A chemical approach for the synthesis of the DNA-binding domain of the oncoprotein MYC

Renata Calo-Lapido, Cristina Penas, Adrián Jiménez-Balsa, M. Eugenio Vázquez, and José L. Mascareñas

Journal Name

COMMUNICATION

A chemical approach for the synthesis of the DNA-binding domain of the oncoprotein MYC

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Renata Calo-Lapido,[a] Cristina Penas,[a] Adrián Jiménez-Balsa,[a] M. Eugenio Vázquez,*[,a] and José L. Mascareñas*[,a]

We describe the first chemical synthesis of a functional mutant of the DNA binding domain of the oncoprotein MYC, using two alternative strategies which involve either one or two Native Chemical Ligations (NCLs). Both routes allowed the efficient synthesis of a miniprotein which is capable of heterodimerizing with MAX, and replicate the DNA binding of the native protein. The versatility of the reported synthetic approach enabled the straightforward preparation of MYC and *Omomyc* analogues, as well as fluorescently labeled derivatives.

The oncogenic MYC transcription factor is a key signaling hub that integrates multiple cellular pathways,[1] and is deregulated and overexpressed in many tumors.[2] As a member of the basic helix–loop–helix-leucine zipper (bHLH-Zip) family of transcription factors,[3] MYC effects are mediated by its association with other proteins of the same family, and in particular with MAX. The resulting MYC/MAX heterodimers are capable of binding selectively to E-box sites (CACGTG) of gene promoters. Despite its pharmacological interest, targeting MYC and affecting its dimerization properties is far from trivial.[4, 5] In this context, recent work by the group of L. Soucek has shown that a protein analogue, *omomyc*, interferes with the activity of MYC and induces growth arrest.[6]

A major drawback for exploring the biological and therapeutic properties of these proteins derives from the lack of efficient methods for their production. Recombinant methods are slow,[7] and more importantly, they do not easily allow the introduction of non-natural elements or purposely designed modifications, such as the introduction of small fluorophores. In this context, alternative chemical approaches are desirable, and they provide practical and versatile syntheses of this protein and designed derivatives.

Herein, we report the total synthesis of a 82-residue miniprotein featuring the DNA interacting domain of natural MYC, with two specific mutations, Asn^{500}→Ala and Asn^{515}→Gln, which do not affect its DNA binding but facilitate the synthesis.[8]

The assembly of such long peptide was carried in a convergent way by using Native Chemical Ligations (NCL). In its standard format, NCL involves the reaction between a C-terminal peptide-thioester and and an*N*-terminal cysteinyl peptide.[9] Since MYC lacks cysteine residues in its sequence, we chose Xaa-Ala as ligation sites because Ala residues can be readily obtained by desulfurization of the Cys required for the NCL.[10]

Preliminary assays of the solid phase assembly of the *N*-terminal peptide fragments revealed the formation of a small proportion of undesired products resulting from spontaneous deamination of asparagine residues Asn^{500} and Asn^{515}.[11] Considering that these residues are not involved in DNA contacts of the natural protein, or in the formation of the MYC/MAX heterodimer, we decided to replace Asn^{500} by Ala, and Asn^{515} by Gln, so that our final target would be a double MYC mutant **[AQ]MYC** (Figure 1). We designed two disconnection schemes using either one or two NCL reactions (Figure 1b). The first case requires the synthesis of two long peptides corresponding to **Ala^{500}–Lys^{536}** (37 residues) and **Ala^{537}–Leu^{581}** (45 residues) fragments (2Fr-NCL, Figure 1b, left), while the second approach involved three shorter peptide segments, **Ala^{500}–Phe^{522}**, **Ala^{523}–Thr^{547}**, and **Ala^{548}–Leu^{581}**, featuring 23, 25 and 34 residues, respectively (3Fr-NCL, Figure 1b, right). The three-fragment approach is more convergent, and hence more appropriate to obtain *N*-labeled **[AQ]MYC** derivatives. Moreover, we anticipated that the double Cys intermediate, **[AQ,C^{523},C^{548}]MYC**, obtained in the 3-Fr-NCL strategy, might be exploited to implement redox- controlled DNA binding.

