ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2014, **50**, 2258

Received 3rd October 2013,
Accepted 19th December 2013

DOI: 10.1039/c3cc47599a

www.rsc.org/chemcomm

The ββα fold of zinc finger proteins as a "natural" protecting group. Chemoselective synthesis of a DNA-binding zinc finger derivative†

Jéssica Rodríguez, Jesús Mosquera, Olalla Vázquez, M. Eugenio Vázquez* and José L. Mascareñas*

We report the selective modification of cysteine residues engineered in peptides that have two additional cysteine residues as part of a Cys_2His_2 zinc finger motif. The chemoselective modification is achieved, thanks to the protecting effect exerted by the zinc cation upon coordination with the native cysteines and histidines of the zinc-finger fold. The strategy allows a straightforward synthesis of DNA binding zinc finger constructs.

The covalent modification of proteins is a powerful strategy to modulate their macromolecular function.[1] Nature accomplishes such alterations through a variety of post-translational modifications that modulate the properties of proteins.[2] Inspired by nature, scientists have long pursued the development of methodologies that could emulate these natural post-conjugation reactions and allow the introduction of desired groups or labels in specific sites of proteins.[1,3] For a reaction to be of general use, it should selectively target the residue of interest in the presence of competing side chains of the unprotected polypeptide, and in aqueous environments.

Many of the strategies developed so far for the site-selective modification of proteins rely on the introduction of designed amino acids equipped with orthogonally reactive groups that can be coupled to external reactants, for instance, by using the well known click chemistry.[4] Although this strategy allows high levels of selectivity, it requires the introduction of non-natural amino acids.[5] Alternatively, one could pursue the modification of specific natural amino acids in proteins, but achieving a good combination of reactivity and chemoselectivity is extremely challenging. In this context, most of the successful modifications of unprotected peptides or proteins have relied on the nucleophilic reactivity window provided by the sulfur atom in cysteine side chains,[6] which allowed a number of selective alkylations with a variety of electrophiles. A major problem of this strategy arises when the peptides or proteins of interest have more than one cysteine residue, because of the competitive formation of polyalkylated products.[7]

Here, we report the selective modification of cysteine residues incorporated into a zinc finger peptide that has two additional cysteines as part of a classical Cys_2His_2 motif. In the presence of Zn(II) these cysteines are trapped in the ββα secondary fold and thereby exhibit a much decreased reactivity (Fig. 1). The ligation provides a straightforward approach to bisbenzamidine–zinc finger conjugates capable of binding to designed DNA target sequences. Since the zinc finger is a natural motif present in many transcription factors, this work opens a door for the chemoselective access to zinc finger derivatives, and for the modification and/or tagging of zinc finger proteins.

The zinc finger family is the largest among eukaryotic transcription factors, and is responsible for regulating the expression of a myriad of genes controlling fundamental cellular programs.[8] The DNA recognition by zinc finger proteins has served as inspiration for the engineering of a large variety of highly appealing artificial gene regulators.[9] Despite the relevance of zinc finger proteins, work towards their selective chemical modification has been very scarce. There are a number of reports on the reactivity of zinc thiolates of different zinc fingers that suggest that, while the cysteines in charged $Zn(Cys)_4$ and $Zn(Cys)_3His$ motifs react with electrophiles or oxidants, neutral $Zn(Cys)_2(His)_2$ are less reactive.[10] Therefore, we reasoned that zinc finger structures might provide a natural protection for the two cysteines involved in the coordination to the metal

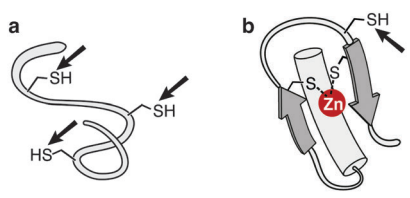

Fig. 1 Schematic illustration of the approach for the chemoselective synthesis of zinc finger derivatives. (a) Unfolded peptide with three nucleophilic sites; (b) zinc-promoted folding and cysteine coordination leaves only a single nucleophilic site for modification.

Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CIQUS), Departamento de Química Orgánica. Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain.
E-mail: joseluis.mascarenas@usc.es; Fax: +34 981 595 012;
Tel: +34 981 576 541-14405

† Electronic supplementary information (ESI) available: Peptide synthesis, full experimental procedures and analytical data of the peptides and products obtained. See DOI: 10.1039/c3cc47599a

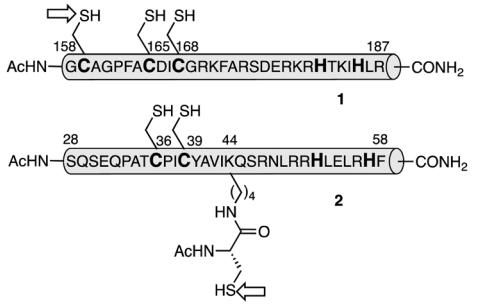

Fig. 2 Structure of model peptides derived from Zif268 (**1**) and GAGA (**2**). Arrows point to the thiols to be selectively modified.

ion, and thereby favor the selective modification of other cysteines present in the peptide. This would provide a convenient and versatile way of making zinc finger derivatives.

To test the approach we synthesized model peptides **1** and **2**, which exhibit the sequence of two different types of Cys_2His_2 fragments, together with an additional cysteine engineered either at the N-terminus (**1**) or in the side chain of an internal lysine (**2**) (Fig. 2). The zinc finger sequence of peptide **1** is based on a fragment of the C-terminal Cys_2His_2 unit of the trimeric Zif268 transcription factor (residues Pro^{162} to Arg^{187} in the crystal structure).[11] Peptide **2** contains the Cys_2His_2 unit of the DNA-binding domain of the GAGA transcription factor of *Drosophila melanogaster* (from residues Ser^{28} to Phe^{58} in the reference structure).[12] Peptide **2** includes a mutation of Arg^{44} to Lys to allow the selective introduction of a cysteine group in its side chain.[13] Both peptides were assembled using standard Fmoc protocols and HBTU as a coupling agent. The Lys residue of peptide **2** was introduced as alloc derivative, for its selective manipulation in the solid phase (see the ESI†).

We tested the selectivity in the alkylation of the sulfur side chains in the peptides with benzyl bromoacetate, a very good electrophile that is commercially available. The reactions were carried out in parallel, in the absence and presence of 1.5 equiv. of $ZnSO_4$, and in a deoxygenated phosphate buffer (pH 7.5). The reactions were monitored by RP-HPLC and the identity of the products was confirmed by mass spectrometry. For peptide **1** the alkylation was carried out at rt in the presence of 4 equiv. of benzyl bromoacetate. After 10 min, LC-MS analysis of an aliquot of the reaction carried out in the absence of Zn(II) showed a mixture of several products, the major being the trialkylated adduct (retention time 22.36 min). The HPLC peak at 19.45 min corresponds to a monoalkylated derivative (Fig. 3, left, and ESI†). Interestingly, when the reaction was carried out in the presence of $ZnSO_4$ we observed a different outcome, with the exclusive formation of a monoalkylated derivative (retention time 18.42 min). The other peaks correspond to the starting peptide **1** (16.73 min) and benzyl bromoacetate (Fig. 3, right).

For the full identification of the monoalkylated product of the Zn containing reaction we did a MS/MS study of the fragments obtained after digestion of the modified peptide with trypsin. After 2 h of digestion, HPLC of the mixture showed two peaks with double-charged signals at 851.36347 and 787.31604 (see the ESI†), which correspond to fragments cleaved at Lys^{171} (y14 fragment) and Arg^{170}. The deconvoluted MS/MS spectra of the first peak show signals corresponding to the y ion series of the partial sequence $GC^{159}AGPFAC^{165}DIC^{168}GRK$, where Cys^{159} is the cysteine alkylated with benzyl bromoacetate, and Cys^{165} and Cys^{168} are carbamidomethyl cysteines,[14] which are formed during the treatment of the sample before digestion (see the ESI†). The mass difference of 251.04275 between the signals at 1602.68829 (y13) and 1351.62554 (y12) can only be accounted for if the cysteine alkylated with benzyl bromoacetate is next to the acetylated glycine (derivative **3**, Fig. 4). Interestingly, the monoalkylated derivatives obtained in the presence or absence of Zn(II) are different, which suggests that the intrinsic reactivity of the terminal cysteine is lower than that of the internal ones (Fig. 3, and ESI†).

