Bioorg. Med. Chem. 41 (2021) 116207

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Full incorporation of the noncanonical amino acid hydroxylysine as a surrogate for lysine in green fluorescent protein

Marc Finkler [a], Sina Ravanbodshirazi [a], Florent Grimm [a], Philip Hartz [b], Albrecht Ott [a,*]

[a] Universität des Saarlandes, Biologische Experimentalphysik, Naturwissenschaftlich-Technische Fakultät, B2 1, Campus, 66123 Saarbrücken, Germany
[b] Universität des Saarlandes, Institut für Biochemie, Naturwissenschaftlich-Technische Fakultät, B2 2, Campus, 66123 Saarbrücken, Germany

ARTICLE INFO

Keywords:
Hydroxylysine
Lysine
Escherichia coli cell extract
Cell-free expression
Noncanonical amino acid

ABSTRACT

The canonical set of amino acids leads to an exceptionally wide range of protein functionality, nevertheless, this set still exhibits limitations. The incorporation of noncanonical amino acids into proteins can enlarge its functional scope. Although proofreading will counteract the charging of tRNAs with other amino acids than the canonical ones, the translation machinery may still accept noncanonical amino acids as surrogates and incorporate them at the canonically prescribed locations within the protein sequence. Here, we use a cell-free expression system to demonstrate the full replacement of L-lysine by L-hydroxylysine at all lysine sites of recombinantly produced GFP. *In vivo*, as a main component of collagen, post-translational L-hydroxylysine generation enables the formation of cross-links. Our work represents a first step towards *in vitro* production of (modified) collagens, more generally of proteins that can easily be crosslinked.

1. Introduction

Production, optimization and modification of proteins and enzymes through protein engineering[1–3] often aim at novel, improved or specialized protein functionalities.[3] Proteins can be obtained by isolation from native organisms, and by chemical or recombinant production (*in vivo* or *in vitro*).[4–7] Proteins can be post-translationally modified by enzymes, or more generally by chemical processes, to alter their amino acids or motifs.[8,9] A different strategy consist of producing the protein directly with the envisioned alterations.[10,11] During chemical protein synthesis the involved substances, in particular the amino acids can be varied, enabling the production of proteins with novel amino acid sequences or non-natural ingredients.[6,12] At the same time, chemical synthesis can be difficult to control and perform, among others because of the production of by-products.[13] Recombinant protein production by an organism or an *in vitro* system[7] enables better control.[13] An *in vitro* system requires DNA carrying the gene to be expressed as well as all regulatory units for expression that will act as a template for protein production.[14–16] The amino acid sequence of a protein can simply be modified by changing the DNA-sequence of the related gene. Unnatural (noncanonical) amino acids can be introduced in a site-specific or residue-specific manner.[17–19]

The site-specific incorporation of noncanonical amino acids into proteins requires an orthogonal pair of tRNA and aminoacyl-tRNA synthetase, and the selection of a desired recognition site for the tRNA (codon).[17,18,20–22] Such an approach can overcome the necessity of amino-acid depletion for the incorporation of analogs into protein. *In vivo,* orthogonal systems were used for instance for site-specific incorporation of lysine/pyrrolysine derivatives.[41] However, such orthogonal systems need to be optimized for each noncanonical amino acid and the number of incorporated noncanonical amino acids in one protein is limited to one or just a few.[19,42]

Residue-specific incorporation of noncanonical amino acids into proteins is simpler.[19] However, it is necessary that the noncanonical amino acid is recognized by the endogenous translation machinery of the expression system,[17,18,23] making the technique suitable for analogs of canonical amino acids.[24,25]

Each noncanonical amino acid presents different affinities to its corresponding aminoacyl-tRNA-synthetases.[17,18,26] Due to proofreading[27,28] performed by the translation machinery, even trace amounts of the canonical amino acids originating from the expression system will be preferred over the surrogate.[17,18,29] The extent of proofreading activity depends on the particular noncanonical amino acid in question.[17] While the successful substitution of tryptophan by a

* Corresponding author.
E-mail addresses: marc.finkler@uni-saarland.de (M. Finkler), sina.ravanbodshirazi@physik.uni-saarland.de (S. Ravanbodshirazi), s8flgrim@stud.uni-saarland.de (F. Grimm), philip.hartz@uni-saarland.de (P. Hartz), albrecht.ott@physik.uni-saarland.de (A. Ott).

https://doi.org/10.1016/j.bmc.2021.116207
Received 20 February 2021; Received in revised form 22 April 2021; Accepted 26 April 2021
Available online 9 May 2021
0968-0896/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

tryptophan analog had to be performed in an expression system devoid of tryptophan, low quantities of arginine did not interfere with incorporation of the arginine analog canavanine.[18,26] The affinity of the arginyl-tRNA-synthetase for canavanine is similar or identical to arginine.[26] However, in the expression system the yield of modified protein was lower than the yield of the wild type.[26] As a general rule, gene expression should be performed in the presence of the surrogate and the absence of the canonical amino acid to be replaced,[18,19] a condition forcing the expression system to accept the noncanonical amino acid.[17,24]

