Article

Novel His-tag Variants for Insertion Inside Polypeptide Chain

Anastasiia G. Tarabarova, Anton Lopukhov, Alexey N. Fedorov, and Maria S. Yurkova*

Read Online

| ACCESS | | Metrics & More | | Article Recommendations | | SI Supporting Information |

ABSTRACT: His-tags are protein affinity tags ubiquitously used due to their convenience and effectiveness. However, in some individual cases, the attachment of His-tags to a protein's N- or C-termini resulted in impairment of the protein's structure or function, which led to attempts to include His-tags inside of polypeptide chains. In this work, we describe newly designed internal His-tags, where two triplets of histidine residues are separated by glycine residues to avoid steric hindrances and consequently minimize their impact on the protein structure. The applicability of these His-tags was tested with eGFP, a multifaceted reference protein, and GrAD207, a modified apical domain of GroEL chaperone, designed to stabilize in soluble form initially insoluble proteins. Both proteins are used as fusion partners for different purposes, and providing them with His-tags introduced into their polypeptide chains should conveniently broaden their functionality without involving the termini. We conclude that the insertable tags may be adjusted for the purification of proteins belonging to different structural classes.
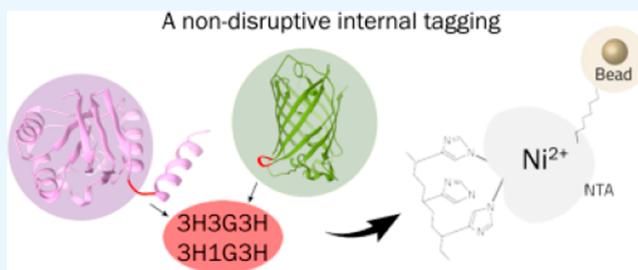
INTRODUCTION

His-tags attached to proteins' N- or C-termini[1] are widely used not only for affinity purification by metal-affinity chromatography, but also for investigations methods like immunoblotting,[2−4] fluorescent dye tagging,[5−9] refolding studying,[10−12] virus vectors development,[13] and others.[14] His-tags are convenient: being short, they allow elution in mild native conditions and provide at the same time a reasonable degree of purity. But in some cases placing His-tags at protein's termini may create problems: a classical hexahistidine sequence can cause protein inclusion body formation[4,15] or oligomerization through histidine residues,[16,17] can impair catalytic enzyme functions[18−21] or protein−protein interaction,[22] or His-tags cannot be used at N- or C-termini because of their functional and/or structural importance. In these cases, His-tags could be used in a noncanonical way by inserting them into the polypeptide chain. There are studies describing such application,[23,24] but the effects of His-tags inserted into the polypeptide chain on a protein's secondary structure and function vary depending on individual proteins and have not been studied in detail to understand the patterns for rational His-tag design. The current work describes the design of novel His-tag variants and their application with two proteins belonging to different structural classes.

MATERIALS AND METHODS

Construction of Genes and Plasmids. All PCR reactions were performed with a T100 thermal cycler BioRad, USA. To obtain pET11cjoe_GrAD207_3H, ET11cjoe_GrAD207_7H, pET11cjoe_GrAD207_3H1G3H, and pET11cjoe_-GrAD207_3H3G3H plasmids, the pET11cjoe_GrAD207 was used as a template.[25] Modification of pET11cjoe_GrAD207 was carried out using the overlapping method with following primers: F1_GrAD207, R1_GrAD207, R_7His, R_3His, R_313His, and R_333His (Table 1).

pET15b_eGFP157_3H3G3H and pET15b_eGFP157_7H plasmids were constructed using site-directed mutagenesis with some modifications[26] with the use of following primers:

F_gfp157_3H3G3H, R_gfp157_3H3G3H, F_gfp157_7H and R_gfp157_7H (Table 1).

Protein Expression and Purification. All GrAD207 variants were expressed in *E. coli* BL21 (DE3) grown in liquid TB medium at 37 °C. Expression was induced by addition of IPTG to a 1 mM final concentration at $OD_{600} = 0.6$. After 3 h, cells were collected by centrifugation at 4000 g at 4 °C in Beckman Coulter (USA) centrifuge.

All eGFP protein variants were expressed in *E. coli* ArcticExpress (DE3) grown in liquid TB medium at 30 °C. Expression was induced by addition of IPTG to 0.5 mM final concentration at $OD_{600} = 0.6$, and the temperature of cultivation decreased to 12 °C. After 24 h, cells were collected.

The cell pellets were resuspended at a ratio of 1 g to 5 mL in buffer containing 50 mM sodium phosphate pH 7.5 and 1 mM PMSF. The biomass was disintegrated using a Q500 Sonicator ultrasonic disintegrator at 4 °C. Sonicating mode was 40%

858

Table 1. Primers Used for Gene and Plasmid Constructions

name	sequence
F1_GrAD207	5′GGGAATTCCATATGGAGACGCTGGAAGCGGTCCTC3′
R1_GrAD207	5′ACGCATCGGTCGACTTAGTTGGTGACGAAGTAGGG GGAG3′
R_7His	5′GTACCCCTTGTCAAACTGGTACCCGTGATGGTGATGGTGATGG-TGGCCGCCCACGATGGTGGT3′
R_3His	5′GTACCCCTTGTCAAACTGGTACCCGTGATGGTGGCCGCCCACGA-TGGTGGT3′
R_313His	5′GTACCCCTTGTCAAACTGGTACCCGTGATGGTGACCGTG-ATGGTGGCCGCCCACGATGGTGGT3′
R_333His	5′GTACCCCTTGTCAAACTGGTACCCGTGATGGTGACCACCA-CCGTGATGGTGGCCGCCCACGATGGTGGT3′
F_gfp157_3H3G3H	5′CATGGCCGACAAGCAGCATCATCATGGTGGTGGTCAT-CATCATAAGAACGGCATCAAGG3′
R_gfp157_3H3G3H	5′CCTTGATGCCGTTCTTATGATGATGACCACCACCA-TGATGATGCTGCTTGTCGGCCATG3′
F_gfp157_7H	5′CATGGCCGACAAGCAGCATCATCATCATCATCACAA-GAACGGCATCAAGG3′
R_gfp157_7H	5′CCTTGATGCCGTTCTTGTGATGATGATGATGATGCTGC-TTGTCGGCCATG3′

