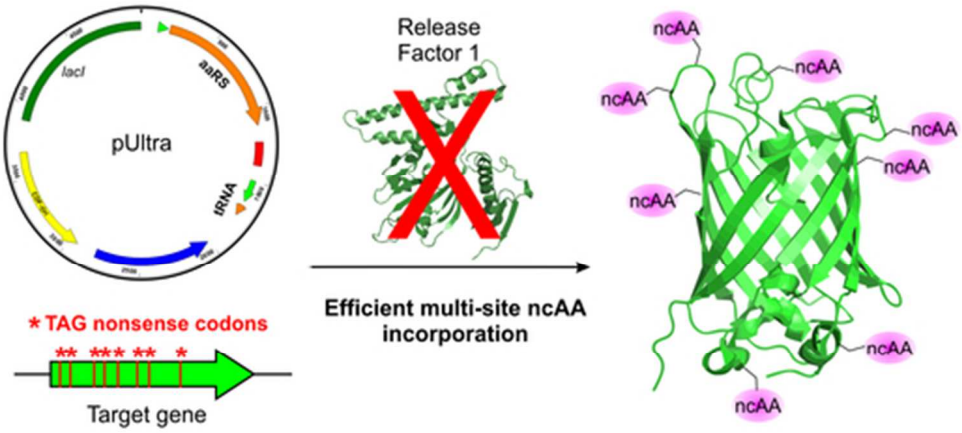
Molecular BioSystems

Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1

Journal:	*Molecular BioSystems*
Manuscript ID	MB-COM-01-2016-000070.R1
Article Type:	Communication
Date Submitted by the Author:	12-Mar-2016

Complete List of Authors:	Zheng, Yunan; Boston College, Chemistry Lajoie, Marc; Harvard Medical School, Department of Genetics; University of Washington , Department of Biochemistry Italia, James; Boston College, Chemistry Chin, Melissa; Boston College, Chemistry Church, George; Harvard Medical School, Genetics Department Chatterjee, Abhishek; Boston College, Chemistry



45x20mm (300 x 300 DPI)

Journal Name

COMMUNICATION

Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1†

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Yunan Zheng,[a] Marc J. Lajoie,[bc] James S. Italia,[a] Melissa Chin,[a] George M. Church,[b] Abhishek Chatterjee*[a]

Site-specific incorporation of noncanonical amino acids (ncAAs) into proteins expressed in *E. coli* using UAG-suppression competes with termination mediated by release factor 1 (RF1). Recently, unconditional deletion of RF1 was achieved in a genomically recoded *E. coli* (C321), devoid of all endogenous UAG stop codons. Here we evaluate the efficiency of ncAA incorporation in this strain using optimized suppression vectors. Even though the absence of RF1 does not benefit the suppression efficiency of a single UAG codon, multi-site incorporation of a series of chemically distinct ncAAs was significantly improved.

The ability to site-specifically incorporate noncanonical amino acids (ncAAs) into a protein presents novel opportunities to investigate and engineer its function.[1-3] Co-translational incorporation of ncAAs into a protein in living cells is achieved using an orthogonal (i.e., does not cross-react with its host counterparts) tRNA/aminoacyl-tRNA synthetase (aaRS) pair, engineered to charge the ncAA of interest in response to a nonsense codon.[1] Among the three nonsense codons, UAG is most frequently used to genetically encode ncAAs in *E. coli*. The efficiency of ncAA-incorporation in response to the UAG codon is negatively affected by competition with release factor 1 (RF1), which is responsible for translation-termination at UAG and UAA stop codons (Fig. 1). Consequently, initial systems for ncAA incorporation afforded poor expression yields and were restricted to the incorporation of a single ncAA per polypeptide.[4] Since UAG-termination is essential for *E. coli* viability, earlier efforts to improve the suboptimal ncAA incorporation efficiency focused predominantly on the directed evolution of the orthogonal tRNA/aaRS pairs to improve its ability to compete with RF1 in *E. coli*, as well as the optimization of their expression levels. Such efforts led to multiple generations of suppression vectors (expressing the orthogonal tRNA/aaRS pair) with increasing efficiency of ncAA incorporation.[5-7] The most advanced *E. coli* suppression vectors are capable of providing near wild-type expression yields for proteins incorporating a single ncAA, and facilitate the incorporation of up to 3 ncAAs into one polypeptide.[7] However, to truly elevate the suppression efficiency of the UAG nonsense codon to the level of sense codons, and to enable the incorporation of unrestricted numbers of ncAAs into proteins in *E. coli*, it is essential to eliminate the competing RF1-mediated termination. Given that RF2 is capable of terminating UAA as well as UGA, the deletion of RF1 only perturbs termination at the UAG stop codon. RF1-deletion from *E. coli* was first made possible by the presence of a mutant RF2, which compensates for the loss of RF1.[8, 9] Another approach relied upon the identification of essential *E. coli* genes containing the UAG termination signal that are most affected by the absence of RF1, and expressing these from a