In this paper, we investigate the incorporation of the noncanonical amino acid hydroxylysine as a surrogate for lysine in GFP protein. Our work is a first step for the incorporation of hydroxylysine in more complex proteins. Hydroxylysine (L-5-Hydroxylysine) is an important component of collagen fibers.[30–32] Its absence can cause several disorders.[33–35] In the organism hydroxylysine is not directly incorporated in the nascent polypeptide chains of collagen,[31,36] rather hydroxylysine results from posttranslational hydroxylation of specific lysine residues by the enzyme lysyl hydroxylase.[31,33,37] Since collagen can only be obtained by isolation from animal sources or cell-cultures originating from animal sources, direct incorporation leads the way to produce collagen and its derivative gelatin without harming animals.[38] Moreover, *in vitro* replacement enables simple, controlled modifications in collagen because the amino acid sequence can easily be changed in the expression system, opening possibilities to adjust the properties of collagen. Modified collagens are important for medical purposes like skin or bone tissue engineering.[39,40] Another purpose of a direct hydroxylysine incorporation is the engineering of proteins with cross-linking activity.[32,35]

We showed earlier[17] that even trace quantities of lysine in the expression system will prevent the efficient incorporation of hydroxylysine as a surrogate. Lysine in the system originates from its preparation. The release of free lysine due to proteolysis is negligible.[29] We call 'residual lysine' the trace quantities of lysine in the virgin extract that can be incorporated into proteins.

Here, as a first step, we investigated the kinetics of green fluorescent protein production as a function of the amount of freshly added lysine. The comparison of the kinetics in a virgin and in a lysine depleted system helped us to see that freshly added and residual lysine are incorporated at different time intervals. In a second step we produced lysine depleted cell-free expression systems that minimize the inhibitory effect of residual lysine on the incorporation of hydroxylysine. For this purpose, we used ssDNA coated beads in order to remove residual lysine from the cell-free expression system.[29] With such a lysine depleted system we were able to produce proteins containing hydroxylysine at all lysine sites.

2. Material and methods

2.1. Expression system

The *E. coli* based cell-free expression system was prepared according to.[7,14]

We used pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500 as a reporter plasmid to visualize gene expression. This plasmid is derived from the plasmid pBest-Or2-Or1-Pr-UTR1-deGFP-T500 (Addgene Cat# 40019).[14–16]

2.2. Lysine depletion

For removing residual lysine from the expression system we used a previously described method with minor modifications.[29] We did not perform the step of producing a dummy protein with high lysine content because the non-specific lysine removal by treatment with ssDNA coated beads was sufficient for our purposes. For treatment, biotinylated ssDNA (5′-Biotin-TEG-CGG CGG GCT TTG CTC GAG TTA GTG GTG ATG GTG ATG-3′, Metabion) was attached to streptavidin-coated magnetic beads T1 (Dynabeads® MyOne™ Streptavidin T1, Invitrogen).[43] Additionally, poly-*T*-coated beads (Dynabeads™ Oligo(dT)25, Invitrogen) were used. 400 µl of each ssDNA coated bead stock solution ($c_{Bead} = 5$ mg/ml) were mixed and prepared according to instructions.[44] 3 µl of GamS (Shortened lambda phage Gam protein, prevents linear DNA degradation[16,45,46]) (at 99 µM) were added to 30 µl of crude *E. coli* cell-free extract. This mixture was then added to the dried beads and incubated for 20 min at 23 °C and 230 rpm. As a reference, 30 µl of crude *E. coli* cell-free extract with 3 µl of GamS were incubated with the same parameters but without ssDNA coated beads.

2.3. Gene expression

During incubation, the following master mix (for 30 µl of crude *E. coli* cell-free extract with 3 µl of GamS) was prepared: to 15 µl amino acids (mixture containing all canonical amino acids at 6 mM, except Leu at 5 mM, but no lysine) we added 3.6 µl magnesium glutamate (at 100 mM), 2.7 µl potassium glutamate (at 3 M), 6.43 µl PGA-buffer (3-Phospho-Glyceric Acid) (14x) and 4.5 µl 40% (v/v) PEG8000. Further we added pBest-p15a-Or2-Or1-Pr-UTR1-deGFP-6xHis-T500 to get a final concentration of 10 nM and filled up with ultrapure water to get a total volume of 39 µl. The final concentration of amino acids, magnesium glutamate and potassium glutamate had to be adjusted for each batch of crude extract to obtain optimal gene expression.

Experiments investigating the incorporation of freshly added and residual lysine into proteins were performed as followed. After 20 min of incubation with (or without) ssDNA coated beads the crude extract including GamS was aliquoted to volumes of 4.4 µl (seven aliquots in one batch). We added 5.2 µl of the master mix to each aliquot. Different quantities of lysine were added to the obtained volume of 9.6 µl. The total reaction volume was adjusted to 12 µl by adding ultrapure water. For investigation of the incorporation of L-5-hydroxylysine (Hyl) (DL-5-Hydroxylysine hydrochloride, Sigma-Aldrich) into proteins, we performed analogous experiments but without aliquoting. The concentrations as given in this paper represent the concentration of the L-enantiomer, that is, half of the concentration of the racemate obtained from the manufacturer.

The incorporation of the freshly added and/or residual amino acid into deGFP was visualized by transferring 10 µl of samples to a microwell plate (Nunc™ 384-Well Optical Bottom Plates # 242764, Thermo Scientific), recording the time course of fluorescence-intensity with a microplate reader (POLARstar OPTIMA, BMG LABTECH) over 16 h at 29 °C.

2.4. Protein purification and mass spectrometry

Mass spectrometry was performed to confirm the incorporation of hydroxylysine into deGFP. For this purpose, we used immobilized metal ion affinity chromatography (His-Spin Protein Miniprep, Zymo Research) to purify the recombinantly produced protein. Protein concentrators (Pierce™ Protein Concentrators PES, 10 K MWCO, 0.5 mL, Thermo Scientific) were used to desalt and concentrate the purified protein solutions. For the latter we washed the purified protein two times with 500 µl ultrapure water followed by two wash steps with 500 µl 20% acetonitrile + 0.5% formic acid. The supernatant containing concentrated purified protein in 20% acetonitrile + 0.5% formic acid was injected into the mass spectrometer (SolariX FTICR-ESI, Bruker) for measurement. The obtained raw mass spectrums were deconvoluted using the Maximum Entropy algorithm of the Compass DataAnalysis software (Bruker Compass DataAnalysis Version 5.0, Bruker). Origin (Origin 2020b, OriginLab) was used for data evaluation.

