De novo helical peptides as target sequences for a specific, fluorogenic protein labelling strategy†

Julia Guy,[a] Roselyne Castonguay,[a] Natalhie B. Campos-Reales Pineda,[a] Valérie Jacquier,[b] Karine Caron,[a] Stephen W. Michnick[b] and Jeffrey W. Keillor*[ab]

Received 3rd September 2009, Accepted 20th January 2010
First published as an Advance Article on the web 23rd February 2010
DOI: 10.1039/b918205e

New methods are needed to selectively label proteins in a manner that minimally perturbs their structures and functions. We have developed a 'small molecule'-based labelling technique that relies on the use of dimaleimide fluorogens that react with a target peptide sequence that presents appropriately spaced, solvent-exposed Cys residues. The thiol addition reaction between target sequence and dimaleimide fluorogen restores the latent fluorescence of the latter and results in the covalent fluorescent labelling of the protein of interest (J. Guy, K. Caron, S. Dufresne, S. W. Michnick, W. G. Skene and J. W. Keillor, *J. Am. Chem. Soc.*, 2007, **129**, 11969–11977). We demonstrated the proof-of-principle of this method previously, using a dicysteine mutant of the helical protein Fos (S. Girouard, M.-H. Houle, A. Grandbois, J. W. Keillor and S. W. Michnick, *J. Am. Chem. Soc.*, 2005, **127**, 559–566). Herein, we present the design of a novel peptide sequence presenting two Cys residues separated by two turns of an α-helix. The secondary structure of this sequence was confirmed by CD spectroscopy, before and after the fluorescent labelling reaction. A new series of di(3-methylmaleimide) fluorogens was prepared and kinetically evaluated, tuning their reactivity toward the target sequence. Attempts were made to increase the reactivity of the parent target sequence by rational design; however, the introduction of basic His residues in the vicinity of one or more Cys residues did not have the desired effect. Finally, epidermal growth factor receptors bearing the *de novo* target sequence were specifically labelled with a di(3-methylmaleimide) fluorescein fluorogen, validating our method for specific cell-surface labelling of proteins. A wide variety of fluorogen and peptide designs can be envisioned with potential applications to multiplexed labelling for the study of temporal and spatial dynamics of protein expression.

Introduction

The fluorescent labelling of proteins is a powerful approach for following the dynamic process of their synthesis and degradation, in addition to determining their localization and protein–protein interactions. Four of the most widely applied methods for specific protein labelling are (1) fluorescent protein fusion, (2) enzyme fusion, (3) enzyme substrate fusion, and (4) small molecule labelling.

The first of these methods, the genetic fusion of a protein of interest with fluorescent proteins such as jellyfish green fluorescent protein (GFP), has seen particularly broad application,[1,2] but presents some practical limitations. For example, the entire sequence of GFP must be properly folded into its 11-stranded β-barrel structure for it to function as a fluorophore, but their folding is slow and the proteins are prone to aggregation.[3] Furthermore, the steric bulk of a 27 kDa β-barrel protein can significantly perturb the stability, interactions and trafficking of the test proteins.[1,3,4]

The second method comprises the genetic fusion of an enzyme to the protein of interest. The pendant enzyme, for example, phosphopantetheine transferase (PPTase),[5,6] *O*6-alkylguanine-DNA alkyltransferase (AGT)[7–9] or mutant haloalkane dehalogenase (HALO)[10,11] can then be irreversibly labelled with a fluorescent ligand. As with the GFP-fusion method, the native biology of the labelled protein of interest may be seriously affected by the bulk of the pendant enzyme.

In the third of these methods, the protein of interest is fused with a short peptide sequence that serves as a substrate for an enzyme that catalyzes the covalent labelling of this peptide tag with a substrate that bears a functional group that can be subsequently attached to a fluorophore through bioorthogonal chemistry. Ting's Q-tag–transglutaminase system[12] and biotin acceptor peptide–biotin ligase system,[13] as well as the farnesylation motif–protein farnesyl transferase system of DiStefano[14] are three examples of this class of labelling methods.

The fourth method embodies the recent use of small organic fluorophores for labelling proteins that harbour a specific, genetically encoded motif, representing the development of powerful alternative labelling methods.[1,15–18] The 'FlAsH'

[a] *Département de chimie, Université de Montréal, Montréal, Canada QC H3C 3J7. E-mail: jw.keillor@umontreal.ca*
[b] *Département de biochimie, Université de Montréal, Montréal, Canada QC H3C 3J7*

† Electronic supplementary information (ESI) available: CD spectra of dC10 peptides, CD deconvolution data and tables of primers used for cloning and mutagenesis. See DOI: 10.1039/b918205e

method developed by Tsien and co-workers employs certain organoarsenic compounds that have been shown to react specifically with a target sequence containing four cysteine residues. Fusion of a small probe protein containing the appropriate sequence allows the test protein to be fluorescently labelled in live cells. The specificity of this 'small molecule' labelling method is dependent on the kinetics and thermodynamics of the interaction between the target peptide sequence and the labelling agent. The target peptide sequence of FlAsH labelling agents presents a tetracysteine motif in a scaffold originally thought to be a helix,[15] but later shown to be a hairpin structure.[18] This hairpin sequence was optimized[18] and minimized to the commonly used CCPGCC tag.

Since then, other groups have developed alternative[19–21] and split[22] tetracysteine binding motifs to bind various biarsenical dyes. Over the course of the widespread application of this labelling method, some drawbacks have been noted. These include the inherent toxicity of organoarsenic compounds,[23] background staining[23] that may persist despite extensive washing,[24] the sensitivity of the tetracysteine motif to oxidizing environments[4] and its tendency to form inactive intermolecular disulfide-linked aggregates.[25]

'Small molecule' labelling methods[26] that overcome some of these drawbacks would be of great value in cell biology. With this goal, we have developed a method based on the use of dimaleimide fluorogens such as **1** (Chart 1) and their fluorogenic addition reaction (FlARe) with the thiol groups of cysteine residues comprised in helical target peptide sequences (Scheme 1).[27–29] Herein we present the design of such a sequence and verification of its helical structure. The helical dicysteine target sequence was subsequently modified through a preliminary point mutation strategy in an attempt to increase its reactivity. The reactivity of complementary fluorogens was tuned to kinetically favour reaction with the target sequence. The use of this new target sequence was validated in the proof-of-principle labelling of a cell-surface protein.

Results and discussion

Target sequence design

As mentioned above, we initially used the helical transcription factor Fos as a scaffold for creating a proof-of-principle target peptide sequence.[28] The fluorescent labelling of our diCys-Fos sequence was studied *in vitro* and demonstrated notably that one equivalent of target peptide was covalently labelled with one equivalent of fluorogen, as expected. Importantly, no

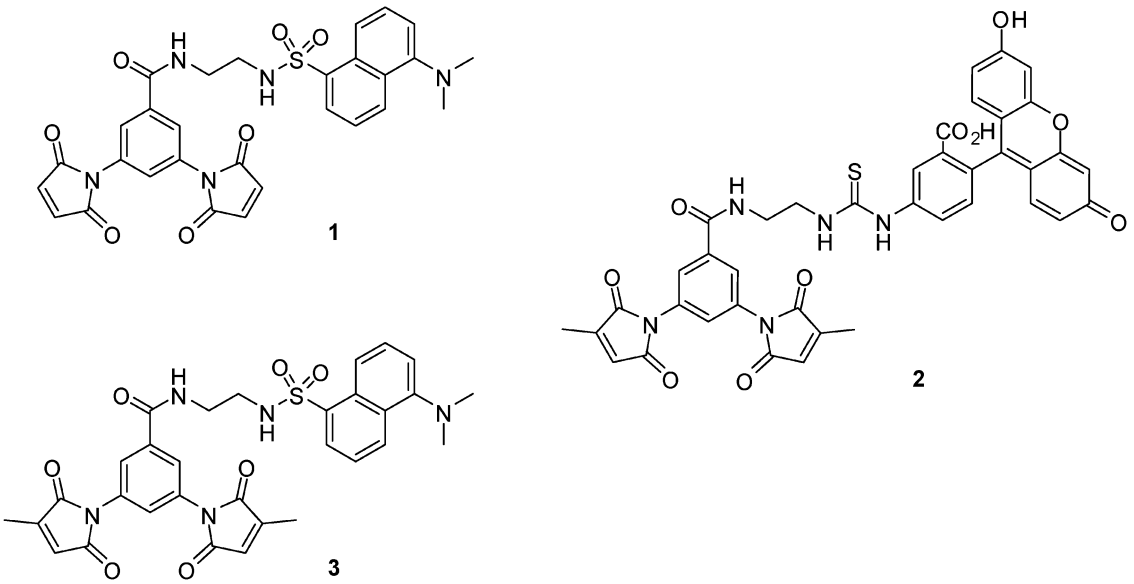
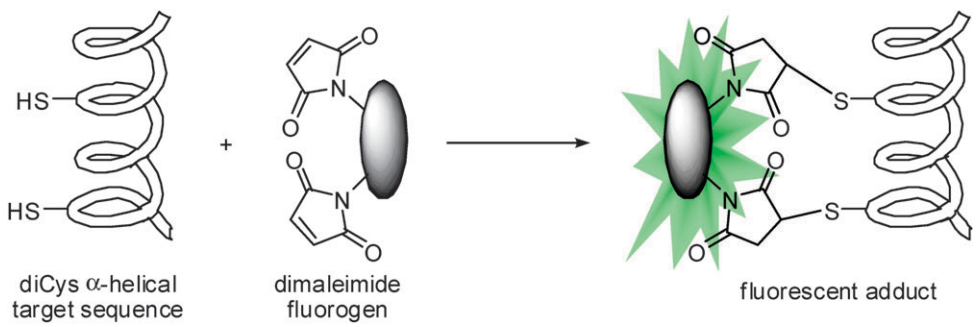