Peptide **2** is analogous to one we previously used to construct DNA binding hybrids containing a zinc finger fragment and a minor groove binder;[13] therefore it seemed specially appropriate to apply the chemoselective strategy for the preparation of DNA binders. Previous synthesis of this and related conjugates required the incorporation of the minor groove-binding agent while the peptide is still protected and attached to the resin, which represents an important drawback in terms of efficiency and versatility.[15] Therefore we first analyzed the reactivity of peptide **2** with benzyl bromoacetate, and found that, in the absence of Zn(II), it is fully converted into the trialkylated adduct after only 5 min at 5 °C (25.09 min, Fig. 5, left). However, in the presence of $ZnSO_4$ the reaction resulted mainly in the formation of a monoalkylated product (22.17 min, for details see the ESI†).

Circular dichroism spectral changes displayed by this monoalkylated product in the presence of Zn(II) are similar to those observed for the precursor peptide **2** (increase in negative ellipticity at 208/222 nm). These changes, consistent with the expected Zn-promoted ββα-folding, confirm the exclusive modification of the cysteine in the side chain of Lys^{44}, since alkylation of other cysteines would prevent such folding (see Fig. S15 of the ESI†).[16]

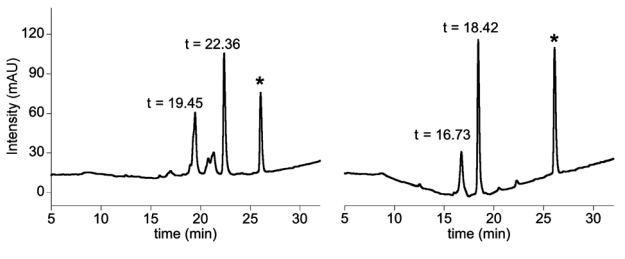

Fig. 3 HPLC traces of the reactions of peptide **1** (200 μM) with benzyl bromoacetate (4 equiv.), in the absence (left) and in the presence (right) of 1.5 equiv. of $ZnSO_4$, after 10 minutes of reaction at rt, in a deoxygenated phosphate buffer (pH 7.5). Peaks with asterisks: benzyl bromoacetate.

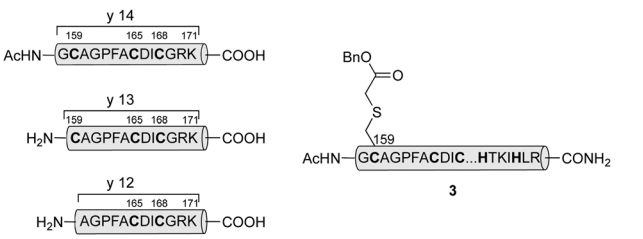

Fig. 4 (left) Schematic representation of the fragmentations of the y ion. (right) Structure of the monoalkylated peptide **3**.

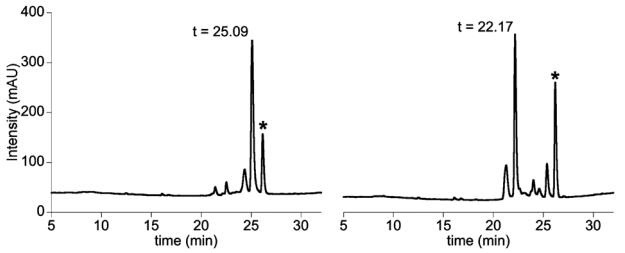

Fig. 5 HPLC chromatograms after 5 minutes of the reaction of peptide **2** (200 μM) with benzyl bromoacetate (4 equiv.), in the absence (left) and in the presence (right) of 1.5 equiv. of $ZnSO_4$ (deoxygenated phosphate buffer, pH 7.5, 5 °C).

