Engineering of the Recombinant Expression and PEGylation Efficiency of the Therapeutic Enzyme Human Thymidine Phosphorylase

Christos S. Karamitros*†, Catrina M. Somody, Giulia Agnello and Scott Rowlinson*†

Aeglea Biotherapeutics, Austin, TX, United States

Human thymidine phosphorylase (HsTP) is an enzyme with important implications in the field of rare metabolic diseases. Defective mutations of HsTP lead to mitochondrial neurogastrointestinal encephalomyopathy (MNGIE), a disease with a high unmet medical need that is associated with severe neurological and gastrointestinal complications. Current efforts focus on the development of an enzyme replacement therapy (ERT) using the *Escherichia coli* ortholog (EcTP). However, bacterial enzymes are counter-indicated for human therapeutic applications because they are recognized as foreign by the human immune system, thereby eliciting adverse immune responses and raising significant safety and efficacy risks. Thus, it is critical to utilize the HsTP enzyme as starting scaffold for pre-clinical drug development, thus de-risking the safety concerns associated with the use of bacterial enzymes. However, HsTP expresses very poorly in *E. coli*, whereas its PEGylation, a crucial chemical modification for achieving long serum persistence of therapeutic enzymes, is highly inefficient and negatively affects its catalytic activity. Here we focused on the engineering of the recombinant expression profile of HsTP in *E. coli* cells, as well as on the optimization of its PEGylation efficiency aiming at the development of an alternative therapeutic approach for MNGIE. We show that phylogenetic and structural analysis of proteins can provide important insights for the rational design of N'-terminus-truncation constructs which exhibit significantly improved recombinant expression levels. In addition, we developed and implemented a criteria-driven rational surface engineering strategy for the substitution of arginine-to-lysine and lysine-to-arginine residues to achieve more efficient, homogeneous and reproducible PEGylation without negatively affecting the enzymatic catalytic activity upon PEGylation. Collectively, our proposed strategies provide an effective way to optimize enzyme PEGylation and *E. coli* recombinant expression and are likely applicable for other proteins and enzymes.

Keywords: enzyme engineering, catalytic activity, recombinant expression, PEGylation, thymidine phosphorylase
INTRODUCTION

Thymidine phosphorylases (E.C. 2.4.2.4.) are enzymes that play a pivotal role in the salvage pathway of pyrimidine nucleoside metabolism in both lower and higher organisms (Desgranges et al., 1981; Okuyama et al., 1996; Pugmire and Ealick, 2002). They catalyze the reversible phosphorylation of deoxycytidine (dThd) and deoxyuridine (dUrd) to 2-deoxyribose-1-phosphate and to their respective bases, with the latter being fluxed towards nucleotide synthesis. Humans have three cytosolic pyrimidine nucleoside phosphorylases, namely a thymidine phosphorylase (HsTP; encoded by TYMP gene) and two uridine phosphorylases (HsUP1 & HsUP2 encoded by UPP1 and UPP2 genes respectively) that have totally distinct structural and biochemical features (Johansson, 2003; Norman et al., 2004). HsTP does not show any significant amino acid sequence similarity against any of the HsUP enzymes whereas HsUP1 and HsUP2 share 66% sequence identity. HsTP is a highly specific N'-ribosyl phosphorylase displaying $k_{cat}/K_M$ values against dThd and dUrd in the range of $10^5 \text{ M}^{-1} \text{ s}^{-1}$ and $10^4 \text{ M}^{-1} \text{ s}^{-1}$ respectively (Deves et al., 2014; Schwartz et al., 2010), whereas HsUP enzymes exhibit higher specificity against uridine (Urd) ($k_{cat}/K_M \sim 10^6 \text{ M}^{-1} \text{ s}^{-1}$) and to a lesser extent towards dUrd and dThd ($k_{cat}/K_M \sim 10^4 \text{ M}^{-1} \text{ s}^{-1}$) (Liu et al., 1998; Renck et al., 2010). Thymidine phosphorylases from lower organisms like Bacillus steaothermophilus (BsTP) and Escherichia coli (EcTP) exhibit similar fold and belong to the same family (family II) of nucleoside phosphorylases with HsTP, while their amino acid sequence identity is ~40%. Particularly, EcTP is highly active against dThd showing a $k_{cat}/K_M \sim 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (Panova et al., 2008; Gbai et al., 2006).

HsTP has attracted the interest of intense research during the last decades primarily due to its implication in tumor cell growth (Matsushita et al., 1999; Moghaddam et al., 1995). Importantly, it has been shown that the platelet-derived endothelial cell growth factor (PD-ECGF) is identical to HsTP and is involved in angiogenic and endothelial cell chemotactic activities (Sumizawa et al., 1993). Overexpression and elevated HsTP activity have been detected and measured in plasma of cancer patients and have been associated with invasive and metastatic profiles of various cancer types including bladder, colorectal, ovarian and pancreatic tumors (Pauly et al., 1977; Takebayashi et al., 1996a; Takebayashi et al., 1996b; Reynolds et al., 1994). HsTP has been shown to degrade the anti-viral drug trifluorothymidine (TFT) as well as other fluoropyrimidines and chemotherapeutic agents (Patterson et al., 1995). These observations prompted the in-depth study of HsTP’s kinetic properties and the elucidation of its catalytic mechanism (Schwartz et al., 2010; Birck and Schramm, 2004; Oh and el Kouni, 2018) to aid the development of potent inhibitors (de Moura Sperotto et al., 2019; Pérez-Pérez et al., 2005). Indeed, the inhibitor 5-chloro-6-[-(2-iminopyridinyl) methyl] uracil hydrochloride (TP1) is one of the most potent HsTP inhibitors (Ki ~ 20 nM), capable of inducing considerable reduction in tumor growth in mice (Matsushita et al., 1999).

Beyond its role and implications in tumor biology, HsTP has also been studied within the context of rare metabolic diseases. Mutations in the HsTP gene (TYMP) which result in either partial or complete loss of enzymatic activity, cause mitochondrial neurogastrointestinal encephalomyopathy (MNGIE) which is an autosomal recessive disease (Slama et al., 2005). HsTP deficiency leads to elevated systemic levels of dThd and dUrd and patients with MNGIE suffer from neurological and gastrointestinal symptoms (Hirano, 1993; Teitelbaum et al., 2002). It has been well-documented that the underlying mechanism for all these symptoms and manifestations is associated with a disorder known as mitochondrial DNA (mtDNA) depletion syndrome (Pacitti et al., 2018; El-Hattab and Scaglia, 2013). That is, build-up of dThd and dUrd leads to an imbalance of nucleotide levels which, in turn, negatively affects the mitochondrial DNA replication that relies on the cytoplasmic pool of nucleotides for proper function. Current management of MNGIE has been primarily supportive with the main goal being to ease the disease manifestations. Thus, MNGIE therapy represents an unmet medical need that requires further attention for the development of effective treatments.

Recent efforts for the development of MNGIE therapies have centered on the autologous hematopoietic stem cell transplantation (HSCT) (Halter et al., 2015), on AAV- and gene-editing-based gene therapy approaches (Torres-Torronteras et al., 2018; Torres-Torronteras et al., 2014; Torres-Torronteras et al., 2016; Parés et al., 2021), as well as on an enzyme replacement therapy (ERT) utilizing EcTP which is encapsulated inside erythrocytes (EE-EcTP) for long plasma persistence and immunogenicity minimization (Bax et al., 2019). However, the HSCT has exhibited a high mortality rate of ~63% (Yadak et al., 2018; Filosto et al., 2012) and the gene therapy approaches, while highly promising, have been explored only in preclinical settings thus far. In addition, the erythrocyte encapsulation strategy that utilizes the bacterial EcTP is associated with a high treatment burden because it requires hospital visits for the administration of the drug. Importantly, bacterial enzymes as human therapeutics are often encumbered with significant safety and efficacy risks because they are recognized as foreign species by the human immune system and, in turn, elicit adverse immune responses and clearing anti-drug antibodies (Pokrovsky et al., 2016; Lipsky et al., 2014). Thus, the use of HsTP for the development of an ERT could represent a more advantageous and potentially safer approach relative to the aforementioned strategies. Unlike in rare diseases where patients completely lack the endogenous protein (e.g., large deletions and intron inversions in haemophilia A) (Levinson et al., 1992) and administration of its functional form can elicit immune responses, in case of MNGIE the vast majority of the mutations result in poor catalytic activity and not in pre-mature termination of the translation. Therefore, most patients would be considered to have natural, cross-reactive immunogenic HsTP. As a result, they are tolerized against HsTP and unlikely to develop significant antibody response to ERT.

