Specific N-terminal protein labelling: use of FMDV 3C^{pro} protease and native chemical ligation†

Gillian K. Busch,[a] Edward W. Tate,[a] Piers R. J. Gaffney,[a] Erika Rosivatz,[b] Rudiger Woscholski[b] and Robin J. Leatherbarrow*[a]

Received (in Cambridge, UK) 21st April 2008, Accepted 14th May 2008
First published as an Advance Article on the web 25th June 2008
DOI: 10.1039/b806727a

We report an effective strategy for generating N-terminal cysteinyl proteins by proteolytic cleavage using the enzyme 3C^{pro}, suitable for a wide range of applications *via* native chemical ligation.

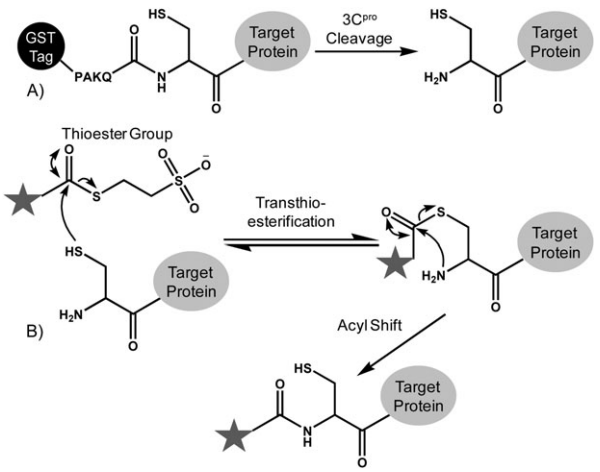
Native chemical ligation (NCL) is a highly selective and facile reaction, allowing a thioester-containing moiety to be joined to an N-terminal cysteine residue *via* a peptide bond.[1] This non-enzymatic, chemoselective reaction occurs at physiological pH and in aqueous solution, making it a useful tool for chemical biology; the thioester being attached can have any physicochemical property that can be tuned by synthesis. NCL is a key tool for a wide range of protein engineering applications including protein labelling and semi-synthesis of difficult to express or post-translationally modified proteins.[2] The only requirement for a protein or peptide to undergo NCL is that it has an N-terminal cysteine residue. However these are rare in wild-type proteins and so to carry out this reaction usually requires engineering of a cysteine at this position. Impressive strides have been made to generate the cysteine *via* solid-phase synthesis,[3] however this method is only applicable to peptides. The most frequent protein-based technique utilises the inducible self-cleavage activity of engineered protein splicing elements known as inteins.[4] However there are disadvantages: *in vivo* cleavage can occur during expression and there is a dependency on the properties of the protein for successful spicing.

An alternative approach involves the use of proteases. By engineering a specific cleavage sequence followed by a cysteine into a recombinant fusion protein, a suitable protease can be used to generate a protein that has an N-terminal cysteine. This strategy has been successfully employed using the protease TEV, which cleaves the sequence ENLYF(Q/S)$^{\downarrow}$G; although mutation of the P1′ glycine to cysteine does result in some loss of proteolytic activity, the ENLYF(Q/S)$^{\downarrow}$C sequence can be used successfully.[5]

In this communication we report the application of a novel protease for N-terminal Cys generation, with cleavage specificity distinct from that of TEV, and demonstrate this application in the synthesis of a new tool for the detection of membrane-bound phosphatidylinsoitol lipids.

The 3C protease (3C^{pro}) from Foot-and-Mouth Disease Virus (FMDV) is a highly selective cysteine protease with a chymotrypsin-like fold. We have recently shown that this protease may be expressed as a soluble recombinant protein in *E. coli* and possesses a consensus cleavage sequence PAKQ$^{\downarrow}$X, where there is no significant preference at the P1′ (X) position.[6] Its high degree of specificity and activity, together with the tolerance for a range of P1′ residues, make 3C^{pro} potentially useful for the production of N-terminal cysteinyl protein fragments. The proposed application of this enzyme to protein labelling *via* NCL is shown in Scheme 1.