amplitude, 1 s on, 2 s off, and working time was 2 min. Total energy per degradation was 400 kJ/ml. Debris was precipitated by centrifugation at 10 000g for 30 min at 4 °C. Then the supernatant was diluted 3-fold by adding 50 mM sodium phosphate at pH 7.5. The solution was loaded onto anion exchange column Hitrap Q XL (GE Healthcare, USA) equilibrated in the same buffer, and the protein was eluted by NaCl gradient. The fractions containing the target protein were combined and loaded onto a HisTrap HP column (GE Healthcare, USA). At the next step, the column was washed with 1 M NaCl to reduce nonspecific interactions. Then the proteins were eluted in an imidazole gradient. The fractions containing the target protein were reloaded to the column and eluted in steps with 50, 100, and 200 mM imidazole. Purified proteins were dialyzed against 10 mM sodium phosphate buffer (pH, 7.5) for further application.

Secondary Structure Study. Circular dichroism spectroscopy was carried out using a Chirascan spectropolarimeter (Applied Photophysics, UK) in wavelength ranges from 180 to 260 nm. The molar ellipticity $[\theta]$ values were calculated from the equation $[\theta] = [\theta]_m \frac{M_{res}}{LC}$, where $[\theta]_m$ is the measured ellipticity (deg), M_{res} is the average molecular weight of protein (Da), L is the optical cuvette path length (mm), and C is the protein concentration (mg/mL). The measurements were taken in 0.1 mm cuvette at protein concentration of 0.2−0.3 mg/mL in 10 mM sodium phosphate buffer, pH 7.5. For calculating the percentage of secondary structure elements, DichroWeb instrument was used.[27]

Surface Plasmon Resonance Spectroscopy. Equilibrium dissociation constants (K_D) for Ni^{2+} complexes formed by His-tagged proteins with immobilized on the chip nitrilotriacetic acid (NTA) were measured using surface plasmon resonance (SPR) spectroscopy. The experiment was performed on Biacore X100 (Biacore AB, Uppsala, Sweden) using an NTA chip. The chip was preloaded with Ni^{2+} by injection of 0.5 mM $NiCl_2$ at a flow rate 20 μL/min, which followed the injection of 20 μL of 350 mM EDTA. Samples were dissolved in running buffer (0.01 M sodium phosphate, 0.5 mM $NiCl_2$, pH 7.4) in serial dilutions up to 5×10^{-8} M. Samples were injected at a flow rate 20 μL/min at 25 °C for 20 s followed by 60 s dissociation. The sensor chip was regenerated by injection of 20 μL of 350 mM EDTA. All solutions were filtered and deoxygenated prior to use. The K_D values were evaluated by using the 1:1 Langmuir association model. Data were analyzed using BIAevaluation 3.0 software. K_D is given as the mean + SD of three independent measurements.

Spectroscopy and Fluorescence Brightness Evaluation. Fluorescence excitation−emission spectra measurements were conducted on Fluoromax-4 spectrofluorometer (HORIBA), and absorbance measurements on Implen Spectrophotometer NanoPhotometer NP80. Measurements for all native proteins were carried out in 150 mM NaCl and 50 mM sodium phosphate buffer, pH 7.5. Determination of molar extinction coefficient was based on measuring the concentration of mature chromophore according to the methodology described in.[28] eGFP and its mutants were alkali-denatured in 1 M NaOH. Under these conditions, the GFP-like chromophore is known to absorb at 447 nm with an extinction coefficient of 44 000 M^{-1} cm^{-1}. Based on the absorbance of native and alkali-denatured proteins, molar extinction coefficients for native states were calculated according to the Beer−Lambert law. The relative method was used to calculate the quantum yield of fluorescence. The area under the curve of the protein emission spectrum was compared with that of the fluorescein-iodoacetamide spectrum with a known quantum yield of 0.5. The fluorescence brightness was calculated as the molar extinction coefficient multiplied by the quantum yield. All the described calculations are illustrated in Figure S1.

