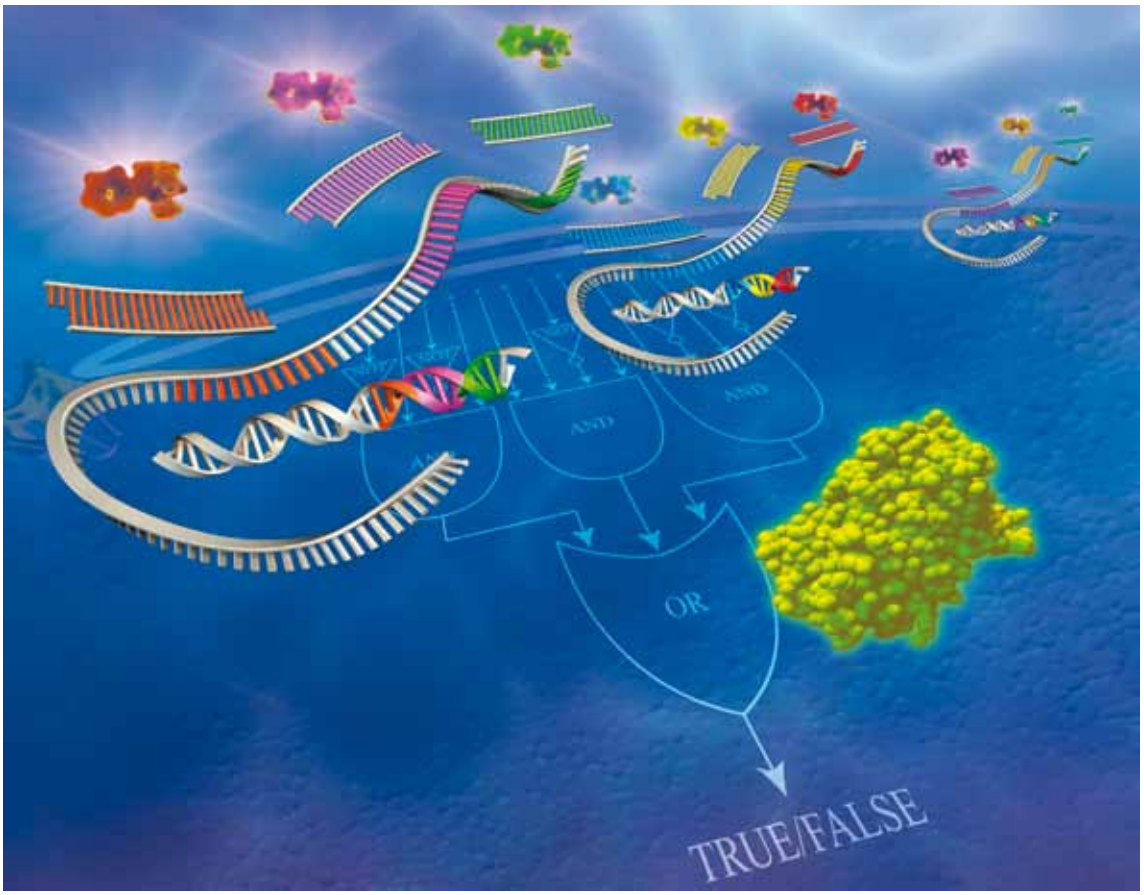
Molecular BioSystems



Expanded chemical diversity sampling through whole protein evolution†‡

Amy J. Baldwin,[ab] James A. J. Arpino,[a] Wayne R. Edwards,[ab] Eric M. Tippmann*[b] and D. Dafydd Jones*[a]

Received 3rd March 2009, Accepted 6th April 2009
First published as an Advance Article on the web 7th May 2009
DOI: 10.1039/b904031e

A directed evolution method has been developed that allows random substitution of a contiguous trinucleotide sequence for TAG throughout a target gene for use in conjunction with an expanded genetic code. Using TEM-1 β-lactamase and enhanced green fluorescent protein as targets, protein variants were identified whose functional phenotype was rescued *in vivo* when co-expressed with orthogonal tRNA–aminoacyl-tRNA synthase pairs that insert *p*-iodophenylalanine in response to UAG. Sequencing of the selected clones that retained the target protein function revealed that >90% of the variants contained in-frame TAG codons distributed throughout the target gene. Such an approach will allow broader sampling of new chemical diversity by proteins, so opening new avenues for studying biological systems and for adapting proteins for biotechnological applications. A common set of reagents allows the method to be used on different protein systems and in combination with an array of different unnatural amino acids, so helping to reveal the true potential for engineering proteins through expanded chemical diversity sampling.