Peptides were assembled using the Nbz method,[12] which involves the postsynthetic generation of a *C*-terminal *N*-acyl urea leaving group (*N*-acyl-benzimidazolinone, Nbz) by on-resin modification of a precursor 3,4-diaminobenzoic acid (Dbz) with *p*-nitrophenyl chloroformate (ESI).

[a] *Departamento de Química Orgánica and Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CIQUS). Universidade de Santiago de Compostela. 15782 Santiago de Compostela, Spain.*

Electronic Supplementary Information (ESI) available: Peptide synthesis, purification and characterization, DNA binding experiments. See DOI: 10.1039/x0xx00000x

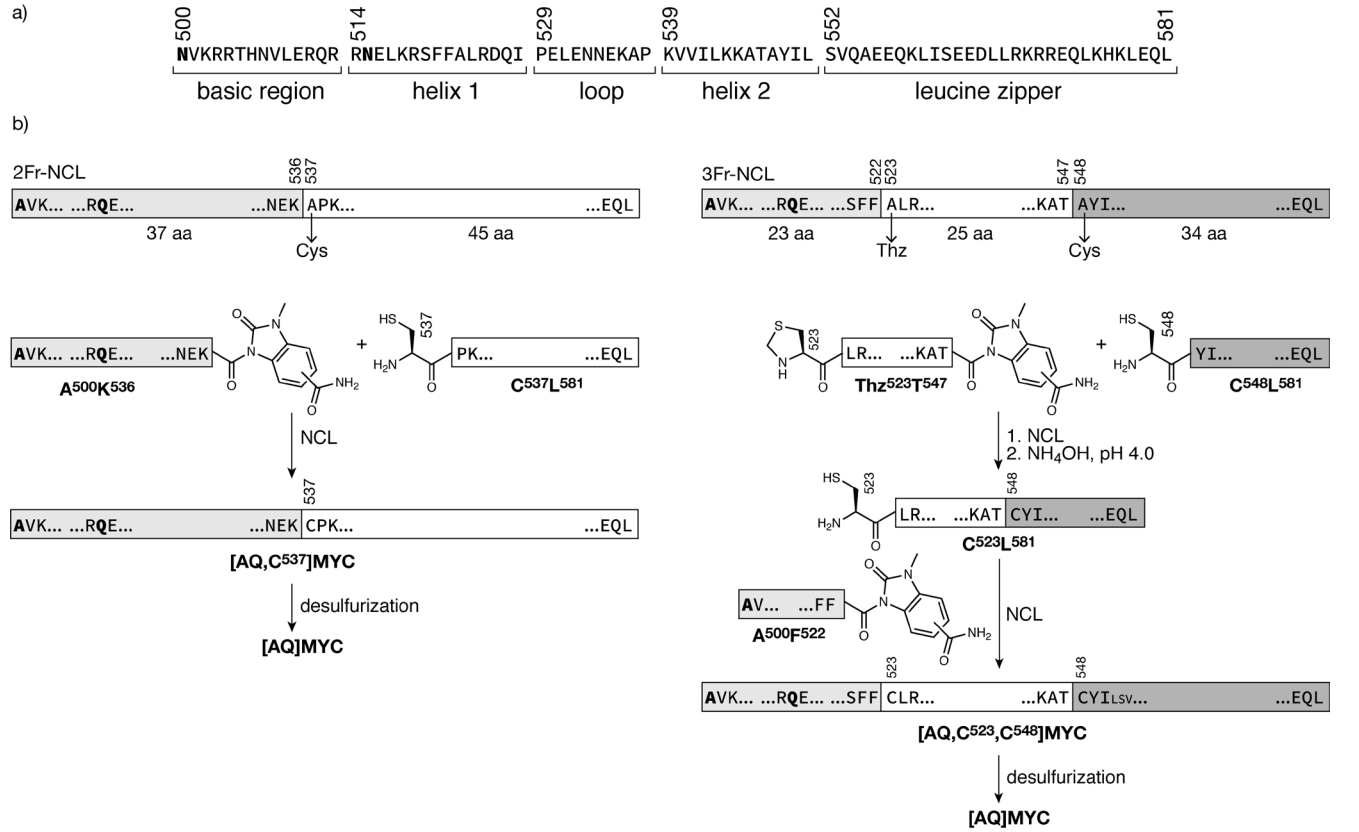

Figure 1. a) Sequence and structural elements of the DNA binding domain of MYC. b) Alternative disconnection schemes for synthesizing the Asn500→Ala and Asn515→Gln MYC mutant (**[AQ]MYC**) from two (2Fr-NCL, left) or three peptide fragments (3Fr-NCL, right), showing the positions selected for the introduction of the reactive Cys (or Thz) residues. NCL conditions: 6 M Gdn HCl, 200 mM phosphate buffer (pH 7.5), 100 mM 4-mercaptophenylacetic acid (MPAA), 40 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl), 4.5 h; Desulfurization: 6 M Gdn HCl, 200 mM phosphate buffer (pH 7.0), 500 mM TCEP (tris(2-carboxyethyl) phosphine), EtSH, VA-044 (2-2'-azobis[2-(2-imidazolin-2-*yl*) propane]dihydrochloride) 1 M, 37º C, 5 h.

After cleavage and deprotection, the Nbz group can be readily exchanged with thiol additives at neutral pH to produce the peptide thioesters required for the NCL (Figure 1b). Briefly, the two-fragment approach allowed the assembly of **[AQ]MYC** using a single NCL by coupling peptide segments **$A^{500}K^{536}$**–Nbz and **$C^{537}L^{581}$** in the presence of 4-mercaptophenyl acetic acid (MPAA) as arylthiol catalyst. The reaction was essentially complete after 4.5 h (54 %, 15 mg scale). The desulfurated target peptide (**[AQ]MYC)** was readily obtained **[AQ,C^{537}]MYC** using standard radical-based methods.