A novel approach to the site-selective dual labelling of a protein via chemoselective cysteine modification†

Ramiz I. Nathani, Paul Moody, Vijay Chudasama, Mark E. B. Smith, Richard J. Fitzmaurice and Stephen Caddick*

Local protein microenvironment is used to control the outcome of reaction between cysteine residues and 2,5-dibromohexanediame. The differential reactivity is exploited to introduce two orthogonal reactive handles onto the surface of a double cysteine mutant of superfolder green fluorescent protein in a regioselective manner. Subsequent elaboration with commonly used thiol and alkyne containing reagents affects site-selective protein dual labelling.

The development of accessible chemical methodologies that enable the site-selective labelling of proteins has transformed the study and utility of these complex biomolecules.† For example, N-acetyl-lysine mimetics can be readily introduced in a defined manner to allow researchers to further investigate this key post-translational modification.‡ Whereas, in the arena of protein therapeutics, the ability to conjugate a drug, probe or lifetime extension technology to a protein in a controlled manner generates homogeneous biologics with a range of functions.§ The site-selective labelling of proteins is typically achieved by exploiting the unique reactivity of a specific functional group in the protein of interest, e.g. thiols,⃣ alkenes,⃣ azides or alkynes.¶

Even more challenging than the site-selective generation of singly modified proteins is the homogenous modification of a protein in multiple distinct positions with different probes. Access to site-selectively dual labelled proteins offers opportunities to perform a range of structural studies, employing techniques such as Förster resonance energy transfer to understand protein structure.∥ Such a methodology would also allow the construction of a range of biologics with dual function, e.g. theranostics,★ or facilitate the optimal conjugation of different probes.★ Alternatively, it might be envisaged that, it should be possible to exploit the differential nucleophilicity of thiols in double cysteine mutant proteins. However, this has brought limited success to date due to the heterogeneity of the modified protein.★

In this report we describe a fundamentally novel approach to the site-selective dual labelling of a protein at non-terminal amino acid positions. Our approach is based on the positioning of two cysteine mutants within a protein sequence such that the two cysteines are cleanly converted into two distinct products upon treatment with a single chemical reagent. More specifically, in the work described herein, a double cysteine mutant of a protein is initially transformed into a bis-sulfonium using a simple small molecule reagent. The fate of each sulfonium thereafter is controlled by the protein microenvironment, i.e. the sulfonium can either persist as a stable entity or eliminate to yield dehydroalanine, thus resulting in site-selective dual functionalisation of a protein based on substrate control. As a
therefore o
sulfonium, using the protein’s microenvironment, could
further factors e
labeling.

elimination to give dehydroalanine
therefore a single cysteine mutant at this position is prone to
preventing elimination to dehydroalanine
which the residue can be selectively introduced

\[\text{Reaction modes of sulfonium 3.}\]

Scheme 1 Reaction modes of sulfonium 3.

proof of concept, we have applied this rationale to generate a
site-selectively dual labelled superfolder green fluorescent protein.

Bioconjugation via the selective modification of free cysteine
in proteins has received a great deal of attention due to the
unique reactivity of its thiol side chain, and the ease with
which the residue can be selectively introduced via site-directed
mutagenesis. For example, Davis has described the modifica-
cation of cysteine to generate dehydroalanine, which can then
be used for bioconjugation in a range of protein substrates,
including a single cysteine mutant (S156C) of the protease
subtilisin. The formation of dehydroalanine from reaction
of cysteine thiol with 2,5-dibromohexanediamide at pH 8
(Scheme 1). However, we have recently discovered that sulfur-
ium 3, derived from a single cysteine mutant of super-
folder green fluorescent protein (S147C), can be isolated and
subsequently used in an alternate bioconjugation strategy, via a
ring opening reaction with azide, to afford an azide functional-
ised protein 5 (Scheme 1, see Fig. S7 and S8 in the ESI†). We
reasoned that the ability to control the stability of a protein
sulfonium, using the protein’s microenvironment, could
therefore offer a new approach to site-selective protein dual
labelling.

Our study began with an attempt to understand the struc-
tural factors effecting the formation of dehydroalanine 2 from a
cysteine derived sulfonium 3 based on our previous observa-
tions and those reported by Davis. As elimination to
dehydroalanine 2 inherently requires loss of the \(\alpha\)-proton of
sulfonium 3 we envisaged that its microenvironment may play
an important role in the fate of the sulfonium species. Inter-
estingly, examination of the crystal structures of subtilisin [PDB
ID: 1GCI,20 see ESI†] and superfolder GFP [PDB ID: 2B3P,21
Fig. 1] did indeed highlight differences in the environments of
the \(\alpha\)-protons of the residues S156 and S147, respectively. In
GFP the structured protein \(\beta\)-barrel appears to shield the \(\alpha\)-
proton of S147 (green) rendering it inaccessible, solvent ac-
sessible surface area = 0.0 Å\(^2\) (calculated using Naccess),\(^\text{21}\)
thus preventing elimination to dehydroalanine 2. However, in
subtilisin the S156 \(\alpha\)-proton appears to be solvent accessible
(see ESI†), solvent accessible surface area = 4.5 Å\(^2\), and there-
therefore a single cysteine mutant at this position is prone to
elimination to give dehydroalanine 2 (Scheme 1).