In an attempt to form the basis for an alternative therapeutic approach for MNGIE, in the present study we report on the engineering of recombinant expression properties of HsTP in E. coli, as well as its PEGylation optimization. Expression of the
mature wild-type HsTP in *E. coli* yielded low levels of total active protein. Analysis of the HsTP crystal structure and direct comparison to its *E. coli* ortholog combined with phylogenetic analysis allowed the design of N'-terminal truncations that did not only improve the recombinant expression levels by several fold, but also improved the enzyme’s catalytic activity against its secondary substrate dUrd. In addition, random lysine PEGylation of wild-type HsTP resulted in poor PEGylation efficiency, as evidenced by the presence of multiple, distinct PEG-mer molecule species while it negatively affected its catalytic activity against its natural substrates. By developing and implementing a rational surface engineering strategy, we identified suitable arginine residues which were substituted with lysine for more efficient, homogeneous and reproducible PEGylation. Conversely, lysine residues whose conjugation with PEG could negatively impact catalytic activity were substituted with arginine to prevent the cross-linking chemical modification, thereby maintaining equal levels of catalytic activity prior and post PEGylation. While ameliorating the biological symptoms is the ultimate goal of therapeutics, an essential component, particularly for protein therapeutics, is the ability to efficiently and reproducibly manufacture them. Our present study proposes two engineering strategies, for the PEGylation and recombinant expression optimization of HsTP, that could be applied towards the development of a parenteral enzyme therapy for MNGIE.

**MATERIALS AND METHODS**

**Cloning of HsTP and EcTP Genes**

Amino acid sequences for wild-type HsTP and EcTP were codon-optimized for *E. coli* using the online tool provided by Integrated DNA Technologies (https://www.idtdna.com/pages/tools/codon-optimization-tool), and gene blocks with the resultant sequences were used as a template for polymerase chain reaction (PCR) amplification. Primers were designed to amplify the HsTP and EcTP gene fragments as follows: the 5′ primer introduced a His6-tag at the N-terminus and an NcoI restriction site; the 3′ primer introduced an EcoRI restriction site. PCR amplification was carried out following the Kapa HiFi protocol (Roche) with an annealing temperature of 60°C and a 2-min extension at 72°C for a total of 25 cycles. The amplicons were digested with NcoI and EcoRI, gel-purified, ligated into pET28a (Novagen) with T4 DNA ligase, and transformed into MC1061 cells (Lucigen). Primers used to generate the variants in this study are provided in Supplementary Table S2.

**FIGURE 1** | The crystal structures and the constructs of HsTP and EcTP. (A) Overlaid crystal structures of HsTP (PDB: 2WK6) and EcTP (PDB: 4LHM) shown in cyan and orange color respectively. Dashed arrow indicates the N'-terminus of each enzyme. (B) Zoomed, overlaid α-helices located at the N'-terminus of HsTP (cyan) and EcTP (orange). Q35 residue shown in red sticks is the first amino acid residue that is structurally-resolved in case of HsTP, with F2 and L3 (shown in orange sticks) occupying the respective structurally-equivalent positions of EcTP. (C) N'-terminal domain of HsTP. Unstructured region without electron density is shown as circles, where each circle represents the respective amino acid residue. Pro-peptide corresponds to the first ten residues and are shown in yellow. Residues A11, S19 and K34 are shown in orange, green and dark pink and represent the positions at which the HsTP constructs were truncated (constructs were truncated at the residue preceding A11, S19 and K34). (D) Cartoon representation of all the HsTP and EcTP constructs tested in the present study. The position of His6-tag as well as the truncation positions are also shown. Starting methionine and a glycine residue (result of the cloning process as described in the Methods section) are preceding the starting position (A11, S19, K34, F2) of all constructs shown in this panel. Structural representations were prepared using PyMol 2.4.1 (Schrödinger, LLC).
TABLE 1 | Summary of the HsTP and EcTP constructs and their expression yields.

| Construct | His®-tag | Range of expression yield/L | Plasmid | Medium | Temperature (°C) | IPTG (mM) |
|-----------|----------|-----------------------------|---------|--------|-----------------|-----------|
| HsTP199   | N’       | 3–5 mg                      | pET28a  | TB     | 30              | 0.5       |
| HsTP215   | C’       | 0.5–1 mg                    | pET28a  | TB     | 30              | 0.5       |
| HsTP216   | C’       | 1–2 mg                      | pET28a  | TB     | 30              | 0.5       |
| HsTP217   | C’       | 4–5 mg                      | pET28a  | TB     | 30              | 0.5       |
| HsTP218   | C’       | 2.5–3 mg                    | pET28a  | TB     | 30              | 0.5       |
| HsTP216   | C’       | 50–70 mg                    | pET28a  | TB     | 16              | 0.1       |
| EcTP      | N’       | 100–120 mg                   | pET28a  | TB     | 30              | 0.5       |

*Expression yield refers to the soluble protein that was recovered after employing the standard 2-step purification scheme described in the Methods section.

clones were screened by colony PCR using as the forward primer, one specific primer for the insert and as the reverse primer, one specific primer for the plasmid backbone.

Expression and Purification of HsTP and EcTP Constructs

Expression

HsTP and EcTP constructs either in BL21 (DE3) or C41 (DE3) cells were cultured in terrific broth (TB) media supplemented with 50 μg/ml of kanamycin at 37°C until OD at 600 nm reached between and 0.8 and 1 units when standard shake flasks were used, or OD of 5 when ultra-yield flasks (Thompson) were used. Then, for HsTP199 and EcTP, 0.5 mM IPTG was added, temperature was reduced to 30°C, and expression was allowed to proceed for 20–22 h. Expression-optimized variants, including HsTP218, were induced with 0.1 mM IPTG and expressed at 16°C for 40 h.

Purification by Immobilized Metal Affinity Chromatography

Cells were pelleted by centrifugation at 4°C and resuspended in lysis buffer consisting of 50 mM Na₂HPO₄, 300 mM NaCl, 10 mM imidazole, 1 mM PMSF protease inhibitors, and 1 μl of 1 mg/ml DNase I for each ml of cell suspension, at pH 8. Resuspended samples were kept on ice and lysed by sonication. After sonication, the samples were centrifuged at 12,000 xg for 1 h at 4°C, the supernatant was decanted, filtered through a 0.2 μm filter, and the pellet discarded. Lysate was mixed with Ni-NTA resin (Qiagen) (Ni-NTA resin volume was in the range of 4–8 ml of slurry mixture depending on the expected protein amount in the cell lystate) which had been previously equilibrated with 20 bed volumes of lysis buffer (50 mM Na₂HPO₄, 300 mM NaCl, 10 mM Imidazole, pH 8). The lysate-resin mixture was incubated at 4°C for 2 h (batch purification) and subsequently was applied by gravity flow to a polypropylene column. The column was washed with 20 bed volumes of washing buffer (50 mM Na₂HPO₄, 300 mM NaCl, 20 mM Imidazole, pH 8) followed by elution with three bed volumes of elution buffer (50 mM Na₂HPO₄, 300 mM NaCl, 300 mM Imidazole, pH 8). The eluted protein was buffer-exchanged against 20 mM Tris-Cl and 20 mM NaCl, pH 7.5 using 10 kDa MWCO protein concentrator tubes (Amicon) and was subjected to ion-exchange chromatography described below.