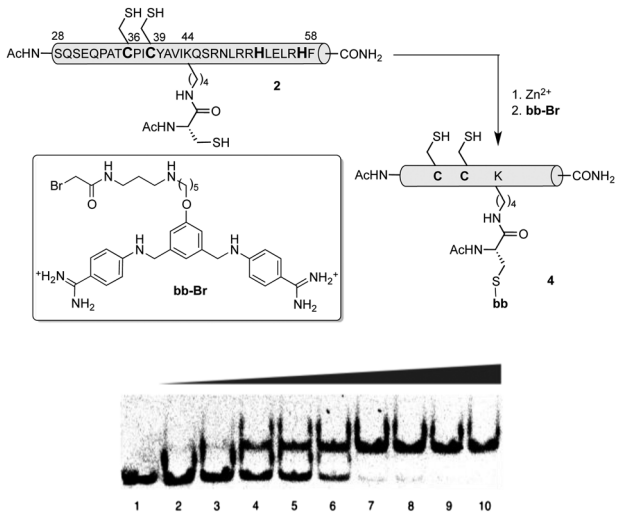

Fig. 6 (top) Synthesis of conjugate **4**. (bottom) EMSA results; lanes 1–10: [**4**] = 0, 100, 200, 300, 400, 500, 600, 800, 900, 1000 nM with 50 nM of **AT·GAGA** dsDNA (SYBR gold staining). Oligonucleotide sequence (binding site underlined, only one strand shown): **AT·GAGA**: 5′-GACGGAATTTGAGAGCGTCG-3′.

With this information at hand we assayed the viability of coupling a more challenging electrophile containing a bisbenzamidine DNA binder (**bb-Br**).[17] The reaction took place efficiently at rt, in the presence of 1.5 equiv. of $ZnSO_4$ and 1 equiv. of the bromide (phosphate buffer, pH 7.5, Fig. 6). Importantly, the resulting conjugate is an efficient and sequence selective DNA binder, as demonstrated by EMSA (Fig. 6).[18] Therefore, the chemoselective ligation provides a very simple approach to functional zinc finger peptides, and opens the possibility of running related reactions with recombinant zinc fingers.

The Zn(II) cation can be exploited both as a folding agent and as a protecting group. This double role allows the selective modification of cysteines present in zinc finger peptides that are not involved in the formation of the Cys_2His_2–Zn fold. The strategy provides a direct approach to functional DNA binding zinc finger conjugates, and promises to be useful for the post-transcriptional modification of zinc finger proteins.

We acknowledge the support from the Spanish grants SAF2010-20822-C02, CTQ2009-14431/BQU, CSD2007-00006, the *Xunta de Galicia* GRC2010/12, GRC2013-041 and INCITE09 209 084PR. J. R. thanks the Fundación Gil Dávila and the *Xunta de Galicia* and J. M. the Spanish MCINN for their PhD fellowships. We specially thank Dr Manuel Marcos for his invaluable help in the MS identification of product **3**.

