Genetic code expansion for multiprotein complex engineering

Christine Koehler[1], Paul F Sauter[2], Mirella Wawryszyn[2], Gemma Estrada Girona[1], Kapil Gupta[3], Jonathan J M Landry[1], Markus Hsi-Yang Fritz[1], Ksenija Radic[1], Jan-Erik Hoffmann[1], Zhuo A Chen[4], Juan Zou[4], Piau Siong Tan[1], Bence Galik[5], Sini Junttila[5], Peggy Stolt-Bergner[5], Giancarlo Pruneri[6], Attila Gyenesei[5], Carsten Schultz[1], Moritz Bosse Biskup[2], Hueseyin Besir[1], Vladimir Benes[1], Juri Rappsilber[4,7], Martin Jechlinger[1], Jan O Korbel[1], Imre Berger[3,8], Stefan Braese[2,9] & Edward A Lemke[1]

We present a baculovirus-based protein engineering method that enables site-specific introduction of unique functionalities in a eukaryotic protein complex recombinantly produced in insect cells. We demonstrate the versatility of this efficient and robust protein production platform, 'MultiBacTAG', (i) for the fluorescent labeling of target proteins and biologics using click chemistries, (ii) for glycoengineering of antibodies, and (iii) for structure–function studies of novel eukaryotic complexes using single-molecule Förster resonance energy transfer as well as site-specific crosslinking strategies.

The generation of sufficient quantities of eukaryotic protein complexes is frequently the first and limiting step for studies of molecular mechanisms using numerous biophysical and biochemical assays. *Escherichia coli* is one of the most popular organisms for recombinant protein production, but many proteins and protein complexes (particularly eukaryotic protein complexes) cannot be expressed in such simple organisms. Over the last decade, the MultiBac system has become a widely used system in basic and applied research for eukaryotic protein complex production[1,2]. A particularly attractive feature of MultiBac is its ability to rapidly shuffle proteins, introduce mutations and generate diverse complexes in a user-friendly format to achieve high-yielding expression in insect cell lines derived from *Spodoptera frugiperda* (Sf) or *Trichoplusia ni*[3]. The power and versatility of the MultiBac platform could be dramatically enhanced by providing this platform with the means to site-specifically engineer diverse custom functionalities into protein complexes.

Genetic code expansion (GCE) allows noncanonical amino acids (ncAAs) with unique functionalities to be encoded site-specifically into a protein of interest (POI). This method was initially developed in *E. coli*, in which there are now more than 200 different ncAAs can be introduced anywhere in a polypeptide chain by simply introducing a rare codon (typically the amber TAG stop codon) in the coding gene of the POI (for reviews, see refs. 4–6). The POI^{TAG} is expressed in an organism that harbors an additional orthogonal tRNA–tRNA synthetase pair (tRNA–RS), in which the enzyme's active site is commonly modified to recognize only a specific ncAA. As such, the amber codon is repurposed as a sense codon only when the ncAA is present in the growth medium.

We set out to implement the GCE system in MultiBac insect cells in order to combine this protein engineering technique with a method for convenient, high-yielding recombinant eukaryotic protein complex generation. We chose to work with the pyrrolysine $tRNA^{Pyl}$–PylRS system from *Methanosarcina mazei* because it had already been adapted for use in a variety of eukaryotic organisms including animals and because most of the available ncAAs have been encoded by this system[4–6].

MultiBac consists of one acceptor and several donor plasmid modules that access a baculoviral genome optimized for multigene expression (**Fig. 1**)[3]. Our test system consisted of plasmids encoding the wild-type (WT) PylRS from *M. mazei*, a gene cassette for the cognate amber suppressor tRNA and a reporter protein, mCherry–$GFP^{39\rightarrow TAG}$. The ratio of GFP signal to mCherry provided a convenient readout of the efficiency of amber suppression as detected by flow cytometry (FC). Subsequently, the system can be tested by transient transfection of Sf21 cells, or it can be used to generate a multigene fusion plasmid following established protocols (**Supplementary Fig. 1** and **Supplementary Note 1**)[3]. We tested various known tRNA expression cassettes driven by external U6 PolIII promoters which were previously used for successful GCE in other eukaryotes, including mammalian cell cultures[7–9] and *Drosophila melanogaster*[10,11]. As PolIII promoters were not documented for Sf21, a tRNA cassette using a U6 promoter from *Bombyx mori*[12], an insect species closely related to *S. frugiperda*, was also tested. Despite critical external PolIII elements largely considered to be conserved across species (see **Supplementary Fig. 2** for a comparison of snRNA U6 genes across species), no reporter POI expression was detected in any of our tests (**Supplementary Fig. 3**).