Q-Sepharose Purification

Following buffer exchange, the IMAC-eluted protein was applied by gravity flow to a polypropylene column loaded with 2 ml of Q-FF resin which had been previously equilibrated with 20 bed volumes of binding and washing buffer (20 mM Tris-Cl, 20 mM NaCl, pH 7.5). The column was washed with 20 bed volumes of binding and washing buffer, then eluted with a 50–500 mM NaCl gradient. Purity of the protein elutions was assessed by SDS-PAGE, and the fractions with the highest concentration and purity were pooled. HsTP was observed to elute between 100 and 200 mM NaCl. Purified HsTP was aliquoted, mixed with 15% (v/v) final glycerol concentration, flash-frozen with liquid nitrogen, and stored at −80°C for future use.

Immunoblotting

Immunoblotting for the detection of expressed HsTP was carried out according to standard protocols (Slieman and Leheste, 2020). Briefly, protein concentration was calculated using NanoDrop (One microvolume UV-Vis spectrophotometer from Thermo Scientific) by recording the A₂₈₀. For crude extracts, A₂₈₀ of 1 was assumed to be equivalent to 1 mg/ml whereas for HsTP and EcTP, the respective theoretical molar extinction coefficient values (ε²₈₀ for HsTP = 23,490 M⁻¹ cm⁻¹ and ε²₈₀ for EcTP = 24,410 M⁻¹ cm⁻¹ as calculated from Expasy’s protparam online tool based on the primary amino acid sequence of the proteins) were used for the conversion of absorbance to molar concentration. Typical protein amount loaded per well was in the range of 10–15 μg. The protein samples were analyzed by SDS-PAGE and subsequently, the separated proteins were transferred onto a nitrocellulose membrane in transfer buffer (25 mM Tris-Cl, 190 mM glycine, 0.1% SDS) at a fixed current of 10 mA overnight at 4°C. Membranes were blocked with 5% skim milk dissolved in TBST buffer for 2 h at room temperature, followed by incubation with monoclonal anti-His® antibodies (Sigma-Aldrich, SAB2702218). Incubation with secondary goat anti-mouse HRP-linked antibodies (ThermoFisher Scientific, G-21040) was carried out for 1 h at room temperature. Upon immunoblotting, bands were detected using an enhanced chemiluminescence
(ECL) kit (SuperSignal West Pico PLUS from ThermoFisher Scientific, 34580).

**Steady-State Kinetic Analysis of HsTP and EcTP**

Steady-state kinetic characterization of HsTP and EcTP against dThd and dUrd was performed by continuously monitoring the decrease in absorbance of dThd and dUrd upon depyrimidination at 290 and 282 nm, respectively. Enzyme concentration in the range of 10–20 nM was used for all the steady-state kinetic measurements. Reactions took place in assay buffer consisting of 20 mM Hepes and 50 mM KH$_2$PO$_4$, pH 7.5, in a final volume of 1 ml placed in UV cuvettes with a pathlength of 1 cm. Prior to the addition of the enzyme, cuvettes were equilibrated at 37°C for 20 min using a heating block (VWR). The reaction progress was typically monitored for 2 min using a Jasco V750 spectrophotometer with a temperature-controlled cuvette holder and the absorbance was converted to concentration using the extinction coefficient of dThd and dUrd ($\Delta\varepsilon_{290} = 1,000$ M$^{-1}$ cm$^{-1}$ and $\Delta\varepsilon_{282} = 1,370$ M$^{-1}$ cm$^{-1}$) (Krenitsky et al., 1976). The obtained $v/[E]$ (initial velocity/total enzyme concentration) values from the linear region of the reaction progress curves with <10% of substrate conversion were plotted against the respective substrate concentrations, and the steady-state kinetic parameters $k_{cat}$ and $k_{cat}/K_M$ were calculated by nonlinear regression using the Michaelis–Menten model (Eq. 1 below) analyzed by the SoftZymics software (Igor Pro, Wavemetrics).

$$v = (k_{cat} \times [S])/(K_M + S)$$  \hspace{1cm} (1)
TABLE 2 | Steady-state kinetic parameters of HsTP199, HsTP218 and EcTP against dThd and dUrd. Assays were performed in assay buffer consisting of 20 mM Hepes, 50 mM KH2PO4, pH 7.5 at the indicated temperature. Parameters are shown as the best fit value ± standard error (Johnson, 2019) upon fitting the experimental data to Michaelis-Menten model (equation 1 in Methods section).

| Enzyme species | dThd | dUrd | Conditions of steady-state kinetics |
|----------------|------|------|-------------------------------------|
|                |      |      |                                     |
| HsTP199        | 5 ± 0.3 | 25 ± 5 | (2 ± 0.4) × 10^5 | 7.75 ± 0.3 | 180 ± 18 | (4.3 ± 0.46) × 10^4 | Assay buffer, 25°C |
| HsTP218        | 4.65 ± 0.1 | 28 ± 2.5 | (1.66 ± 0.15) × 10^5 | 4.4 ± 0.2 | 50 ± 8 | (8.8 ± 1.4) × 10^4 | Assay buffer, 25°C |
| EcTP           | 402 ± 20 | 390 ± 50 | (1 ± 0.15) × 10^6 | 575 ± 24 | 390 ± 36 | (1.5 ± 0.15) × 10^6 | Assay buffer, 25°C |

FIGURE 4 | mPEG5kDa-conjugation reaction with HsTP199 (A) Scheme of the conjugation reaction of HsTP199 with mPEG5kDa. The N-hydroxysuccinimide ester (NHS ester) reactive group of mPEG5kDa reacts with primary amines (lysine side group and N’-terminus amino-group) leading to the final conjugation adduct with methoxy-PEG and the simultaneous liberation of NHS. (B) Lane 1: Purified HsTP199 (2-step purification including IMAC and anion exchange as described in Methods section) used for PEGylation; lanes 2 and 3 show mPEG5kDa-conjugated HsTP218 at 1:20 and 1:30 protein:mPEG5kDa ratio respectively. Each lane contains ∼15 μg of protein.
Analytical Size-Exclusion Chromatography of HsTP

The purified protein samples were diluted in PBS buffer to a final concentration of approximately 2 mg/ml. A total of 2 µl of diluted samples was injected on a size exclusion column (TSKgel G4000SWxL, 7.8 × 300 mm, 8 µ particle size, from Tosoh). Isocratic elution was performed at a flow rate of 0.3 ml/min, using a Thermo ultimate HPLC system. The mobile phase contained 90% of phosphate buffer (50 mM sodium phosphate, 200 mM sodium chloride, adjusted to pH 7.2) and 10% ethanol. The column eluent was monitored by UV detection at 214 nm. The BioRad gel filtration standard was used as molecular weight markers.