We first examined the efficiency of cleavage using peptide substrates, synthesised by Fmoc/*t*Bu solid phase peptide synthesis. The peptide APAKQ$^{\downarrow}$CLDFDLLK **1**, corresponding to the protease's optimum cleavage sequence in the region P4-P1 with a cysteine incorporated at the P1′ position,‡[6] was incubated with 3C^{pro} at 37 °C for 16 h (200 μL; 700 nmol peptide **1**, 3 nmol 3C^{pro}, 100 mM sodium phosphate pH 7.4, 1 mM EDTA, 1 mM TCEP, 5% v/v glycerol). Analysis by HPLC and MALDI-TOF showed that the enzyme cleaved the peptide stoichiometrically and specifically, yielding two products CLDFDLLK **2a** and APAKQ **2b**, each with the expected mass (Fig. 1(A) and (B)). **2a** was then reacted with a carboxyfluorescein thioester, *S*-2-mercaptoethylsulfonate



Scheme 1 Strategy for N-terminal labelling of proteins: (A) proteolysis of a GST-PAKQC-fusion protein by 3C^{pro}. (B) Labelling of the protein *via* native chemical ligation.

[a] *Department of Chemistry and the Chemical Biology Centre, Imperial College London, South Kensington Campus, London, UK SW7 2AZ. E-mail: r.leatherbarrow@imperial.ac.uk; Fax: (+44) (0)20 7594 1139*
[b] *Department of Life Sciences and the Chemical Biology Centre, Imperial College London, South Kensington Campus, London, UK SW7 2AZ*

† Electronic supplementary information (ESI) available: Engineering the pGEX_3C_PH plasmid and plasmid map. See DOI: 10.1039/b806727a

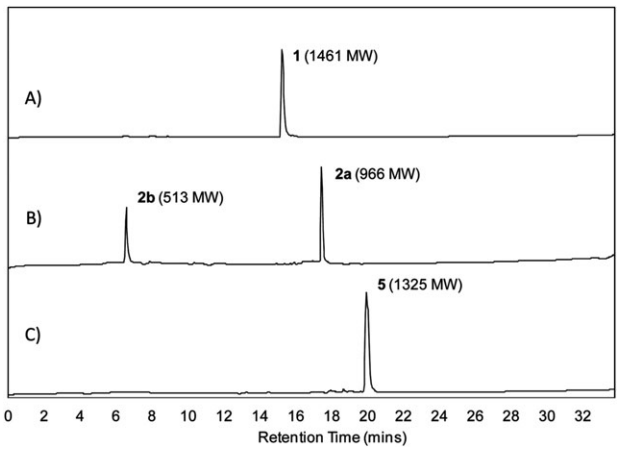


Fig. 1 HPLC elution profiles of: (A) Starting peptide **1**, incorporating the $3C^{pro}$ cleavage sequence. (B) After incubation with $3C^{pro}$ at 37 °C for 16 h, showing cleavage into two fragments **2a** and **2b**. (C) Fragment **2a** following NCL reaction with the thioester **3**, at room temperature, for 5 h to introduce fluorescein at the N-terminus of the peptide **5**. In each case the fragment was identified by mass.