RESULTS AND DISCUSSION

In the case of proteins used as fusion partners for different purposes, as it is the case in the current work, it is convenient to have free N- and C-termini available for target proteins; moreover, using any terminus for attaching His-tag might create steric impediments for either the folding of the whole construct with the target or its affinity purification in native conditions. The reasonable option is to introduce the tag into the polypeptide chain with the aim to preserve protein's structure and to have the tag exposed to solution. When inserted into the polypeptide chain, the polyhistidine sequence is thought to form helical elements, especially in the context of negatively and positively charged amino acids.[29] This point is important to consider when incorporating His-tags into a loop of a polypeptide chain because the formation of helical and any other secondary structure elements, other than unstructured, may not be preferable. To avoid the introduction of unbroken polyhistidine stretch, other amino acid residues are sometimes added to the tag, usually alanine residues.[30] In two novel alternative His-tags described in this work, we have used glycine residues to separate triplets of histidine residues (3H1G3H and 3H3G3H). It is known that glycine residues break down helical elements due to their high conformational mobility; because of it, glycine residues are usually found at the end of helical elements, as well as inside unstructured regions, and we assigned them just this role. Histidine residues, in turn,
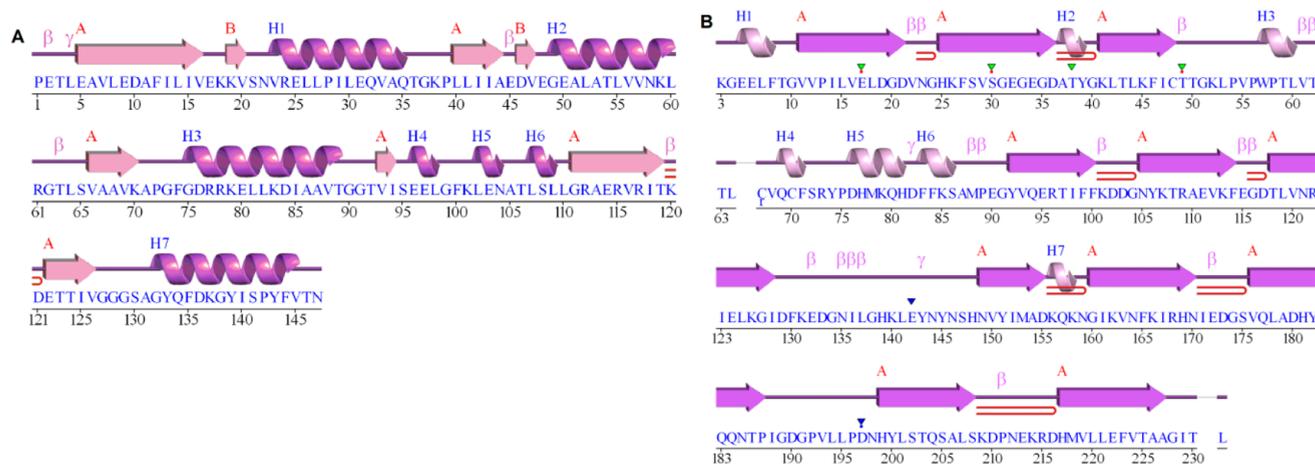
Figure 1. Diagram of the secondary structure elements aligned with primary structure (A) for GrAD207 based on the predicted GrAD207 model and (B) for eGFP based on the spatial structure of eGFP (PDB: 6YLQ). The diagram was obtained using the PDBsum server.[35] Arrows indicate β-strands, purple spirals indicate α-spirals, and red hairpins indicate the β-hairpin motif. The inscriptions β and γ indicate corresponding turns.
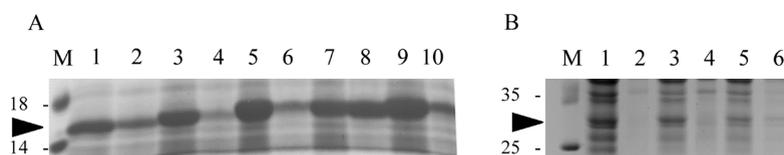
Figure 2. Electropherogram of supernatant and debris of *E. coli* producer strains of (A) GrAD207 and (B) eGFP variants. (M) Marker of protein molecular weight. (A) GrAD207: 1, supernatant; 2, debris; GrAD207_3H: 3, supernatant; 4, debris; GrAD207_3H1G3H: 5, supernatant; 6, debris; GrAD207_7H: 7, supernatant; 8, debris; GrAD207_3H3G3H: 9, supernatant; 10, debris. (B) eGFP: 1, supernatant; 2, debris; eGFP157_3H3G3H: 3, supernatant; 4, debris; eGFP157_7H: 5, supernatant; 6, debris. In each panel, the black arrow indicates the position of the target protein.

were joined in triplets; triplets were chosen based on the following considerations. Modeling of interactions of polyhistidine peptides with immobilized transition metal ions revealed that residues occupying positions i and $i+2$ or i and $i+5$ are preferentially involved in this process.[30,31] These data, valid for peptides, may presumably also be true for His-tagged proteins in the absence of significant steric effects. Hence, we proposed noncanonical peptides 3H1G3H and 3H3G3H for embedding into polypeptide chains, besides more common polyhistidine tags. The tags were tested on two proteins belonging to different structural classes, and the obtained results allowed us to propose reasons for the rational design of His-tags to include in different proteins. Two proteins to which the idea of newly designed internal His-tags was applied are both used as fusion partners for different purposes. GrAD207, a permutated apical domain of *T. thermophilus* GroEL, was designed for stabilization in solution of proteins prone to aggregation, and allowed obtaining two initially insoluble proteins in stable soluble constructs.[32,33] The opportunity to purify these and other possible fusions with GrAD207 under mild native conditions should be a precious addition to its features. GrAD207 belongs to α and β structural classes, three-layer ββα fold (3.50 CATCH index). The introduced His-tags were used instead of the linker connecting the native N- and C-termini of permutated protein; thus, the insertion site of His-tags fell between amino acid residues 128 and 132 at the turn between the β-strand and the α-helix close to the C-terminus of GrAD207. The site of insertion is located away from the functionally significant substrate-binding surface of

GrAD207 and does not affect structural elements involved in the formation of the hydrophobic core of the protein. GrAD207 was constructed with the insertions of 3H1G3H and 3H3G3H tags as well as 7H and 3H tags.