Chart 1



Scheme 1 Fluorogenic addition reaction (FlARe) labelling strategy.

cross-linked peptide adducts were observed, and the labelled monoCys-Fos mutant was only weakly fluorescent,[28] in keeping with recent studies from our laboratories that demonstrate that one unreacted maleimide residue is sufficient to quench fluorescence through a photoinduced electron transfer mechanism.[29] Furthermore, the recombinant diCys-Fos peptide could be labelled, with some selectivity, in the lysate of the *Escherichia coli* culture in which it was expressed.[27] Subsequent application of the diCys-Fos peptide sequence to intracellular labelling demonstrated the feasibility of the approach, but a tendency of the labelled protein to localize in the nucleus, probably due to the role of the parent protein Fos as a transcription factor. This result demonstrated the need to develop a short, stable helical target sequence that will not play a biological role in future cellular labelling applications.

A brief survey of the literature revealed a synthetic peptide, Ac-EAAAREAAAREAAARQ-NH_2, designed to fold into a short, monomeric α-helix comprising four complete turns.[30,31] The N-terminal acetyl group and the C-terminal carboxamide group were then replaced with known N-cap and C-cap sequences,[32] preserving the intrinsic dipole of the helical structure in the resulting genetically encodable sequence LSAAEAAAREAAAREAAAKGGK. This sequence was extended slightly to form an integral number of turns in order to further stabilize the helix,[32] giving a final 'parent sequence' of LSAAEAAAREAAAREAAARAGGK.

Two cysteine residues were then introduced as (i, i + 7) point mutants at different positions in this base sequence. The resulting dicysteine sequences were tested for predicted helicity according to the AGADIR algorithm.[33,34] The final sequence retained from this sampling, LSAAECAAREAACREAAARAGGK, was named dC10, in reference to the fact that two cysteine residues separated by two turns of an α-helix would be ~10 Å apart (Table 1).

Target sequence expression and characterization

Woolley and co-workers have designed peptide sequences containing two cysteine residues in the past, and shown them to be helical.[35] In contrast, Tsien and co-workers have shown that the introduction of four cysteine residues into a purported α-helical sequence was sufficient to provoke the sequence to adopt a β-hairpin motif.[15,18] Similarly, Mayer *et al.* have recently reported a short tetracysteine motif designed to be helical,[19] but an empirical investigation of its secondary structure has not yet been reported.

In order to independently determine the secondary structure of our dC10 sequence, it was expressed as a fusion product with an arbitrary test protein. Maltose binding protein (MBP) was chosen as a test protein for structural and kinetic characterization (*vide infra*) since it can be easily expressed in significant quantities and in soluble form. Consequently, our dC10 sequence was genetically fused to MBP *via* a thrombin-cleavable spacer and the resulting fusion protein, MBP-dC10, was expressed and purified as reported in the Experimental section. The helicity of the dC10 sequence, both before and after labelling with fluorogen $dM10^{M}$-FITC (**2**, Chart 1) was then confirmed by CD spectroscopy (see ESI† for the CD spectra (Fig. S1 and S2) and mean ellipticity data (Table S1)).

Table 1 Sequences of proof-of-principle protein diCys-Fos and de novo peptide dC10

Helix name	Amino acid sequence
DiCys-Fos[a]	MRGSHHHHHHGIHGRAQSIGRRGKVEQ LSP EEEEKRRIRRERNKMAAAKCRNRRRECTDT LQAETDQLEDEKSALQTEIANLLKEKEKLEFI LAAHRPASKIPNDLG
dC10	LSAAECAAREAACREAAARAGGK

[a] See ref. 28.

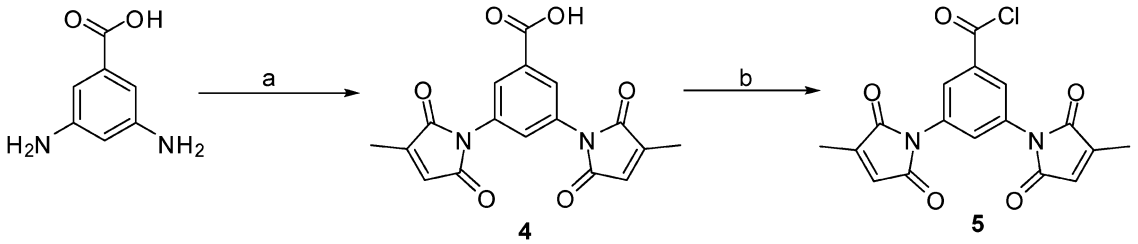
Kinetic evaluation and design of new fluorogens

In the envisaged application of our method, the dC10 helix would be genetically fused to a protein of interest, and the resulting fusion protein would react with a complementary fluorogen. The most relevant context for determining the reactivity of our new dC10 sequence was therefore deemed to be as a fusion protein with test protein MBP. Initial kinetic experiments were conducted with dansyl fluorogen **1**, the model compound of our recent photophysical studies.[29] By way of comparison, the reaction of **1** with diCys-Fos[28] and glutathione (GSH) was also evaluated kinetically. As described in the Experimental section, the fluorescence increase under equimolar conditions was measured using a stopped-flow fluorometer. The similarity of the size of the observed fluorescence increase, regardless of the identity of the thiol reactant, confirms our previous conclusion[29] that the fluorescence increase is due to the reaction of the maleimide group, rather than any potential non-covalent interactions. The fluorescence increase (Fig. 1) was then fit to a second order equation to give the rate constants shown in Table 2. Our new dC10 sequence reacts nearly 3-fold faster than our original proof-of-principle diCys-Fos protein. No reaction was observed with unmodified MBP, used as a negative control. After confirmation that the reaction with GSH is first order in thiol, the second order rate constant for the reaction of GSH with fluorogen **1** was determined by the initial rates method and found to be over an order of magnitude lower than that of MBP-dC10 (Table 2). While this preliminary result was strongly encouraging, we were concerned that the high reactivity of dansyl dimaleimide **1**, derived from the inherent electrophilicity of the maleimide group, may render it impractical for application in biological milieu. We therefore sought to attenuate its electrophilicity slightly in the design of a second generation of dimaleimide fluorogens.

To this end, the 3-methylmaleimide group was incorporated into a dimaleimide scaffold as shown in Scheme 2, *via* a

Table 2 Rate constants of reaction of dansyl fluorogens **1** and **3** with diCys-Fos, MBP-dC10 and glutathione (GSH) at 20 °C (pH 7.5)

	k_2/M^{-1} s^{-1}		
	Fluorogen **1**	Fluorogen **3**	Rate constant attenuation
DiCys-Fos	4000	53	75
MBP-dC10	11 130	190	59
GSH	1097	1.44	762



Scheme 2 (a) (i) citraconic anhydride, acetone, rt, 2 h (ii) HMDS, $ZnCl_2$, toluene–DMF, reflux toluene, 1.5 h, 86% for two steps; (b) $SOCl_2$, reflux, 3 h, 80%.

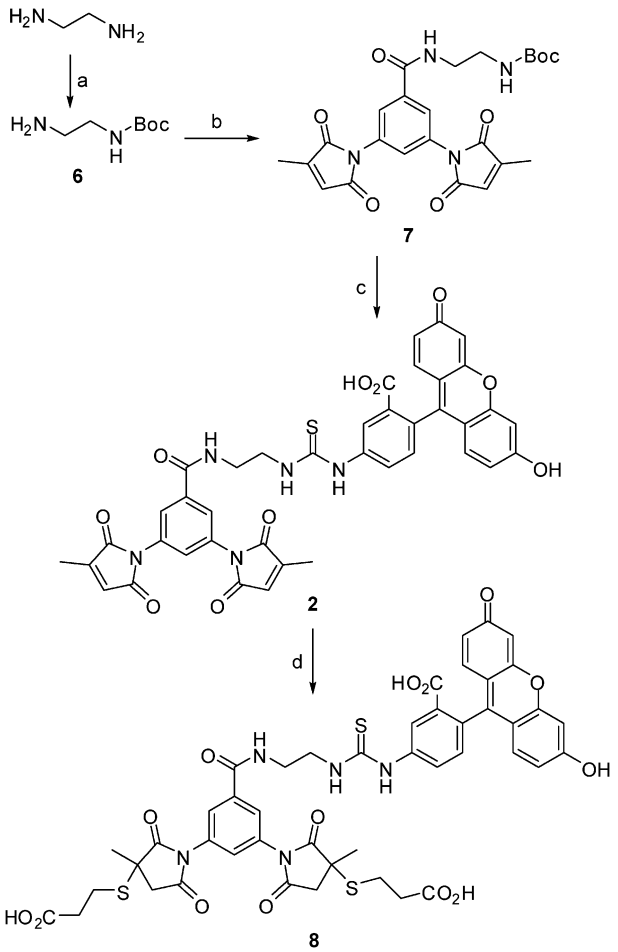
synthetic route similar to that developed previously.[29] Briefly, the 3-methylmaleimide groups of dimaleimide **4** were added in excellent yield by an efficient two-step reaction involving addition of citraconic anhydride and subsequent cyclization of the corresponding maleamic acid in the presence of a Lewis acid and HMDS. The resulting dimaleimide acid **4** was heated to reflux in thionyl chloride to give the corresponding acid chloride **5** in excellent yield.

