNA fragments originated from the same restriction

digestion they may be presumed to be equimolar, so use appropriate volumes of each fragment in the ligation reaction to sample from 1:1 to 3:1 insert to vector ratios.

8. To calculate the total number of clones containing MuDel that comprise the library, plate a small proportion of transformed cells (~1%) on LB agar plates supplemented with 20 μg/ml chloramphenicol. After overnight incubation at 37°C, extrapolate the number of colonies back to the number present in the total transformation mixture (*see* **Note 13**).

9. Use the remaining portion of transformed cells to inoculate 1L LB broth supplemented with 20 μg/ml chloramphenicol and incubate in a shaker at 37°C for 12-16 h. Purify the pooled MuDel target gene insertion plasmid DNA library from liquid culture and store at -20°C (*see* **Notes 14 & 15**).

3.2 Construction of trinucleotide deletion library (TND)

The initial MuDel library constructed as outlined in **Subheading 3.1** is the basis for TND library generation. At this stage, all the clones present in the MuDel library are assumed to have MuDel inserted within the target gene.

1. Remove MuDel from transposon-insert library by digestion with MlyI (NE Biolabs). Perform digestion with 1 U of enzyme per μg DNA in NEBuffer 4 supplemented with 0.1 mg/ml BSA in a total reaction volume of 50 μl. Incubate the reaction at 37°C for 1 h/μg DNA (*see* **Note 16**).

2. Separate the resulting DNA fragments by agarose gel electrophoresis. Extract the DNA fragment equivalent to the size of the original starting plasmid (pNOM or derivative) plus the target gene using a QIAquick gel extraction kit (QIAGEN). Discard the MuDel band.

3. Re-circularise the purified linear library DNA (50 ng) by intramolecular ligation with the Quick Ligation™ Kit (NE Biolabs). Incubate ligation reactions at 25°C for 20 min (*see* **Note 17**).

4. Use 1 μl of the ligation reaction mixture to transform *E. coli* electrocompetent or chemically competent cells. Grow transformed cells on LB agar plates supplemented with 100 μg/ml ampicillin (*see* **Note 18**).

5. The library can now be screened to identify protein variants with desired properties.

3.3. Construction of a Trinucleotide exchange (TriNEx) library

The initial MuDel library constructed as outlined in **3.1** will be the basis for generating the TriNEx library. At this stage, all the variants present will be deemed to have MuDel inserted within the target gene.

1. Generate the SubSeq DNA cassette (**Fig. 2**) by PCR as outlined in **2.1.4**. Phosphorylate the SubSeq PCR product using 10 U T4 polynucleotide kinase (NE Biolabs) in 1 x T4 polynucleotide kinase reaction buffer (NE Biolabs) and 1 mM ATP. Incubate the reaction at 37°C for 1 h then heat inactivate at 65°C for 20 min. Purify the resulting DNA using the QIAquick PCR purification kit.

2. Remove MuDel from the pooled plasmid DNA by digestion with MlyI as outlined in **Subheading 3.2.1**. Dephosphorylate digested DNA by adding APex heat-labile alkaline phosphatase (1 unit/μg DNA) (Epicentre Biotechnologies) to the restriction endonuclease reaction and incubate at 37°C for the duration of the restriction digest. Heat inactivate the phosphatase and MlyI by incubating the reaction mixture at 70°C for 5-10 min.

3. Separate the resulting DNA digest by agarose gel electrophoresis and extract the fragment equivalent to the size of the original starting plasmid (pNOM or derivative) plus the target gene from the gel.

4. Ligate the SubSeq DNA cassette into linear plasmid DNA with T4 DNA ligase for 30 min at room temperature. In a total volume of 20 μl, use a molar ratio of 1:3 (pNOM:SubSeq) with a total of 200-250ng DNA in each reaction (*see* **Note 19**).