Introduction

The shared genetic code allows most organisms to incorporate the same 20 amino acids into proteins, restricting the chemical functionality that can be sampled by proteins. Expansion of the genetic code allowing proteins to sample new chemistry *via* incorporation of unnatural amino acids in response to a designated codon during protein synthesis has recently become a reality.[1–3] Given our limited knowledge concerning the influence of unnatural amino acids on the co-operative network of weak interactions that comprise a functional protein, it is currently difficult to predict the consequence of substituting one of the 20 natural amino acids for an unnatural one. This may restrict the usefulness of rational or semi-rational protein engineering approaches based on site-specific mutagenesis, so unduly limiting the potential of expanded genetic code systems. Directed evolution[4–6] is a powerful protein engineering approach based on the concept of Darwinian evolution that, amongst other things, has recently been used to optimise proteins containing non-canonical amino acids introduced by the auxotrophic replacement strategy[7,8] and for generating proteins with a reduced amino acid alphabet.[9] Development of directed evolution approaches capable of diversifying the position of unnatural amino acid incorporation throughout a protein quickly and efficiently would be of benefit. To address this, we describe a directed evolution method (Fig. 1a and ESI‡, Fig. S1), based on the trinucleotide exchange (TriNEx) approach[10] that can deliver global codon replacement throughout a target gene for use in conjunction with an expanded genetic code. While traditional TriNEx involves the random replacement of one contiguous trinucleotide sequence for another irrespective of coding frame and is reliant on the existing genetic code, we have shown here that it can be adapted to label randomly a target gene with in-frame TAG

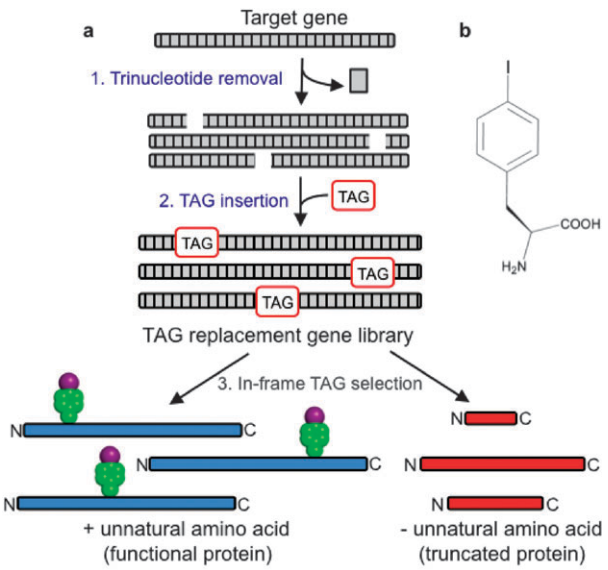


Fig. 1 (a) Schematic outline for generating a library of in-frame TAG codon replacements. 1. A contiguous trinucleotide sequence is randomly removed from the target gene using an engineered transposon.[13] 2. The TAG trinucleotide sequence is then inserted into the break in the gene using a synthetic DNA cassette[10] to replace the trinucleotide sequence deleted in step 1. This generates a TAG codon replacement library. A detailed mechanism of the trinucleotide exchange process is given in ESI‡. 3. Selection for in-frame TAG. Selection is based on the dependence of the protein's activity on the presence of the unnatural amino acid incorporated into the growing polypeptide in response to TAG. In the absence of an unnatural amino acid, TAG will be recognised as a stop codon and protein synthesis prematurely terminated. (b) Structure of *p*-iodo-L-phenylalanine.

[a] *Cardiff University, School of Biosciences, Main Building, Park Place, Cardiff, UK CF10 3AT. E-mail: jonesdd@cf.ac.uk*
[b] *Cardiff University, School of Chemistry, Main Building, Park Place, Cardiff, UK CF10 3AT. E-mail: tippmann@cardiff.ac.uk*

† This article is part of a *Molecular BioSystems* themed issue on Synthetic Biology.

‡ Electronic supplementary information (ESI) available: Detailed experimental procedures and information regarding construction of libraries, sequence information for each variant. See DOI: 10.1039/b904031e

codons reprogrammed to incorporate the unnatural amino acid *p*-iodo-L-phenylalanine (Fig. 1b).