eGFP is an enhanced green fluorescent protein (GFP) of jellyfish *Aequorea victoria* that has an increased fluorescence intensity induced by random mutagenesis. This makes eGFP suitable for a number of applications such as a reporter gene, *ex vivo* imaging, or for determining transfection efficiency.[34] eGFP's commercially available variants are usually purified by N-terminal His tag: without the benefit of affinity purification, eGFP is a difficult protein prone to nonspecific interactions with chromatographic matrix, which we fully experienced when purifying it as a control, without tags. If the case of designing a fusion construct with eGFP is in consideration, then it is preferable to have its termini free. Structurally, eGFP belongs to the mainly β-sheet proteins, β-can fold (2.40 CATCH index). The site for His-tag insertion was chosen between 157 and 158 amino acid residues at the turn of β-hairpin. The choice of the insertion site for eGFP was based on the analysis of the published 3D structure (PDB 6YLQ) and literature data, taking into account its structural and functional properties.[2−6] A turn in the β-hairpin between amino acid residues 157 and 158 was chosen based on 3D structure, because it is exposed to solution, and on data that mutations in this place should not affect the maturation of the GFP chromophore. eGFP was constructed with insertions of the 3H3G3H and 7H tags. A diagram of the secondary structure elements aligned with primary structure for both proteins is shown in Figure 1.
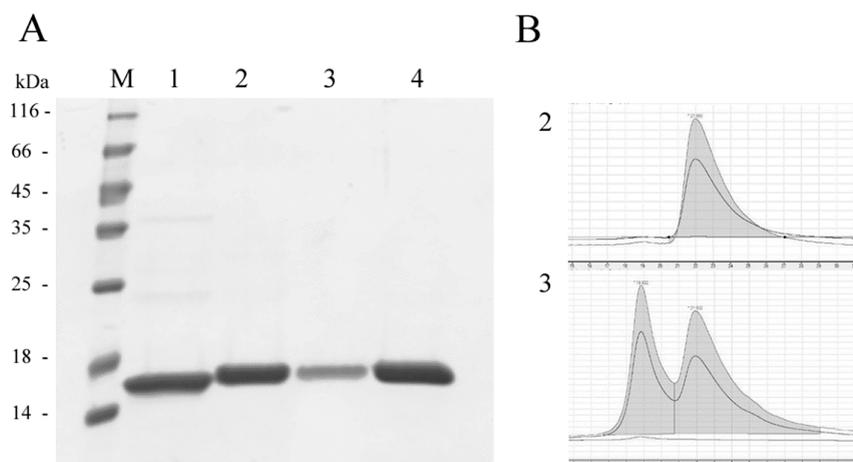
Figure 3. Stability study of GrAD207 variants in solution: (A) purified proteins; (B) elution profiles. For both panels: (1) GrAD207_3H, (2) GrAD207_3H1G3H, (3) GrAD207_7H, and (4) GrAD207_3H3G3H.
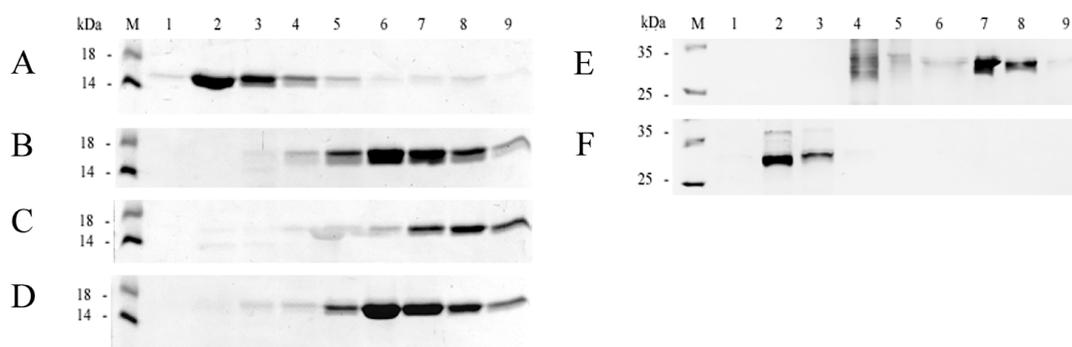
Figure 4. Study of affinity properties of His-tagged proteins: electropherograms of elution peaks of (A) GrAD207_3H, (B) GrAD207_3H1G3H, (C) GrAD207_7H, (D) GrAD207_3H3G3H, (E) eGFP157_7H, and (F) eGFP157_3H3G3H. For all panels, the line numbers correspond to fraction numbers and M is a protein molecular weight marker.

All obtained constructs were expressed in corresponding *E. coli* strains: in BL21(DE3) for GrAD207 constructs (GrAD207, GrAD207_3H, GrAD207_7H, GrAD207_3H1G3H, and GrAD207_3H3G3H) and in ArcticExpress(DE3) for eGFP constructs (eGFP, eGFP157_7H, and eGFP157_3H3G3H). The distribution of expressed proteins between soluble and insoluble fractions is shown in Figure 2. All of the proteins were expressed predominantly in the soluble form, except in the case of GrAD207_7H, for which the ratio of soluble and aggregated forms was about 1:1 (Figure 2A, lines 7 and 8). Thus, the insertion of 7H-tag into GrAD207 led to inclusion body formation. Such a situation was not detected for eGFP tagged either with 7H or 3H3G3H (Figure 2B).

For further work, proteins were purified and tested for stability (Figure 3). In the GrAD207_7H sample, a white slurry appeared after 1 week of storage at +4 °C. No visually detectable changes were observed in other samples under the same conditions. Examination of the GrAD207_7H sample by size-exclusion chromatography and comparison of the elution profile with that of GrAD207_3H1G3H revealed the presence of several forms of GrAD207_7H in solution (Figure 3B). These results indicate that 7H-tag insertion into GrAD207 not only increased the probability of inclusion body formation but also promoted the aggregation of purified protein in solution.

In the case of 7H introduced into eGFP, such effects were not detected.

Determination of relative retention time on Ni^{2+}-affinity resin served as a qualitative characteristic of the constructs' affinity to immobilized Ni^{2+} ions. Electrophoretic analysis of obtained elution profiles (Figure 4) showed that proteins with all the variants of His-tags were able to bind to Ni^{2+}-affinity resin, but had different ability to retain on the resin depending on inserted His-tag (Figure 4)

For His-tagged GrAD207 variants, their relative retention times are summarized in Figure 5.