Phylogenetic Analysis of Thymidine Phosphorylase Enzymes

Phylogenetic analysis was performed by using the program Geneious prime. The full-length amino acid sequences of HsTP and EcTP deposited in the Uniprot database (P19971 and P07650) were used as query sequences to perform massive BLAST alignments and thus, determine conserved residues. Geneious is directly connected to the NCBI database and BLAST alignments and thus, determine conserved residues. For the generation of the PEGylation variants HsTP 240 and HsTP241 site-directed mutagenesis was performed following the overlap extension PCR methodology (Nelson and Fitch, 2011). Briefly, this method comprises three successive amplification steps and involves four primers as follows: two external ones that cover the 5′- and 3′- ends of the parental sequence and two additional ones which carry the desired mutations to be incorporated in the final sequence. Two independent PCR reactions (PCR1: forward external primer covering the 5′ combined with the reverse carrying the terminator sequence was subjected to analysis based on the ribosome binding site (RBS) predict-mode algorithm using E. coli as the host organism for recombinant expression. The RBS calculator in the predict-mode identifies all the starting codons in the mRNA sequence and calculates a translation initiation rate (TIR) in arbitrary units (a.u.; the larger the value, the higher the TIR) for each of them according to its statistical thermodynamics algorithm as described elsewhere (Salis et al., 2009). The TIR calculations are accompanied by ribosome binding Gibbs free energy calculations that represent the quantification of the interaction between the ribosome and the mRNA that, in turn, affects the TIR. The reported $\Delta G_{total}$ is defined as follows: $\Delta G_{total} = \Delta G_{mRNA-rRNA} + \Delta G_{spacing} + \Delta G_{start} - \Delta G_{standby} - \Delta G_{mRNA}$, where $\Delta G_{mRNA-rRNA}$ is the energy released when the last nine nucleotides of the E. coli 16S rRNA hybridizes and co-folds to the mRNA sub-sequence. $\Delta G_{start}$ is the energy released when the start codon hybridizes to the initiating tRNA anticodon loop. $\Delta G_{spacing}$ is the free energy cost caused by a non-optimal physical distance between the 16S rRNA binding site and the start codon. $\Delta G_{mRNA}$ is the work required to unfold the mRNA sub-sequence when it folds to its most stable secondary structure and $\Delta G_{standby}$ is the work required to unfold any secondary structures sequestering the standby site after the 30S complex assembly.

Site-Directed Mutagenesis and Generation of PEGylation Variants

For the generation of the PEGylation variants HsTP240 and HsTP241 site-directed mutagenesis was performed following the overlap extension PCR methodology (Nelson and Fitch, 2011). Briefly, this method comprises three successive amplification steps and involves four primers as follows: two external ones that cover the 5′- and 3′- ends of the parental sequence and two additional ones which carry the desired mutations to be incorporated in the final sequence. Two independent PCR reactions (PCR1: forward external primer covering the 5′ combined with the reverse carrying the
mutations and PCR2: reverse external at the 3′ combined with forward carrying the mutations) were performed to amplify two fragments which overlap at the regions flanking the mismatches. The two amplified fragments were agarose gel-purified and in a final third step, they were combined in equal molar quantities with the initial external primers and were subjected to the last PCR reaction resulting in the final amplicon that carries the desired point mutations. The PCR amplicons along with the pET28a plasmid were digested overnight at 37°C with NcoI/EcoRI and ultimately were cloned using T4 DNA ligase. The incorporated mutations were sequence-verified by Sanger sequencing.

**Preparation of PEGylated HsTP and EcTP**

Conjugation reactions with methoxy-5-kDa-PEG-succinimidy-succinate (mPEG<sup>5</sup>kDa, NOF Corporation) were performed in 100 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 8.5 using

**TABLE 3** dThd steady-state kinetic parameters of HsTP<sup>199</sup>, HsTP<sup>218</sup>, HsTP<sup>240</sup> and HsTP<sup>241</sup> prior and post-PEGylation. Conjugation reactions took place at either 1:10, 1:20 or 1:30 protein:PEG molar ratio as described in the Methods section. Parameters are shown as the best fit value ± standard error (Johnson, 2019) upon fitting the experimental data to Michaelis-Menten model (equation 1 in Methods section).

| Enzyme species | dThd | Conditions of steady-state kinetics |
|----------------|------|-----------------------------------|
|                |      |                                   |
| HsTP<sup>199</sup> | k<sub>catalytic</sub> (s<sup>−1</sup>) | K<sub>M</sub> (μM) | k<sub>catalytic</sub>/K<sub>M</sub> (M<sup>−1</sup>s<sup>−1</sup>) |
|                | 5 ± 0.3 | 25 ± 5 | (2 ± 0.4) × 10<sup>5</sup> |
| PEGylated HsTP<sup>199</sup> (1:20) | 3.75 ± 0.25 | 33 ± 9 | (1.1 ± 0.3) × 10<sup>5</sup> |
| PEGylated HsTP<sup>199</sup> (1:30) | 1.8 ± 0.1 | 20 ± 8 | (0.9 ± 0.36) × 10<sup>5</sup> |
| HsTP<sup>218</sup> | 11.5 ± 0.5 | 65 ± 8 | (1.7 ± 0.23) × 10<sup>5</sup> |
| PEGylated HsTP<sup>218</sup> (1:10) | 6.3 ± 0.25 | 52 ± 6 | (1.2 ± 0.13) × 10<sup>5</sup> |
| PEGylated HsTP<sup>218</sup> (1:20) | 6 ± 0.2 | 65 ± 9 | (0.9 ± 0.13) × 10<sup>5</sup> |
| HsTP<sup>240</sup> | 11.5 ± 0.4 | 59 ± 7 | (2 ± 0.25) × 10<sup>5</sup> |
| PEGylated HsTP<sup>240</sup> (1:10) | 9.3 ± 0.3 | 65 ± 6 | (1.43 ± 0.14) × 10<sup>5</sup> |
| PEGylated HsTP<sup>240</sup> (1:20) | 10 ± 0.5 | 90 ± 10 | (1 ± 0.135) × 10<sup>5</sup> |
| HsTP<sup>241</sup> | 13 ± 0.35 | 74 ± 9 | (1.75 ± 0.22) × 10<sup>5</sup> |
| PEGylated HsTP<sup>241</sup> (1:10) | 11.7 ± 0.55 | 95 ± 8 | (1.23 ± 0.12) × 10<sup>5</sup> |
| PEGylated HsTP<sup>241</sup> (1:20) | 10 ± 0.3 | 67 ± 7 | (1.5 ± 0.18) × 10<sup>5</sup> |

**SCHEME 1** | Engineer-in and engineer-out selection criteria for arginine and lysine residues respectively. SASA: solvent-accessible surface area.

**SCHEME 1** | Engineer-in and engineer-out selection criteria for arginine and lysine residues respectively. SASA: solvent-accessible surface area.

**SCHEME 1** | Engineer-in and engineer-out selection criteria for arginine and lysine residues respectively. SASA: solvent-accessible surface area.
purified HsTP and EcTP as described above in the respective section. Final enzyme concentration in the reaction was ∼115 μM and depending on the protein: PEG ratio, the respective concentration of PEG was adjusted accordingly: 1,150, 2,300, and 3,450 μM for 1:10, 1:20, and 1:30 M ratio, respectively. The lyophilized PEG was weighed
and was added directly into a 2 ml tube containing either purified HsTP \(^{199}\) or EcTP at a volume of 1 ml. Immediately after the addition of the PEG, the tube was vortexed continuously for 30 s, followed by incubation at room temperature for 30 min under rotating conditions. Subsequently, the mixture was exhaustively buffer-exchanged to remove excess of unreactive PEG using protein concentrators with a 50-kDa MWCO (Thermo Scientific). HsTP \(^{240}\) and HsTP \(^{241}\) PEGylation variants and HsTP \(^{218}\) were conjugated following the same process except for the buffer, which was 50 mM Na₂HPO₄ and 50 mM NaCl, pH 8.0, and the final enzyme concentration was 200 μM. Following the final buffer exchange step, PEGylation homogeneity was assessed by SDS-PAGE and steady-state kinetics was performed using dThd as the substrate. PEGylated enzymes were buffer-exchanged against 50 mM Na₂HPO₄ and 50 mM NaCl, pH 7.0, mixed with a 15% final concentration of glycerol, flash-frozen in liquid nitrogen, and stored at -80°C.