The methylated dimaleimide moiety $dM10^M$ was then incorporated into fluorogens comprising either a fluorescein group (**2**) or a dansyl group (**3**), as shown in Schemes 3 and 4. This convergent synthetic strategy, previously validated for the synthesis of dM10-dansyl **1**,[29] allows the rapid preparation of fluorogens comprising different quenched fluorophores that exhibit different spectral properties after their thiol addition labelling reactions. For the synthesis of fluorescein fluorogen **2**, acid chloride **5** was condensed with ethylenediamine, initially mono-protected as Boc derivative **6**,[29] to afford amide **7** (Scheme 3). After removal of the Boc protecting group and condensation with fluorescein isothiocyanate (FITC), fluorogen **2** was obtained in good yield.

In the synthesis of dansyl fluorogen **3**, the mono-protected diamine **6** was first condensed with dansyl chloride to give sulfonamide **9** (Scheme 4). Removal of the Boc group followed by immediate reaction of the resulting primary amine with acid chloride **5** afforded the desired dansyl fluorogen **3** quantitatively. Both fluorogens **2** and **3** were treated with an excess of mercaptopropionic acid (MPA) to give the corresponding dithiolated derivatives **8** and **10**, respectively, in good yields.

The product of the reaction of $dM10^M$ fluorogen **3** with the dC10 peptide sequence was then confirmed by mass spectrometry. The mass of MBP-dC10 was measured by ESI-MS before and after FlARe labelling with $dM10^M$-dansyl (**3**). The mass of the resulting adduct was greater than that of MBP-dC10 by 652 mass units (Table 3). Compared to the theoretical increase of 615 mass units, this result clearly confirms the 1 : 1 stoichiometry expected for the FlARe labelling reaction of all fluorogens bearing similar dimaleimide fragments displaying similar reactivity (see below).

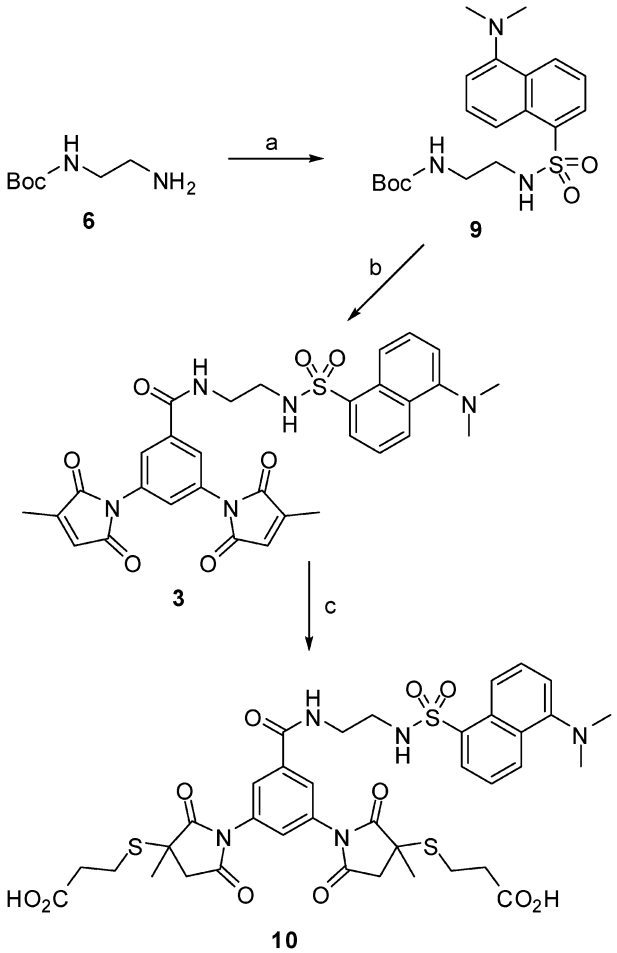
The observed mass also suggests that in the sample submitted for MS analysis, the adduct imides may have subsequently reacted with two equivalents of water (+36 mass units) to form the ring-opened hydrolysis product. The hydrolysis of maleimides has been studied kinetically[36–39] and is much slower than their thiol addition reactions.[36,39] Adducts resulting from the thiol addition reaction have also been shown to



Scheme 3 (a) Boc_2O, CH_2Cl_2, rt, 16 h, 89%; (b) **5**, NEt_3, CH_2Cl_2–CH_3CN, rt, 2 h, 94%; (c) (i) TFA, rt, 2.5 h (ii) FITC, NEt_3, DMF, 18 h, 80%; (d) 3-mercaptopropionic acid, acetone, rt, 18 h, 71%.

undergo hydrolysis, but even more slowly than the parent maleimide.[36] This putative hydrolytic side reaction apparently does not affect the C–S bond formed upon thiol addition, nor the fluorescence of the addition product.[27,28,40]

The introduction of a methyl group to each maleimide group of dM10-dansyl **1** was found to attenuate the reactivity of the resulting fluorogen **3**, as expected. The second order rate constants for the reaction of $dM10^M$-dansyl **3** with proof-of-principle protein diCys-Fos, generic target protein MBP-dC10 and control thiol GSH are shown in Table 2.



Scheme 4 (a) Dansyl chloride, NEt_3, CH_2Cl_2, rt, 6 h, 100%; (b) (i) TFA, rt, 1.5 h (ii) **5**, NEt_3, CH_3CN–CH_2Cl_2, 18 h, 100%; (c) 3-mercaptopropionic acid, CH_3CN–DMSO, rt, 24 h, 83%.

Table 3 Experimental ESI-MS masses measured for MBP-dC10, before and after FlARe labelling with dansyl fluorogen **3**

Protein	Observed mass
MBP-dC10	46 387[a]
MBP-dC10-dM10^M-dansyl	47 039[b]

[a] Theoretical mass of 46 371 (monoisotopic) or 46 399 (average) was calculated using Compute PI/MW on Expasy.org. [b] Theoretical mass of 47 014 expected from labelling with one equivalent of **3**.

The same relative reactivity was observed as in the case of fluorogen **1**, namely, MBP-dC10 > diCys-Fos ≫ GSH. Interestingly, while the second order rate constants for reaction with protein targets were lowered by roughly two orders of magnitude, the rate constant for reaction with GSH was lowered by nearly three orders of magnitude (Table 2). As a consequence, fluorogen **3** shows even greater kinetic selectivity than fluorogen **1** for reaction with the target sequence dC10 compared to GSH. Fig. 1 clearly illustrates the striking distinction between target sequence reactivity relative to the reactivity of a representative adventitious thiol; after less than 5 minutes, the fluorescence of labelled protein is optimally distinguished over that of the di-GSH adduct.

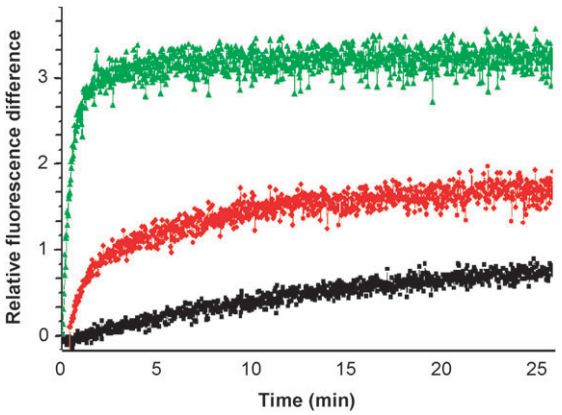


Fig. 1 Reaction of 100 μM of fluorogen dM10^M-dansyl **3** with 100 μM of test protein MBP-dC10 (green), 100 μM of diCys-Fos (red) or 200 μM GSH (black) as a representative adventitious thiol (λ_{exc} = 495 nm, λ_{fl} = 525 nm). Fluorescence intensity was normalized to that of **3**, which was then subtracted as background.

Further attenuation of maleimide reactivity may enhance this distinction even further.