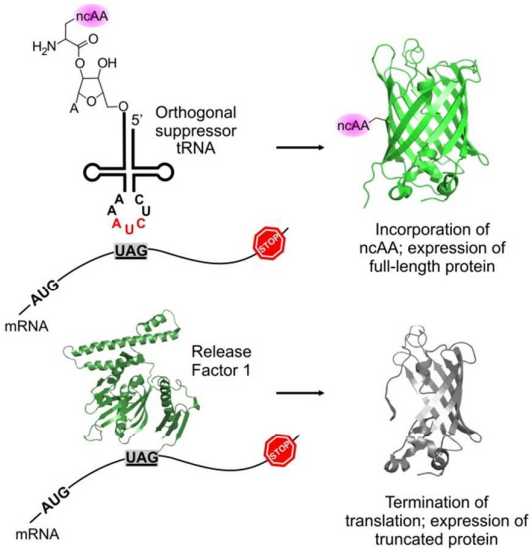


Fig. 1: Release factor 1 mediated termination competes with the incorporation of ncAA in response to the UAG nonsense codon

[a] Department of Chemistry, Boston College, 2609 Beacon Street, Chestnut Hill, MA 02467, USA. Email: abhishek.chatterjee@bc.edu; Tel: +1-617-552-1778.
[b] Department of genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA
[c] Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.

† Electronic Supplementary Information (ESI) available: Additional experimental details, sequence of the reporters, expression analysis, and ESI-MS analysis of reporter proteins. See DOI: 10.1039/x0xx00000x

plasmid with an alternative stop codon.[10, 11] The absence of RF1 in these strains was shown to enhance the efficiency of ncAA incorporation in response to UAG.

Using high-throughput genome engineering, all endogenous UAG codons from a strain of *E. coli* (C321) was recently replaced with UAA, leaving RF1 functionally redundant.[12] Unlike previous instances, unconditional deletion of RF1 from this strain truly reassigns the meaning of the UAG codon. However, evaluation of ncAA incorporation efficiency in the resulting strain (C321.ΔRF1) in response to a single UAG revealed surprisingly little improvement upon RF1-deletion.[12] Unlike previous characterizations of RF1-deficient strains, this experiment used an optimized suppression vector for ncAA incorporation, and it was proposed that the associated high efficiency led to the observed lack of improvement upon RF1 deletion. To better understand the performance of optimized suppression vectors in this strain, here we evaluate ncAA incorporation by two such vectors, pEVOL[6] and pUltra,[7] into reporter proteins designed to offer increasing levels of challenge.

Development of the highly efficient pEVOL and pUltra suppression vectors (Fig. S1) involved the optimization of the expression levels of the engineered tRNA/aaRS pair, as well as other parameters such as plasmid copy number.[6, 7] Both vectors express the orthogonal UAG-suppressing tRNA from a strong *proK* promoter. In pEVOL, expression of the aaRS is achieved from two different promoters: a weak constitutively active (*glnS*), and a strong inducible promoter (*araBAD*) – a pattern that is mimicked in pUltra by the use of a single strong inducible promoter (*tacI*) that allows low levels of leaky expression in the absence of induction. Additionally, pUltra harbors a slightly higher copy number CloDF13 origin of replication relative to pEVOL (p15a). While both suppression vectors perform well, pUltra shows enhanced efficacy when less active tRNA/aaRS pairs are employed.[7]