Therefore, to identify a potentially useful promoter, we resorted to sequencing and annotating the genome of Sf21 cells (**Supplementary Note 2** and **Supplementary Table 1**). We identified eight snRNA U6 genes and a dicistronic tRNA expression cassette with a gene architecture analogous to that previously used for efficient GCE in

[1]European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. [2]Karlsruhe Institute of Technology (KIT), Institute of Organic Chemistry, Karlsruhe, Germany. [3]European Molecular Biology Laboratory (EMBL), Grenoble, Grenoble, France. [4]Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom. [5]Vienna Biocenter Core Facilities (VBCF GmbH), Vienna, Austria. [6]Division of Pathology and Laboratory Medicine, European Institute of Oncology, Milan, Italy. [7]Chair of Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany. [8]The School of Biochemistry, University of Bristol, Bristol, United Kingdom. [9]Karlsruhe Institute of Technology (KIT), Institut für Toxikologie und Genetik, Eggenstein-Leopoldshafen, Germany. Correspondence should be addressed to E.A.L. (lemke@embl.de).

RECEIVED 24 APRIL; ACCEPTED 2 SEPTEMBER; PUBLISHED ONLINE 17 OCTOBER 2016; DOI:10.1038/NMETH.4032

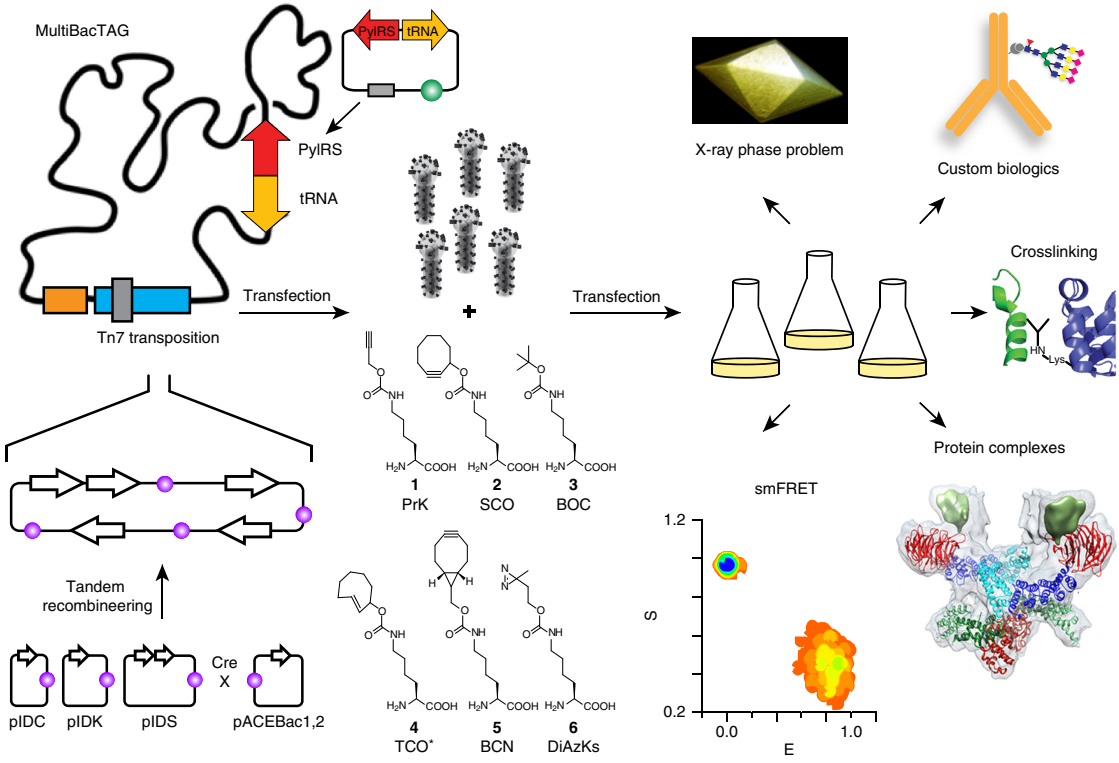

Figure 1 | An overview of the MultiBacTAG system for the expression of multidomain protein complexes in insect cells with different ncAAs for diverse applications. Several POIs can be combined using tandem recombineering of several donor plasmids and one acceptor plasmid (pIDC, pIDK, pIDS, and pACEBac1,2) via Cre–loxP sites (violet sphere, more details given in corresponding **Supplementary Fig. 1**), and then plasmids are inserted into the Tn7 site in the bacmid DNA, which contains the $tRNA^{Pyl}$–PylRS pair. After production of the baculovirus, insect cells are transduced and the ncAA of choice is added. Structures of ncAAs used in this work are shown: propargyl–lysine (1, PrK), cyclooctyne–lysine (2, SCO), Boc–lysine (3, BOC), *trans*-cyclooctene–lysine (4, TCO*), BCN–lysine (5, BCN), and diaziridine-lysine (6, DiAzKs). HN, nitrogen-hydrogen bond; Lys, lysine residue.

Saccharomyces cerevisiae (**Supplementary Figs. 4** and **5**)[13]. As identified by FC analysis, only six U6-driven tRNA constructs allowed for efficient amber suppression (**Supplementary Figs. 5** and **6**).