Fluorometric characterization of fluorogens and their dithiolated adducts

Fluorescence spectra were recorded for parent fluorophores FITC and dansylamide, as well as their dM10^M-fluorogens (**2** and **3**). Comparison of excitation and emission wavelengths (Table 4) reveals very little red- or blue-shifts for a given fluorophore upon coupling to a dimaleimide moiety. However, comparison of quantum yields demonstrates the sharp decrease in fluorescence due to the presence of an intramolecular 3-methylmaleimide group. Recently we showed that the mechanism for fluorescence quenching by a maleimide group involves non-radiative electron transfer from the excited state of the fluorophore to the LUMO of the maleimide group.[29]

Fluorogens **2** and **3** were then allowed to undergo two thiol addition reactions with mercaptopropionic acid (MPA) (see Schemes 3 and 4). Fluorometric characterization of the resulting dithiolated adducts **8** and **10** is summarized in Tables 4 and 5. The absorptivities and quantum yields measured for dithiolated adducts **8** and **10** (Table 4) compare favourably to those measured for their respective parent

Table 4 Spectroscopic data of reference compounds, synthetic fluorogens and their dithiolated adducts

Compound	λ_{exc}/nm	ε_{max}/M^{-1} cm^{-1}	λ_{fl}/nm	ϕ_{fl}[c]
FITC[a]	495	5800	525	0.93[d]
dM10^M-FITC (**2**)[a]	495	12 700	525	0.38
8[a]	495	8000	525	0.60
Dansylamide[b]	335	3200	511	0.37[e]
dM10^M-dansyl (**3**)[b]	335	3600	518	0.07
10[b]	335	2500	525	0.15

[a] Measurements performed in basic ethanol solution. [b] Measurements performed in acetonitrile. [c] Quantum yields were calculated according to the method shown in ref. 42. [d] Value taken from ref. 43. [e] Value taken from ref. 44.

Table 5 Fluorescence enhancement of synthetic fluorogens upon reaction with two equivalents of thiol in aqueous media

Fluorogen	λ_{exc}/nm	λ_{fl}/nm	FE[a]
dM10-dansyl (**1**)	331	530	2.1
$dM10^{M}$-FITC (**2**)	494	517	5.8
$dM10^{M}$-dansyl (**3**)	331	551	2.0

[a] Fluorescence enhancement was determined from the ratio of fluorescence intensity after reaction with two equivalents of MPA in 50 mM HEPES (pH 7.5) with 5% DMSO (dithiolated adduct) and the fluorescence intensity before reaction (non-thiolated fluorogen).

fluorophores, confirming the brightness of the fluorescent adducts. Additionally, the fluorescence enhancement (FE) values of fluorogens **1–3** were measured upon reaction with two equivalents of MPA (Table 5). These measurements were made in aqueous media in order to be as pertinent as possible to envisaged labelling applications.

From these values, it is clear that the quenching efficiency of the $dM10^{M}$ moiety varies significantly from one fluorogen to another. The value of 5.8 measured for FITC-based fluorogen **2** is comparable to those measured previously for various fluorogens bearing maleimide groups attached directly to their fluorescent cores (FE = 8 to 50)[28] and bodes well for practical applications (*vide infra*). By way of comparison, the value of 2.0 measured for dansyl-based fluorogen **3** is significantly lower, so it was not selected for further application herein (see also Fig. S3, ESI†).

The Rehm–Weller equation predicts that quenching would be more thermodynamically favoured for higher energy excited states,[29] and the greater fluorescent enhancement recently measured for a blue $dM10^{M}$ fluorogen confirms this prediction.[41] This may explain why the green fluorogen **2** (λ_{fl} = 517 nm) is more effectively quenched than the yellow-green fluorogen **3** (λ_{fl} = 551 nm). However, it is highly probable that solvation and conformational effects are at least as important. For example, the FE value of dansyl fluorogen **1** was found to be 2.1 in aqueous media (Table 5), whereas a value of 5.5 was measured previously in acetonitrile.[29] It is possible that in the latter case, fluorogen **1** adopts a conformation that allows more efficient PET quenching. To probe this hypothesis, we are currently exploring the effect of the conformational rigidity and length of the spacer group between the fluorophore and the $dM10^{M}$ moiety with respect to the observed fluorescent enhancement.[41]

Point mutation of dC10 sequence

The reactivity of the first Cys residue of the dC10 sequence is pertinent to the efficiency of a given labelling reaction, whereas the second Cys residue will react rapidly in a virtually instantaneous intramolecular addition reaction, since the effective concentrations determined for intramolecular nucleophilic addition reactions range from 10^{6} M to 10^{10} M,[45–47] with an average value of 10^{8} M reported for systems whose conformations are not constrained.[48] The parent sequence of our dC10 target peptide was therefore modified by a point mutation strategy with the goal of improving the efficiency of its labelling reaction.

Since the mechanism of the thiol addition reaction in question involves nucleophilic attack by the thiolate, we attempted to increase the relative proportion of this ionization state by lowering the pK_a of the cysteine sulfhydryl group. Basic histidine residues were therefore substituted into the sequence, in the spatial proximity of at least one of the cysteines of dC10 (Table 6). Thus, a histidine residue was introduced at position 2 (separated from Cys6 of the sequence dC10-H2 by one turn of the α-helix), at position 7 (adjacent to Cys6 in dC10-H7), at position 9 (on the turn of the α-helix between Cys6 and Cys13 in dC10-H9) and at positions 2 and 17 (one turn distant from each of the two Cys residues in dC10-H2H17). The cloning, expression and purification of these mutants are reported in the Experimental section. Measurement of the second order rate constants of their reaction with fluorescein-based fluorogen **2** is also described in the Experimental section and the values are reported in Table 6.

From this table, the rate constant of 256 M^{-1} s^{-1}, measured for the reaction of **2** with MBP-dC10, can be used to make three instructive comparisons. Firstly, compared to the rate constant measured for the reaction of **2** with GSH (0.3 M^{-1} s^{-1}) it is apparent that the test protein MBP-dC10 sequence is at least two orders of magnitude more reactive, as was observed for fluorogen **3** (Table 2). Secondly, compared to the rate constant measured for the reaction of MBP-dC10 with fluorogen **3** (Table 2) it is apparent that both fluorogens, bearing identical $dM10^{M}$ moieties, have very similar reactivities towards the same test protein. This suggests the fluorophore attached to the other end of the molecule does not have a strong influence on fluorogen reactivity. Thirdly, it is also apparent from Table 6 that introduction of histidine residues into the dC10 parent sequence offers no kinetic

Table 6 Rate constants for reaction of $dM10^{M}$-FITC (**2**) with series of dC10-helices

Helix name	Peptide sequence of MBP C-terminal fused helix	k_2/M^{-1} s^{-1}[a]
dC10	L_1SAAEC̲AAREAAC̲REAAARAGGK$_{23}$	256
dC10-H2	L_1**H$_2$**AAEC̲AAREAAC̲REAAARAGGK$_{23}$	43
dC10-H7	L_1SAAEC̲**H$_7$**AREAAC̲REAAARAGGK$_{23}$	33
dC10-H9	L_1SAAEC̲AA**H$_9$**EAAC̲REAAARAGGK$_{23}$	N.R.[b]
dC10-H2H17	L_1**H$_2$**AAEC̲AAREAAC̲REA**H$_{17}$**ARAGGK$_{23}$	92
dC10-A9	L_1SAAEC̲AAA$_9$EAAC̲REAAARAGGK$_{23}$	32
dC10-R8A9	L_1SAAEC̲A**R$_8$A$_9$**EAAC̲REAAARAGGK$_{23}$	85

[a] All kinetic studies performed at 20 °C. To a solution of 100 μM peptide in 50 mM HEPES (pH 7.5), with 10% DMSO, was added 100 μM of fluorogen. The sample was excited at 495 nm and the fluorescence emission was followed at 525 nm as a function of time. [b] No reaction detected.

advantage and is in fact generally deleterious. This suggests there is little cooperativity between the ionization states of adjacent imidazole and thiol groups, perhaps owing to their orientation or solvation.

We were particularly intrigued by the lack of activity of mutant dC10-H9 and wondered if the presence of the aromatic histidine side-chain at the $i + 3$ position in between the cysteine residues sterically hindered the reaction with fluorogen **2**, or if the resulting sequence showed less propensity to fold into a helix. To test the sensitivity of the reaction to mutations at that position, the sequences dC10-A9 and dC10-R8A9 were prepared. Upon kinetic evaluation, both showed modest activity. The point mutant dC10-A9 presents an alanine residue in the place of the histidine residue of sequence dC10-H9; the fact that it shows some activity is consistent with the hypothesis that the histidine may sterically hinder reaction with a 3-methylmaleimide group. Since the Arg9 residue of the parent dC10 sequence may also cause some steric hindrance, it was exchanged with Ala8 to give the sequence dC10-R8A9. This sequence showed greater reactivity than the dC10-A9 sequence, underlining the overall importance of the presence of the arginine residue and suggesting it may play an important structural role. However, the reduced reactivity of dC10-R8A9 relative to the parent dC10 sequence suggests that the putative structural role of arginine is served more efficiently at position 9, potentially forming a stabilizing ionic interaction with Glu5 on the adjacent turn of the helix.

Detection of cell-surface expression of EGF receptors labelled by FlARe

In its current state, the FlARe method presented herein demonstrates potential for development into a protein labelling strategy. Preliminary results indicate that dimaleimide fluorogens have low cytotoxicity.[27] The two-point, robust covalent labelling of a secondary structural motif of limited conformational flexibility may allow the orientation of a protein of interest to be determined in living cells using polarized fluorescence microscopy.[49] The modular design[29] of the fluorogens allows for facile variation of their colour or other properties, such as solubility, for monitoring protein externalization.[50] Furthermore, the distance between the cysteine residues in the dC10 sequence should render them less susceptible to the oxidation and disulfide bond formation associated with short tetracysteine motifs.[4,25] By this reasoning, the dC10 helix may be a more appropriate peptide target for the labelling of proteins expressed in an oxidizing environment on the outer surface of a cell.