[13] Assembly of **[AQ]MYC** following the three-fragment approach (3Fr-NCL) was carried out in a similar way. In this case, one of the peptide fragments used in the first NCL step features the Cys in the form of thiazolidine (Thz) to avoid oligomerization and unwanted self-couplings (**$Thz^{523}T^{547}$**, Figure 1).[14] Following the first NCL, the Thz was removed by adding hydroxylamine to the reaction mixture and adjusting the pH to 4, and the resulting peptide (**$C^{523}Q^{581}$**) was isolated by HPLC (over 20 % for this NCL). The second NCL afforded the expected product (in this case containing two Cys residues in positions 548 and 523), which was isolated by reverse phase HPLC (Figure S18). After desulfurization and purification, we obtained **[AQ]MYC** (over 4 % global yield). It must be highlighted that both synthetic approaches (2Fr-NCL and 3Fr-NCL) can be carried out at a milligram scale to give the desired miniprotein **[AQ]MYC**, which presented a good water solubility.

With **[AQ]MYC** at hand, we studied its DNA binding properties in comparison with the wild-type DNA binding domain of MYC (^{wt}MYC).[15] In agreement with earlier reports, MAX is capable of binding as homodimer to the E-box site, although with a low association constant (Figure 2a and b, lanes 2 and 3). In contrast, ^{wt}MYC does not appear to dimerize and bind to the DNA (Figure 2a, lanes 4 and 5). As expected, incubation of MAX with increasing concentrations of ^{wt}MYC, resulted in the appearance of a slightly less retarded band (Figure 2a, lanes 6-9) that is consistent with the formation of the DNA complex of the MAX/^{wt}MYC heterodimer.[16] Importantly, our synthetic MYC analog, **[AQ]MYC**, reproduces the MAX recognition and DNA binding of ^{wt}MYC, although with some differences: **[AQ]MYC** shows a slightly more tendency to homodimerize (Figure 2b, lane 5), and the heterodimer of **[AQ]MYC** with MAX binds the target E-DNA at lower concentrations than that of ^{wt}MYC (Figure 2a and 2b, compare lanes 7). Control EMSA experiments with mutated DNAs demonstrated that the **[AQ]MYC**/**MAX** heterodimer binds the E-DNA with high selectivity (Figure S30). Circular Dichroism studies were also consistent with the expected binding/folding processes (Figure S39).[17]