Notes and references

1 (*a*) C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts and Co., Englewood, Colorado, 2005; (*b*) C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto Jr., *Angew. Chem., Int. Ed.*, 2005, **44**, 7342; (*c*) T. L. Foley and M. D. Burkart, *Curr. Opin. Chem. Biol.*, 2007, **11**, 12; (*d*) D. Qi, C. M. Tann, D. Haring and M. D. Distefano, *Chem. Rev.*, 2001, **101**, 3081; (*e*) I. S. Carrico, *Chem. Soc. Rev.*, 2008, **37**, 1423.
2 D. P. Gamblin, S. I. van Kasteren, J. M. Chalker and B. G. Davis, *FEBS J.*, 2008, **275**, 1949.
3 (*a*) E. Baslé, N. Joubert and M. Pucheault, *Chem. Biol.*, 2010, **17**, 213; (*b*) M. Rashidian, J. K. Dozier and M. D. Distefano, *Bioconjugate Chem.*, 2013, **24**, 1277.
4 (*a*) H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004; (*b*) S. S. Van Berkel, M. B. van Eldijk and J. C. M. van Hest, *Angew. Chem., Int. Ed.*, 2011, **50**, 8806.
5 (*a*) V. Muralidharan and T. W. Muir, *Nat. Methods*, 2006, **3**, 429; (*b*) A. J. de Graaf, M. Kooijman, W. E. Hennink and E. Mastrobattista, *Bioconjugate Chem.*, 2009, **20**, 1281.
6 (*a*) G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, London, 2008; (*b*) J. M. Chalker, G. J. L. Bernardes, Y. A. Lin and B. G. Davis, *Chem.–Asian J.*, 2009, **4**, 630.
7 (*a*) M. W. Crankshaw and G. A. Grant, *Modification at Cysteine*, Wiley, Unit 15.1, 1996; (*b*) J. J. Smith, D. W. Conrad, M. J. Cuneo and H. W. Hellinga, *Protein Sci.*, 2005, **14**, 64; (*c*) J. M. Kuiper, R. Pluta, W. H. C. Huibers, F. Fusetti, E. R. Geertsma and B. Poolman, *Protein Sci.*, 2009, **18**, 1033.
8 C. O. Pabo, E. Peisach and R. A. Grant, *Annu. Rev. Biochem.*, 2001, **70**, 313.
9 (*a*) A. Klug, *Annu. Rev. Biochem.*, 2010, **79**, 213; (*b*) F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang and P. D. Gregory, *Nat. Rev. Genet.*, 2010, **11**, 636; (*c*) N. Holt, J. Wang, K. Kim, G. Friedman, X. Wang, V. Taupin, G. M. Crooks, D. B. Kohn, P. D. Gregory, M. C. Holmes and P. M. Cannon, *Nat Biotechnol.*, 2010, **28**, 839.
10 M. Isaac, J. M. Latour and O. Sénèque, *Chem. Sci.*, 2012, **3**, 3409.
11 M. Elrod-Erickson, M. A. Rould, L. Nekludova and C. O. Pabo, *Structure*, 1996, **4**, 1171 (PDB ID: 1AAY).
12 J. G. Omichinski, P. V. Pedone, G. Felsenfeld, A. M. Gronenborn and G. M. Clore, *Nat. Struct. Biol.*, 1997, **4**, 122 (PDB ID: 1YUI).
13 (*a*) O. Vázquez, M. E. Vázquez, J. B. Blanco, L. Castedo and J. L. Mascareñas, *Angew. Chem., Int. Ed.*, 2007, **46**, 6886; (*b*) M. I. Sánchez, O. Vázquez, M. E. Vázquez and J. L. Mascareñas, *Chem.–Eur. J.*, 2013, **19**, 9923.
14 Positions correspond to the numbering scheme in the Zif268 structure 1AAY in ref. 11.
15 (*a*) J. B. Blanco, V. I. Dodero, M. E. Vázquez, M. Mosquera, L. Castedo and J. L. Mascareñas, *Angew. Chem., Int. Ed.*, 2006, **45**, 8210; (*b*) J. B. Blanco, M. E. Vázquez, L. Castedo and J. L. Mascareñas, *ChemBioChem*, 2005, **6**, 2173; (*c*) J. B. Blanco, O. Vazquez, J. Martinez-Costas, L. Castedo and J. L. Mascareñas, *Chem.–Eur. J.*, 2005, **11**, 4171; (*d*) J. B. Blanco, M. E. Vázquez, J. Martinez-Costas, L. Castedo and J. L. Mascareñas, *Chem. Biol.*, 2003, **10**, 713; (*e*) M. E. Vázquez, A. M. Caamaño, J. Martínez-Costas, L. Castedo and J. L. Mascareñas, *Angew. Chem., Int. Ed.*, 2001, **40**, 4723.
16 A. D. Frankel, J. M. Berg and C. O. Pabo, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 4841.
17 (*a*) M. I. Sánchez, O. Vazquez, J. Martinez-Costas, M. E. Vazquez and J. L. Mascareñas, *Chem. Sci.*, 2012, **3**, 2383; (*b*) O. Vázquez, M. I. Sánchez, J. Martínez-Costas, M. E. Vázquez and J. L. Mascareñas, *Org. Lett.*, 2010, **12**, 216.
18 As shown in the ESI† (Fig. S12), hybrid **4** does not bind to DNA sequences mutated in the peptide binding site.