Using the same analysis, we identified residue T230 of
superfolder GFP (Fig. 1), close to the C-terminus, as having a
solvent accessible \(\alpha\)-proton (green), solvent accessible surface
area = 4.3 Å\(^2\). Thus, in order to evaluate our hypothesis on
the significance of \(\alpha\)-proton accessibility on the fate of sulfo-
nium, we expressed and treated a single cysteine mutant at
position 230, GFP(T230C, 233D) 6, with 2,5-dibromohexane-
diamide (4, 50 eq., 2 h, 37 °C). Gratifyingly, this generated
GFP(T230Dha, 233Δ) 7, as determined by LCMS (observed
28543, expected 28541), cleanly, as a single product after 2.5 h
at 37 °C (Scheme 2).†

Equipped with these findings we sought to explore the use
of controlled sulfonium elimination as an approach for the site-
selective dual labelling of a protein. Thus, we generated double
mutant GFP(S147C, T230C, 233D) 8 with 2,5-dibromohexanediamide (4,
50 eq., 2 h, 37 °C) to see if the results observed on the single
mutants at positions 147 and 230 would be translated. To our
delight, we observed formation of dual modiﬁed GFP(S147Sulf, T230Dha, 233Δ) 9 as the sole identiﬁable product, as deter-
determined by SDS-PAGE and LCMS (observed 28 699, expected 28 697, Fig. 2). Furthermore, direct treatment of GFP(S147Sulf, T230Dha, 233Δ) 9 with azide (NaN₃, >1000 eq., 37 °C, 2 h) resulted in the addition of a single azide to give GFP(S147Azide, T230Dha, 233Δ) 10, a site-selectively dual functionalised protein bearing orthogonal reactive handles, as determined by SDS-PAGE and LCMS (observed 28 736, expected 28 739, Fig. 2).

The regioselectivity of the reaction of GFP(S147Sulf, T230Dha, 233Δ) 9 with azide was ascertained by incubation of a fresh sample of GFP(T230Dha, 233Δ) 7 with sodium azide (>1000 eq., 37 °C, 4 h). Cross reaction of azide with dehydroalanine in 7 was not observed, conﬁrming selective ring opening of position 147 sulfonium in GFP(S147Sulf, T230Dha, 233Δ) 9 generating GFP(S147Azide, T230Dha, 233Δ) 10 (see ESI, Fig. S12†).

We then sought to demonstrate the utility of azide/dehy-
droalanine constructs through elaboration of the protein scaf-
fold via reaction with these two orthogonal groups (Scheme 3). Thus, we treated GFP(S147Azide, T230Dha, 233Δ) 10 with commercially available strained alkyne dye, dibenzylclooctyne PEG4-Fluor 545 (Jena Biosciences), to affect the formation of the expected dye-GFP conjugate (observed 29 679, expected 29 675) through a chemoselective strain-promoted alkyne-azide cycloaddition. It has previously been demonstrated that protein dehydroalanines undergo facile conjugation with free thiols, albeit not in a stereodeﬁned manner, allowing introduction of various thiolated probes.¹⁹ Thus, subsequent addition of a simple thiol, 2-mercaptoethanol, to the interme-
diate Dye-GFP conjugate gave dual labelled GFP 11 (observed 29 755, expected 29 753) decorated with a dye, tetrame-
ethylrhodamine, and a thiol in a regioselective manner.†

**Conclusions**

A facile and generally accessible methodology for the site-
selective dual modiﬁcation of proteins is a hitherto unmet need
and offers a wide range of possibilities for development of new
approaches to chemical biology and for therapeutic develop-
ment. Although other techniques are available to aﬀect the site-
selective dual labelling of a protein, they tend to be limited and
each suffer from a number of drawbacks. The strategy for che-
moselective protein dual labelling described herein enables
dual functionalisation using orthogonal reactivity of natural
amino acid side chains under mild conditions, underpinned by
two sequential chemoselective reactions. We are hopeful that
this strategy can be applied to any pair of cysteine residues
wherein one residue has an accessible α-proton, and thus
readily forms dehydroalanine, and the α-proton of the second
residue is suﬃciently shielded, and thus persists as a sulfonium
that can undergo chemoselective ring opening by the addition
of azide. Moreover, the derived azide/dehydroalanine proteins
can be further modiﬁed in a chemoselective manner using well
established methods and commercial reagents to affect site-selective dual labelling in a facile manner.

**Acknowledgements**

The authors gratefully acknowledge the EPSRC, BBSRC, Wellcome Trust, UCL, UCLB and GSK for support of our programme.

**Notes and references**


23 Images generated using PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.