Choosing U6 promoter 2, we generated a new MultiBac baculoviral genome, termed MultiBacTAG, in which the $tRNA^{Pyl}$–PylRS pair was directly integrated into the viral backbone at the Cre–loxP site (**Fig. 1** and **Supplementary Fig. 1**; superscripts ^{WT} or ^{AF} used to indicate two different PylRS mutants enabling incorporation of different ncAAs shown in **Fig. 1**)[14–16]. The resulting baculovirus maintained the advantageous features of the MultiBac insect cell system, including modularity, protease deficiency and delayed insect cell lysis[3] (further details in **Supplementary Fig. 1**).

We first performed an expression test using different reporters and ncAAs (**Fig. 2**). Expression of the bulky ncAA cyclooctyne–lysine (SCO) using $MultiBacTAG^{AF}$ in a 1-l culture yielded approximately 2 mg of $GFP^{39\rightarrow SCO}$ (**Fig. 2a**), an amount only five-fold lower than the average yield of this simple reporter in state-of-the-art *E. coli* GCE systems for the same tRNA–RS and ncAA[14–16] (see **Supplementary Fig. 7** for mass spectrometry (MS) validation, **Supplementary Fig. 8** for full-size SDS–PAGE, and **Supplementary Table 2** for an overview and comparison of all expression yields in this study). The corresponding flow cytometry analysis of mCherry–$GFP^{39\rightarrow TAG}$ is shown in **Figure 2b**, indicating high efficiency of the GCE MultiBacTAG system (**Supplementary Fig. 8** for complete flow cytometry analysis).

MultiBacTAG was further used to engineer Herceptin, a monoclonal antibody and major protein biologic against breast cancer that selectively associates with cancer cells overexpressing the Her2 tumor marker (**Fig. 2** and **Supplementary Fig. 8**)[17]. Amber mutants (A121TAG and A132TAG) were introduced into known permissive sites of the heavy chain of Herceptin[17], and the genes coding for light and heavy chains were inserted into $MultiBacTAG^{WT}$ and $MultiBacTAG^{AF}$. Herceptin was produced intracellularly and containing different ncAAs that permitted further bioconjugation 'click' reactions with diverse substrates ranging from fluorescent dyes to novel glycosyl groups to underline the potential for glycoengineering (**Fig. 2c–f**, **Supplementary Figs. 8–10**, **Supplementary Table 2** for analytics and yields; and **Supplementary Note 3** for details on glycan used). In particular, *trans*-cyclooctyne–lysine derivatives (TCO*) can undergo particularly fast strain-promoted Diels–Alder cycloadditions with tetrazines (SPDAC) and thus allow for exceptionally mild labeling conditions[14–16]. Indeed, TAMRA tetrazine-labeled $Herceptin^{121\rightarrow TCO^*\rightarrow TAMRA}$ showed a characteristic positive staining pattern of paraffin-embedded human patient samples (**Fig. 2g**,**h**; **Supplementary Fig. 11**; and **Supplementary Table 3** for tumor characteristics and IEO database identifiers (HistoIDs).