**RESULTS**

**HsTP Is Poorly Expressed in *E. Coli***

Successful recombinant expression of HsTP in *E. coli* has been previously described by other groups with soluble expression levels reported in the range of ~3 mg/L of culture (Schwartz et al., 2010). Available information deposited in public research databases (e.g., NCBI and Uniprot) suggests that HsTP harbors a N-terminal 10-residue long pro-peptide (Uniprot entry P19971) which is cleaved during a post-translational modification and processing step, thereby yielding the final mature polypeptide chain. While the biological role of this pro-peptide, as well as its impact on the enzyme’s biochemical features remain largely unclear, all prior studies focusing on the biochemical characterization of HsTP used the putatively mature domain at its N-terminal site truncated (Schwartz et al., 2010; Deves et al., 2014). Accordingly, we designed a HsTP gene carrying a His\(^{6}\)-tag at the N-terminus, codon-optimized for *E. coli* expression, and devoid of the first ten residues corresponding to the pro-peptide sequence, cloned into pET28-a plasmid (constructed termed HsTP\(^{199}\), Figure 1D) and tested its recombinant expression in the *E. coli* BL21 (DE3) strain. Indeed, the levels of the recovered active enzyme were in the range of 3–5 mg/L of TB culture medium (yield is defined as the amount of purified protein obtained after two purification steps as described in the Methods section, Table 1), while a considerable amount of the produced protein was detected in an insoluble inclusion body fraction (~50% more than the soluble) (Supplementary Figures S1, S2). Multiple attempts to increase the expression titers (both soluble and total amount) by 1) switching to C41 (DE3) *E. coli* expression strain which has been shown to benefit soluble expression levels (Kwon et al., 2015) (Supplementary Figures S1, S2), 2) altering expression conditions (e.g., Isopropyl-β-D-thiogalactopyranoside; IPTG concentrations, temperature, OD\(^{600}\) at the induction, culture media) (Rosano and Ceccharelli, 2014; Gutiérrez-González et al., 2019), and 3) testing a high-copy-number plasmid (pJC-20) (Clos and Brandau, 1994) (Supplementary Figure S3) failed to improve expression levels of HsTP\(^{199}\). Taken together, these data showed that achieving high levels of HsTP bacterial expression would necessitate a more in-depth analysis and additional extensive experimental efforts.

**HsTP and EcTP Have Distinct N’-Terminal Domains and Recombinant Expression Profiles**

Considering the fact that we used a synthetic, codon-optimized gene for optimal expression in *E. coli*, it was curious as to why all our efforts failed to yield large quantities of HsTP. High level expression of recombinant human proteins in *E. coli* can be very challenging to optimize as multiple molecular factors such as mRNA stability and structure (Simonetti et al., 2009; de Smit and van Duin, 1994), translation initiation rates (Espah Borujeni et al., 2014) and structural features of the protein (particularly at the N’-terminus) (Wruck et al., 2017) can play a pivotal role. The latter evidence motivated us to inspect the crystal structure of HsTP (PDB entry 2WK6) (Mitsiki et al., 2009) and compare it to the respective structure of the bacterial ortholog EcTP (PDB entry 4LHM) (Timofeev et al., 2014). The two enzymes belong to the family II of nucleoside phosphorylases, they are relatively close structural homologs (Cα-RMSD ~5.2 Å) (Figure 1A) and they share 38.4% amino acid sequence identity. Interestingly, we found that none of the available HsTP crystal structures deposited in the protein data bank (PDB) (2WK6, 1UOU, 2J0F) show a well-resolved secondary structure of the N’-terminal 10-residue pro-peptide. In fact, a longer sequence consisting of the first thirty-three residues of HsTP is structurally missing, possibly due to its large intrinsic flexibility which can result in low electron density as has been shown in other studies (Palamini et al., 2016). The first, structurally-visible residue of HsTP is Q35 which is located at the top of an unapped α-helix (Figures 1A,B), whereas EcTP has a leucine (L3) at the respective structurally equivalent position and is lacking an unstructured domain at its N’-terminus (Figure 1B). Amino acid sequence alignments of family II eukaryotic and prokaryotic nucleoside phosphorylases revealed that, while eukaryotic species harbor a N’-terminus sequence of variable length which is very poorly conserved (part of which corresponds to the pro-peptide), the prokaryotic orthologs totally lack such a sequence (Supplementary Figure S4).

Thus, we reasoned that a direct expression comparison of HsTP against its *E. coli* ortholog EcTP could shed more light on the expression difficulties encountered with HsTP. The EcTP construct was designed and cloned in a manner similar to HsTP (i.e. codon optimized gene was cloned in pET28a with a His\(^{6}\)-tag at the N’-terminus) (Figure 1D) using the predicted amino acid sequence encoded by the deoA gene from the K-12 strain as has been deposited in the EcoGene 3.0 database (Zhou and Rudd, 2013). Prior to the experimental expression assessment of EcTP, analysis of the translation initiation rate (TIR) for both constructs (HsTP\(^{199}\) & EcTP) using the algorithm developed by Salis lab (Salis et al., 2009; Salis, 2011) yielded identical TIR values (11,643 arbitrary units, Supplementary Table S1). This algorithm uses a statistical-thermodynamics-derived model to calculate the free energy of
different interactions between the mRNA and the ribosome and the sum of all those binding free energies (shown in Supplementary Table S1 as $\Delta G_{\text{total}}$) determines how likely it is for the ribosome to bind to an mRNA molecule and initiate translation (the higher the translation rate in Supplementary Table S1, the more likely for the ribosome to initiate translation). Particularly, calculations have shown that the secondary structure of the mRNA plays a determining role in those interactions and, in turn, in the translation rate (Espah Borujeni et al., 2017). The modelled mRNA structures of HsTP199 and EcTP including the 5’-UTR and the first ten codons of each construct were predicted to be identical as both were cloned into the same plasmid (identical ribosome binding sites) harboring a His6-tag at the N’-terminus along with the same linkers. Thus, based on the TIR analysis very similar expression levels of HsTP199 and EcTP were anticipated. However, surprisingly, expression of EcTP in BL21 (DE3) under the same conditions used for HsTP199 yielded $\sim$100–120 mg of protein per L of culture medium (entire amount in the soluble fraction) which is a $\sim$28-fold larger quantity relative to HsTP199 (3–5 mg/L, Table 1). These findings led us to hypothesize that the primary and secondary structure of HsTP’s N’-terminal domain, could possibly impede efficient protein expression.

N’-Terminal Truncation of HsTP Significantly Improves Its Expression in E. Coli

Primary and secondary protein structure can have a profound effect on the translation, as well as protein solubility and toxicity. For example, if the N-terminal region contains many prolines or positively charged amino acids, it can slow down ribosome translocation and elongation (lower protein synthesis and higher ribosome sequestration) (Wruck et al., 2017; Charneski and Hurst, 2013; Lu and Deutsch, 2014). Indeed, close inspection of the HsTP199 construct revealed that the first five residues at the N’-terminus include three proline residues i.e. M1-G2-A3-P4-P5-A6-P7 (Figure 1C, A11-P12-P13-A14-P15 based on the full-length numbering; G2 is a scar residue as a result of the cloning process using NcoI site for the cloning of all constructs in the present study). This could negatively affect co- translational folding and in turn, explain the presence of a considerable amount of HsTP in the insoluble fraction. In addition, the N’-terminal highly flexible and unstructured polypeptide fragment of HsTP199 spanning the range A11-K34 could further impede co-translational folding (Waudby et al., 2019; Ciryam et al., 2013).