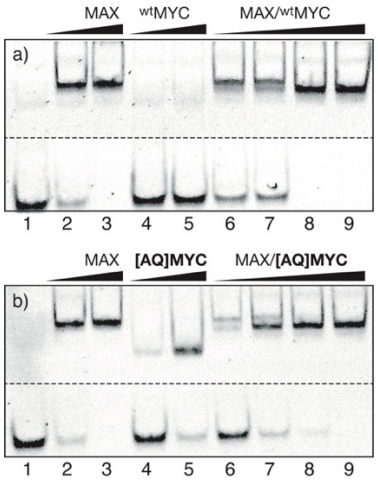

Figure 2. Non-denaturing EMSA showing the binding of MAX, the expressed ^{wt}MYC, and the synthetic **[AQ]MYC** to a double stranded oligo containing the target E-box sequence (**E-DNA**, 5´-CTGTAGGC CACGTG ACCGGATG-3´, only one chain shown). a) Lanes 1-3: 0, 500, and 1000 nM MAX; lanes 5-6: 500 and 1000 nM ^{wt}MYC; lanes 7-10: 250, 500, and 750 nM ^{wt}MYC with constant 500 nM MAX. b) same as in gel a, but with the synthetic **[AQ]MYC** instead of ^{wt}MYC. All lanes contain 50 nM **E-DNA** in 20 mM Tris-HCl pH 7.5, 90 mM KCl, 1.8 mM $MgCl_2$, 1.8 mM EDTA, 9% glycerol, 0.11 mg/mL BSA, 2.25% NP-40 (20 ºC). The gels were visualized by staining with SYBR^{TM} Gold.

These experiments confirm that the synthetic product **[AQ]MYC** heterodimerizes with MAX to bind the E-box site. One of the advantages of the synthetic approach is that, in addition to the final product **[AQ]MYC**, we also obtained the synthetic intermediate **[AQ,C^{523},C^{548}]MYC**, featuring Cys at the ligation sites. Inspection of the X-ray structure of the cMYC/cMAX/DNA complex reveals that the positions corresponding to these cysteine residues are not involved in DNA binding (Figure 3a), and therefore we predicted that this mutant might show similar DNA binding properties than **[AQ]MYC**. We also reasoned that engaging these cysteines in an intramolecular disulfide bond should abolish the DNA binding. For simplicity, from now on we will refer to the double Cys mutant, **[AQ,C^{523},C^{548}]MYC**, as **[AQ]MYC(SH)**. Treatment of **[AQ]MYC(SH)** with DTNB (5,5'-dithiobis-(2-nitrobenzoic acid), in 100 mM Tris·HCl buffer (pH 7.5) for 1 h,[18] led to the corresponding oxidized peptide **[AQ]MYC(S-S)**, featuring a disulfide crosslink between residues Cys^{523} and Cys^{548} (isolated and identified by ESI-MS, see the ESI).

Addition of increasing amounts of **[AQ]MYC(SH)** to a mixture of MAX and **E-DNA** resulted in two slow-migrating bands, the first, predominant at low **[AQ]MYC(SH)** concentrations, must correspond to the homodimeric MAX_2/**E-DNA** complex, (Figure 3a, lanes 1-4, band a); the second band, observed at higher concentrations of **[AQ]MYC(SH)**, is consistent with the heterodimeric **[AQ]MYC(SH)**/MAX/**E-DNA** complex (Figure 3a, band b). As expected, oxidation provides a disulfide that is unable to cooperate with MAX and bind DNA, and therefore we only observe the band arising from the MAX homodimer (Figure 3b, lane 1). However, *in situ* addition of a reducing agent (TCEP) regenerates the parent reduced peptide **[AQ]MYC(SH)**, and the heterodimeric DNA binding (Figure 4b, lane 2). These preliminary results propose a strategy to develop externally activatable MYC surrogates. Using the three-fragment NCL we also synthesized the labeled analogs *TMR*-**[AQ]MYC(SH)** and *TMR*-**[AQ]MYC**, featuring a TMR (5-Carboxytetramethylrhodamine) fluorophore at the *N*-terminal positions (see the ESI).