GamS is a necessary ingredient for our method of lysine removal. It was extracted by His-Tag purification as well, producing strong peaks in the mass spectrum that interfered with deconvolution ($c(\text{GamS}) \gg c(\text{deGFP18Hyl})$). Consequently, the spectrum of GamS had to be

subtracted before deconvolution.

2.5. Sds-PAGE

We performed SDS-PAGE to strengthen the results obtained by mass spectrometry. For this purpose, we produced deGFP and purified, desalted and concentrated as described in section 2.4 except that washing exclusively consisted of four steps using 500 µl ultrapure water. The concentrated and purified protein was denatured in sample buffer (TruPAGE™ LDS Sample Buffer, Sigma-Aldrich). For SDS-PAGE we used precast gels (TruPAGE™ Precast Gels 10%, 10 × 10 cm, 12-well, Sigma-Aldrich) and MOPS buffer (TruPAGE™ Tris-MOPS SDS Express Running Buffer, Sigma-Aldrich).

3. Results and discussion

3.1. Kinetics of green fluorescent protein production as a function of supplemented lysine

We looked at the kinetics of protein production as a function of the amount of freshly added lysine in a virgin expression system and a lysine depleted one (Fig. 1). We were able to monitor real time protein production by choosing green fluorescent protein (deGFP) as a model protein and measuring the fluorescence-intensity of deGFP produced in our cell-free expression system.

Fig. 1 shows the time course of the fluorescence-intensity of the cell-free expression system resulting from deGFP production. We take the fluorescence-intensity as proportional to the amount of produced deGFP. Only for the highest concentrations of freshly added lysine we observed a single regime only. In all other cases we distinguished two different regimes: a first plateau of deGFP production was reached after a few hours. Then a conversion point emerged where deGFP production restarted to saturate again only later. The restart in production was more prominent in Fig. 1A than in Fig. 1B where an important fraction of the residual lysine was removed from the cell-free expression reaction ($c_{residual\ Lys,\ 1A} = 0.009$ µg/µl; $c_{residual\ Lys,\ 1B} = 0.002$ µg/µl, concentrations deduced from deGFP fluorescence-intensity as previously described[29]). Compared to Fig. 1A, in Fig. 1B the conversion point appeared much later, and the second increase was substantially weaker. Moreover, in both cases the addition of fresh lysine had an immediate effect on the increase in fluorescence-intensity during the first hours of the experiment. The position of the first plateau differed, while the difference between the first and the second saturation level remained almost constant ($\Delta_{S1S2,\ 1A} = 5445.33 \pm 347.36$; $\Delta_{S1S2,\ 1B} = 1715.7 \pm 409.86$, $c_{added\ Lys} = 0$ mM, …, 0.2 mM).

Note that the first saturation levels of fluorescence-intensity of the virgin cell-free expression system (Fig. 1A) and the final saturation levels of lysine depleted system (Fig. 1B) were comparable for different concentrations of lysine added (Fig. 2).

We interpret the observations above by the first plateau of the fluorescence-intensity representing the incorporation of the freshly added lysine into deGFP while the second plateau corresponds to the incorporation of residual lysine. The collapse from two plateaus to only one at high concentrations of freshly added lysine may well correspond to a suppression of the incorporation of residual lysine due to the high ratio of freshly added to residual lysine.

To verify our interpretation, we performed the same experiments using the stable lysine isotope $^{13}C_6\ ^{15}N_2$ L-lysine (LysIso) (L-Lysine-2HCl, 13C6, 15 N2 for SILAC, Thermo Scientific). This isotope has the same chemical properties as the non-isotopic isomer but differs in mass.[47–49] The label enabled us to detect the incorporation of isotopic lysine within the deGFP (18 lysine sites) since $M_{LysIso} - M_{Lys} = 8$ Da. For our purpose,

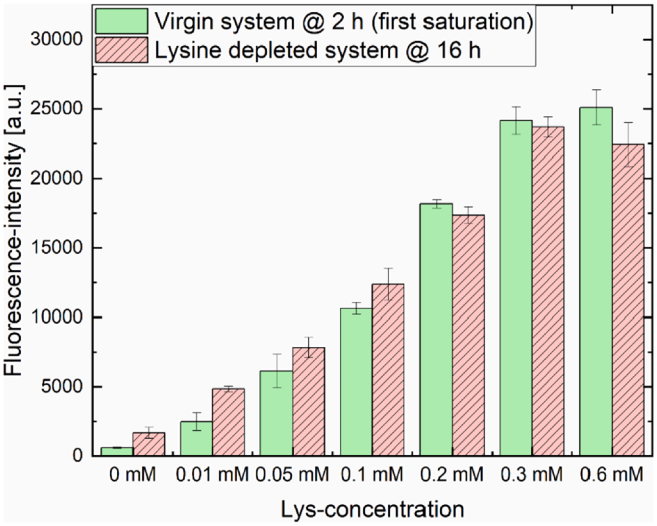

Fig. 2. Comparison of the first saturation level of deGFP production in a virgin system ($c_{residual\ Lys} = 0.009$ µg/µl) and the final saturation level of deGFP production in a lysine depleted system ($c_{residual\ Lys} = 0.002$ µg/µl). The graph shows the ratios of the values from Fig. 1A @ t = 2 h and Fig. 1B @ t = 16 h for different concentrations of freshly added lysine.