For both proteins with the 7H tag, the retention time was the longest, and it was the shortest for GrAD207_3H. The chromatographic method revealed only a slight difference between the retention times of GrAD207_3H1G3H and GrAD207_3H3G3H, and this difference became more prominent as the resin was saturated with the protein. Increasing the amount of the protein on the resin reduces the likelihood of rebinding of the protein to the matrix during the elution process and also reduces the influence of nonspecific interactions with the resin matrix. These considerations led to suggestion that for GrAD207_3H3G3H some nonspecific interactions contributed to its binding to Ni^{2+} resin, which is further discussed later. For eGFP157_3H3G3H, the elution profile was similar to that of GrAD207_3H but not to that of GrAD207_3H3G3H,
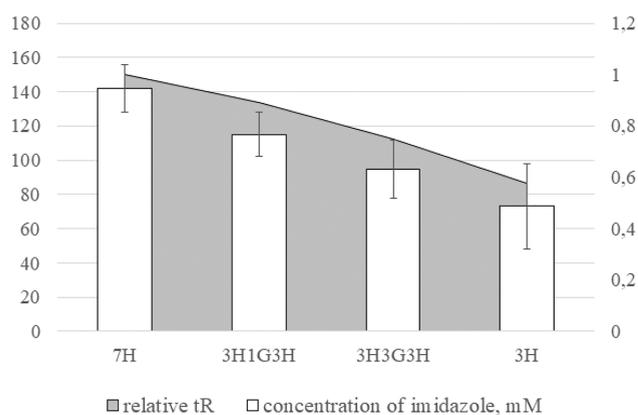
Figure 5. Relative retention time of His-tagged GrAD207 variants by the Ni^{2+}-containing resin. Imidazole concentration (mM) is shown on the left scale, and relative retention time on the right.

suggesting that the 3H3G3H tag has a different impact on the two proteins' structures.

Surface plasmon resonance spectroscopy was used for quantitative assessment of interactions between His-labeled GrAD207 variants and NTA upon complexation with Ni^{2+}. Raw sensograms are given in the Supporting Information (Figures S2−S14). As a necessary control, in the absence of $NiCl_2$ complex formation was not observed, and the signal was at the noise level (Figure S2). The obtained equilibrium dissociation constants are listed in Table 2.

Table 2. Equilibrium Dissociation Constants (K_D) for Ni^{2+} Complexes Formed by His-Labeled GrAD207 Variants and Immobilized on Chip Nitrilotriacetic Acid (NTA), Measured by Surface Plasmon Resonance (SPR) Spectroscopy

Protein sample	$K_D{}^a$ (μM)
GrAD207_3H	8.1 ± 1.3
GrAD207_7H	2.5 ± 0.9
GrAD207_3H1G3H	2.7 ± 0.8
GrAD207_3H3G3H	6.0 ± 1.6

$^a K_D$ is given as mean + SD of three independent measurements.

The K_D value for GrAD207_3H was in good agreement with the previously obtained data for classical His-tagged proteins.[36,37] The formation of the most stable complex was observed in the case of GrAD207_7H, which was correlated with its longest retention time. With the substitution of one of the His residues for Gly one, the strength of the complex practically did not change. A slight decrease in the strength of the complex was observed with further increase in the number of Gly residues between His triplets, which agrees with the data for binding of oligohistidine and mixed oligohistidine/ oligoalanine peptides with Ni^{2+}-NTA.[30]

Changes in the secondary structures of all obtained constructs were assessed by circular dichroism spectroscopy (Figure 6). Among GrAD207 variants, the greatest changes in the shape of the spectrum occurred in GrAD207_7H (Figure 6A), for which the prevalence of beta-sheets and irregular structures was detected (Table 3). For other GrAD207

Table 3. Results of the Quantitative CD Spectra Analysis^a

Sample		Helices (%)	β-strands (%)	Turns (%)	Unordered (%)
GrAD207	3H	24 ± 2	28 ± 3	22 ± 1	26 ± 3
	3H1G3H	22 ± 2	31 ± 3	22 ± 1	25 ± 3
	7H	0	33 ± 3	18 ± 1	48 ± 5
	3H3G3H	21 ± 1	31 ± 3	23 ± 1	25 ± 3
	untagged	43 ± 4	29 ± 3	6 ± 1	22 ± 2
eGFP	3H3G3H	14 ± 1	34 ± 3	21 ± 2	31 ± 3
	7H	10 ± 1	37 ± 4	21 ± 2	32 ± 3
	untagged	10 ± 1	46 ± 5	12 ± 1	32 ± 3

^{a}For calculating the percentage of secondary structure elements, DichroWeb instrument was used.[27]

variants, the changes were minimal, the spectrum of GrAD207_3H3G3H being the closest to that of the native structure (Figure 6A). As for the spectra of eGFP variants (Figure 6B), His-tagged eGFP variants showed a clear shift of the minimum to the left compared to untagged structure. However, the insertion of His-tags did not result in denaturation or in a disordered structure.

Quantitative analysis of the spectra for the content of secondary structure types for GrAD207 constructs (Table 3) showed the highest percentage, 24%, of helical elements and the highest percentage, 26%, of irregular structures in
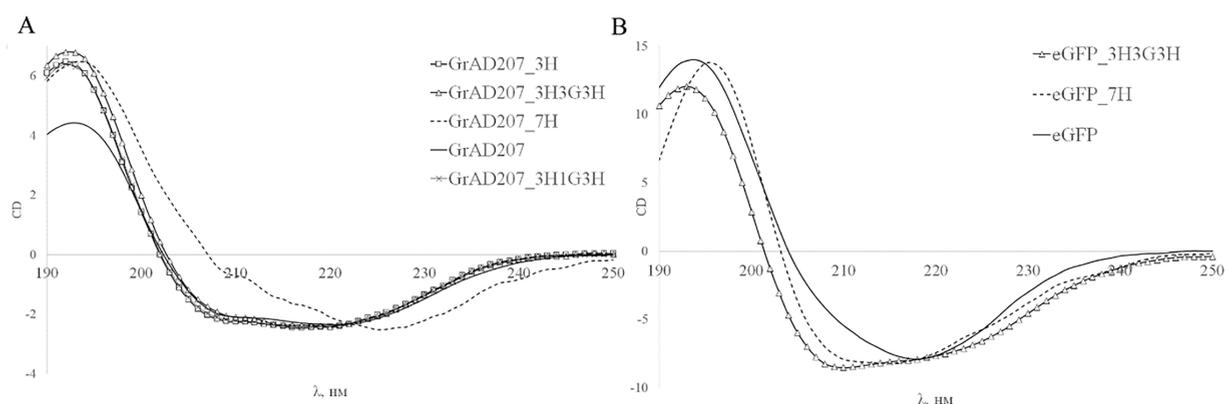
Figure 6. Investigation of the Influence of His-tags on the secondary structure of the target proteins: (A) overlay of GrAD207_3H, GrAD207_7H, GrAD207_3H1G3H, and GrAD207_3H3G3H CD spectra; (B) overlay of eGFP157_7H, eGFP157_3H3G3H CD spectra, and eGFP spectrum predicted by PDBMD2CD software.[38]