5. Electroporate 2 μl of the ligation reaction (1/10th) into *E. coli* DH5α cells. Calculate the number of transformed cells containing SubSeq incorporated within the host plasmid, by plating the equivalent of 2-3% of the electroporated cells on LB agar plates supplemented with 25 μg/ml kanamycin. Incubate the plates overnight at 37°C and then count the number of colonies observed to calculate ligation and transformation efficiency. The observed transformation efficiency is expected to be 3-4 x 10^4 cfu/μg DNA (*see* **Note 20**).

6. Add the remaining portion of transformed cells to 50 ml of LB broth containing 200 μg/ml kanamycin and incubate in a shaking incubator at 37°C for 16 h.

7. Isolate plasmid DNA from the culture using a QIAprep spin miniprep kit (QIAGEN) and store at -20°C (*see* **Note 21**).

8. Digest the purified plasmid DNA (2 μg) with MlyI for 1 h at 37°C (**Subheading 3.2.1**) to remove the SubSeq section, with the terminal trinucleotide sequence of SubSeq donated to the target gene. Agarose gel electrophoresis was used to separate excised SubSeq from linearised DNA corresponding to the host plasmid containing the target gene. Extract the latter from the gel as outlined in **Subheading 3.2.2**.

9. Re-circularise the linear plasmid DNA (~50-100 ng) using T4 DNA ligase as outlined in **Subheading 3.2.3**.

10. Transform the resulting library as outlined in **Subheading 3.2.4**.

11. The library can now be screened to identify protein variants with the required properties.

3.4. Construction of a Domain Insertion library

The MuDel library constructed as outlined in **Subheading 3.1** is the basis for generating the Domain Insert library. At this stage, all the variants present are assumed to have MuDel inserted within the target gene.

1. Construct the DNA cassette comprising the protein segment/domain insert by PCR using the Phusion High Fidelity DNA polymerase system (NE Biolabs). The general strategy for constructing the domain insert is outlined in **Fig. 3**. Design terminal oligonucleotides to incorporate linker sequence and open reading frame (ORF) requirements. With regards to the latter, design different ORF versions of the DNA insert cassette to take into account cross-codon trinucleotide deletions inherent on MuDel excision. This increases DNA cassette insertion position diversity within the target gene (*see* **Note 22**).

2. Phosphorylate the different ORF DNA insertion cassettes as outlined in **Subheading 3.3.1**.

3. Remove MuDel from the pooled plasmid DNA containing the target gene and dephosphorylate as outlined in **Subheading 3.3.2.**

4. Ligate the DNA insert cassette into linear plasmid DNA using T4 DNA ligase for 30 min at room temperature as described in **Subheading 3.3.4**. Perform individual ligations for each ORF version of the DNA insert cassette.

5. Transform the ligation reactions into *E. coli* cells by electroporation. Plate transformed cells on LB agar supplemented with 100 μg/ml ampicillin and incubate at 37°C for at least 16 h.

6. Assess DNA cassette insertion efficiency by screening colonies by colony PCR using the GoTaq DNA polymerase (Promega) system and primers flanking the target gene. Positive clones will produce an amplified DNA product corresponding to the size of the original target gene plus the DNA cassette insert.

7. Further analyze positive clones by PCR to assess frequency of correct DNA cassette insertion orientation and diversity of the insertion position. Use a primer pair comprising a target gene flanking primer and a DNA cassette insert construction primer from **Subheading 3.4.1** (*see* **Note 23**).

8. The libraries are now ready to be screened to identify domain/segment insertion variants with desired properties.

4. Notes

1. Plasmids pNOM [10] and pNOM-XP3 [15] are bespoke plasmids available from the authors for use with MuDel and host the target gene. pNOM is the original vector and is based on pUC18 backbone with pNOM-XP3 being an inducible target gene over-expression version of pNOM. pNOM (**p**lasmid **NO M**lyI) was constructed to remove all MlyI sites present in the original pUC18 plasmid (-1 to 1979 bp region) and to incorporate a useful multiple cloning site with the common restrictions sites NdeI and XhoI at either end. Removal of MlyI sites notably changed 2 bp in the origin of replication and introduced a silent mutation in the *bla* ampicillin resistance gene; neither appeared to be detrimental to the use of the plasmid. pNOM-XP3 is a derivative of pNOM containing the inducible T7 promoter upstream of the target gene. The original T7 promoter was mutated to remove an MlyI recognition sequence. The NdeI recognition sequence is present downstream from the T7 promoter to aid in-frame initiation codon cloning.