Moreover, following the same three-fragment approach, we synthesized two *Omomyc* analogues: (*TMR*-**[AQS]OMOMYC** and the bis-cysteine derivative (*TMR*-**[AQS]OMOMYC(SH)** (ESI).[19] Preliminary uptake studies revealed that the cysteine-containing miniproteins are more efficiently internalized (over 4 times in A549 and 7-8 times in HeLa) than the desulfurized analogues (Figure 4, top vs bottom micrographies and Figure S37).[20, 21]

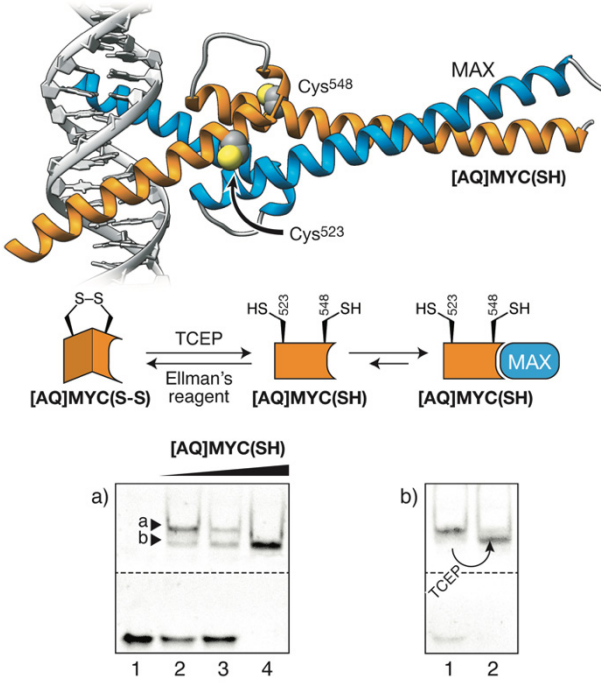

Figure 3. *Top*: X-ray structure of the complex MYC-MAX/DNA, highlighting the positions of the cysteines in **[AQ]MYC(SH)** (shown in orange). *Middle*: Schematic representation of the reduction/activation process: The crosslinked **[AQ]MYC(S-S)** adopts a conformation incompatible with MAX binding. *Bottom*: a) EMSA analysis of the DNA binding of **[AQ]MYC(SH)** and MAX to **E-DNA**: Lanes 1-4: 50 nM **E-DNA**; lanes 2-4: 500 nM MAX and 250, 500, and 1000 nM of **[AQ]MYC(SH)**; b) Redox switching: 50 nM **E-DNA**, lane 1: 500 nM MAX and 1000 nM **[AQ]MYC(S-S)**; lane 2: after addition of TECP. All lanes contain 50 nM DNA in 20 mM Tris-HCl pH 7.5, 90 mM KCl, 1.8 mM $MgCl_2$, 1.8 mM EDTA, 9 % glycerol, 0.11 mg/mL BSA, 2.25 % NP-40 (20 ºC). Dashed line indicates that the top and bottom of the gel with the bands was mounted in the final image.

Additionally, the *Omomyc* analogue is also better uptaken than *TMR-***[AQ]MYC**; and tends to accumulate in the perinuclear space (Figure S36). Our results confirm that small variations in the sequence of MYC can have very important consequences in solubility, aggregation and internalization properties.

In summary, we have reported the first synthesis of a functional, DNA-binding analog of the bHLHzip domain of MYC using chemical methods. The flexibility of the methodology allowed a practical synthesis of a redox-activatable, and fluorescently tagged derivatives not only of MYC but also of its therapeutically-relevant analog *Omomyc*. These derivatives could not be assembled using expression approaches. Cell

uptake studies revealed that all the miniproteins are internalized, albeit with different efficiencies, depending on the sequences.