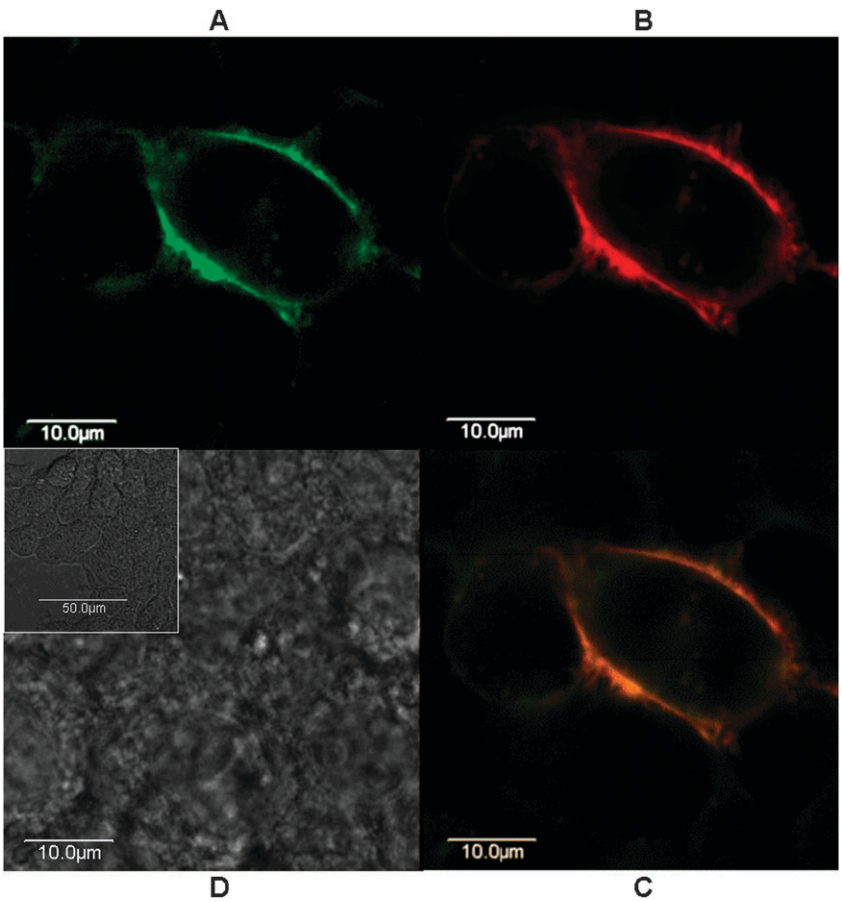


Fig. 2 Labelling of EGF receptors. A: Confocal images of HEK 293 cells expressing EGFR tagged with dC10 sequence and labelled with $dM10^{M}$-FITC (**2**). B: Same cell as A, with EGFR bound to rhodamine-labelled receptor ligand EGF. C: Superposition of A and B showing overlap of rhodamine-fluorescein signals. D: Phase-contrast image of the same field.

We sought to exploit this aspect of our current method in an immediate application of FlARe to the study of cell membrane protein expression, a process key to normal cell differentiation as well as pathological processes such as cancers. Existing large probes for the dynamics of membrane protein expression, including GFP-tagged receptors or fluorescently labelled membrane protein-specific antibodies, could modify coupling of membrane proteins to the endocytic machinery that controls protein membrane insertion, internalization and recycling. For instance, the latter can cause an artifactual distribution of the membrane proteins by cross-linking them on the cell-surface.[51,52] By using membrane-impermeable FlARe probes against membrane proteins, it should be possible to selectively label the proteins without background of intracellular pools of already internalized proteins and to monitor the fate of protein under conditions that change their turnover.

As a proof-of-principle experiment, we then tested our labelling strategy on a cell-surface protein. For this simple demonstration, we chose to label the oncogene epidermal growth factor receptor (EGFR).[53–55] The coding sequence for EGFR was fused to that coding for dC10 at its N-terminus and the fusion was expressed in a human fibroblast cell line, HEK 293. Since the product of this fusion is expressed with the N-terminus oriented extracellularly, the dC10 sequence is accessible to fluorogen. Addition of the membrane-impermeable fluorogen **2** to cells in culture allowed for the direct visualization of this receptor on the surface of living cells (Fig. 2A). In contrast, addition of fluorogen **2** to cells expressing EGFR that lacked the dC10 sequence did not result in any fluorescent cell-surface labelling (ESI†, Fig. S4). The specificity of the labelling was further demonstrated by the co-localization of the signal emanating from the adduct of **2** with that of the ligand EGF conjugated to rhodamine (Fig. 2B).

This simple labelling experiment serves as a clear demonstration that specific labelling of membrane proteins is feasible with the FlARe strategy. This allows us to envisage temporal studies of externalization, internalization, recycling and degradation of individual proteins. With the development of different fluorogen peptide pairs (in progress), the labelling of multiple membrane proteins may also be possible.

Experimental

Materials

All synthetic starting materials were purchased from Sigma Aldrich and used without further purification. Solvents were dried using GlassContour System (Irvine, CA) columns.

The restriction enzymes *SalI* and *HindIII*, SAP (shrimp alkaline phosphatase) and all reagents used for PCR and primer phosphorylation were purchased from New England Biolabs and Fermentas. All buffer components were purchased from Sigma and used without further purification. The components of the bacterial cultures, IPTG and ampicillin were purchased from BioShop (Montreal). Primers were ordered from BioCorp (Montreal, Canada) or IDT DNA (Iowa, USA). All plasmid constructs were sequenced by the Genomic platform of IRIC (Université de Montréal, Canada).

Protein concentration was determined by Bradford assay (Bio-Rad) and the concentration of the cleaved dC10 helix by BCA assay (using BSA as a standard in each case). Cleavage of the dC10 moiety was performed with thrombin–biotin (Calbiochem). The CD spectra were recorded on an Applied Photophysics instrument.

Cell culture medium and the mammalian expression vector were obtained from Invitrogen; transfection reagent was obtained from Mirus. Cells were grown in plates purchased from Costar, Corning Inc., NY. Rhodamine EGF was purchased from Molecular Probes. Labelled cells were imaged on a Olympus FV3000 inverted confocal microscope.

Methods

Synthesis

Generalities. Reactions requiring anhydrous conditions were carried out under dry nitrogen atmosphere employing conventional bench-top techniques. ^{13}C-NMR and ^{1}H-NMR spectra were recorded on AMXR400 or AMX300 spectrometers and were referenced to the residual proton or ^{13}C signal of the solvent. Mass spectra were determined by FAB+ ionization on an AutoSpec Q spectrometer at the Regional Mass Spectrometry Centre at the Université de Montréal. Melting points (uncorrected) were determined on a Thomas-Hoover Unimelt melting point apparatus.

Synthetic procedures

*3,5-Di(3-methylmaleimido)benzoic acid (**4**).* A solution of 3,5-diaminobenzoic acid (563 mg; 3.70 mmol; 1 eq.) and citraconic anhydride (1.0 mL; 11.1 mmol; 3 eq.) in acetone (20 mL) was stirred at room temperature for 3 h. The resulting maleamic acid was produced as a yellowish precipitate that was filtered, washed with acetone, dried and used immediately without further purification. HMDS (2.31 mL; 11.1 mmol; 3 eq.) and $ZnCl_2$ (1.0 g; 7.4 mmol; 2 eq.) were added to a stirred suspension of the maleamic acid in toluene and the mixture was heated to reflux for 1.5 h. After cooling to room temperature, EtOAc was added (30 mL) and the organic layer was washed successively with a 10% solution of HCl (20 mL) and brine (20 mL), dried over $MgSO_4$ and evaporated to give acid **4** (1.18 g; 3.47 mmol) as a beige solid in 94% yield. Mp: >180 °C. ^{1}H-NMR (DMSO-d_6): 7.96 (d, J = 1.7 Hz, 2H), 7.63 (d, J = 1.8 Hz, 1H), 6.82 (d, J = 1.8 Hz, 2H), 3.29 (br s, –OH), 2.10 (s, 6H). ^{13}C-NMR (DMSO-d_6): 170.4, 169.5, 166.2, 146.2, 132.8, 132.2, 128.2, 127.8, 126.0, 11.0. MS: $[M + H]^+$: 341.07681, $[M + Na]^+$: 363.05876.

*3,5-Di(3-methylmaleimido)benzoyl chloride (**5**).* A stirred suspension of acid **4** (911 mg; 2.68 mmol; 1 eq.) in thionyl chloride (11 mL) was heated to reflux until a clear solution had formed, after which the excess $SOCl_2$ was evaporated *in vacuo* and the resulting powder was dried under vacuum to obtain 756 mg (2.11 mmol; 79%) of **5** as a yellow solid which was used without further purification. Mp: >180 °C. ^{1}H-NMR (DMSO-d_6): 7.95 (d, J = 1.9 Hz, 2H), 7.63 (t, J = 1.9 Hz, 1H), 6.84 (d, J = 1.8 Hz, 2H), 2.08 (d, J = 1.8 Hz, 6H). ^{13}C-NMR (DMSO-d_6): 170.3, 169.3, 166.0, 146.0, 132.7, 132.1, 128.8, 127.7, 125.9, 10.9. MS: $[M + Na]^+$: 381.02487.