GrAD207_3H. GrAD207_3H1G3H and GrAD207_3H3G3H had almost the same ratio of secondary structure elements, but the variant with 3G showed a decrease in the proportion of helical elements and an increase in turns by 1%. For eGFP variants, quantitative spectrum analysis for secondary structure types showed an increase in helical elements of up to 14% in eGFP157_3H3G3H compared to 10% in eGFP157_7H.

Among the studied GrAD207 variants, GrAD207_3H3G3H, GrAD207_3H1G3H, and GrAD207_3H had the form of CD spectra very close to that of GrAD207 (Figure 6A). For GrAD207_3H3G3H, besides that, qualitative and quantitative analyses of its interaction with Ni^{2+} ions suggested that the insertion of the 3H3G3H tag into GrAD207 maintained or even strengthened hydrophobic interactions of GrAD207_3H3G3H. This suggestion is significant for GrAD207, which, being the modified apical domain of GroEL, retains the chaperonin function by binding substrate proteins. Probably, the GrAD207_3H3G3H construct conserved its substrate-binding surface in the most intact form, but this speculation needs experimental proof, which was out of scope of the current study. Still, the CD spectra of GrAD207 constructs, except that of GrAD207_7H, were all very similar to that of untagged control, which supports the conservation of substrate-binding functions. In the case of eGFP, preservation of its function is less clear from the resulting CD spectra (Figure 6B), and for eGFP constructs, fluorescence studies were conducted. A comparison of the shape of emission spectra (Figure S1A) showed them to be identical.

To compare the brightness of the eGFP constructs (Figure S1B, Table 4), their molar extinction coefficients (Figure S1C,

Table 4. Fluorescent Characteristics of eGFP Variants

Sample	CE, M^{-1} cm^{-1}	QY	Brightness	Relative brightness
eGFP157_3H3G3H	47 067	0.55	26.08	0.78
eGFP157_7H	53 354	0.70	37.57	1.12
eGFP	54 016	0.67	36.21	1.08
lit. eGFP[39]	55 900	0.60	33.54	1.00

Table 4) and the relative quantum yields (Figure S1D, Table 4) were obtained. The molar extinction coefficients of eGFP and eGFP157_7H had very close values of 54 016 and 53 354 $M^{-1}cm^{-1}$ respectively, while for eGFP157_3H3G3H it was 47 067 $M^{-1}cm^{-1}$. Calculation of the relative quantum yield resulted in 0.55 unit for eGFP157_3H3G3H, whereas for eGFP157_7H and eGFP it was 0.70 and 0.67 units, respectively. Literature data show the quantum yield for eGFP to be 0.60 units.[39]

The calculated brightness values of the eGFP constructs were quite consistent, with eGFP_3H3G3H being the least bright (Table 4).

The brightness of the variant with 7H tag was 1.36 units higher than that of untagged eGFP, while the brightness of eGFP157_3H3G3H was 10.13 units lower. Thus, the construct eGFP157_7H remained fully functional, with free N- and C-termini, and acquired affinity for Ni^{2+} resin.

Considering all the results taken together, it is possible to advance some arguments concerning His-tags used in the context of proteins with different structural organization.

For GrAD207, belonging to α and β structural class (three-layer $\beta\beta\alpha$ fold, 3.50 CATCH index), the 3H3G3H tag had the least influence on its secondary structure; the influence of the

3H1G3H tag was also minimal. The K_D values of GrAD207_3H3G3H and GrAD207_3H1G3H differed 2-fold, being 6.02×10^{-6} M and 2.75×10^{-6} M, respectively, while in the chromatographic study, the difference was less pronounced; the metal-affinity chromatographic method revealed only a slight difference in retention time between GrAD207_3H3G3H and GrAD207_3H1G3H. The reason for similar chromatographic behavior of the two constructs with different K_D values could lay in stronger nonspecific interactions of GrAD207_3H3G3H with the resin matrix, due to the absence of an effect of 3H3G3H on the secondary structure of GrAD207; this is also confirmed by the similarity of their CD spectra. Insertion of 7H tag had a more dramatic effect on the GrAD207 secondary structure, and it also caused aggregation and inclusion body formation, so we do not suppose this construct retains chaperone functions as the other ones do.

The effect of introduced His-tags on eGFP, belonging to the mainly β-sheet proteins (β-can fold, 2.40 CATCH index), was quite the opposite. The difference in CD spectra between untagged and His-tagged eGFP variants can be explained by the contribution of secondary structure of α-helix type, whose signal is generally more prominent than that from β-sheets. Still, as the shape of the obtained spectra did not correspond to unstructured proteins, it can be argued that the insertion of 7H and 3H3G3H tags did not lead to significant changes in the overall structural organization of eGFP. Functionally, insertion of 3H3G3H tag in the position between 157 and 158 amino acid residues of eGFP reduced quantum yield and molar extinction coefficient of the protein's chromophore, while the insertion of the 7H tag at the same position left these characteristics unchanged if not enhanced.