To evaluate the efficiency of ncAA incorporation facilitated by these vectors in the C321.ΔRF1 strain, we used a *t5-lac* driven superfolder GFP reporter,[13, 14] whose high structural stability should resist potential perturbation from the incorporation of multiple ncAAs. Additionally, the full-length reporter protein – but not the truncated products – emits the characteristic fluorescence that enables convenient evaluation of its intracellular expression levels.[14] We introduced an increasing number of contiguous UAG codons at the N-terminus of sfGFP to generate a range of reporters that provide an increasing degree of challenge for full-length protein expression (Figure 2A). To evaluate the efficiency of ncAA incorporation into these reporters we chose pEVOL and pUltra plasmids harboring a highly polyspecific *M. jannaschii* tyrosyl-tRNA/aaRS derived orthogonal pair.[15] This tRNA/aaRS pair is capable of charging a number of different ncAAs, while discriminating against endogenous amino acids. The suppression vectors were individually co-transformed with these reporters into the C321.ΔRF1 strain, as well as control strains harboring a functional RF1 (ECNR1, the progenitor strain for C321.ΔRF1; and DH10B), and the expression levels of full-length sfGFP were monitored in the presence or absence of 1 mM O-methyltyrosine (**1**, Fig. 3C) in the media, a known substrate for the aforementioned polyspecific aaRS. As our preliminary experiments revealed comparable reporter-expression in all strains harboring a functional RF1, only DH10B was used in subsequent experiments to represent them.

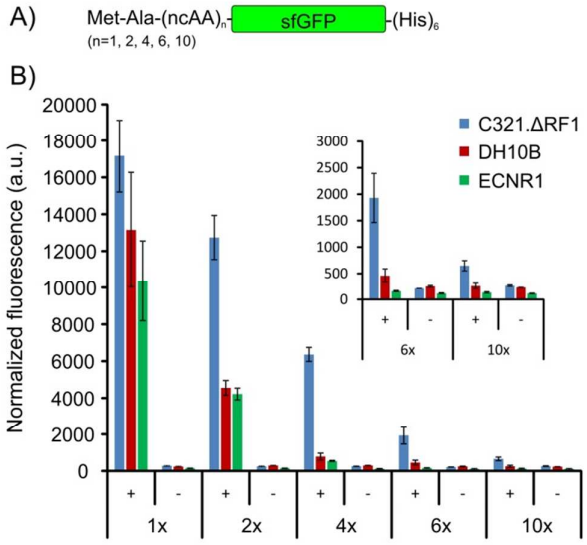


Fig. 2: Expression of sfGFP reporters with increasing number of contiguous UAG stop codons using a pUltra suppression vector that encodes an engineered tRNA/aaRS pair with relaxed substrate specificity. A) The design of the sfGFP reporters harboring contiguous N-terminal UAG codons. B) Expression level of full-length sfGFP in the C321.ΔRF1 strain (blue), as well as two other *E. coli* strains (green and red) harboring a functional RF1, in the presence and absence of 1 mM O-methyltyrosine (**1**, Fig 3C), as measured by the sfGFP-specific fluorescence in the bacterial lysate. Data for reporters with 6 or 10 TAG codons are magnified at inset. The number of UAG codons in a reporter is indicated along the x-axis, and the expression levels in the presence or the absence of the ncAA is denoted by "+" and "-", respectively.

We observed that the expression levels of the sfGFP reporter harboring a single UAG were comparable in the presence or absence of RF1 (Figure 2B), corroborating previous observations. However, the absence of RF1 was found to be increasingly beneficial for the expression of reporters with increasing number of contiguous UAG codons. While the expression levels of the $(UAG)_4$-sfGFP in strains harboring a functional RF1 dropped to ~6% of $(UAG)_1$-sfGFP, the corresponding ratio was >36% in C321.ΔRF1 (Figure 2B). Furthermore, the $(UAG)_6$-sfGFP reporter failed to express any full-length protein in strains harboring a functional RF1, while yielding fluorescent sfGFP at an approximately 10% level relative to $(UAG)_1$-sfGFP in C321.ΔRF1 (Figure 2B). Even the $(UAG)_{10}$-sfGFP reporter exhibited a small but measurable expression of full-length sfGFP in the RF1-deficient strain. Use of the pEVOL suppression vector encoding the same tRNA/aaRS pair yielded similar levels of sfGFP expression as observed with pUltra (Fig. S2), and in both cases expression was strictly dependent upon the presence of the ncAA in the media. Incorporation of the appropriate numbers of O-