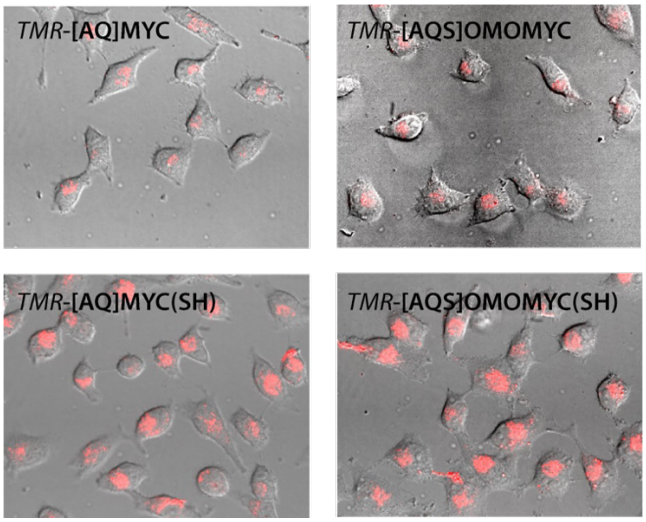

Figure 4. Overlay of both red channel and DIC fluorescence micrographies of A549 cells incubated for 4 h with 250 nM of the synthetic peptides in DMEM (without FBS). Cells were then washed twice with PSB solution and three more times with DMEM. Pictures were taken at 60X and λ_{exc} = 561 nm for the red channel, in a confocal microscope.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Financial support from the MINECO (CTQ2015-70698-R and SAF2016-76689-R, orfeo-cinqa CTQ2016-81797-REDC), Xunta de Galicia (2015-CP082, ED431C 2017/19, ED431B 2018/04, Centro singular de investigación de Galicia accreditation 2016–2019, ED431G/09) and the European Union (European Regional Development Fund - ERDF), *Fundación AECC* (IDEAS197VAZQ), and the European Research Council (Advanced Grant No. 340055) are acknowledged. R.C.L. thanks Xunta de Galicia and European Social Fund for her fellowship. A.J.B. thanks Xunta de Galicia for his Post-doctoral fellowship. We are grateful to Prof. Dr. Bernhard Lüscher, who generously shared with us the pBluescript p21 Max, pGEX Myc c92, pGEX-2T-Myc C176, and pGEX 3X plasmids for expression of the MYC and MAX protein

Notes and references

1 a) C. Bouchard, P. Staller and M. Eilers, *Trends Cell Biol*., 1998, *8*, 202–206; b) C. J. Poole and J. van Riggelen, *Genes*, 2017, **8**, 142.

2 M. Gabay, Y. Li and D. W. Felsher Cold Spring Harb *Perspect Med*. 2014, **4**, a014241.

3 a) C. W. Garvie and C. Wolberger, *Mol. Cell*, 2001, **8**, 937–946; b) E. Pazos, J. Mosquera, M. E. Vázquez and J. L. Mascareñas, *ChemBioChem* 2011, **12**, 1958–1973; c) M. E. Vázquez, A. M. Caamaño and J. L. Mascareñas, *Chem. Soc. Rev*. 2003, **32**, 338–349; d) S. Boga, D. Bouzada, D. García Peña, M. Vázquez López and M. E. Vázquez, *Eur. J. Org. Chem*. 2018, **2018**, 249–261.

4 a) V. N. Uversky, *Expert Opin. Drug Discov*., 2012, **7**, 475–488; b) S. J. Metallo, *Curr. Opin. Chem. Biol*., 2010, **14**, 481–488.

5 a) J. S. Lazo and E. R. Sharlow, *Annu. Rev. Pharmacol. Toxicol*., 2016, **56**, 23–40; b) C. V. Dang, E. P. Reddy, K. M. Shokat and L. Soucek, *Nat. Rev. Cancer*, 2017, **17**, 502–508.

6 a) L. Soucek, R. Jucker, L. Panacchia, R. Ricordy, F. Tato and S. Nasi, *Cancer Res*., 2002, **62**, 3507–3510; b) M.-E. Beaulieu, T. Jauset, D. Massó-Vallés, S. Martínez-Martín, P. Rahl, L. Maltais, M. F. Zacarias-Fluck, S. Casacuberta-Serra, E. Serrano del Pozo, C. Fiore, L. Foradada, V. Castillo Cano, M. Sánchez-Hervás, M. Guenther, E. Romero Sanz, M. Oteo, C. Tremblay, G. Martín, D. Letourneau, M. Montagne, M. Á. Morcillo Alonso, J. R. Whitfield, P. Lavigne and L. Soucek, *Science Trans. Med.* 2019, **11**, eaar5012.DOI: 10.1126/scitranslmed.aar5012.