Next, we applied the MultiBacTAG system to discover hitherto unidentified protein complex dynamics. Genetic and biochemical data suggested the existence of a pentameric transcription factor complex formed between the human TATA-box binding protein (TBP), cognate DNA containing a TATA box, the general transcription factor TFIIA, and the histone-fold-containing TBP-associated

factors TAF11 and TAF13, which constitute a histone-fold pair[18,19]. We used MultiBacTAG to modify TAF13 in a coexpression experiment with WT TAF11 by using a dual-expression cassette inserted into MultiBacTAG virus. Single-molecule (sm) Förster resonance energy transfer (FRET) allows the measuring of distances in proteins between a site-specifically installed donor and acceptor dye pair[20]. We generated a TAF13$^{20\rightarrow SCO}$ mutant and labeled this in an SPDAC reaction with a smFRET-suitable tetrazine derivative of the donor dye Alexa488. We also labeled a reactive cysteine in TAF13$^{20\rightarrow SCO}$ with a maleimide-derivative acceptor dye Alexa 594 (detailed in **Supplementary Fig. 12**). We then performed smFRET measurements of the TAF11–TAF13$^{20\rightarrow A488,37\rightarrow A594}$ complex. As shown in **Figure 3a**, we detected a population at FRET efficiency (E_{FRET}) = 0.8, a value that provides an important distance constraint for further structural model building.