methyltyrosine residues into these reporters was verified by mass-spectrometry following the purification of the full-length proteins by Ni-NTA affinity chromatography using a C-terminal polyhistidine tag (Table S1). Yields of purified sfGFP with 1, 2, 4, and 6 UAGs were 450 mg/L, 416 mg/L, 204 mg/L, and 68 mg/L, respectively (approximately 65%, 60%, 30%, and 10% relative to wild type sfGFP, respectively), and their SDS-PAGE analysis ensured over 90% purity (Fig. S10). Weak expression, as well as poor solubility of the protein expressed from the $(UAG)_{10}$-sfGFP reporter precluded its isolation and further characterization.

Taking advantage of the substrate promiscuity of the tRNA/aaRS pair used in this assay, we further tested the incorporation efficiency of a number of additional ncAAs (**2**-**5**, Fig. 3C) using the same set of reporters. In the presence of *p*-iodo-phenylalanine (**2**), *p*-azido-phenylalanine (**3**), and *p*-acetyl-phenylalanine (**4**), which are charged efficiently by this tRNA/aaRS pair, expression profiles of the aforementioned reporters were similar to what was observed using O-methyltyrosine (Fig. S3-S5). However, when O-propargyltyrosine (**5**) was used, which is charged less efficiently, the advantage of the ΔRF1 strain for expressing reporters with multiple UAG codons was less pronounced (Fig. S6). We further tested the expression of these reporters using two additional *M. jannaschii* tyrosyl-tRNA/aaRS derived orthogonal pairs specific for two polar ncAAs, *p*-boronophenylalnine (**6**, Fig 3C)[16] and O-sulfotyrosine (**7**, Fig 3C).[17] In both cases, the tRNA/aaRS pairs were expressed from the pEVOL suppression vector. Reporter expression with *p*-boronophenylalnine followed a pattern similar to what was observed before (Fig. S8), enabling robust expressions of $(UAG)_4$-sfGFP and $(UAG)_6$-sfGFP in the C321.ΔRF1 strain, but not in strains with a functional RF1. Interestingly, when incorporating O-sulfotyrosine, expression of reporters with multiple contiguous UAG codons were sharply diminished, despite strong expression of the $(UAG)_1$-sfGFP reporter (Fig. S7). The expression of the $(UAG)_2$-sfGFP reporter incorporating O-sulfotyrosine was substantially lower relative to the corresponding experiments using other ncAAs, and no expression was observed for reporters with additional UAG codons (Fig. S7). This reduction is presumably due to the strong negative charge associated with the side chain of O-sulfotyrosine, which may adversely affect translation if incorporated successively.

The series of reporters used so far employs an increasing number of contiguous UAG codons at the N-terminus of sfGFP. Suppression of such contiguous stop codons can be more challenging than if these are well-separated by sense codons. Furthermore, incorporation of a large number contiguous ncAAs can negatively affect translation (as observed with O-sulfotyrosine), or protein behavior (e.g., the poor solubility of $(UAG)_{10}$-sfGFP expressed with O-methyltyrosine). We created an additional sfGFP reporter, where all of the surface exposed threonine residues (eight in total) were replaced with UAG codons (sfGFP-$(UAG)_8$-scattered; Fig. 3A). Scattered distribution of the UAG codons may alleviate potential challenges associated with earlier reporters. Expression of this