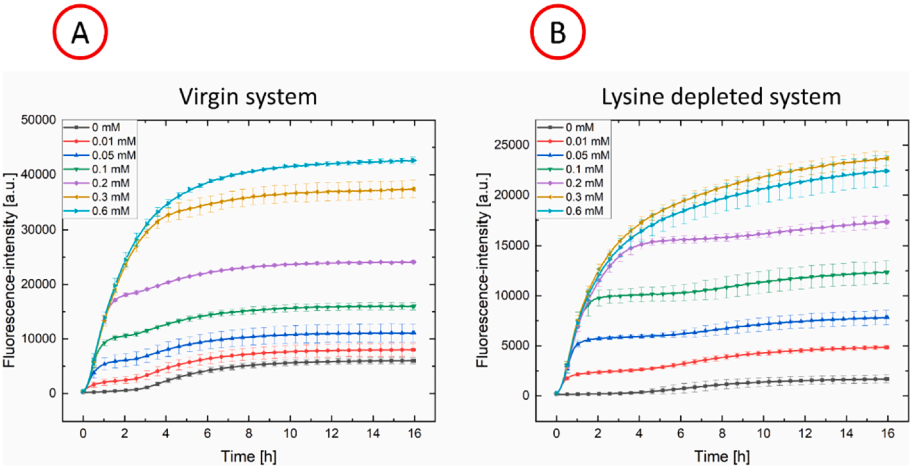

Fig. 1. Time course of the fluorescence-intensity of green fluorescent protein (deGFP) in our cell-free expression system for different concentrations of freshly added lysine (insert). The experiment was performed using a virgin system (A, residual lysine present) as well as a lysine depleted system (B, amount of residual lysine reduced by treatment with ssDNA coated beads).

we recorded a mass spectrum of deGFP from a virgin cell-free expression reaction with freshly added isotopic lysine (c = 0.1 mM) after 2 h of incubation (first plateau) and after 16 h (second plateau) at 29 °C. We also determined the mass spectrum after 16 h if no isotopic lysine (c = 0 mM) or a high amount of isotopic lysine (c = 0.6 mM) was added. The four mass spectra are shown in Fig. 3.

The results shown in Fig. 3 are in excellent agreement with our assumption regarding the kinetics of freshly added vs. residual lysine.

From the different kinetics discussed above, we conclude that there is a threshold in the concentration of freshly added lysine at which the incorporation of residual lysine is suppressed. In the following we show how to use this observation to achieve efficient hydroxylysine incorporation in green fluorescent protein.

3.2. Hydroxylysine incorporation

We investigated the incorporation of hydroxylysine into deGFP. Fig. 4 shows the chemical structure of L-hydroxylysine compared to L-lysine.

In a first step we tried to force our virgin cell-free expression system

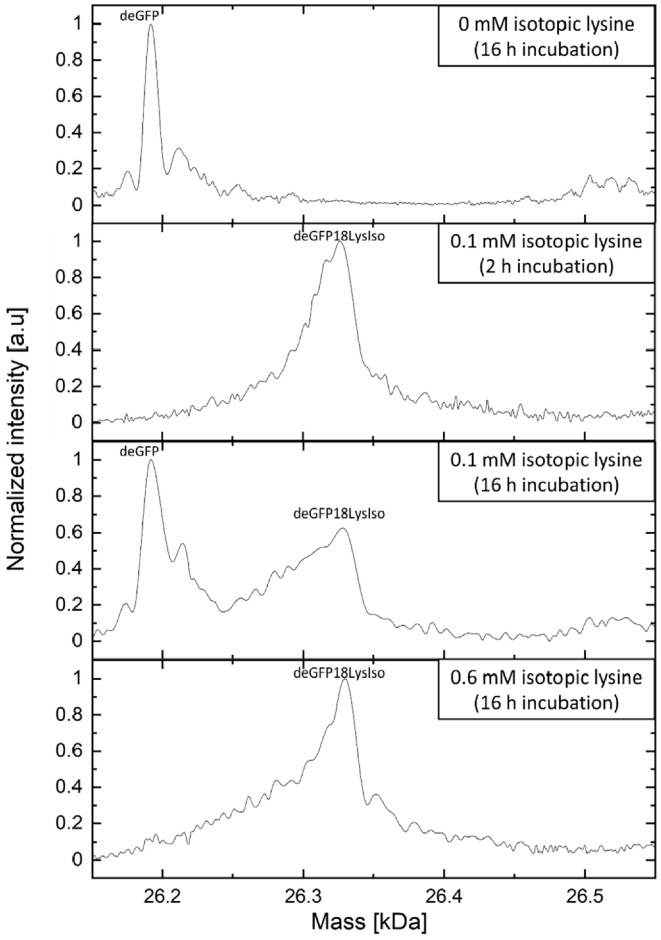

Fig. 3. Mass spectra of deGFP from different cell-free expression reactions, taken after 16 h of gene expression at 29 °C with 0 mM, 0.1 mM, or 0.6 mM of isotopic lysine added. Additionally, for 0.1 mM of isotopic lysine, a mass spectrum was recorded after 2 h of expression. Comparing the cases of no isotopic lysine added (c = 0 mM) and isotopic lysine at 0.6 mM, we understand that the incorporation of residual lysine (M_{deGFP} = 26.193 kDa) was suppressed by the presence of the isotope ($M_{deGFP18LysIso}$ = 26.337 kDa). In the case of 0.1 mM of isotopic lysine and 2 h of incubation we detected solely the isotopic peak, M = 26.333 kDa (the difference between measured and literature value of 4 Da is due to non-isotopic impurities resulting in broadened peaks). After 16 h, however, both peaks (isotopic and non-isotopic) were present.