N-*[2-*tert*-Butoxycarbonylaminoethyl]-3,5-di(3-methylmaleimido)-benzamide (***7***)*. To a solution of **5** (400 mg; 1.12 mmol; 1 eq.) in a mixture of dry CH_2Cl_2–CH_3CN (1 : 1, v/v, 60 mL) were added amine **6**[29] (232 mg; 1.50 mmol; 1.3 eq.) in CH_2Cl_2 (6 mL) and 0.31 mL of triethylamine (2.24 mmol; 2.0 eq.) under a nitrogen atmosphere. After stirring at room temperature for 2 h 45 min, 100 mL of CH_2Cl_2 were added and the organic layer was washed with a saturated solution of $NaHCO_3$ (4 × 30 mL), dried over $MgSO_4$ and removed *in vacuo* to give the desired compound **7** (469 mg; 0.97 mmol) as a brown solid. Mp: 138–140 °C. ^{1}H-NMR ($CDCl_3$): 7.83 (d, J = 1.9 Hz, 2H), 7.62 (t, J = 1.9 Hz, 1H), 6.50 (d, J = 1.8 Hz, 2H), 5.11 (br s, 1H), 3.56–3.52 (m, 2H), 3.42–3.38 (m, 2H), 2.18 (d, J = 1.8 Hz, 6H), 1.40 (s, 9H). ^{13}C-NMR ($CDCl_3$): 170.1, 169.0, 166.2, 146.2, 136.1, 132.7, 127.8, 125.1, 123.1, 80.0, 42.1, 40.3, 28.5, 11.3. MS: $[M + Na]^+$: 505.16937.

N-*[2-(5-Fluoresceinthiocarbamoyl)aminoethyl]-3,5-di(3-methyl-maleimido)benzamide (***2***)*. Amide **7** (218 mg; 0.45 mmol; 1 eq.) was combined with trifluoroacetic acid (12 mL) under a nitrogen atmosphere at room temperature. The reaction mixture was stirred for 2.5 h, then solvent was removed *in vacuo* to afford the corresponding primary amine quantitatively as its trifluoroacetate salt in the form of a viscous brown oil (223.2 mg; 0.45 mmol). To a solution of this amine (223.2 mg; 0.45 mmol; 1.1 eq.) in dry DMF (20 mL) was added a solution of FITC (160 mg; 0.41 mmol; 1.0 eq.) in dry DMF (8 mL), followed by anhydrous Et_3N (0.35 mL; 2.52 mmol; 5.6 eq.). The mixture was stirred at room temperature overnight under nitrogen atmosphere. The solvent was removed *in vacuo* and the crude product dissolved in EtOAc (100 mL). The organic layer was washed with distilled water (4 × 50 mL), dried over $MgSO_4$ and removed *in vacuo* to yield 243.4 mg (0.32 mmol; 77%) of the desired product **2** as an orange powder. Mp: >180 °C. ^{1}H-NMR (DMSO-d_6): 10.13 (br s, H), 10.04 (br s, NH, 1H), 8.83 (br s, NH, 1H), 8.20–8.19 (m, 2H), 7.91 (d, J = 1.8 Hz, 2H), 7.74 (br s, 1H, NH), 7.52 (t, J = 1.8 Hz, 1H), 7.19 (d, J = 8.4 Hz, 1H), 6.84 (d, J = 1.8 Hz, 2H), 6.68 (d, J = 1.9 Hz, 2H), 6.63–6.54 (m, 4H), 3.73 (br s, 2H), 3.57–3.54 (m, 2H), 2.08 (s, 6H). ^{13}C-NMR (DMSO-d_6): 169.5, 162.5, 159.8, 159.0, 158.2, 154.9, 149.9, 143.0, 138.8, 137.5, 132.9, 127.9, 126.8, 125.0, 122.7, 122.0, 120.6, 120.3, 119.7, 117.9, 117.5, 106.8, 104.1, 97.3, 79.6, 43.2, 34.5, 13.2. MS: $[M + H]^+$: 772.17079, $[M + Na]^+$: 794.15273.

N-*[2-(5-Fluoresceinthiocarbamoyl)aminoethyl]-3,5-di-(3-methyl-3-(2-carboxyethylthio)-1-succinimidyl)benzamide (***8***)*. To a solution of compound **2** (27.8 mg; 0.04 mmol; 1 eq.) in acetone (5 mL), mercaptopropionic acid (7.86 μL; 0.09 mmol; 2.5 eq.) was added and the mixture was stirred at room temperature overnight. Et_2O (10 mL) was added and the resulting precipitate was filtered and washed with Et_2O (3 × 30 mL) to give **8** (25 mg; 0.03 mmol) as yellow solid in 71% yield. Mp: >180 °C. ^{1}H-NMR (DMSO-d_6): 8.11–8.10 (m, 1H), 7.81 (d, J = 1.9 Hz, 2H), 7.69 (br s, 1H, NH), 7.67 (br s, 1H, NH), 7.49 (t, J = 1.9 Hz, 1H), 7.11 (d, J = 8.2 Hz, 1H), 6.62–6.46 (m, 7H), 4.07 (br s, 2H), 3.80–3.72 (m, 2H), 3.58–3.49 (m, 3H), 2.05 (s, 6H). ^{13}C-NMR (DMSO-d_6): 169.5, 162.6, 159.8, 159.0, 158.3, 154.9, 150.0, 143.0, 138.8, 137.5, 132.9, 127.9, 126.8, 125.0, 122.7, 122.0, 120.7, 120.3, 119.8, 118.5, 117.9, 117.5, 106.8, 104.1, 97.2, 79.6, 62.9, 45.3, 43.2, 36.1, 34.5, 17.2, 13.2. MS: $[M - C_3H_6O_2S]^+$: 878.1781, $[M - C_6H_{12}O_4S_2]^+$: 772.1712.

N-*[2-*tert*-Butoxycarbonylaminoethyl]-5-(dimethylamino)-naphthalene-1-sulfonamide (***9***)*. To a solution of dansyl chloride (613 mg; 2.30 mmol; 1 eq.) in dry CH_2Cl_2 (20 mL) were added a solution of amine **6**[29] (400 mg; 2.5 mmol; 1.1 eq.) in CH_2Cl_2 (4 mL) and Et_3N (0.64 mL; 4.6 mmol; 2 eq.) under a nitrogen atmosphere. After stirring at room temperature for 6 h the organic layer was washed with a saturated solution of $NaHCO_3$ (3 × 30 mL), dried over $MgSO_4$ and removed *in vacuo* to yield compound **9** quantitatively as yellow foam. ^{1}H-NMR ($CDCl_3$): 8.57 (d, J = 8.5 Hz, 1H), 8.33 (d, J = 8.6 Hz, 1H), 8.25 (dd, J = 7.3 Hz, J = 1.1 Hz, 1H), 7.63–7.51 (m, 2H), 7.21 (d, J = 7.5 Hz, 1H), 5.83 (br s, 1H, NH), 5.03 (br s, 1H, NH), 3.32–3.14 (m, 2H), 3.06–2.98 (m, 2H), 2.91 (s, 6H), 1.40 (s, 9H). ^{13}C-NMR ($CDCl_3$): 157.3, 152.7, 135.4, 134.2, 131.3, 130.7, 130.4, 129.3, 124.1, 119.6, 116.1, 80.5, 46.3; 44.4, 41.1, 29.2. MS: $[M + H]^+$: 394.17950, $[M + Na]^+$: 416.16145.

N-*[2-(5-(Dimethylamino)naphthalene-1-sulfonyl)aminoethyl]-3,5-di(3-methylmaleimido)benzamide (***3***)*. Amide **9** (709.3 mg; 1.80 mmol; 1 eq.) and trifluoroacetic acid (10 mL) were combined under a nitrogen atmosphere at room temperature. The reaction mixture was stirred for 1.5 h, then solvent was removed *in vacuo* to give an oily residue. The resulting crude product was used without further purification and dissolved in dry CH_2Cl_2 (10 mL) and a solution of **5** (302.4 mg; 0.84 mmol; 1.2 eq.) in a mixture of dry CH_2Cl_2–CH_3CN (1 : 1, v/v, 40 mL) was added followed by 0.66 mL of triethylamine (4.92 mmol; 7.0 eq.). After stirring overnight at room temperature under nitrogen atmosphere the solvent was removed *in vacuo*. The resulting crude product was dissolved in $CHCl_3$ and the organic phase was washed with a saturated solution of $NaHCO_3$ (1 × 30 mL), distilled water (1 × 30 mL), then dried over $MgSO_4$ and removed *in vacuo* to give **3** quantitatively as a yellow powder. Mp: 155–157 °C. ^{1}H-NMR (DMSO-d_6): 8.62 (bt, J = 5.5 Hz, 1H, NH), 8.45 (d, J = 8.6 Hz, 1H), 8.29 (d, J = 8.6 Hz, 1H), 8.13 (d, J = 7.2 Hz, 2H), 7.82 (d, J = 1.9 Hz, 2H), 7.62–7.57 (m, 2H), 7.52 (d, J = 1.9 Hz, 1H), 7.21 (d, J = 7.5 Hz, 1H), 6.85 (d, J = 1.8 Hz, 2H), 3.32–3.29 (m, 2H), 2.98–2.94 (m, 2H), 2.81 (s, 6H), 2.09 (s, 6H). ^{13}C-NMR (DMSO-d_6): 170.2, 169.3, 164.6, 151.4, 146.0, 135.8, 135.4, 132.3, 129.5, 129.1, 128.3, 127.9, 127.7, 127.2, 124.5, 123.6, 119.0, 115.1, 45.0, 41.6, 40.2, 10.9. MS: $[M + H]^+$: 616.18605, $[M + Na]^+$: 638.16799.