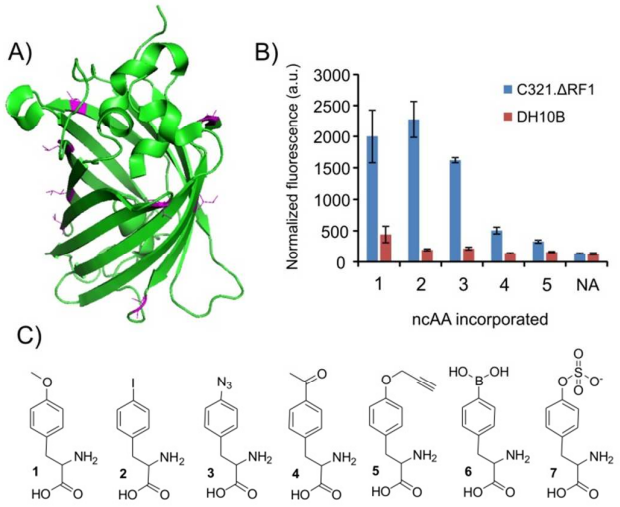


Fig. 3: Expression of a sfGFP reporter harboring 8 UAG codons replacing all surface exposed threonine residues in *E. coli* strains with (DH10B) or without (C321.ΔRF1) a functional RF1, mediated by the aforementioned pUltra suppression vector encoding the polyspecific tRNA/aaRS pair. A) Structure of sfGFP (pdb ID 4lqt) demonstrating the positions of the eight threonine residues (in magenta) mutated to ncAAs. B) Expression levels of full-length sfGFP in C321.ΔRF1 as well as DH10B in the presence of 5 different ncAAs (**1**-**5**, Fig 3C), as well as in the absence of any added ncAA on the media (NA), as measured by the sfGFP-specific fluorescence in the bacterial lysate. C) Structures of the ncAAs used in this report.

reporter was first evaluated using the aforementioned polyspecific orthogonal tRNA/aaRS pair expressed from the pUltra vector and the ncAAs **1**-**5** described in Fig. 3C. We found little to no protein expression in the DH10B strain, while expression in the C321.ΔRF1 strain led to robust sfGFP fluorescence in the presence of ncAAs **1**, **2**, and **3** (Fig. 3B). Incorporation efficiencies of ncAAs **4** and **5** were found to be relatively weaker. We further evaluated the expression of this reporter using pEVOL vectors encoding tRNA/aaRS pairs specific for O-sulfotyrosine (**7**) and *p*-boronophenylalnine (**6**), and observed robust expression of full-length protein in both cases (Fig. S9). For most ncAAs, the expression of full-length sfGFP from this reporter matched the corresponding levels obtained from the contiguous-$(UAG)_6$-sGFP reporter. The successful expression of sfGFP-$(UAG)_8$-scattered incorporating O-sulfotyrosine is worth noting, given the incorporation of the same ncAA into reporters harboring more than two successive UAG codons was unsuccessful, further suggesting a possible negative impact from successive incorporation of strongly charged ncAAs on translation. The expressed protein was isolated by Ni-NTA chromatography in each case (isolated yields ranged between 15-70 mg/L; approximately 2-10% of wild-type sfGFP expression level), and was analyzed by mass-spectrometry to confirm the incorporation of appropriate ncAAs (Table S1).

While our results demonstrate the benefits associated with the unconditional deletion of RF1 when expressing proteins incorporating multiple ncAAs, it is also clear that the absence of RF1 alone is not sufficient to elevate the suppression

efficiency of UAG to a level similar to sense codons. The presence of increasing numbers of UAG codons in a reporter, contiguous or scattered, does lead to decreased levels of protein expression in the C321.ΔRF1 strain. The non-native orthogonal tRNA/aaRS pair used for ncAA incorporation could be responsible for the observed lack of efficiency. These pairs are typically imported into *E. coli* from archaea or eukaryotes to ensure non-cross-reactivity with host counterparts, and may not have optimal interaction with *E. coli* translational machinery. Even though the *M. jannaschii* tyrosyl tRNA has been evolved for improved performance in *E. coli* for ncAA incorporation,[18] there may still be room for further improvement. Suboptimal interaction between the engineered tRNA and its cognate aaRS may also contribute to reduced efficiency.[19] A recent report demonstrating that the evolution of improved tRNA/aaRS variants leads to enhanced ncAA incorporation suggests that this factor is at least in part responsible for the lower efficiency of UAG.[20] However, the involvement of additional endogenous factors limiting the translational efficiency of the UAG codon cannot yet be ruled out. The C321.ΔRF1 strain provides a great platform to further investigate the mechanism responsible for the suboptimal efficiency of multiple UAG suppression in the absence of RF1.