Results and discussion

TEM-1 β-lactamase encoded by the *bla* gene and enhanced green fluorescent protein (eGFP) encoded by the *eGFP* gene were used as two separate targets. These two proteins have disparate structures and functions, so provide contrasting models for investigating the effects of unnatural amino acids on the properties of a protein. The amino acid, *p*-iodo-L-phenylalanine (*p*-iodoPhe; Fig. 1b), was incorporated by constructing the plasmids and appropriate mutations reported for the *Methanococcus jannaschii* derived amber suppressor $tRNA^{Tyr}_{CUA}$–tyrosyl-tRNA synthetase pair engineered to incorporate *p*-iodoPhe in response to the UAG codon in the transcribed gene mRNA.[11,12]

The *bla* and *eGFP* genes were processed by the TriNEx method (Fig. 1a) to generate libraries with single TAG codons at random positions within the gene. Initially an engineered transposon[13,14] is used to remove a contiguous trinucleotide sequence, followed by insertion of the TAG trinucleotide sequence donated from a synthetic DNA cassette (see ESI‡ for mechanism and experimental details).[10,15] The resulting libraries were termed bla^{TAG} and $eGFP^{TAG}$. To identify gene variants in the libraries with an in-frame TAG, an *in vivo* selection process in *Escherichia coli* was used (Fig. 2a and b; ESI‡). In the absence of *p*-iodoPhe, TAG will be recognised solely as a stop codon and prematurely terminate translation generating a truncated inactive protein. In the presence of *p*-iodoPhe and the *M. jannaschii* amber suppressor pair, the unnatural amino acid will be incorporated in response to the UAG codon and a full-length protein synthesised, which folds to a functional form if the *p*-iodoPhe is tolerated at that particular residue in the protein (Fig. 2a and b).

The bla^{TAG} and $eGFP^{TAG}$ libraries represent two disparate phenotypes and for each target we used an appropriate approach to select variants tolerant to the insertion of *p*-iodoPhe. To select TEM-1 variants tolerant to *p*-iodoPhe substitution, *E. coli* cells containing a plasmid harbouring the *M. jannaschii* amber suppressor pair were transformed with the bla^{TAG} library and cells were grown on LB agar in the presence of 1 mM *p*-iodoPhe and carbenicillin. A small selection comprising 456 clones was then randomly chosen (from $>10^4$ clones) and replica plated onto LB agar containing 100 μg ml^{-1} carbenicillin in the absence and presence of 1 mM *p*-iodoPhe. Those clones that displayed resistance only in the presence of the unnatural amino acid were deemed to contain an in-frame TAG in *bla* that translates to a position in TEM-1 tolerant to the incorporation of *p*-iodoPhe. The location of in-frame TAG codons and hence site of *p*-iodoPhe incorporation within TEM-1 for all 57 variants identified from the screen whose *in vivo* activity was dependent on *p*-iodoPhe (Fig. 2a) were determined by gene sequencing (Fig. 2c and ESI‡, Table S1).

A similar procedure was used for the $eGFP^{TAG}$ library. *E. coli* containing a plasmid harbouring the amber suppressor pair was transformed with the $eGFP^{TAG}$ library and grown on LB agar in the absence of *p*-iodoPhe to remove variants whose fluorescence was not contingent on amber suppression. A total of 846 $eGFP^{TAG}$ clones displaying no observable fluorescence were selected at random (from a total of $>10^4$ clones), providing a variant pool enriched with in-frame TAG substitutions (ESI‡). The variant *eGFP* was sequenced from 26 clones (from a total $>10^3$) that were observed to fluoresce in the presence of 1 mM *p*-iodoPhe and the amber suppressor pair (Fig. 2b) to determine the trinucleotide exchange position and hence site of *p*-iodoPhe incorporation within eGFP (Fig. 2c and ESI‡, Table S2).