N-*[2-(5-(Dimethylamino)naphthalene-1-sulfonyl)aminoethyl]-3,5-di(3-methyl-3-(2-carboxyethyl)thio-1-succinimidyl)benzamide (***10***)*. Compound **3** (67.6 mg; 0.11 mmol; 1 eq.) was dissolved in a CH_3CN–DMSO solution (7 : 1, v/v, 8 mL). Mercapto-propionic acid (24 μL; 0.27 mmol; 2.5 eq.) was added and the mixture was stirred at room temperature for 24 h. The CH_3CN was removed *in vacuo* and Et_2O (10 mL) was added. The resulting precipitate was filtered and washed with Et_2O (3 × 10 mL) to give **10** (74.1 mg; 0.091 mmol; 83%) as a yellow solid. Mp: 79–81 °C. ^{1}H-NMR (acetone-d_6): 8.41 (d, J = 8.0 Hz, 1H), 8.15 (d, J = 7.0 Hz, 1H), 8.08 (d, J = 5.6 Hz, 1H),

7.81–7.73 (m, 2H), 7.43–7.41 (m, 3H), 7.03 (d, J = 5.6 Hz, 1H), 4.25 (br s, acid), 2.97–2.92 (m, 8H), 2.82–2.75 (m, 8H), 2.54 (s, 12H). ^{13}C-NMR (acetone-d_6): 176.9, 174.1, 173.9, 172.8, 151.8, 135.6, 134.6, 132.6, 130.4, 129.8, 129.5, 129.2, 128.4, 125.1, 123.2, 118.8, 115.3, 46.4, 45.3, 43.6, 42.5, 33.9, 33.2, 22.2, 15.0. MS: $[M + H]^+$: 828.20375, $[M + Na]^+$: 850.18569.

Cloning and expression

Cloning of MBP-dC10. The helix dC-10 was subcloned from a previous construct (not shown) by cleaving with the restriction enzymes *XhoI* and *HindIII* and insertion into plasmid pMAL-3C, prepared by cleavage with restriction enzymes *SalI* and *HindIII* (the *XhoI* and *SalI* have cohesive ends, but this restriction site is subsequently lost). The resulting plasmid was sequenced and the final nucleotide sequences are shown in Table S2 (ESI†). The *XhoI* restriction site immediately precedes the helix (CTCGAG **CTG**...) and the *HindIII* restriction site is found downstream of the stop codon (...**TAG** GAGCTCGGTACCCCGGGTCGACCTGCAGCC AAGCTT).

Cloning of MBP-dC10-H2, -H7, -H9 and -H2H17. Site-directed mutagenesis was performed with the QuikChange Kit (Stratagene) in order to incorporate His residues into the dC10 helical sequence. The primers used in the site-directed mutagenesis reaction are shown in Table S3 (ESI†). For mutant dC10-H2H17, two rounds of site-directed mutagenesis were performed to introduce the two His mutations into the parent dC10 sequence.

Cloning of MBP-dC10-R8A9 and MBP-dC10-A9. Test proteins containing helices dC10-R8A9 (featuring inversion of the Arg9 and Ala8 residues) and dC10-A9 (with an Ala replacing Arg9) were constructed by inserting the reconstituted desired helix (encoding the cohesive end of *SalI* and *HindIII* restriction sites) in pMAL-3C. The helix was reconstituted from two long primers whose sequences are reported in Table S4 (ESI†). Each primer was first phosphorylated by mixing 3 μL of a 100 μM stock solution of primer 1 with the same quantity of primer 2, 3 μL of 10× PNK buffer, 2 μL of a 10 mM ATP stock solution, and 2 μL of T4 PNK in a final volume of 30 μL at 37 °C for 90 min. The annealing was performed afterwards by adding 4 μL of a 0.5 M NaCl solution, then placing the tubes in a boiling water bath for 2 minutes. The tubes were then allowed to cool to room temperature. The plasmid pMAL-3C (encoding for the MBP protein) was cleaved with the restriction enzymes *SalI* and *HindIII* and dephosphorylated with SAP at 37 °C for 30 minutes. After cleaning the plasmid and the primers, a ligation was performed and chemically competent *E. coli* DH5α cells were subsequently transformed. Mini-preps were performed on the resulting colonies and DNA sequencing was carried out in house.

Expression and purification of the MBP constructs. Each plasmid was transformed into competent *E. coli* BL21 (DE3) pREP4 cells, then a culture of the transformed cells was grown overnight and used for inoculation of a 1 L culture of a rich medium (10 g tryptone, 5 g yeast extract, 5 g NaCl) supplemented with 0.2% D-glucose and ampicillin, as described in the manufacturer's standard procedures[56] at 37 °C. When the culture reached an OD (600 nm) of around 0.6, the protein expression was induced by addition of IPTG to a final concentration of 0.3 mM followed by incubation at the same temperature for 2 hours (or at 28 °C for 16 hours). The cells were then centrifuged at 6000 rpm for 10 minutes and washed by centrifugation in a lysis buffer (20 mM Tris–HCl (pH 7.4), 0.2 M NaCl, 1 mM EDTA) at 3000 rpm for 15 minutes. Cell lysis was performed by sonication on ice (3 × 30 s, 1 minute between each cycle) and the membranes were removed by centrifugation at 9000 g for 20 minutes. The supernatant was mixed with the previously equilibrated amylose resin and slowly agitated for at least one hour at 4 °C. The resin was then poured into a fritted column and washed with at least 10 CV of lysis buffer, followed by 5 CV of elution buffer (lysis buffer + 10 mM maltose). The protein was concentrated by ultrafiltration with Amicon tubes (30 000 MWCO) to a concentration of around 12–15 mg mL^{-1}. A general yield of approximately 8–10 mg L^{-1} was obtained and the purity of the protein was verified by SDS-PAGE, as well as ESI-MS or MALDI-TOF.

Cleavage of the dC10 helix and CD studies. The protein MBP-dC10 was buffer-exchanged for 5 mM sodium phosphate, pH 7.0, and around 1.5 mg was then mixed with thrombin (1 U) in the thrombin buffer (supplied with the enzyme) for digestion at room temperature for 3.5 hours. An ultrafiltration was done with Microcon tubes (30 000 MWCO) to remove the MBP, the uncleaved MBP-dC10 and the thrombin from the cleaved dC-10 helix. The peptide dC-10 was then used for circular dichroism studies. A final concentration of approximately 35–40 μM was necessary for a strong signal. Spectra were recorded at 23 °C in a cuvette of 1 mm pathlength, with 4 s of acquisition for each point between 185 and 260 nm (0.7 nm steps), using 5 mM sodium phosphate (pH 7.0) as a blank. Each experiment was performed in duplicate.

Labelling of MBP-dC10 with fluorogen dM10^{M}-FITC (2). A concentration of 100 μM of MBP-dC10 was labelled with 1 equivalent of dM10^{M}-FITC (**2**) in MBP elution buffer at room temperature for 40 minutes. At this point the labelled protein can be visualized by SDS-PAGE. For CD studies, the protein was extensively washed by repeated ultrafiltration with Microcon tubes (30 000 MWCO) in 5 mM sodium phosphate (pH 7.0) until free fluorogen was no longer visible in the flow-through. The labelled protein was then diluted down to 100 μM and thrombin digestion was performed as described above.

Kinetics

Kinetic experiments were carried out using a Varian Eclipse fluorimeter thermostated at 20 °C. All reactions were prepared in 50 mM HEPES buffer (pH 7.5) in 10% DMSO for **3** and 10% basic ethanol for **2**. The final concentration of dicysteine proteins was 100 μM whereas the final concentration of GSH was 200 μM. For dansyl fluorogen **3**, samples were excited at 355 nm and the fluorescence emission was followed at 520 nm

as a function of time. For fluorescein fluorogen **2**, fluorescence was observed at 525 nm upon excitation at 495 nm. Rate constants were determined either by the initial rates method or by non-linear regression to the following second order equation:

$$[\mathrm{P}] = [\mathrm{dM}]_0 - \frac{1}{k_2 t + \frac{1}{[\mathrm{dM}]_0}}$$

where P is the fluorescent dithiolated adduct, dM is the dimaleimide fluorogen, k_2 is the second order rate constant and t is the time.

Cellular labelling

To tag EGFR at its extracellular N-terminus, the restriction sites *Age*I and *Xho*I were inserted by site-directed mutagenesis after the signal sequence (first 24 aa) of EGFR (ErbB1) in the vector pcDNA3.1, and the dC10 fragment was subcloned between these sites.