Conclusions

In conclusion, our report presents a detailed investigation of ncAA incorporation efficiency in response to UAG codons in the C321.ΔRF1 strain using optimized suppression vectors pEVOL and pUltra. We demonstrate that the use of these vectors in the C321.ΔRF1 strain enables the production of proteins with multiple ncAAs at a substantially higher level than strains that harbor a functional RF1, even though the suppression efficiency of a single UAG is not significantly improved. Improved multi-site incorporation of a number of ncAAs harboring chemically distinct functionalities further underscores the utility of this system. The ability to reliably express proteins incorporating multiple ncAAs will enable the generation of new protein functions that can lead to the development of novel materials and therapeutics, among other applications.

Acknowledgement

The pEVOL and pUltra plasmids were gifts from Prof. Peter Schultz (TSRI). A.C. is thankful for support from the Richard and Susan Smith Family Foundation, Newton, MA.

References

1 C. C. Liu and P. G. Schultz, *Annual review of biochemistry*, 2010, **79**, 413-444.
2 M. J. Lajoie, D. Soll and G. M. Church, *Journal of molecular biology*, 2015, DOI: 10.1016/j.jmb.2015.09.003.
3 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis, *Chemical Science*, 2015, **6**, 50-69.
4 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498-500.
5 Y. Ryu and P. G. Schultz, *Nature methods*, 2006, **3**, 263-265.
6 T. S. Young, I. Ahmad, J. A. Yin and P. G. Schultz, *Journal of molecular biology*, 2010, **395**, 361-374.
7 A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao and P. G. Schultz, *Biochemistry*, 2013, **52**, 1828-1837.
8 D. B. Johnson, C. Wang, J. Xu, M. D. Schultz, R. J. Schmitz, J. R. Ecker and L. Wang, *ACS chemical biology*, 2012, **7**, 1337-1344.
9 D. B. Johnson, J. Xu, Z. Shen, J. K. Takimoto, M. D. Schultz, R. J. Schmitz, Z. Xiang, J. R. Ecker, S. P. Briggs and L. Wang, *Nature chemical biology*, 2011, **7**, 779-786.
10 T. Mukai, A. Hayashi, F. Iraha, A. Sato, K. Ohtake, S. Yokoyama and K. Sakamoto, *Nucleic acids research*, 2010, **38**, 8188-8195.
11 K. Ohtake, A. Sato, T. Mukai, N. Hino, S. Yokoyama and K. Sakamoto, *Journal of bacteriology*, 2012, **194**, 2606-2613.
12 M. J. Lajoie, A. J. Rovner, D. B. Goodman, H. R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church and F. J. Isaacs, *Science*, 2013, **342**, 357-360.
13 J. D. Pedelacq, S. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo, *Nature biotechnology*, 2006, **24**, 79-88.
14 A. Chatterjee, M. J. Lajoie, H. Xiao, G. M. Church and P. G. Schultz, *Chembiochem*, 2014, **15**, 1782-1786.
15 D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon and P. G. Schultz, *Biochemistry*, 2011, **50**, 1894-1900.
16 E. Brustad, M. L. Bushey, J. W. Lee, D. Groff, W. Liu and P. G. Schultz, *Angewandte Chemie International ed. in English*, 2008, **47**, 8220-8223.
17 C. C. Liu and P. G. Schultz, *Nature biotechnology*, 2006, **24**, 1436-1440.
18 J. Guo, C. E. Melancon, 3rd, H. S. Lee, D. Groff and P. G. Schultz, *Angewandte Chemie International ed. in English*, 2009, **48**, 9148-9151.
19 B. J. Rauch, J. J. Porter, R. A. Mehl and J. J. Perona, *Biochemistry*, 2016, **55**, 618-28.
20 M. Amiram, A. D. Haimovich, C. Fan, Y. S. Wang, H. R. Aerni, I. Ntai, D. W. Moonan, N. J. Ma, A. J. Rovner, S. H. Hong, N. L. Kelleher, A. L. Goodman, M. C. Jewett, D. Soll, J. Rinehart and F. J. Isaacs, *Nature biotechnology*, 2015, **33**, 1272-1279.