Summing up, the novel noncanonical His-tags, 3H1G3H and 3H3G3H, can be recommended for insertion into proteins to separate conformationally mobile structures and helical elements, while the 7H tag may be more appropriate in cases of conformationally rigid structures. The 3H1G3H tag has higher affinity to Ni^{2+}-resin than the 3H3G3H tag and may be used instead of the 7H tag if the latter causes protein damage.

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c06682.

Calculations of fluorescent characteristics of eGFP variants (Figure S1) and raw sensograms for surface plasmon resonance spectroscopy for interactions between His-labeled GrAD207 variants with Ni^{2+}NTA (Figures S2−S14) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Maria S. Yurkova − *A N Bach Institute of Biochemistry of the Russian Academy of Sciences, Moscow 119071, Russian Federation;* ⓞ orcid.org/0009-0007-5658-0891; Email: m.yurkova@fbras.ru

Authors

Anastasiia G. Tarabarova − *A N Bach Institute of Biochemistry of the Russian Academy of Sciences, Moscow 119071, Russian Federation;* ⓞ orcid.org/0009-0000-4752-0220

Anton Lopukhov − *Chemistry Department, Lomonosov Moscow State University, Moscow 119991, Russian Federation;* ● orcid.org/0000-0002-3517-505X

Alexey N. Fedorov − *FSI Federal Research Centre Fundamentals of Biotechnology of the Russian Academy of Sciences, Moscow 119071, Russian Federation;* ● orcid.org/0000-0002-7642-2360

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c06682

Notes

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Hochuli, E. Large-scale chromatography of recombinant proteins. *Journal of Chromatography A* **1988**, *444*, 293−302.

(2) Sano, K.; Atarashi, R.; Satoh, K.; Ishibashi, D.; Nakagaki, T.; Iwasaki, Y.; Yoshida, M.; Murayama, S.; Mishima, K.; Nishida, N. Prion-Like Seeding of Misfolded α-Synuclein in the Brains of Dementia with Lewy Body Patients in RT-QUIC. *Molecular Neurobiology* **2018**, *55*, 3916−3930.

(3) Sekiya, Y.; Sawasato, K.; Nishiyama, K.-i. Expression of Cds4/5 of Arabidopsis chloroplasts in E. coli reveals the membrane topology of the C-terminal region of CDP-diacylglycerol synthases. *Genes to Cells* **2021**, *26*, 727−738.

(4) Zhao, D.; Huang, Z. Effect of His-Tag on Expression, Purification, and Structure of Zinc Finger Protein, ZNF191(243−368). *Bioinorganic Chemistry and Applications* **2016**, *2016*, No. e8206854.

(5) Bartoschik, T.; Galinec, S.; Kleusch, C.; Walkiewicz, K.; Breitsprecher, D.; Weigert, S.; Muller, Y. A.; You, C.; Piehler, J.; Vercruysse, T.; Daelemans, D.; Tschammer, N. Near-native, site-specific and purification-free protein labeling for quantitative protein interaction analysis by MicroScale Thermophoresis. *Sci. Rep.* **2018**, *8*, 4977.

(6) Chao, A.; Jiang, N.; Yang, Y.; Li, H.; Sun, H. A Ni-NTA-based red fluorescence probe for protein labelling in live cells. *J. Mater. Chem. B* **2017**, *5*, 1166−1173.

(7) Gatterdam, K.; Joest, E. F.; Gatterdam, V.; Tampé, R. The Scaffold Design of Trivalent Chelator Heads Dictates Affinity and Stability for Labeling His-tagged Proteins in vitro and in Cells. *Angew. Chem.* **2018**, *130*, 12575−12579.

(8) Mahalingam, M.; Fessenden, J. D. Methods for Labeling Skeletal Muscle Ion Channels Site-Specifically with Fluorophores Suitable for FRET-Based Structural Analysis. In *Methods in Enzymology*; Shukla, A. K., Ed.; Academic Press, 2015; pp 455−474.

(9) Okitsu, K.; Misawa, T.; Shoda, T.; Kurihara, M.; Demizu, Y. Development of an ON/OFF switchable fluorescent probe targeting His tag fused proteins in living cells. *Bioorg. Med. Chem. Lett.* **2017**, *27*, 3417−3422.

(10) Ban, B.; Sharma, M.; Shetty, J. Optimization of Methods for the Production and Refolding of Biologically Active Disulfide Bond-Rich Antibody Fragments in Microbial Hosts. *Antibodies* **2020**, *9*, 39.

(11) Dagar, V. K.; Adivitiya; Khasa, Y. P. High-level expression and efficient refolding of therapeutically important recombinant human Interleukin-3 (hIL-3) in E. coli. *Protein Expression Purif.* **2017**, *131*, 51−59.

(12) Mao, S.; Song, Z.; Wu, M.; Wang, X.; Lu, F.; Qin, H. M. Expression, Purification, Refolding, and Characterization of a Neverland Protein From Caenorhabditis elegans. *Frontiers in Bioengineering and Biotechnology* **2020** *8*, DOI: 10.3389/fbioe.2020.593041.

(13) Chung, Y. H.; Volckaert, B. A.; Steinmetz, N. F. Development of a Modular NTA:His Tag Viral Vaccine for Co-delivery of Antigen and Adjuvant. *Bioconjugate Chem.* **2023**, *34*, 269−278.

(14) Chavan, S. G.; Kim, D.; Hwang, J.; Choi, Y.; Hong, J. W.; Kim, J.; Lee, M. H.; Hwang, M. P.; Choi, J. Enhanced Detection of Infectious Pancreatic Necrosis Virus via Lateral Flow Chip and Fluorometric Biosensors Based on Self-Assembled Protein Nanoprobes. *ACS sensors* **2019**, *4*, 2937−2944.

(15) Zhu, S.; Gong, C.; Ren, L.; Li, X.; Song, D.; Zheng, G. A simple and effective strategy for solving the problem of inclusion bodies in recombinant protein technology: His-tag deletions enhance soluble expression. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 837−845.