HEK293 cells were plated into Special Optics 96-well plates and grown in DMEM supplemented with 10% FBS. Transient transfections were performed using the TransIT-LT1 transfection reagent. Twenty-four or 48 h after transfection, the cells were briefly washed with OPTI-MEM and incubated with 10 μM dM10-FITC in OPTI-MEM for 10 min at room temperature. The cells were then washed 3 times over 5 min with OPTI-MEM and imaged on an inverted confocal microscope (excitation: 488 nm; emission: 505–525 nm).

For co-localization studies of dC10-EGFR with its ligand, 400 ng mL^{-1} of the tetramethylrhodamine conjugate of epidermal growth factor (rhodamine EGF, molecular probes) was added to the cells and images were taken after 2 or 3 min (excitation: 488 nm and 543 nm; emission: 505–525 nm and 575–630 nm; dichroic mirror at 570 nm).

Conclusions

We have designed new peptides, shown to be α-helical, that serve as target sequences for our dimaleimide fluorogens. The convergent synthesis and kinetic evaluation of a new series of these 3-methylmaleimide fluorogens demonstrate their tuned reactivity toward the target sequence relative to glutathione. According to this method, a protein of interest bearing the short target sequence could be covalently, fluorescently labelled with small molecules in a fluorogenic addition reaction (FlARe). The FlARe method was validated in a cellular context by specifically labelling the receptor protein EGFR expressed on the surface of HEK293 cells. The scope and limitations of other biological applications of this labelling method are currently under investigation.

Acknowledgements

The authors acknowledge the financial support of the Canadian Institutes of Health Research (CIHR) and the Natural Sciences and Engineering Research Council (NSERC). VJ thanks the Swiss National Science Foundation for a postdoctoral fellowship and KC is grateful to the Fonds Québecois de Recherche sur la Nature et les Technologies (FQRNT) for a postgraduate scholarship.

References

1 J. Zhang, R. E. Campbell, A. Y. Ting and R. Y. Tsien, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 906–918.
2 W.-K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman and E. K. O'Shea, *Nature*, 2003, **425**, 686.
3 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.
4 N. Soh, *Sensors*, 2008, **8**, 1004–1024.
5 L. Vivero-Pol, N. George, H. Krumm, K. Johnsson and N. Johnsson, *J. Am. Chem. Soc.*, 2005, **127**, 12770–12771.
6 N. Johnsson, N. George and K. Johnsson, *ChemBioChem*, 2005, **6**, 47–52.
7 A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson, *Nat. Biotechnol.*, 2003, **21**, 86–89.
8 A. Keppler, M. Kindermann, S. Gendreizig, H. Pick, H. Vogel and K. Johnsson, *Methods*, 2004, **32**, 437–444.
9 A. Keppler, H. Pick, C. Arrivoli, H. Vogel and K. Johnsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9955–9959.
10 Technical Manual, Promega, "HaloTag Interchangeable Labeling Technology"; http://www.promega.com, 2006.
11 G. V. Los and K. Wood, *Methods Mol. Biol. (Totowa, N. J.)*, 2007, **356**, 195–208.
12 C. W. Lin and A. Y. Ting, *J. Am. Chem. Soc.*, 2006, **128**, 4542.
13 I. Chen, M. Howarth, W. Y. Lin and A. Y. Ting, *Nat. Methods*, 2005, **2**, 99.
14 B. P. Duckworth, Z. Zhang, A. Hosokawa and M. D. Distefano, *ChemBioChem*, 2007, **8**, 98.
15 B. A. Griffin, R. S. Adams and R. Y. Tsien, *Science*, 2000, **281**, 269–272.
16 B. A. Griffin, S. R. Adams, J. Jones and R. Y. Tsien, *Methods Enzymol.*, 2000, **327**, 565–578.
17 G. Gaietta, T. J. Deerinck, S. R. Adams, J. Bouwer, O. Tour, D. W. Laird, G. E. Sosinsky, R. Y. Tsien and M. H. Ellisman, *Science*, 2002, **296**, 503–507.
18 S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis and R. Y. Tsien, *J. Am. Chem. Soc.*, 2002, **124**, 6063–6076.
19 H. Cao, Y. Xiong, T. Wang, B. Chen, T. C. Squier and M. U. Mayer, *J. Am. Chem. Soc.*, 2007, **129**, 8672.
20 B. Chen, H. Cao, P. Yan, M. U. Mayer and T. C. Squier, *Bioconjugate Chem.*, 2007, **18**, 1259–1265.
21 T. Wang, P. Yan, T. C. Squier and M. U. Mayer, *ChemBioChem*, 2007, **8**, 1937–1940.
22 N. W. Luedtke, R. J. Dexter, D. B. Fried and A. Schepartz, *Nat. Chem. Biol.*, 2007, **3**, 779–784.
23 M. F. Langhorst, S. Genisyuerek and C. A. O. Stuermer, *Histochem. Cell Biol.*, 2006, **125**, 743–747.
24 K. Stroffekava, C. Proenza and K. G. Beam, *Pflugers Arch. Eur. J. Physiol.*, 2001, **442**, 859–866.
25 B. Chen, M. U. Mayer, L. M. Markillie, D. L. Stenoien and T. C. Squier, *Biochemistry*, 2005, **44**, 905–914.
26 T. F. Halo, J. Appelbaum, E. M. Hobert, D. M. Balkin and A. Schepartz, *J. Am. Chem. Soc.*, 2009, **131**, 438–439.
27 S. Girouard, *PhD thesis*, Université de Montréal, 2005.
28 S. Girouard, M.-H. Houle, A. Grandbois, J. W. Keillor and S. W. Michnick, *J. Am. Chem. Soc.*, 2005, **127**, 559–566.
29 J. Guy, K. Caron, S. Dufresne, S. W. Michnick, W. G. Skene and J. W. Keillor, *J. Am. Chem. Soc.*, 2007, **129**, 11969–11977.
30 S. Marqusee, V. H. Robbins and R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 5286–5290.
31 G. Merutka and E. Stellwagen, *Biochemistry*, 1990, **29**, 894–898.
32 S. Penel, R. G. Morrison, R. J. Mortishire-Smith and A. J. Doig, *J. Mol. Biol.*, 1999, **293**, 1211–1219.
33 V. Muñoz and L. Serrano, *Biopolymers*, 1997, **41**, 495–509.
34 The AGADIR algorithm is available on-line at http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html.
35 D. G. Flint, J. R. Kumita, O. S. Smart and G. A. Woolley, *Chem. Biol.*, 2002, **9**, 391–397.
36 J. Kalia and R. T. Raines, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 6286.
37 J. E. T. Corrie, *J. Chem. Soc., Perkin Trans. 1*, 1994, 2975.
38 M. Machida, M. I. Machida and Y. Kanaoka, *Chem. Pharm. Bull.*, 1977, **25**, 2739.
39 J. D. Gregory, *J. Am. Chem. Soc.*, 1955, **77**, 3922.

40 M.-H. Houle, *MSc thesis*, Université de Montréal, 2003.
41 K. Caron, *MSc thesis*, Université de Montréal, 2010.
42 J. N. Demas and G. A. Crosby, *J. Phys. Chem.*, 1971, **75**, 991.
43 L. Wang, A. K. Gaigalas, F. Abbasi, G. E. Marti, R. F. Vogt and A. Schwartz, *J. Res. Natl. Inst. Stand. Technol.*, 2002, **107**, 339.
44 Y.-H. Li, L.-M. Chan, L. Tyer, R. T. Moody, C. M. Himel and D. M. Hercules, *J. Am. Chem. Soc.*, 1975, **97**, 3118.
45 A. J. Kirby and G. J. Lloyd, *J. Chem. Soc., Perkin Trans. 2*, 1976, 1753.
46 R. T. Borchardt and L. A. Cohen, *J. Am. Chem. Soc.*, 1972, **94**, 9166.
47 R. T. Borchardt and L. A. Cohen, *J. Am. Chem. Soc.*, 1973, **95**, 8308.
48 A. J. Kirby and P. W. Lancaster, *J. Chem. Soc., Perkin Trans. 2*, 1972, 1206–1214.
49 D. Zenisek, J. Steyer, M. Feldman and W. Almers, *Neuron*, 2002, **35**, 1085–1097.
50 J. Boehm, M. G. Kang, R. C. Johnson, J. Esteban, R. L. Huganir and R. Malinow, *Neuron*, 2006, **51**, 213–225.
51 M. Spaargaren, L. H. Defize, J. Boonstra and S. W. de Laat, *J. Biol. Chem.*, 1991, **266**, 1733–1739.
52 Z. Fan, Y. Lu, X. Wu and J. Mendelsohn, *J. Biol. Chem.*, 1994, **269**, 27595–27602.
53 A. Citri and Y. Yarden, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 505–516.
54 Y. Yarden and M. X. Sliwkowski, *Nat. Rev. Mol. Cell Biol.*, 2001, **2**, 127–137.
55 H. S. Wiley, *Exp. Cell Res.*, 2003, **284**, 78–88.
56 See http://www.neb.com/nebecomm/ManualFiles/manualE8000.pdf.