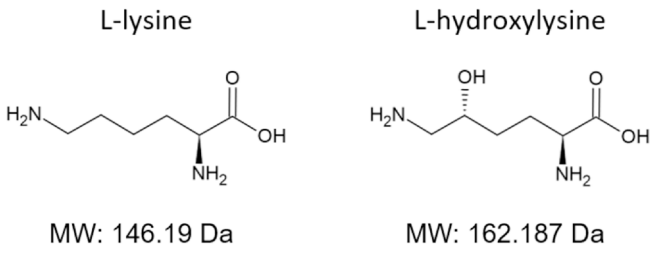

Fig. 4. Chemical structure of L-lysine (left) and L-hydroxylysine (right).

to accept hydroxylysine by increasing the amount of hydroxylysine. Fig. 5 shows the mass spectra for different, initial concentrations of hydroxylysine after 16 h of incubation. (Measurements after 2 h did not lead to any spectrum because of low protein content (data not shown).) We just observed peaks of correctly produced deGFP with different ratios of lysine to hydroxylysine. Peaks of ill-produced deGFP were not observed.

By increasing the concentration from 1 mM of hydroxylysine to 16.8 mM ($c_{Hyl}/c_{residual\ Lys} \approx 154$) we achieved comparable peak intensities for what we attribute to all possible combinations of lysine/hydroxylysine (deGFP, deGFP1Hyl, deGFP2Hyl, …, deGFP18Hyl) (Fig. 5). Even if the ratio of incorporated lysine to hydroxylysine can be controlled statistically, there is a huge spread in the distribution, and we were not able to achieve the production of proteins with a defined ratio. The peak corresponding to full replacement of lysine exhibited very low intensity. Further increase of the concentration of hydroxylysine to 50 mM ($c_{Hyl}/c_{residual\ Lys} \approx 458$) led to a spectrum showing different lysine/hydroxylysine combinations but with lower signal-to-noise ratio (figure S 2). We understand that high concentrations of hydroxylysine diminish the probability of protein production. This was confirmed by SDS-PAGE

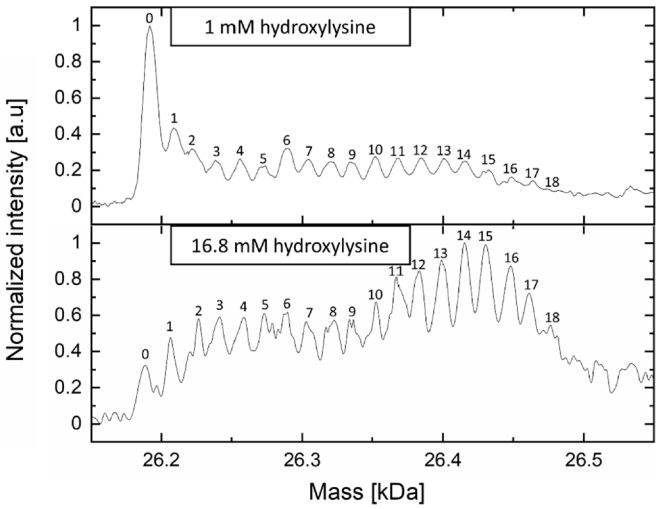

Fig. 5. Mass spectra from deGFP as obtained at two different concentrations of hydroxylysine (insert) supplemented to a virgin cell-free expression system. The spectra were recorded after 16 h of incubation at 29 °C. The numbers above the peaks indicate the number n of incorporated hydroxylysines. The number of incorporated lysines is 18-n. The mass spectrum for a concentration of 1 mM of hydroxylysine (top) exhibits a most intense peak from deGFP containing 18 lysines (M_{deGFP} = 26.193 kDa). This lysine corresponds to residual lysine from the virgin extract. Lysine was present at much lower concentrations than hydroxylysine ($c_{Hyl}/c_{residual\ Lys} \approx 9$, where c_{Hyl} corresponds to the amount of added hydroxylysine, $c_{residual\ Lys}$ determined as previously described[29]). All peaks corresponding to 1, 2, 3, …., 17, (18) lysines replaced by hydroxylysine ($M_{Hyl} - M_{Lys}$ = 16 Da; $M_{deGFPnHyl} = M_{deGFP} + n \cdot 16$ Da, n = 1, …, 18) were also present, albeit with clearly lower intensity. For higher concentrations of hydroxylysine (16.8 mM), the probability of hydroxylysine incorporation into deGFP is seen to increase.

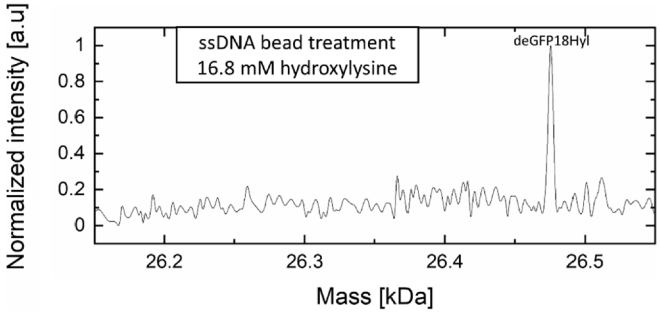

Fig. 6. Mass spectrum of deGFP produced from a cell-free expression system depleted in lysine, containing 16.8 mM of hydroxylysine. The spectrum was recorded after 16 h of incubation at 29 °C. We observed a peak at M = 26.476 kDa that corresponds to deGFP with 18 hydroxylysines instead of lysines ($M_{deGFP18Hyl}$ = 26.481 kDa, the difference between the measured and the literature value of 5 Da for $M_{deGFP18Hyl}$ was likely due to deconvolution errors caused by the background noise).