We designed and generated HsTP truncated constructs (Figure 1D) within the first 33 residues (Figure 1C) which are not visible in its crystal structure (Figures 1A,B). Among the four new constructs tested, one (HsTP218) yielded large amounts of expressed protein, yet the protein was almost exclusively observed in the insoluble fraction when expressed at 30°C (Figure 2). Expression of HsTP218 under low temperatures (16°C and 22°C) and low IPTG conditions (Figure 3A; Table 1, Supplementary Figure S5) allowed the protein to fold properly and achieve soluble expression levels in a range more similar to those of EcTP (up to 70 mg/L, Table 1). The HsTP218 construct is missing the entire unstructured N’-terminal region comprising the first thirty-three amino acid residues of the full-length HsTP and structurally, it resembles the respective N’-terminus of EcTP. That is, downstream of the first K34 residue of HsTP218 the enzyme’s secondary structure forms an α-helix similar to EcTP (Figure 1B). Worth highlighting, the steady-state kinetic analysis of purified HsTP218 (Figure 3B) and HsTP199 (mature form of HsTP) showed very similar kinetic properties against dThd but a 2-fold increase in $k_{cat}/K_{M}$ for dThd in favor of HsTP218 (Table 2). In addition, analytical size exclusion chromatography (aSEC) confirmed the oligomeric state of HsTP218 exhibiting a dimer conformation (Supplementary Figure S6), which is in agreement with the active form of the enzyme as suggested by its crystal structure and previous studies (Norman et al., 2004).

Interestingly, TIR analysis of HsTP218 revealed that this construct is characterized by the highest TIR ($\sim$18,179 arbitrary units) among all the constructs tested in the present study (Supplementary Table S1). This high value of TIR is indicative of a high propensity for translation initiation as a result of the strong interactions between the HsTP218 mRNA’s secondary structure and the respective ribosomal binding sites essential for proper translation. Although the interactions between the mRNA and ribosomal RNA (that are governed and mediated by the mRNA’s structure) affect the TIR via a totally different mechanism than the protein’s primary and secondary structure at the N’-terminus, which was our initial hypothesis, HsTP218 expression may overall have been benefitted by both mechanisms. That is, from one side, the first ten codons of the new HsTP218 truncated construct (His6-tag present at the C’-terminus) altered the secondary structure of its mRNA in a way that enhanced its binding to the ribosome and consequently the translation as evidenced by the higher TIR relative to the initial construct HsTP199. On the other hand, the lack of multiple successive proline residues in the new N’-terminus likely facilitated the co-translational folding and improved soluble expression. Of note, an analysis of 200 highly expressed endogenous E. coli proteins showed that the most abundant amino acid residue at the 3rd position of the polypeptide chain is lysine (Bivona et al., 2010), which is in agreement with the HsTP218’s sequence (M1-G2-K3-Q4--). HsTP216 and HsTP217 constructs appeared to have very similar TIRs, which is in-line with their low expression levels, but HsTP216 showed $\sim$4-fold lower soluble protein, possibly attributed to the successive proline residues at its N’-terminus that could stall translation (Starosta et al., 2014; Peil et al., 2013; Huter et al., 2017).

PEGylation Optimization of HsTP

Upon improvement of the recombinant expression of HsTP, we undertook efforts to optimize its PEGylation. PEG has been increasingly used in the field of biologics during the recent decades as a time extension strategy aiming at the improvement of the thermodynamic stability (Santos et al., 2019; Moreno-Pérez et al., 2016; immunogenicity (Gefen et al., 2013) and the pharmacokinetic (PK) properties of therapeutic proteins (Harris et al., 2001; Hamidi et al., 2006). Initial efforts to conjugate HsTP199 with methoxy-5-kDa-PEG-succinimidyl-succinate (mPEG5kDa) targeting primary amines on the surface (primarily lysines and N’-terminus-amino group) resulted in an apparent heterogeneous mixture of three distinct PEGylated enzyme species (Figures 4A,B).
indicative of an inefficient PEGylation reaction. Indeed, HsTP contains eleven lysine residues per monomer (Figure 5A), however analysis of its crystal structure revealed that only four (K43, K139, K253, K275) are adequately surface-exposed and thus accessible to react with mPEG5kDa (Figure 5B). In fact, as shown in Figure 5B, perhaps the only lysine residue that can react with mPEG5kDa with high certainty and efficiency due to its high solvent accessibility is K253. In contrast, K43, K139 and K275 are less exposed, thereby likely yielding a heterogeneous mixture of PEGylated species. Indeed, based on the PEGylation pattern shown in Figure 4B, up to five sites appear to be modified with mPEG5kDa as this would result in a ~25 kDa increase in the molecular weight (MW) yielding a total of ~75 kDa per monomer (upper band in the gel). It appears reasonable to assume that those five sites consist of the four lysine residues K43, K139, K253 and K275 in addition to the N’-terminus. Further, upon PEGylation, HsTP199 exhibits a 2.8-and 2.2-fold decrease in $k_{cat}^{dThd}$ and $k_{cat}/K_M^{dThd}$ respectively (Table 3) suggesting that chemical modification with mPEG5kDa of one or more lysine residues, negatively affects the enzyme’s catalytic activity.

To address the limited PEGylation efficiency of HsTP and decrease in catalytic activity, we employed a rational surface engineering approach where arginine residues, located at surface-exposed positions suitable for PEGylation, were identified and were mutated to lysine. Conversely, lysine residues whose modification with mPEG5kDa could hamper the catalytic activity were engineered-out and mutated to arginine. Relevant enzymology and enzyme-PEGylation literature was thoroughly reviewed, and the respective information was integrated into certain criteria (summarized in Scheme 1) that guided our engineering efforts. HsTP has thirty-two arginine residues widely distributed around its surface (Figure 6A). By analyzing and evaluating all arginine residues, we identified five residues (R329, R342, R345, R358, R453) that satisfied all our criteria and are shown in Figures 6B,C. These arginine residues 1) are not highly conserved among thymidine phosphorlyases that belong to family II (Supplementary Figure S7A) and thus, are less likely to negatively impact the enzyme’s stability and catalytic activity, 2) they have very large solvent-accessible surface area (SASA) with the only exception being R342 which barely met our SASA criterion (>75 Å², Supplementary Figure S7B), 3) they are not located on loops whose dynamics may be altered upon PEGylation and, in turn, negatively impact catalytic activity (rigidity issue (Rodriguez-Martínez et al., 2009)), 4) do not form salt bridges with aspartate or glutamate residues as this could alter the pKa of the newly introduced lysine residue, requiring the PEGylation reaction to be performed at highly alkaline pH and 5) they are not located nearby the active site and therefore, they do not prevent the substrate molecules from diffusing into it. Furthermore, analysis of the four lysine residues (K43, K139, K253, K275) based on our “engineer-out” criteria (Scheme 1) led us to mutate K139 and K275 to arginine and thus, prevent their PEGylation. Both lysine residues are located on flexible loops (flexibility was assessed by the B-factor values of those domains) which connect 2nd-shell domains nearby the active site (Figure 5A). Previous studies have shown that PEGylation of enzymes correlates with reduced structural dynamics and this in turn, may impair catalytic activity which can be directly dependent on conformational dynamics (Rodriguez-Martínez et al., 2008; Hsieh and Lin, 2015; Morgenstern et al., 2017; Santos et al., 2019). Of note, PEGylation of EcTP whose 8/22 lysine residues are located on surface-exposed flexible loops led to even greater reduction of its $k_{cat}^{dThd}$ relative to HsTP (400 s⁻¹ vs 90 s⁻¹ for un-PEGylated and PEGylated EcTP species respectively) (Supplementary Figure S8).