7 M.-E. Beaulieu, F.-O. McDuff, V. Frappier, M. Montagne, J.-F. Naud and P. Lavigne, *J. Mol. Recognit*., 2012, **25**, 414–426. High efficiencies have been recently reported: M. E. Beaulieu, F. O. McDuff, M. Bédard, M. Montagne, P. Lavigne P., *Methods Mol. Biol*. 2013, **1012**, 7-20.

8 Residues numbered according to the X-ray structure of the MYC-MAX heterodimer bound to the DNA E-box (PDB ID: 1NKP). S. K. Nair and S. K. Burley, Cell, 2003, **112**, 193–205.

1 9 a) P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, *Science*, 1994, **266**, 776–779; b) C. P. R. Hackenberger and D. Schwarzer, *Angew. Chem. Int. Ed Eng*., 2008, **47**, 10030–10074; c) S. Bondalapati, M. Jbara and A. Brik, *Nat. Chem*. 2016, **8**, 407–418; d) V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J-C. M. Monbaliu, and O. Melnyk, *Chem. Rev.* 2019, DOI: 10.1021/acs.chemrev.8b00712.

10 L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc*., 2001, **123**, 526–533.

11 T. Geiger and S. Clarke, *J. Biol. Chem*., 1987, **262**, 785–794.

12 J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem. Int. Ed*., 2008, **120**, 6957–6961.

13 S. Liu, B. L. Pentelute and S. B. H. Kent, *Angew. Chem. Int. Ed*., 2011, **51**, 993–999.

14 D. Bang and S. B. H. Kent, *Angew. Chem. Int. Ed*., 2004, **43**, 2534–2538.

15 This protein was obtained through bacterial expression of the pGEX Myc c92 plasmid. Sequence: GSPSSDTEEN VKRRTHNVLE RQRRNELKR SFFALRDQI PELENNEKA PKVVILKKA TAYILSVQA EEQKLISEE DLLRKRREQ LKHKLEQLR NSCA; wild-type MAX: GIQMSDNDDI EVESDADKRA HHNALERKRR DHIKDSFHSL RDSVPSLQGE KASRAQILDK ATEYIQYMRR KNHTHQQDID DLKRQNALL EQQVRALEK ARSSAQLQTN YPSSDNSLYT NAKGSTISAF DGGSDSSSES EPEEPQSRK KLRMEAS.

16 C. Muhle-Goll, M. Nilges and A. Pastore, *Biochemistry*, 1995, **34**, 13554–13564.

17 a) C. Muhle, T. Gibson, P. Schuck, D. Schubert, D. Nalis, and A. Pastore, *Biochemistry*, 1994, **33**, 11296–11306; b) O. Vázquez, M. E. Vázquez, J. B. Blanco, L. Castedo and J. L. Mascareñas, *Angew. Chem. Int. Ed*., 2007, **46**, 6886–6890; c) M. E. Vázquez, A. M. Caamaño, J. Martínez-Costas, L. Castedo and J. L. Mascareñas, *Angew. Chem. Int. Ed*., 2001, **40**, 4723–4725.

18 J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vázquez and J.L. Mascareñas, *Nat. Commun*., 2013, **4**, 1–8.

19 These derivatives contain key mutations present in *omomyc* with respect to MYC, as well as the A and Q mutations used in the case of our MYC analogue (see the ESI†).
20 N. Chuard, G. Gasparini, D. Moreau, S. Lörcher, C. Palivan, W. Meier, N. Sakai and S. Matile, *Angew. Chem. Int. Ed Engl*. 2017, **56**, 2947–2950.
21 a) X. Meng, T. Li, Y. Zhao and C. Wu, *ACS Chem. Biol*. 2018, **13**, 3078–3086; b) N. Chuard, G. Gasparini, D. Moreau, S. Lörcher, C. Palivan, W. Meier, N. Sakai and S. Matile, *Angew. Chem. Int. Ed Eng*. 2017, **56**, 2947–2950; c) S. Learte-Aymamí, N. Curado, J. Rodríguez, M. E. Vázquez and J. L. Mascareñas, *J. Am. Chem. Soc*., 2017, **139**, 16188–16193.