(16) Ayoub, N.; Roth, P.; Ucurum, Z.; Fotiadis, D.; Hirschi, S. Structural and biochemical insights into His-tag-induced higher-order oligomerization of membrane proteins by cryo-EM and size exclusion chromatography. *J. Struct. Biol.* **2023**, *215*, 107924.

(17) López-Laguna, H.; Rueda, A.; Martínez-Torró, C.; Sánchez-Alba, L.; Carratalá, J. V.; Atienza-Garriga, J.; Parladé, E.; Sánchez, J. M.; Serna, N.; Voltà-Durán, E.; Ferrer-Miralles, N.; Reverter, D.; Mangues, R.; Villaverde, A.; Vázquez, E.; Unzueta, U. Biofabrication of Self-Assembling Covalent Protein Nanoparticles through Histidine-Templated Cysteine Coupling. *ACS Sustainable Chem. Eng.* **2023**, *11*, 4133−4144.

(18) Aslantas, Y.; Surmeli, N. B. Effects of N-Terminal and C-Terminal Polyhistidine Tag on the Stability and Function of the Thermophilic P450 CYP119. *Bioinorganic Chemistry and Applications* **2019**, *2019*, No. e8080697.

(19) Elgharbi, F.; Ben Hlima, H.; Ben Mabrouk, S.; Hmida-Sayari, A. Expression of a Copper Activated Xylanase in Yeast: Location of the His-Tag in the Protein Significantly Affects the Enzymatic Properties. *Molecular Biotechnology* **2023**, *65* (7), 1109−1118.

(20) Esen, H.; Alpdağtaş, S.; Mervan Çakar, M.; Binay, B. Tailoring of recombinant FDH: effect of histidine tag location on solubility and catalytic properties of Chaetomium thermophilum formate dehydrogenase (CtFDH). *Preparative Biochemistry & Biotechnology* **2019**, *49*, 529−534.

(21) Majorek, K. A.; Kuhn, M. L.; Chruszcz, M.; Anderson, W. F.; Minor, W. Double trouble—Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Sci.* **2014**, *23*, 1359−1368.

(22) Flores, S. S.; Clop, P. D.; Barra, J. L.; Argaraña, C. E.; Perillo, M. A.; Nolan, V.; Sánchez, J. M. His-tag β-galactosidase supramolecular performance. *Biophys. Chem.* **2022**, *281*, 106739.

(23) Deng, A.; Boxer, S. G. Structural Insight into the Photochemistry of Split Green Fluorescent Proteins: A Unique Role for a His-Tag. *J. Am. Chem. Soc.* **2018**, *140*, 375−381.

(24) Michel-Souzy, S.; Hamelmann, N. M.; Zarzuela-Pura, S.; Paulusse, J. M. J.; Cornelissen, J. J. L. M. Introduction of Surface Loops as a Tool for Encapsulin Functionalization. *Biomacromolecules* **2021**, *22*, 5234−5242.

(25) Sharapova, O. A.; Yurkova, M. S.; Fedorov, A. N. A minichaperone-based fusion system for producing insoluble proteins in soluble stable forms. *Protein engineering, design & selection: PEDS* **2016**, *29*, 57−64.

(26) Edelheit, O.; Hanukoglu, A.; Hanukoglu, I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnology* **2009**, *9*, 61.

(27) Miles, A. J.; Ramalli, S. G.; Wallace, B. A. DichroWeb, a website for calculating protein secondary structure from circular dichroism

spectroscopic data. *Protein Science: A Publication of the Protein Society* **2022**, *31*, 37−46.

(28) Mamontova, A. V.; Shakhov, A. M.; Lukyanov, K. A.; Bogdanov, A. M. Deciphering the Role of Positions 145 and 165 in Fluorescence Lifetime Shortening in the EGFP Variants. *Biomolecules* **2020**, *10*, 1547.

(29) Noble Jesus, C. On the self-assembly of pH-sensitive histidine-based copolypeptides. *Doctoral Thesis*; University College London, 2020.

(30) Knecht, S.; Ricklin, D.; Eberle, A. N.; Ernst, B. Oligohis-tags: mechanisms of binding to Ni2+-NTA surfaces. *Journal of Molecular Recognition* **2009**, *22*, 270−279.

(31) Chen, C. W.; Liu, H. L.; Lin, J. C.; Ho, Y. Molecular Dynamics Simulations of Metal Ion Binding to the His-tag Motif. *Journal of the Chinese Chemical Society* **2005**, *52*, 1281−1290.

(32) Yurkova, M. S.; Sharapova, O. A.; Zenin, V. A.; Fedorov, A. N. Versatile format of minichaperone-based protein fusion system. *Sci. Rep.* **2019**, *9*, 15063.

(33) Yurkova, M. S.; Fedorov, A. N. GroEL—A Versatile Chaperone for Engineering and a Plethora of Applications. *Biomolecules* **2022**, *12*, 607.

(34) Cormack, B. P.; Valdivia, R. H.; Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996**, *173*, 33−38.

(35) Laskowski, R. A.; Jabłońska, J.; Pravda, L.; Vařeková, R. S.; Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Science: A Publication of the Protein Society* **2018**, *27*, 129−134.

(36) Fischer, M.; Leech, A. P.; Hubbard, R. E. Comparative assessment of different histidine-tags for immobilization of protein onto surface plasmon resonance sensorchips. *Anal. Chem.* **2011**, *83*, 1800−1807.

(37) Nieba, L.; Nieba-Axmann, S. E.; Persson, A.; Hämäläinen, M.; Edebratt, F.; Hansson, A.; Lidholm, J.; Magnusson, K.; Karlsson, A. F.; Plückthun, A. BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Anal. Biochem.* **1997**, *252*, 217−228.

(38) Drew, E. D.; Janes, R. W. PDBMD2CD: providing predicted protein circular dichroism spectra from multiple molecular dynamics-generated protein structures. *Nucleic Acids Res.* **2020**, *48*, W17−W24.

(39) Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.