2. Versions of pNOM-XP3 are available with NcoI in place of NdeI.

3. Target protein expression can be leaky when using pNOM-XP3. We assume this is due to the mutated T7 promoter.

4. MuDel was generated from the commercially available Cam^R containing Mu Entranceposon present within the pEntranceposon plasmid (http://www.thermoscientificbio.com/mutagenesis/entranceposons/). MuDel was recloned back into the pEntranceposon plasmid within the BglII site. MuDel was removed from the plasmid for use in the mutagenesis procedure using BglII and separated from the plasmid backbone by agarose gel electrophoresis followed by gel excision and purification using the Qiagen QIAquick gel extraction kit.

5. We have tried using PCR amplified forms of MuDel but have found that BglII digested transposon from plasmid gives by far the best results.

6. The SubSeq^{NNN} DNA cassette was constructed by PCR using the Phusion™ high fidelity DNA polymerase system (NE Biolabs). The Kan^R gene region of the Entranceposon™ (kan^R-3) (Finnzymes) was amplified as two separate fragments to remove an MlyI site in the middle of the gene and introduce MlyI sites at strategic positions towards the termini of the cassette. The nature of the SubSeq used was determined by a terminal primer 5'[Phos]-**NNN**GAAAG<u>GACTC</u>AGTGTGTCGGCGGCCGCGGATCCT-3' (MlyI site underlined) where the 5' end 3 bp sequence (NNN) that is ultimately donated to the target gene can be determined by the user.

7. It is important that any MlyI sites are removed in the target gene. This can be achieved through the use of silent mutations.

8. Although manufacturers state *in vitro* transposition into linear DNA with MuA is feasible, we have had no success with this approach so have always used circular plasmid DNA as the target for MuDel insertion.

9. Chemically competent cells with high transformation efficiencies (>10^9 cfu/µg DNA) should be used at this stage due to the relatively high concentrations of DNA.

10. We recommend that the calculated number of colonies should be ~2-3 times the size of the target plasmid (plasmid backbone plus target gene) to ensure maximal transposon insertion coverage.

11. Colony PCR (using primers that flank the target gene) can be performed on randomly selected colonies as an alternative way to calculate the percentage of the library containing MuDel insertions within the target gene. Analysis of amplified products by agarose gel electrophoresis should give two amplified products corresponding to: (i) the target gene alone; (ii) the target gene plus MuDel (1307 bp). The percentage of analysed colonies with the larger size band is representative of the percentage of variants in the library with MuDel inserted within the target gene region of the plasmid.

12. As an alternative option, the remaining cells may be spread onto 150 mm diameter LB agar plates supplemented with 20 µg/ml chloramphenicol. Collecting cells from single colonies of roughly equal size straight from plates can help reduce bias which may result from different growth rates in liquid media. Ensure single colony resolution by pipetting 150-200 µl per plate and dispersing cells with a glass spreader. Following overnight incubation at 37°C, isolate the MuDel library by adding LB broth (5 ml) to each plate and gently shaking the plates for 10-15 mins to suspend cells. Collect the LB broth and purify the MuDel insertion plasmid DNA library from liquid culture using Qiagen Plasmid Midi Kit and store at -20°C.

13. We recommend that the calculated number of colonies should be >3 times the size of the target plasmid (plasmid backbone plus target gene) to ensure maximal transposon insertion coverage.