We generated two HsTP variants (HsTP240 and HsTP241) containing lysine-to-arginine and arginine-to-lysine substitutions using as a template the expression-optimized HsTP218 enzyme. HsTP240 contains K139R-K275R-R342K-R345K-R358K while HsTP241 has all seven identical mutations, K139R-K275R-R329K-R342K-R345K-R358K-R453K. PEGylation efficiency as well as steady-state kinetics were subsequently determined. Indeed, both HsTP240 and HsTP241 variants showed a significant improvement in PEGylation homogeneity relative to their parental HsTP218, with HsTP241 performing the best and clearly showing only one visible band around 100 kDa (Figure 7) when 1:20 (protein:PEG) molar ratio conjugation reaction conditions were used. In contrast, HsTP199 showed an almost identical profile for both PEGylation treatments (1:10 and 1:20 M ratio) which consists of five distinct bands on the SDS-PAGE (Figure 7) suggesting considerably lower conjugation efficiency. Very importantly, the steady-state kinetic characterization of HsTP241 revealed that the PEGylated enzyme retained its $k_{cat}/K_M^{dThd}$ at the same levels as prior to PEGylation while PEGylation of HsTP218 caused a 2-fold decrease similar to what observed for HsTP199 (Table 3). Interestingly, upon PEGylation, the HsTP240 variant, which is missing the R329K and R453K substitutions, exhibited a decreased $k_{cat}/K_M^{dThd}$ at the same levels as prior to PEGylation while PEGylation of HsTP218 caused a 2.2-fold decrease similar to what observed for HsTP199 (Supplementary Table 3). Taken together, our data demonstrate that by rational selection of suitable arginine and lysine residues for mutagenesis, PEGylation efficiency and consistency can be significantly improved while preserving the catalytic activity post-PEGylation. Our experimentally-validated rational PEGylation design may be applicable and benefit other protein and enzyme therapeutic molecules including asparaginase (Karamitros and Konrad, 2016; Menegucci et al., 2019), cysteinase (Cramer et al., 2017) and kynureninase (Tripplet et al., 2018; Karamitros et al., 2020).

**DISCUSSION**

Efficient recombinant production of biologics plays a crucial role in the viability of the drug development process. Availability of large quantities of a protein of interest facilitates the activities associated with all the stages of the drug development trajectory including biochemical and biophysical characterization, downstream processing, manufacturing and formulation optimizations, as well as pharmacological and toxicity assessments. Despite the
immense progress in the field of molecular biology with the development of engineered *E. coli* strains and improved plasmids capable of expressing complex proteins of mammalian origins, the efficient expression of certain proteins can still be very challenging, and in turn, have a profound impact on cost and time. Thus, achieving high expression levels of high-quality protein that retains its biological activity (e.g., binding or catalytic properties) represents an important early focus in the field of biologics.

Similarly, the PEGylation of therapeutic proteins receives equal attention during the early phase of drug candidate selection and optimization. The conjugation of PEG to proteins has emerged as a powerful strategy to overcome significant limitations of protein pharmaceuticals such as sub-optimal in-vivo half-life and efficacy (Cheng et al., 2019), thermodynamic and serum instability (Lawrence and Price, 2016), and immunogenicity (Gefen et al., 2013). For example, there have been reported cases where protein PEGylation improved its pharmacokinetic properties by 4,200% (4 h vs 7 days for unPEGylated and PEGylated species respectively) (Holsberg et al., 2002). Although alternative time-extension approaches have been recently developed (Despanie et al., 2016; Haeckel et al., 2016; Ding et al., 2014) and hold high promise, their wide clinical efficacy with multiple large biomacromolecules such as proteins has not yet been demonstrated. However, it needs to be stressed that as it has been demonstrated in some cases, PEG can be immunogenic leading to rapid blood clearance of the therapeutic protein due to the generation of anti-PEG antibodies (Yang and Lai, 2015). In those cases, further optimization of the PEGylation process (e.g., different type of PEG, sites of PEGylation) or alternative time extension strategies must be explored and evaluated. In addition, Nevertheless, the FDA has approved twenty-four PEGylated macromolecular drugs as of 2020, including six enzymes (adenosine deaminase, asparaginase, uricase, phenylalanine ammonia lyase). Therefore, PEGylation still represents a very attractive and potentially safe approach for the enhancement of the pharmacological features of therapeutic proteins, and particularly enzymes. However, the covalent attachment of PEG to the surface of enzymes (either targeting N-term, lysine or cysteine residues) often suffers by two main limitations: 1) poor PEGylation efficiency due to the low number of surface exposed residues, thereby resulting in a polydisperse population of species with distinct biochemical properties attributed to the different degrees of PEGylation that complicates the biochemical and biophysical characterization, and 2) conjugation of PEG can negatively impact the enzyme’s catalytic activity and thus, higher doses of a drug (i.e. higher cost) may be required to achieve a desirable pharmacological effect.

In the present study, we employed protein engineering strategies to optimize the bacterial expression and PEGylation of the human enzyme thymidine phosphorylase, an enzyme that could be used as a scaffold for the treatment of MNGIE, an inborn error of metabolism with high unmet medical need. While the mature HsTP displays very promising catalytic activity against the toxic metabolites which drive the MNGIE disease, its recombinant expression in *E. coli* was very poor, yielding ~5 mg/L of culture medium. Multiple attempts to improve the expression of the enzyme by scouting different conditions, *E. coli* strains and plasmids failed to benefit protein production. Literature suggests that protein expression in *E. coli* can be governed by several underlying mechanisms with protein’s primary and secondary structure at the N-terminus and mRNA’s secondary structure being the most important factors (Wruck et al., 2017; de Smit and van Duin, 1994). Careful inspection of the HsTP’s crystal structure revealed that the first thirty-four amino acid residues were missing from the crystal structure strongly suggesting that this domain is highly flexible and disordered. Of note, this domain includes a 10-residue pro-peptide (Figures 1C,D) with unknown function in the case of HsTP. While such pro-peptides have well-established inhibitory functions in other human enzymes like proteases (Boon et al., 2020), this did not appear to be the case for HsTP as the full-length species was active when assayed against dThd. In addition, the initially-tested HsTP199 (mature form of HsTP without the pro-peptide at the N-terminus) carries three proline residues within the first five residues showing a APPAP motif which could have a detrimental effect on ribosomal translation causing strong stalling (Peil et al., 2013; Huter et al., 2017). Phylogenetic analysis of thymidine phosphorylase enzymes and direct comparison of HsTP against its bacterial ortholog EcTP, which expresses at very high levels, led to the design and generation of the HsTP218, which expresses at very high levels, led to the design and generation of the HsTP218 construct that structurally resembles EcTP at the N-terminus and improved expression up to ~14-fold. Interestingly, the translation initiation rate (TIR) analysis suggested that the mRNA of HsTP218 binds more strongly to the ribosome than the other HsTP truncated constructs tested. Our findings suggest that sequence and structural differences among orthologs from different organisms can provide resourceful information that aid the rational engineering of desirable biochemical proteins including recombinant expression levels.

In this study we also proposed and experimentally validated a surface engineering strategy that enabled PEGylation improvement of HsTP at two different levels. Through rationally-designed arginine-to-lysine substitutions we managed to significantly improve PEGylation efficiency and homogeneity using a protein:PEG molar ratio as low as 1:20. This can be very advantageous on multiple levels: 1) additional extensive ion exchange purification steps are not required to resolve and isolate the different enzyme PEG-mers which are generated upon partial PEGylation, 2) the more homogeneous PEGylated enzyme species allow for a more accurate biochemical assessment, 3) low protein:PEG molar ratios can have a hugely beneficial impact on the costly stage of manufacturing process and equally importantly 4) a homogeneously-PEGylated protein is likely to have more consistent pharmacodynamics effects as well as a safer profile. On a second level, by mutating lysine-to-arginine residues at specific locations, the catalytic activity of HsTP was not negatively impacted by PEGylation. Enzyme PEGylation can be accompanied by a loss in catalytic activity as it has been shown in several studies (Zaghmi et al., 2019; Harris and Chess, 2003; Rodríguez-Martínez et al., 2009). While it is a very challenging and complex task to determine the exact underlying molecular mechanisms for this activity loss, increasing experimental (Rodriguez-Martinez et al., 2008) and computational (Yang
et al., 2011; Wilding et al., 2018) evidence suggest that the loss can be largely attributed to increased conformational rigidity and restricted dynamics upon PEGylation. Based on these observations, we hypothesized and showed that PEGylation of lysine residues located on loops can be responsible for such activity loss. Enzyme loops are usually characterized by increased conformational plasticity which can be critical for catalytic activity (Malabanan et al., 2010; Yu et al., 2017). By mutating two loop-located lysine to arginine residues, the catalytic activity of HsTP was retained at the same levels observed prior to PEGylation, thus further supporting our initial hypothesis. Nevertheless, PEG conjugation at highly flexible and unstructured loops may be associated with increased stability (Dozier and Distefano, 2015; Wilding et al., 2018; Lawrence and Price, 2016) and therefore, the PEGylation goals must be assessed and defined on a case-specific manner. Taken together, our proposed strategies provide an effective way to optimize enzyme PEGylation and E. coli recombinant expression and are likely applicable for other proteins and enzymes for both therapeutic and industrial applications.