Sequence analysis of the selected clones that retained either *in vivo* β-lactamase activity or fluorescence when grown in the presence of *p*-iodoPhe revealed that 100% of *bla* and 92% of *eGFP* gene variants contained an in-frame TAG (Fig. 2c; ESI‡, Tables S1 and S2). Even from the comparatively small sampling set analysed and the bulky nature of the *p*-iodoPhe,

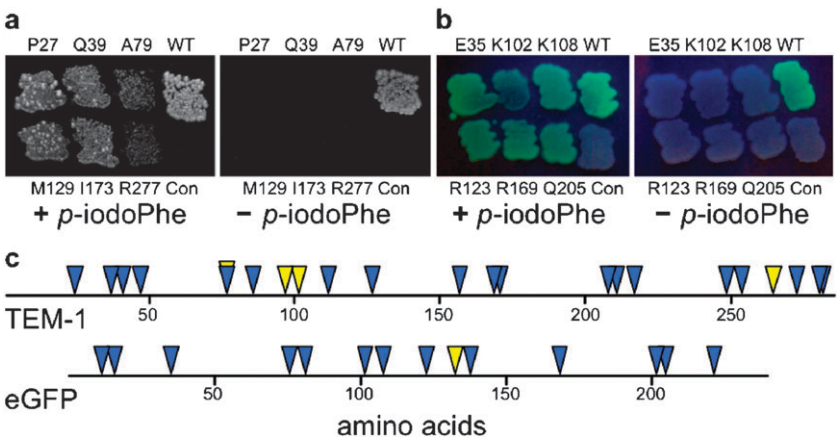


Fig. 2 Incorporation of *p*-iodoPhe in response to TAG codon. (a) TEM-1 *in vivo* activity in the presence (+ *p*-iodoPhe) and absence (– *p*-iodoPhe) of a few selected variants. Con and WT represent the negative (cells with no *bla* containing plasmid) and the positive controls (cells transformed with plasmid harbouring wt *bla*), respectively. (b) eGFP *in vivo* fluorescence of several selected variants in the presence (+ *p*-iodoPhe) and absence (– *p*-iodoPhe). Con and WT represent the negative (cells transformed with empty expression plasmid) and the positive controls (cells transformed with *eGFP* containing expression plasmid), respectively. The residues indicated are predicted to be replaced by *p*-iodoPhe. (c) Location of *p*-iodoPhe incorporation into TEM-1 and eGFP based on gene sequence analysis. Incorporation of *p*-iodoPhe in response to a single codon replacement by TAG (blue triangles) or by known cross-codon replacement (yellow triangles).

the predicted unnatural amino acid substitutions were distributed throughout TEM-1 and eGFP (Fig. 2c and 3), equating to the replacement of 22 and 14 different residues, respectively. Many of the amino acid residues whose codons were replaced by TAG occurred close to the active centres of TEM-1 and eGFP (Fig. 3). For example, L15, E34, R122, Q204 and L221 all lie close in space to the eGFP chromophore;[16] E104, M129, E171 and I173 are close to the active site of TEM-1.[17]

The majority (86%) of in-frame TAG replacements were generated by alteration of a single codon. The remaining 14% were generated by cross-codon replacement with the TAG antisense sequence, CTA (Fig. 2c and ESI‡, Tables S1 and S2); the C replaces the final nucleotide of one codon and TA replaces the first two nucleotides of the next codon that when combined with a G at the third nucleotide position reconstitutes an in-frame TAG. Only one case was observed where the mutation adjacent to the TAG codon resulted in an amino acid substitution (eGFP K132N, E133iF); the rest generated silent mutations (ESI‡, Tables S1 and S2). Only two of the sequenced *eGFP* variants did not contain an in-frame TAG (ESI‡, Table S2). To access the TAG codon, 70% of the variants required the replacement of at least 2 nucleotides per codon. Thus, established directed evolution methods such as error-prone PCR that introduce predominately non-contiguous point mutations are unlikely to sample a single TAG codon.