14. An additional step may be included at this point to confirm MuDel has inserted randomly and uniformly across the length of the target gene. The pool of plasmid DNA can be digested with MlyI and one of the restriction endonucleases originally used to clone the target gene into the host plasmid (e.g. NdeI or XhoI). Analysis of the digestion mixture by agarose gel electrophoresis should produce a distinctive band at ~1310 bp corresponding to excised MuDel and two smears: one ranging from the size of the host plasmid (*e.g.* 2146 bp for pNOM-XP3) to the size of the original plasmid plus the target gene; another smear ranging from the 0 bp to the size of the target gene. We generally use XhoI to clone our target gene into pNOM or derivative. Thus we normally analyse MuDel insertion diversity with XhoI and MlyI. XhoI serves to linearise the plasmid DNA; the resulting variable length DNA fragments are generated by MlyI digestion and depend on the placement of MuDel within the target.

15. An alternate selection strategy to **Subheading 3.1 items 5-9** can be used if a negative selection or screen is feasible. All the transformed *E. coli* cells can be plated on LB agar supplemented with chloramphenicol to select for clones containing MuDel within the host

plasmid which are subsequently screened for target protein activity. Those colonies that **do** display target protein activity are deemed not to have a disrupted gene due to MuDel insertion. Those colonies that **do not** display target protein activity are deemed to contain MuDel inserted within the gene due to disruption of the protein coding region.

16. It is important to use relatively fresh MlyI at this stage as in some rare instances we have observed downstream problems when using older stocks of the restriction endonuclease.

17. DNA ligated using the Quick Ligation™ Kit protocol must be column purified prior to transformation by electroporation. PEG in the Quick Ligation™ buffer inhibits transformation by electroporation reducing transformation efficiency and in some cases cause arcing in the electroporation cuvettes. We have also successfully used standard T4 ligase approaches with extended incubation times.

18. If using the pNOM-XP3 plasmid system, we recommend using *E. coli* strains compatible with T7 promoter systems such as BL21-Gold (DE3) or Tuner™ (DE3).

19. We have used different T4 DNA ligase systems, including the Quick Ligation™ kit by NE Biolabs and the Fast-Link™ ligation kits by Epicentre. All have been successful.

20. It is important to maximise the efficiency of ligation and the resulting transformation so as to maintain high molecular diversity. The SubSeq-plasmid DNA is a blunt end ligation so efficiency may be lower compared to ligation involving base overhangs.

21. At this stage it is wise to isolate plasmid DNA using a "miniprep" approach to prevent total loss of the library DNA should large scale "maxi or midiprep" purification fail. Any unused culture can be divided into aliquots, frozen and stored until required.

22. To facilitate downstream selection for DNA cassette insertion within the target gene, a Kan^R (Kanamycin) gene can be inserted within the DNA cassette between convenient restriction sites. We have used this approach successfully *(14)* although it does add extra steps to the process.

23. An additional step using the DNA cassette insert with Kan^R placed within the coding sequence (**Subheading 3.4.1**) has also been used to select solely for clones containing the DNA cassette insert. The additional step involves transforming *E. coli* with the ligation products of **Subheading 3.4.4**. The transformed cells were used to inoculate 50 ml LB broth supplemented with 25 μg/ml kanamycin and incubated at 37°C in a shaking incubator for ~16 h. Plasmid DNA was isolated from the cultures and the Kan^R cassette removed by digestion with the appropriate restriction endonuclease. Restriction digests were separated by agarose gel electrophoresis and DNA bands corresponding to linear library plasmid DNA isolated and purified by gel extraction. Purified linear plasmid library DNA was recircularised by intramolecular ligation (**Subheading 3.2.3**) producing the domain/segment insert libraries, which was then used to transform *E. coli* Tuner^{TM} (DE3) cells by electroporation. The transformed cells were grown on LB agar supplemented with 100 μg/ml ampicillin and 150 μM IPTG (if using the pNOM-XP3 plasmid), and incubated for 24 h at 37°C.

Acknowledgements

References

1.   Cobb RE, Si T and Zhao H (2012) Directed evolution: an evolving and enabling synthetic biology tool. Curr Opin Chem Biol 16:285-91.