**DATA AVAILABILITY STATEMENT**

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

**REFERENCES**

Bax, B. E., Levene, M., Bain, M. D., Fairbanks, L. D., Filoisto, M., Uçar, S., et al. (2019). Erythrocyte Encapsulated Thymidine Phosphorylase for the Treatment of Patients with Mitochondrial Neurogastrointestinal Encephalomyopathy: Study Protocol for a Multi-Centre, Multiple Dose, Open Label Trial. Jcm 9, 1096. doi:10.3390/jcm8081096

Birck, M. R., and Schramm, V. L. (2004). Nucleophilic Participation in the Transition State for Human Thymidine Phosphorylase. J. Am. Chem. Soc. 126, 2447–2453. doi:10.1021/ja039260h

Bivona, L., Zou, Z., Stutzman, N., and Sun, P. D. (2010). Influence of the Second Amino Acid on Recombinant Protein Expression. Protein Expr. Purif. 74, 248–256. doi:10.1016/j.pep.2010.06.005

Boon, L., Ugarte-Berzal, E., Vandooren, J., and Opdenakker, G. (2020). Protease Proppeptide Structures, Mechanisms of Activation, and Functions. Crit. Rev. Biochem. Mol. Biol. 55, 111–165. doi:10.1080/10409238.2020.1742090

Charneski, C. A., and Hurst, L. D. (2013). Positively Charged Residues Are the Major Determinants of Ribosomal Velocity. Plos Biol. 11, e1001508. doi:10.1371/journal.pbio.1001508

Cheng, F., Yang, J., Schwanenberg, U., and Zhu, L. (2019). Rational Surface Engineering of an Arginine Deiminase (An Antitumor Enzyme) for Increased PEGylation Efficiency. Biotechnol. Bioeng. 116, 2156–2166. doi:10.1002/bit.27011

Cirvany, P., Morimoto, R. I., Vendruscolo, M., Dobson, C. M., and O’Brien, E. P. (2013). In Vivo translation Rates Can Substantially Delay the Cotranslational Folding of the Escherichia coli Cytosolic Proteome. Proc. Natl. Acad. Sci. 110, E132–E140. doi:10.1073/pnas.1213624110

Clos, J., and Brandau, S. (1994). pC20 and pC40 - Two High-Copy-Number Vectors for T7 RNA Polymerase-dependent Expression of Recombinant Genes in Escherichia coli. Protein Expr. Purif. 5, 133–137. doi:10.1006/prep.1994.1020

Cramer, S. L., Saha, A., Liu, J., Tadi, S., Tzizian, S., Yan, W., et al. (2017). Systemic depletion of L-cyst(e)ine with cyst(e)inase increases reactive oxygen species and suppresses tumor growth. Nat. Med. 23, 120–127. doi:10.1038/nm.4322

de Moura Sperotto, N. D., Deves Roth, C., Rodrigues-Junior, V. S., Ev Neves, C., Reisdorfer Paula, F., da Silva Dadda, A., et al. (2019). Design of Novel Inhibitors of Human Thymidine Phosphorylase: Synthesis, Enzyme Inhibition, In Vitro Toxicity, and Impact on Human Glioblastoma Cancer. J. Med. Chem. 62, 1231–1245. doi:10.1021/acs.jmedchem.8b01305

de Smit, M. H., and van Duin, J. (1994). Control of Translation by mRNA Secondary Structure in Escherichia coli. J. Mol. Biol. 244, 144–150. doi:10.1006/jmbi.1994.1714

Desgranges, C., Razaka, G., Rabaud, M., and Bricaud, H. (1981). Catalolism of Thymidine in Human Blood Platelets Purification and Properties of Thymidine Phosphorylase. Biochim. Biophys. Acta (Bba) - Nucleic Acids Protein Synth. 654, 211–218. doi:10.1016/0005-2787(81)90174-x

Despanje, J., Dhandhukia, J. P., Hamm-Alvarez, S. F., and MacKay, J. A. (2016). Elastin-Like Polypeptides: Therapeutic Applications for an Emerging Class of Nanomedicines. J. Controlled Release 240, 93–108. doi:10.1016/j.jconrel.2015.11.010

Deves, C., Rostirolla, D. C., Martinelli, L. K. B., Bizarro, C. V., Santos, D. S., and Basso, L. A. (2014). The Kinetic Mechanism of Human Thymidine Phosphorylase - a Molecular Target for Cancer Drug Development. Mol. Biosyst. 10, 592–604. doi:10.1039/c3mb704533

Ding, S., Song, M., Sim, B.-C., Gu, C., Podust, V. N., Wang, C.-W., et al. (2014). Multivalent Antiviral XTEN-Peptide Conjugates with Long In Vivo Half-Life and Enhanced Solubility. Bioconjug. Chem. 25, 1351–1359. doi:10.1021/bc502155m

Dozier, J., and Distefano, M. (2015). Site-Specific PEGylation of Therapeutic Proteins. Ijms 16, 25831–25864. doi:10.3390/ijms161025831

El-Hattab, A. W., and Scaglia, F. (2013). Mitochondrial DNA Depletion Syndromes: Review and Updates of Genetic Basis, Manifestations, and Therapeutic Options. Neurotherapeutics 10, 186–198. doi:10.1007/s13311-013-0177-6

**AUTHOR CONTRIBUTIONS**

CSK was responsible for conceptualizing, performing experiments, analyzing, interpreting and discussing data and drafting the manuscript. CS performed experiments and wrote part of the manuscript. GA and SR were involved in the discussion of the results and contributed to the editing of the manuscript.

**FUNDING**

The work presented in this manuscript was entirely funded by Aeglea Biotherapeutics.

**ACKNOWLEDGMENTS**

We are thankful to Associate Scientist Jing Meil for technical assistance with the analytical size-exclusion chromatography analysis.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.793985/full#supplementary-material
Yu, H., Yan, Y., Zhang, C., and Dalby, P. A. (2017). Two Strategies to Engineer Flexible Loops for Improved Enzyme Thermostability. Sci. Rep. 7, 41212. doi:10.1038/srep41212

Zaghmi, A., Mendez-Villuendas, E., Greschner, A. A., Liu, J. Y., de Haan, H. W., and Gauthier, M. A. (2019). Mechanisms of Activity Loss for a Multi-PEGylated Protein by experiment and Simulation. Mater. Today Chem. 12, 121–131. doi:10.1016/j.mtchem.2018.12.007

Zhou, J., and Rudd, K. E. (2013). EcoGene 3.0. Nucleic Acids Res. 41, D613–D624. doi:10.1093/nar/gks1235

**Conflict of Interest:** CSK, CS, GA and SR were employed by Aeglea Biotherapeutics.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Karamitros, Somody, Agnello and Rowlinson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.