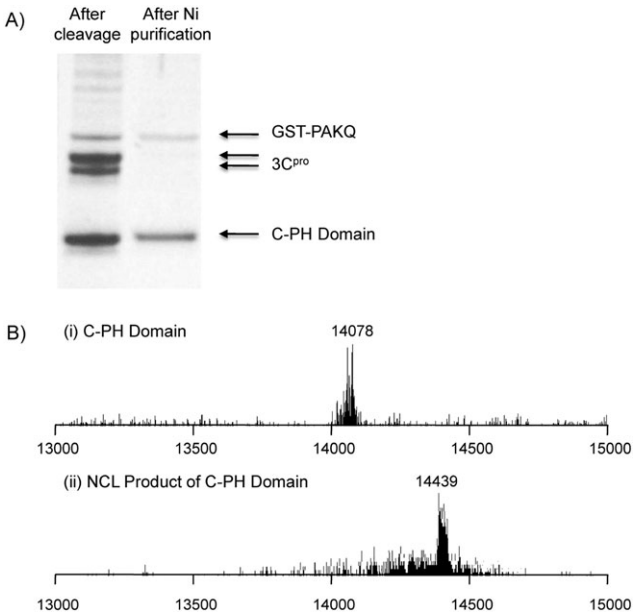


Fig. 2 Labelling of the PH domain. (A) SDS-PAGE analysis of the N-terminal C-PH domain liberated into the supernatant after cleavage of GST-PAKQC-PH domain fusion protein bound to glutathione-Sepharose. Final purification on Ni Sepharose removes the His-tagged $3C^{pro}$. The gel was visualised by Coomassie staining. (B) MALDI-TOF analysis of (i) the isolated C-PH domain, (ii) the purified fluorescein-labelled C-PH domain following NCL reaction with **3**, at room temperature, for 5 h.

carboxyfluoresceinthioate **3**,[7] under standard NCL conditions for 5 h at room temperature (0.1 M phosphate buffer, pH 7.4, with 20-fold excess of mercaptoethansulfonate **4** as the thiol catalyst). This produced an N-terminally fluorescein tagged peptide **5** in quantitative yield (Fig. 1(C)), confirming the use of $3C^{pro}$ for generating peptides suitable for NCL.

To demonstrate the application of $3C^{pro}$ for NCL-mediated labelling, we have added a fluorescent tag specifically to the N-terminus of the PH domain of Protein Kinase B (PKB). PKB, or Akt, is a serine/threonine kinase that is a key component in a number of signalling cascades. A vital part of PKB is the PH domain, which targets the protein to specific membrane-bound phosphatidylinositol polyphosphates (PIP_n). This binding opens up the kinase domain for activation, continuing the signalling cascade. With numerous downstream factors involved in apoptosis, expression and cellular proliferation, alterations of PKB can result in cancer. Therefore it is important to understand fully the activation and signalling pathways of this enzyme.[8]

To enable facile expression of this protein bearing an N-terminal cysteine, an expression–purification strategy was developed (Scheme 1) where a vector containing the PAKQC motif was introduced immediately downstream of a GST affinity tag. The target PH domain protein was then inserted directly after the cysteine codon, see ESI.† The fusion protein expression was induced overnight in *E. coli* BL21 (DE3) cells containing pGEX_3C_PH, using 100 μM IPTG at 18 °C, and purified using glutathione–Sepharose 4B beads.

Optimum conditions for $3C^{pro}$ cleavage of the GST-PAKQC-PH domain fusion protein were found by a time course experiment. Digestion at a ratio of 1 : 20 (enzyme : protein) was essentially complete after 16 h, producing C-PH domain (PH domain with an N-terminal cysteine residue). To purify the C-PH domain, digestion was performed on material bound to glutathione beads, allowing the cleaved C-PH domain to be recovered into the supernatant. Removal of $3C^{pro}$ (which appears as a double band after incubation) was achieved by binding the His-tagged protease to Ni Sepharose beads (GE Healthcare) giving the C-PH domain product, with a small residual amount of GST (Fig. 2(A)).

The C-PH domain was incubated with a four-fold excess of **3** for 5 h under standard NCL conditions as before. MALDI-TOF analysis, after sample desalting, showed quantitative labelling with a single fluorescein per protein molecule (Fig. 2(B); C-PH domain, expected mass 14 099, found 14 078; fluorescein-C-PH domain, expected mass 14 458, found 14 439. The observed and expected masses are within the error for MALDI-TOF measurements on proteins of this size.).