(Fig. 7). Our observation agrees well with the fact that for each batch of the expression system, the concentrations of amino acids need to be adjusted to optimize protein production rates. Moreover, at higher concentrations of hydroxylysine ($c_{Hyl} > 50$ mM), deGFP related peaks vanished from the mass spectrum (data not shown). We conclude that elevated concentrations of hydroxylysine are inhibitory to gene expression. To achieve full incorporation, it is essential to decrease the amount of residual lysine within the cell-free expression system.

For this reason, we added hydroxylysine to a lysine depleted expression system. At a concentration of 16.8 mM ($c_{Hyl}/c_{residual\ Lys} \approx$ 665), after 16 h of gene expression, we observed the presence of hydroxylysine only (Fig. 6).

Peaks representing deGFP species with different amounts of hydroxylysine and lysine were hardly visible, indicating the efficient incorporation of hydroxylysine into deGFP. Further increase of the hydroxylysine concentration entailed the absence of any related peaks (data not shown). We understand that high levels of hydroxylysine are, again, inhibitory.

The noise in the mass spectra was caused by GamS, present in the sample, interfering with deconvolution. Therefore, we performed SDS-PAGE to strengthen our results (Fig. 7).

Although the mass difference of 288 Da between native deGFP and deGFP18Hyl is difficult to detect by SDS-PAGE, we see that the deGFP indicating band in lane 2 (Dep Hyl) is positioned slightly higher than the band in lane 3 (Lys) strengthening the results presented in Fig. 6. Comparing the intensities of the corresponding bands, we observe that the amount of deGFP produced per cell-free expression reaction was much higher than the amount of deGFP18Hyl, produced under otherwise identical conditions.

Our method for lysine removal also depletes residual arginine. In the case of arginine, we observed similar kinetics as for lysine (see figures S 1A and S 1B). Our work represents a first step towards the incorporation of arginine surrogates as well.

4. Conclusion

In this paper we presented a method to fully replace lysine by hydroxylysine in green fluorescent protein. The expression system was designed to produce proteins just by adding DNA. It seems reasonable to assume that if our method works for GFP, it will also work for many other proteins. However, verification will be mandatory if one plans to realize a dedicated construct. To the best of our knowledge, the full replacement of lysine by hydroxylysine in proteins had not been achieved previously. Our method is based on the observation that freshly added lysine is incorporated earlier than residual lysine, which is incorporated only after a lag of a few hours (Fig. 1). Consequently,

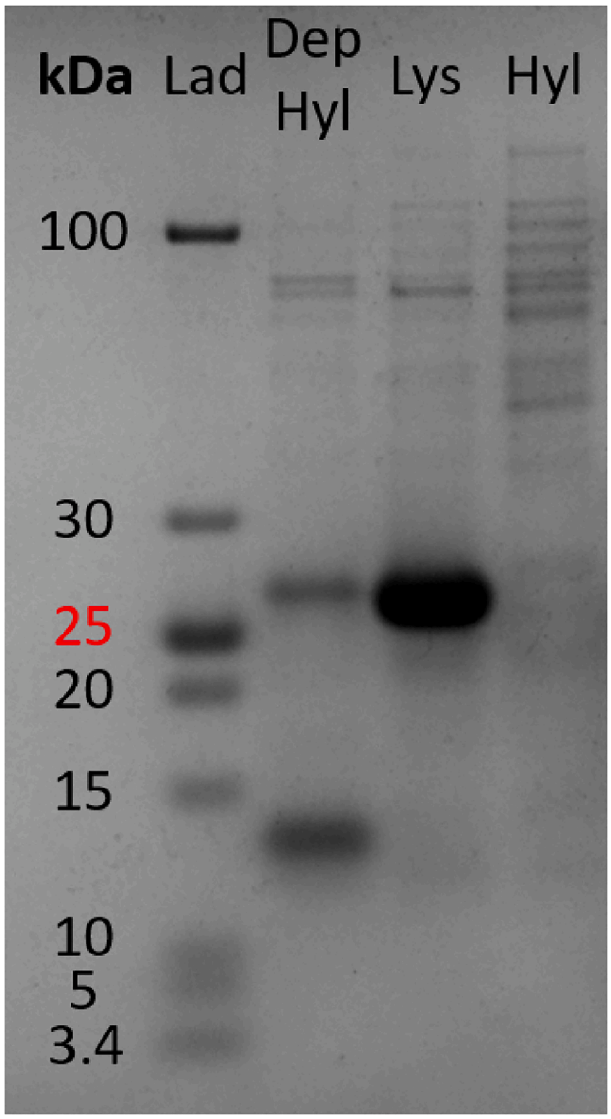

Fig. 7. SDS-PAGE of deGFP extracted from different preparations of the cell-free expression system after 16 h of incubation at 29 °C. The first lane (Lad) shows the ladder (PageRuler™ Unstained Low Range Protein Ladder, Thermo Scientific). The numbers left to this lane indicate the weight of the bands in kDa. The second lane (Dep Hyl) results from a lysine depleted cell-free expression system containing 16.8 mM of hydroxylysine. The third lane (Lys) is from a cell-free expression system containing 1 mM lysine and the fourth lane (Hyl) results from a cell-free expression system containing 50 mM hydroxylysine. In lane 2 (Dep Hyl) two bands indicate deGFP18Hyl ($M_{deGFP18Hyl}$ = 26.481 kDa) and GamS (M_{GamS} = 12.634 kDa), while lane 3 (Lys) contains a band indicating deGFP (M_{deGFP} = 26.193 kDa). Lane 4 (Hyl) shows no bands indicating deGFP production (high molecular weight bands are residues from the cell-free expression system). This confirms our assumption that high concentrations of hydroxylysine diminish the probability of protein production.

protein synthesis can be stopped sufficiently early (after 2 h) to prevent the incorporation of residual lysine. We showed that further depletion of residual lysine avoids its incorporation even at a later stage of protein synthesis (16 h), increasing the yield. We believe that the here proposed method represents a powerful tool that can be useful for the incorporation of other lysine and arginine surrogates as well.