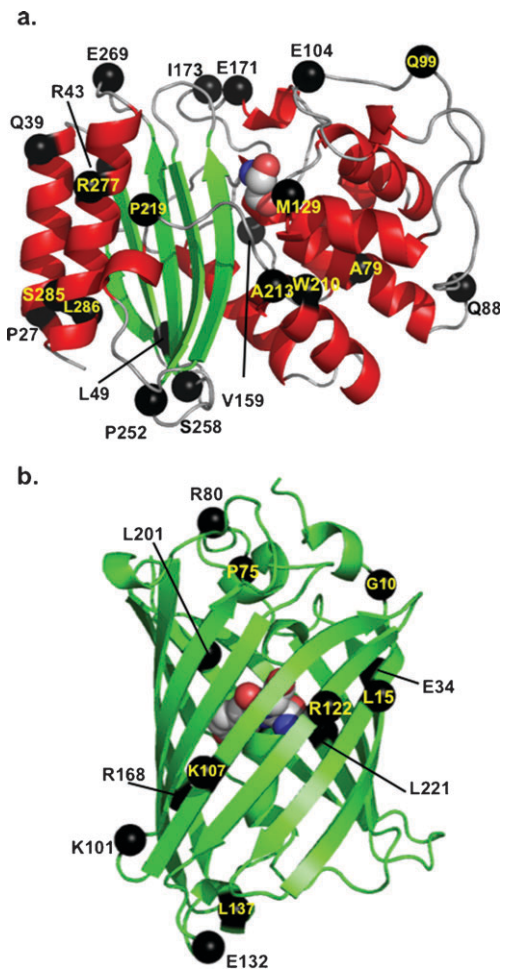


Fig. 3 Position of residues tolerant to substitution with *p*-iodoPhe. (a) TEM-1 β-lactamase. The catalytic residue Ser70 is shown as a space-filling representation and coloured by CPK. (b) eGFP. The chromophore comprising residues 65-TYG-67 is shown as a space-filling representation and coloured by CPK. The sphere corresponding the Q204 is hidden behind the chromophore.

Conclusions

In summary, we have demonstrated a directed evolution approach that allows a gene to be randomly labelled with in-frame TAG codons for the incorporation of unnatural amino acids at many different positions within a target protein *in vivo*. Combining the TriNEx approach with an expanded genetic code system provides a new approach for engineering proteins through the incorporation of unnatural amino acids that was previously unavailable. When used in conjunction with *in vitro* recombination,[18] multiple residues in a protein can be sampled if required. This in turn will allow a broader sampling of new chemical diversity by proteins, so opening new avenues for studying biological systems and for adapting proteins for biotechnological applications. For example, we are currently screening the $eGFP^{TAG}$ library with alternative unnatural amino acids to identify GFP variants with novel physicochemical properties. The use of a common set of reagents allows the method to be used on many different protein systems and in combination with an array of different unnatural amino acids, so helping to reveal the true potential for engineering proteins through expanded chemical diversity sampling.

Acknowledgements

This work was supported by the BBSRC (BB/E007384 and BB/E001084), EPSRC (EP/D001641) and a Wellcome Trust VIP award to AJB.

Notes and references

1 D. L. Beene, D. A. Dougherty and H. A. Lester, *Curr. Opin. Neurobiol.*, 2003, **13**, 264–270.
2 A. J. Link, M. L. Mock and D. A. Tirrell, *Curr. Opin. Biotechnol.*, 2003, **14**, 603–609.
3 J. Xie and P. G. Schultz, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 775–782.
4 C. Jackel, P. Kast and D. Hilvert, *Annu. Rev. Biophys.*, 2008, **37**, 153–173.
5 C. M. Yuen and D. R. Liu, *Nat. Methods*, 2007, **4**, 995–997.
6 J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan and F. H. Arnold, *Curr. Opin. Struct. Biol.*, 2005, **15**, 447–452.
7 T. H. Yoo, A. J. Link and D. A. Tirrell, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13887–13890.
8 J. K. Montclare and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2006, **45**, 4518–4521.
9 M. T. Reetz, D. Kahakeaw and J. Sanchis, *Mol. BioSyst.*, 2009, **5**, 115–122.
10 A. J. Baldwin, K. Busse, A. M. Simm and D. D. Jones, *Nucleic Acids Res.*, 2008, **36**, e77.
11 J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon and P. G. Schultz, *Nat. Biotechnol.*, 2004, **22**, 1297–1301.
12 Y. Ryu and P. G. Schultz, *Nat. Methods*, 2006, **3**, 263–265.
13 D. D. Jones, *Nucleic Acids Res.*, 2005, **33**, e80.
14 A. M. Simm, A. J. Baldwin, K. Busse and D. D. Jones, *FEBS Lett.*, 2007, **581**, 3904–3908.
15 W. R. Edwards, K. Busse, R. K. Allemann and D. D. Jones, *Nucleic Acids Res.*, 2008, **36**, e78.
16 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.
17 A. Matagne, J. Lamotte-Brasseur and J. M. Frere, *Biochem. J.*, 1998, **330**(Pt 2), 581–598.
18 W. P. Stemmer, *Nature*, 1994, **370**, 389–391.