The PKB PH domain binds to the lipids $PI(3,4,5)P_3$ and $PI(3,4)P_2$ that are found on the inner leaflet of the cell membrane, organelles, vesicles and in the nucleus. A current method of investigating their distribution within cells is to use fixed cell imaging. This involves either stimulating $PI(3,4,5)P_3$ production, which is subsequently metabolised into $PI(3,4)P_2$, in cells with the growth factor insulin or inhibiting production with the synthetic molecule LY294002 (LY). The cells are then fixed to glass slides and incubated with variants of fluorescently tagged PH domain to probe for the lipids. The PH domain is typically labelled *via* an AlexaFluor succinimidal ester, which will react with any available primary amines on a protein's surface.[9]

To demonstrate the utility of our specific NCL-mediated labelling, we compared the NCL labelled fluorescein protein with one labelled non-specifically by AlexaFluor. Insulin and LY treated NIH 3T3 fibroblasts were incubated with fluorescein-PH domain or AlexaFluor-PH domain counter-stained with the nuclear probe DAPI and imaged.§

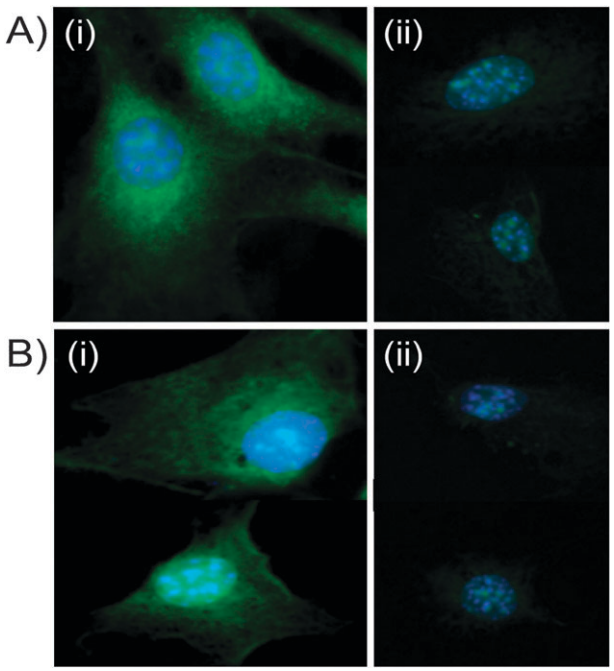


Fig. 3 Fixed cell imaging of labelled PH-domain binding to PIP_3 in fibroblast cells, counter-stained with DAPI to show the nucleus (blue). (A) AlexaFluor labelled PH domain in (i) insulin stimulated cells; (ii) LY inhibited cells. (B) NCL fluorescein labelled PH domain in (i) insulin stimulated cells; (ii) LY inhibited cells.

The AlexaFluor-PH domain visualisation of $PI(3,4,5)P_3$ and $PI(3,4)P_2$ is shown in Fig. 3(A). In Fig. 3(A)(i), the cells were stimulated with insulin to generate $PI(3,4,5)P_3$; the green coloration of the Alexa-Fluor PH domain shows the characteristic localization of the $PI(3,4,5)P_3$ and $PI(3,4)P_2$, with the nucleus stained blue using DAPI. Fig. 3(A)(ii) is from cells that are treated to inhibit $PI(3,4,5)P_3$ production, resulting in far less PH domain binding. In this figure the green coloration is almost absent, leaving just the nucleus visible. This is the expected PH domain binding behaviour under these stimulation conditions and the differential staining confirms that the binding is specific. Fig. 3(B) shows the results with our new fluorescein-C-PH domain. Cells in Fig. 3(B)(i) are stimulated whereas those in Fig. 3(B)(ii) are inhibited. As is apparent, the images obtained with the controlled NCL labelled material are fully comparable to the AlexaFluor ones, validating this NCL labelling method and proving that the product of this procedure retains biological activity and specificity.