Author contributions

AO and MF designed the research. MF performed the research with the help of SR and FG. MF analyzed data with the help of AO, SR and PH.

MF and AO wrote the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the collaborative research center SFB 1027 funded by the Deutsche Forschungsgemeinschaft (DFG), and by the Human Frontier Science Program (HFSP, RGP0037/2015). We thank Emanuel Worst for help and fruitful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bmc.2021.116207.

References

1 Ulmer K. Protein engineering. *Science*. 1983;219:666–671. https://doi.org/10.1126/science.6572017.
2 Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol*. 1992;224:771–782. https://doi.org/10.1016/0022-2836(92)90561-W.
3 Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol*. 1996;6:178–182. https://doi.org/10.1016/s0960-9822(02)00450-5.
4 Tan SC, Yiap BC. DNA, RNA, and protein extraction: the past and the present. *J Biomed Biotechnol*. 2009;2009:1–10. https://doi.org/10.1155/2009/574398.
5 Chen R. Bacterial expression systems for recombinant protein production: E. coli and Beyond. *Biotechnol Adv*. 2012;30:1102–1107. https://doi.org/10.1016/j.biotechadv.2011.09.013.
6 Fang G-M, Li Y-M, Shen F, et al. Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chemie*. 2011;50:7645–7649. https://doi.org/10.1002/anie.201100996.
7 Sun ZZ, Hayes CA, Shin J, Caschera F, Murray RM, Noireaux V. Protocols for implementing an Escherichia Coli based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp*. 2013;79:1–14. https://doi.org/10.3791/50762.
8 Spicer CD, Davis BG. Selective chemical protein modification. *Nat Commun*. 2014;5:1–14. https://doi.org/10.1038/ncomms5740.
9 Rashidian M, Dozier JK, Distefano MD. Enzymatic labeling of proteins: techniques and approaches. *Bioconjug Chem*. 2013;24:1277–1294. https://doi.org/10.1021/bc400102w.
10 Zoller MJ. New molecular biology methods for protein engineering. *Curr Opin Biotechnol*. 1991;2:526–531. https://doi.org/10.1016/S0959-440X(05)80084-3.
11 Lu Y. Cell-free synthetic biology: engineering in an open world. *Synth Syst Biotechnol*. 2017;2:23–27. https://doi.org/10.1016/j.synbio.2017.02.003.
12 Muir TW, Sondhi D, Cole PA. Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA*. 1998;95:6705–6710. https://doi.org/10.1073/pnas.95.12.6705.
13 Gavrilescu M, Chisti Y. Biotechnology - a sustainable alternative for chemical industry. *Biotechnol Adv*. 2005;23:471–499. https://doi.org/10.1016/j.biotechadv.2005.03.004.
14 Shin J, Noireaux V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *J Biol Eng*. 2010;4:1–9. https://doi.org/10.1186/1754-1611-4-8.
15 Schenkelberger M, Shanak S, Finkler M, et al. Expression regulation by a methyl-CpG binding domain in an E. coli based, cell-free TX-TL system. *Phys Biol*. 2017;14:1–11. https://doi.org/10.1088/1478-3975/aa5d37.
16 Finkler M, Ott A. Bead-based assay for spatiotemporal gene expression control in cell-free transcription-translation systems. *Biotechniques*. 2019;66:29–33. https://doi.org/10.2144/btn-2018-0097.
17 Worst EG, Exner MP, de Simone A, et al. Residue-specific incorporation of noncanonical amino acids into model proteins using an escherichia coli cell-free transcription-translation system. *J Vis Exp*. 2016;114:1–14. https://doi.org/10.3791/54273.
18 Singh-Blom A, Hughes RA, Ellington AD. An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. *J Biotechnol*. 2014;178:12–22. https://doi.org/10.1016/j.jbiotec.2014.02.009.
19 Johnson JA, Lu YY, Van Deventer JA, Tirrell DA. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr Opin Chem Biol*. 2010;14:774–780. https://doi.org/10.1016/j.cbpa.2010.09.013.
20 Sakamoto K, Hayashi A, Sakamoto A, et al. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res*. 2002;30:4692–4699. https://doi.org/10.1093/nar/gkf589.
21 Goerke AR, Swartz JR. High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnol Bioeng*. 2009;102:400–416. https://doi.org/10.1002/bit.22070.
22 Wan W, Huang Y, Wang Z, et al. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia Coli. *Angew Chemie*. 2010;49:3211–3214. https://doi.org/10.1002/anie.201000465.
23 Gao W, Cho E, Liu Y, Lu Y. Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. *Front Pharmacol*. 2019;10:1–8. https://doi.org/10.3389/fphar.2019.00611.
24 Yang B, Ayyadurai N, Yun H, et al. In vivo residue-specific dopa-incorporated engineered mussel bioglue with enhanced adhesion and water resistance. *Angew Chemie*. 2014;53:13360–13364. https://doi.org/10.1002/anie.201406099.