2.   Jackel C, Kast P and Hilvert D (2008) Protein design by directed evolution. Annu Rev Biophys 37:153-73.

3.   Lutz S and Patrick WM (2004) Novel methods for directed evolution of enzymes: quality, not quantity. Curr Opin Biotechnol 15:291-7.

4.   Neylon C (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Res 32:1448-59.

5.   Brustad EM and Arnold FH (2011) Optimizing non-natural protein function with directed evolution. Curr Opin Chem Biol 15:201-10.

6.   Koide S (2009) Generation of new protein functions by nonhomologous combinations and rearrangements of domains and modules. Curr Opin Biotechnol 20:398-404.

7.   Chothia C, Gough J, Vogel C and Teichmann SA (2003) Evolution of the protein repertoire. Science 300:1701-3.

8.   Taylor MS, Ponting CP and Copley RR (2004) Occurrence and consequences of coding sequence insertions and deletions in Mammalian genomes. Genome Res 14:555-66.

9.   Shortle D and Sondek J (1995) The emerging role of insertions and deletions in protein engineering. Curr Opin Biotechnol 6:387-93.

10.   Jones DD (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:e80.

11.   Simm AM, Baldwin AJ, Busse K and Jones DD (2007) Investigating protein structural plasticity by surveying the consequence of an amino acid deletion from TEM-1 β-lactamase. FEBS Lett 581:3904-8.

12.   Baldwin AJ, Arpino JA, Edwards WR, Tippmann EM and Jones DD (2009) Expanded chemical diversity sampling through whole protein evolution. Mol Biosyst 5:764-6.

13.   Baldwin AJ, Busse K, Simm AM and Jones DD (2008) Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). Nucleic Acids Res 36:e77.

14.   Arpino JA, Czapinska H, Piasecka A, Edwards WR, Barker P, Gajda MJ, Bochtler M and Jones DD (2012) Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold. J Am Chem Soc 134:13632-40.

15.   Edwards WR, Busse K, Allemann RK and Jones DD (2008) Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. Nucleic Acids Res 36:e78. 10.1093/nar/gkn363

16.   Edwards WR, Williams AJ, Morris JL, Baldwin AJ, Allemann RK and Jones DD (2010) Regulation of β-lactamase activity by remote binding of haem: Functional coupling of unrelated proteins through domain insertion. Biochemistry.

17.   Haapa S, Taira S, Heikkinen E and Savilahti H (1999) An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. Nucleic Acids Res 27:2777-84.

18.   Fastrez J (2009) Engineering allosteric regulation into biological catalysts. Chembiochem 10:2824-35.

19.   Ferraz RM, Vera A, Aris A and Villaverde A (2006) Insertional protein engineering for analytical molecular sensing. Microb Cell Fact 5:15.