These results confirm that proteolytic cleavage by $3C^{pro}$ is an effective means of generating an N-terminal cysteinyl protein suitable for NCL. The enzyme cleaves proteins and peptides with good specificity for its consensus sequence PAKQ-X. Protocols have been developed for the generation of an N-terminal cysteine protein *via* the site-specific cleavage of a GST-PAKQC-fusion protein with $3C^{pro}$. To demonstrate an application of NCL, a single molecule of fluorescein was stoichiometrically ligated onto the C-PH domain; the labelled protein was then used as a probe for $PI(3,4,5)P_3$ and $PI(3,4)P_2$, proving as effective in fixed cell imaging as the non-specifically labelled PH domain. Our biologically active C-PH domain can now be ligated to alternative fluorescent tags, environmentally sensitive probes, biotin, or even immobilised on a surface. As NCL is a generic method, requiring only a thioester and a cysteine N-terminus, the versatility of this system should allow it to be exploited for a wide range of proteins and applications.

This research was supported by an EPSRC Doctoral Training award to the Chemical Biology Centre, Imperial College London. We would like to thank Stephen Curry of Imperial College for helpful discussion.

Notes and references

‡ The paper reports that using alanine scanning there appeared to be a preference for phenylalanine over alanine at the P4′ position. However, in more extensive recent work we find that a range of residues are well-tolerated at this locus, meaning that prime-side interactions are not very significant for this enzyme (R. J. Leatherbarrow, S. R. Knox and S. Curry, in preparation).

§ *Fixed cell imaging*:[9] Starved NIH 3T3 fibroblasts, seeded onto poly-L-lysine coated glass coverslips, were either stimulated with 1 μg ml^{-1} of insulin for 5 min or inhibited with 50 μM LY294002 for 30 min, fixed in 4% (w/v) paraformaldehyde, freeze–thaw permeabilised and blocked with 3% (w/v) BSA/PBS for 1 h. Cells were incubated with the labelled PH domain (1 : 1000 diluted in blocking solution) for 1 h, washed with 3 × PBS and counter-stained with 300 nM DAPI. Thoroughly washed coverslips were mounted on glass slides using Mowiol supplemented with 0.6% (w/v) DABCO, sealed, dried overnight and observed under a Nikon TE2000 fluorescence microscope using DAPI and FITC filters.

1. P. E. Dawson, T. W. Muir, I. Clarklewis and S. B. H. Kent, *Science*, 1994, **266**, 776–779.
2. V. Muralidharan and T. W. Muir, *Nat. Methods*, 2006, **3**, 429–438.
3. C. Haase and O. Seitz, *Angew. Chem., Int. Ed.*, 2008, **47**, 1553–1556.
4. M. Q. Xu and T. C. Evans, *Methods*, 2001, **24**, 257–277.
5. T. J. Tolbert and C. H. Wong, *Angew. Chem., Int. Ed.*, 2002, **41**, 2171–2174.
6. J. R. Birtley, S. R. Knox, A. M. Jaulent, P. Brick, R. J. Leatherbarrow and S. Curry, *J. Biol. Chem.*, 2005, **280**, 11520–11527.
7. I. E. Gentle, D. P. De Souza and M. Baca, *Bioconjugate Chem.*, 2004, **15**, 658–663.
8. K. M. Nicholson and N. G. Anderson, *Cell. Signalling*, 2002, **14**, 381–395.
9. E. Rosivatz, J. G. Matthews, N. Q. McDonald, X. Mulet, K. K. Ho, N. Lossi, A. C. Schmid, M. Mirabelli, K. M. Pomeranz, C. Erneux, E. W. F. Lam, R. Vilar and R. Woscholski, *ACS Chem. Biol.*, 2006, **1**, 780–790.