25 Yoo TH, Tirrell DA. High-throughput screening for methionyl-TRNA synthetases that enable residue-specific incorporation of noncanonical amino acids into recombinant proteins in bacterial cells. *Angew Chemie*. 2007;46:5340–5343. https://doi.org/10.1002/anie.200700779.
26 Worst EG, Exner MP, De Simone A, et al. Cell-free expression with the toxic amino acid canavanine. *Bioorg Med Chem Lett*. 2015;25:3658–3660. https://doi.org/10.1016/j.bmcl.2015.06.045.
27 Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci*. 1974;71:4135–4139. https://doi.org/10.1073/pnas.71.10.4135.
28 Ninio J. Kinetic amplification of enzyme discrimination. *Biochimie*. 1975;57:587–595. https://doi.org/10.1016/S0300-9084(75)80139-8.
29 Finkler M, Kurt Ö, Grimm F, Hartz P, Ott A. A bead-based method for the removal of the amino acid lysine from cell-free transcription-translation systems. *J Biotechnol X*. 2020;8:1–6. https://doi.org/10.1016/j.btecx.2020.100024.
30 Dunn JA, McCance DR, Thorpe SR, Lyons TJ, Baynes JW. Age-Dependent Accumulation of Nε-(Carboxymethyl)Lysine and Nε- (Carboxymethyl)Hydroxylysine in Human Skin Collagen. *Biochemistry*. 1991;30:1205–1210. https://doi.org/10.1021/bi00219a007.
31 Cudic M, Lauer-Fields JL, Fields GB. Improved synthesis of 5-hydroxylysine (Hyl) derivatives. *J Pept Res*. 2005;65:272–283. https://doi.org/10.1111/j.1399-3011.2005.00215.x.
32 Cudic M, Fields GB. Modulation of receptor binding to collagen by glycosylated 5-hydroxylysine: chemical biology approaches made feasible by Carpino's Fmoc Group. *Pept Sci*. 2020;112:1–12. https://doi.org/10.1002/pep2.24156.
33 Krane SM, Pinnell SR, Erbe RW. Lysyl-protocollagen hydroxylase deficiency in fibroblasts from siblings with hydroxylysine-deficient collagen. *Proc Natl Acad Sci USA*. 1972;69:2899–2903. https://doi.org/10.1073/pnas.69.10.2899.
34 Burkhardt A, Grupe G. Hydroxylysine deficiency, conspicuous skeletal lesions and a strange burial practice in a historical german island population. *Int J Osteoarchaeol*. 2018;28:227–236. https://doi.org/10.1002/oa.2649.
35 Gjaltema RAF, Bank RA. Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. *Crit Rev Biochem Mol Biol*. 2017;52:74–95. https://doi.org/10.1080/10409238.2016.1269716.
36 Popenoe EA, Aronson RB, Van Slyke DD. Hydroxylysine formation from lysine during collagen biosynthesis. *Proc Natl Acad Sci*. 1966;55:393–397. https://doi.org/10.1073/pnas.55.2.393.
37 Makris EA, MacBarb RF, Responte DJ, Hu JC, Athanasiou KA. A copper sulfate and hydroxylysine treatment regimen for enhancing collagen cross-linking and biomechanical properties in engineered neocartilage. *FASEB J*. 2013;27:2421–2430. https://doi.org/10.1096/fj.12-224030.
38 Deane CS, Bass JJ, Crossland H, Phillips BE, Atherton PJ. Animal, plant, collagen and blended dietary proteins: effects on musculoskeletal outcomes. *Nutrients*. 2020;12:1–35. https://doi.org/10.3390/nu12092670.
39 Maisani M, Ziane S, Ehret C, et al. A new composite hydrogel combining the biological properties of collagen with the mechanical properties of a supramolecular scaffold for bone tissue engineering. *J Tissue Eng Regen Med*. 2018;12:e1489–e1500. https://doi.org/10.1002/term.2569.
40 Ma L, Gao C, Mao Z, et al. Collagen/chitosan porous scaffolds with improved biostability for skin tissue engineering. *Biomaterials*. 2003;24:4833–4841. https://doi.org/10.1016/S0142-9612(03)00374-0.
41 Gerrits M, Budisa N, Merk H. Site-specific chemoselective pyrrolysine analogues incorporation using the cell-free protein synthesis system. *ACS Synth Biol*. 2019;8:381–390. https://doi.org/10.1021/acssynbio.8b00421.
42 Agostini F, Völler JS, Koksch B, Acevedo-Rocha CG, Kubyshkin V, Budisa N. Biocatalysis with unnatural amino acids: enzymology meets xenobiology. *Angew Chemie - Int Ed*. 2017;56:9680–9703. https://doi.org/10.1002/anie.201610129.
43 Invitrogen. Manual for DynabeadsTM MyOneTM Streptavidin T1; 2017.
44 Ambion. Manual for Dynabeads® Oligo (DT)25; 2012.
45 Garamella J, Marshall R, Rustad M, Noireaux V. The all E. coli TX-TL Toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth Biol*. 2016;5:344–355. https://doi.org/10.1021/acssynbio.5b00296.
46 Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM. Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system. *ACS Synth Biol*. 2014;3:387–397. https://doi.org/10.1021/sb400131a.
47 Ong S-E, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002;1:376–386. https://doi.org/10.1074/mcp.M200025-MCP200.
48 Ong S-E, Kratchmarova I, Mann M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res*. 2003;2:173–181. https://doi.org/10.1021/pr0255708.
49 Ong S-E. The expanding field of SILAC. *Anal Bioanal Chem*. 2012;404:967–976. https://doi.org/10.1007/s00216-012-5998-3.