20.   Ostermeier M (2005) Engineering allosteric protein switches by domain insertion. Protein Eng Des Sel 18:359-64.
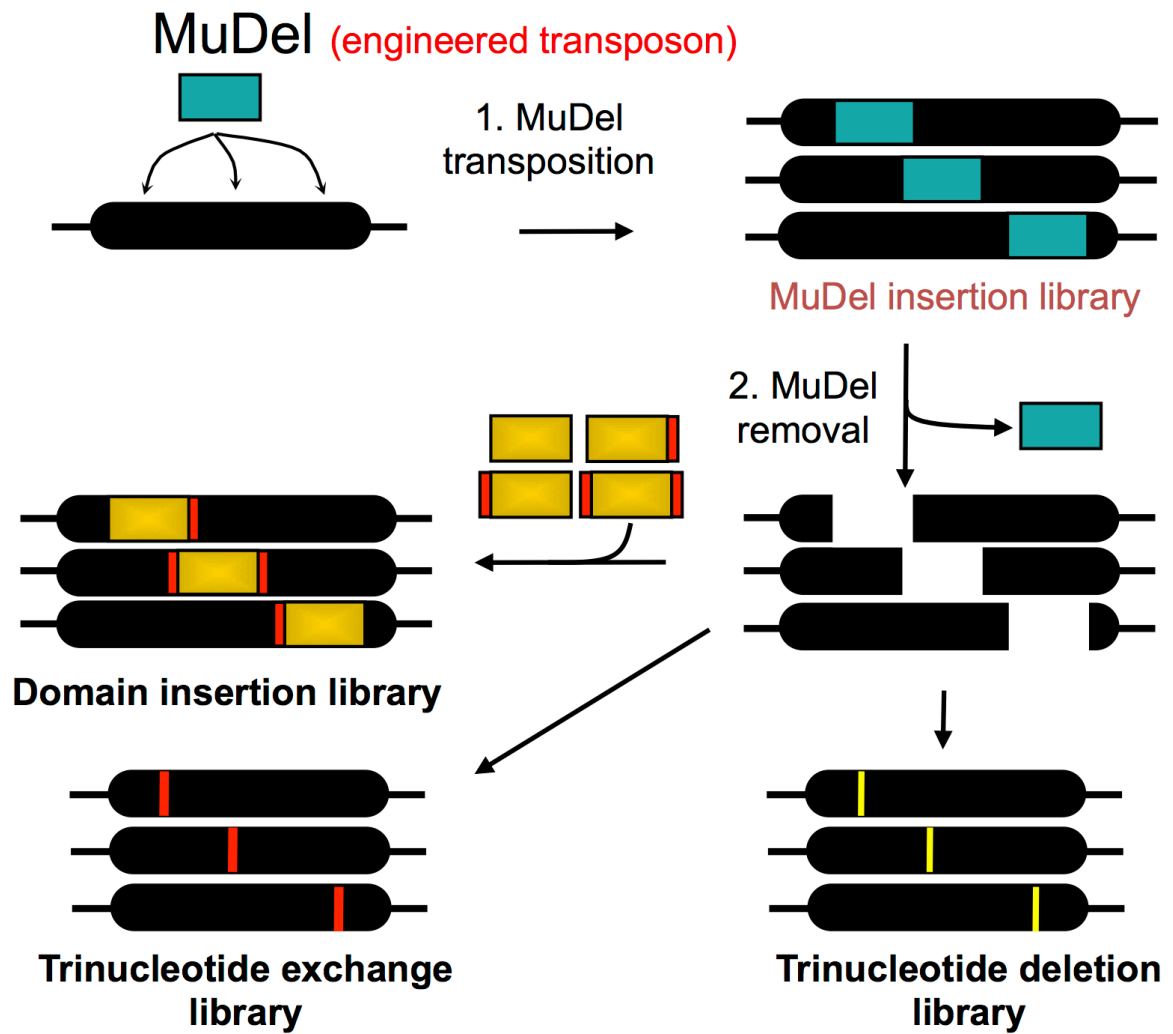
Fig. 1. Simple outline of the transposon-based random mutagenesis approach. The common first step involves the generation of a MuDel whole gene insertion library followed by its removal to generate random breaks in the target DNA. From this point, the methods diverge so that different mutational events can be sampled, namely trinucleotide deletion, trinucleotide exchange and domain insertion.
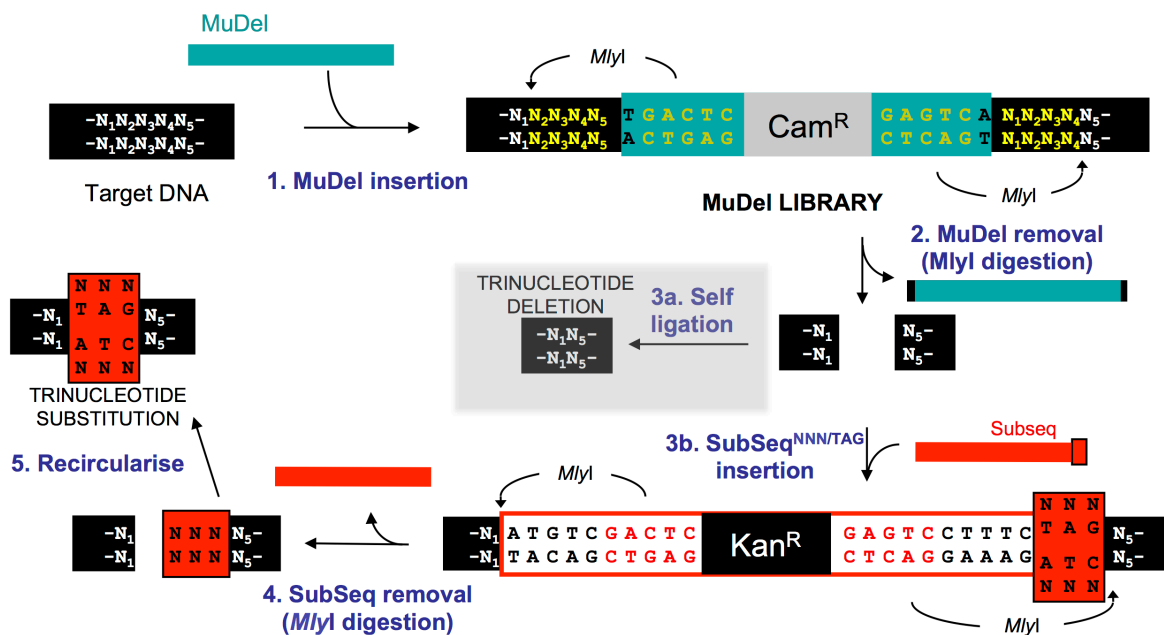
Fig. 2. Mechanism of MuDel-based mutagenesis. Step 1. MuDel Insertion. Two MlyI

recognition sites (5'GAGTC(N)$_5$↓) are placed 1 bp away from the site of transposon insertion.

Insertion of MuDel results in the duplication of 5 bp ($N_1N_2N_3N_4N_5$) of the target gene at the

insertion point. Step 2. Digestion with MlyI removes MuDel together with 8 bp of the target

DNA (4 bp at each end), which equates to removal of a contiguous 3 bp sequence from the

starting target gene ($N_2N_3N_4$). Step 3a. TND. Self-ligation at this stage results in the

reconstitution of the full-length gene but with 3 bp ($N_2N_3N_4$) removed. Step 3b. TriNEx.

SubSeq is ligated into the gap vacated by MuDel. SubSeq contains two MlyI recognition sites

strategically placed towards the ends of the cassette. One site is located so that MlyI will cut

at the exact point where the target DNA joins SubSeq. The second site will cut 3 bp into

SubSeq so donating 3 bp ($N_XN_YN_Z$) to the target DNA. Step 4. Digestion with MlyI removes

SubSeq but with 3 bp of its sequence now replacing the 3 bp deleted from the target gene.

Step 5. Intramolecular ligation reforms the target gene but with one contiguous trinucleotide

sequence replaced with another.

Cassette Library	ORF	Cassettes
None-1	I	GCA DOMAIN AGG
None-2	II	CGCA DOMAIN AG
None-3	III	CTGCA DOMAIN A
GGS-1	I	GGCGGTAGCGCA DOMAIN AGGGGTGGCAGC
GGS-2	II	CGGCGGTAGCGCA DOMAIN AGGGGTGGCAG
GGS-3	III	CTGGCGGTAGCGCA DOMAIN AGGGGTGGCA
Ran-1	I	NNSGAT DOMAIN CGTNNS
Ran-2	II	SNNSGAT DOMAIN CGTNNSNN
Ran-3	III	NSNNSGAT DOMAIN CGTNNSN

Fig. 3. Example permutations of the DNA insert cassette for protein domain/segment insertion. ORF refers to the different open reading frame libraries used to reconstitute an in-frame chimeric gene on cassette insertion. Additional nucleotides required for frame correction are underlined. The first and last codons of the protein insertion coding region are shown on a black background. The rest of the intervening sequence is represented by DOMAIN. Three domain linking sequence variations are shown: (1) no linking sequences (None-X); (2) GlyGlySer linker (GGS-X); (3) Random short linker (Ran-X).