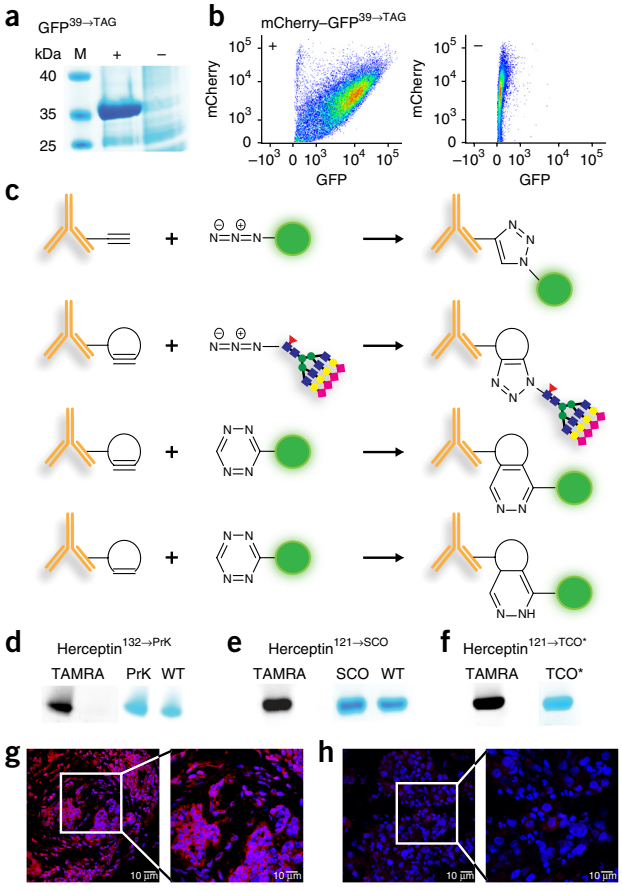

Figure 2 | Characterization of MultiBacTAG, click labeling of Herceptin, and detection of human cancer. (**a**) SDS–PAGE after purification of GFP$^{39\rightarrow TAG}$ expressed in Sf21 cells transfected with MultiBacTAG^{AF} grown in the presence (+) and absence (–) of 1 mM SCO (see **Supplementary Fig. 8** for full-size gels and other ncAAs). (**b**) The corresponding flow cytometry analysis of mCherry–GFP$^{39\rightarrow TAG}$ with (+) and without (–) PrK (representative of three experiments shown). (**c**) Possible labeling reactions between antibody and dye (green dot) or glycan. From top to bottom: (i) Copper-catalyzed click -labeling reaction between a terminal alkyne and an azide. (ii) Copper-free strain promoted azide alkyne cycloaddition between BCN and an azide-containing glycan structure (see **Supplementary Fig. 10** for experimental data). (iii,iv) Different SPDAC reactions. (**d**–**f**) UV scans of the different labeling reactions on the left and Coomassie-stained SDS–PAGE gels on the right of each panel (full-size gels in **Supplementary Fig. 8**). (**d**) Copper-based click chemistry of Herceptin$^{132\rightarrow PrK}$ with fluorescein-azide. (**e**) SPDAC reaction between Herceptin$^{121\rightarrow SCO}$ with TAMRA-tetrazine (Herceptin WT used as negative control). (**f**) SPDAC reaction between Herceptin$^{121\rightarrow TCO*}$ and TAMRA-tetrazine. (**g**,**h**) Herceptin$^{121\rightarrow TCO*\rightarrow TAMRA}$ –based detection of cancer cells in human patient samples (n = 3 for positive and negative tissue samples shown here and in **Supplementary Fig. 11**). Human tumor sections included Her2$^+$ (**g**, cancer tissue overexpressing Her2) and Her2$^-$ (**h**, healthy tissue; for HistoIDs see **Supplementary Table 3**) samples. Images shown are maximum projections of 35 planes spanning 5 µm total. Expansions show a two-fold zoomed image of the squared areas. Blue channel, DAPI; red channel: Herceptin$^{121\rightarrow TCO*}$ labeled with TAMRA-tetrazine.

To directly probe protein–protein binding, we designed another mutant that we speculated was located at a binding interface. We inserted the ncAA DiAzKs (**Fig. 3b** and **Supplementary Note 4** for synthesis of DiAzKs), which harbors an efficient diazirine protein crosslinker[7,21], to generate a TAF11–TAF13$^{34\rightarrow DiAzKs}$ complex. We then performed a set of photocrosslinking experiments followed with subsequent SDS–PAGE and western blot (WB) analysis, as summarized in **Figure 3c** (detailed in **Supplementary Fig. 13**). While TAF11–TAF13$^{34\rightarrow DiAzKs}$ yielded a single-band crosslink product, a double band appeared in a TBP-dependent fashion after UV excitation. SDS–PAGE and WB analysis showed that none of the double bands contained TBP, but that they had an electrophoretic mobility expected for the TAF11–TAF13 complex. As this indicated a conformational change induced by TBP, we used crosslinking and MS to reveal the actual residues involved. As shown in **Supplementary Figure 14** and **Supplementary Table 4**, we detected five regions of TAF11 to link with TAF13$^{34\rightarrow DiAzKs}$. One region, TAF11$^{146–149}$, showed marked reduction in linkage in the presence of TBP (Mann–Whitney U test, $P < 0.05$ in both biological replica; **Fig. 3d,e**). In contrast, crosslinks in region TAF11$^{151–155}$ shown in **Figure 3d** stayed largely unaffected, indicating that TBP induces specific conformational dynamics at the interface to the TAF11$^{146–149}$ region when a TAF11–TAF13–TBP complex is formed (a trimeric complex was also confirmed using size-exclusion chromatography; **Supplementary Fig. 15**). Our results hint at different modes of assembly involving TAF11, TAF13, and TBP in the absence of cognate DNA and TFIIA (**Fig. 3f**), and they set the stage for structure–function determination of the TAF11–TAF13–TBP complex in an integrative approach.

In summary, MultiBacTAG combines the advantages of high-level expression of even very large eukaryotic protein assemblies offered by the MultiBac system with efficient site-specific incorporation of functionalized amino acids as a means to engineer and analyze these complexes and their interactions. As the components of the GCE system are inserted into the backbone of MultiBac, the system can be applied readily by the user without prior experience or training in GCE, so existing MultiBac insect cell users should be able to move their system to MultiBacTAG without encountering many hurdles. We showed here a selection of applications for MultiBacTAG, ranging from fluorescence labeling of specific targets to engineering therapeutic protein biologics compatible with human tissue studies and glycoengineering. Additionally, the 99.6%-completed genome of Sf21 presented here will allow for further genetic engineering (e.g., release factor or tRNA expression tuning) of this cell line for protein production using GCE[4–6]. We anticipate that MultiBacTAG in insect cells will enable a wide range of possibilities for custom protein design for biotechnological and pharmaceutical applications, and that it will be useful in the understanding of protein complexes and their functional interactions by unlocking these biological assemblies.

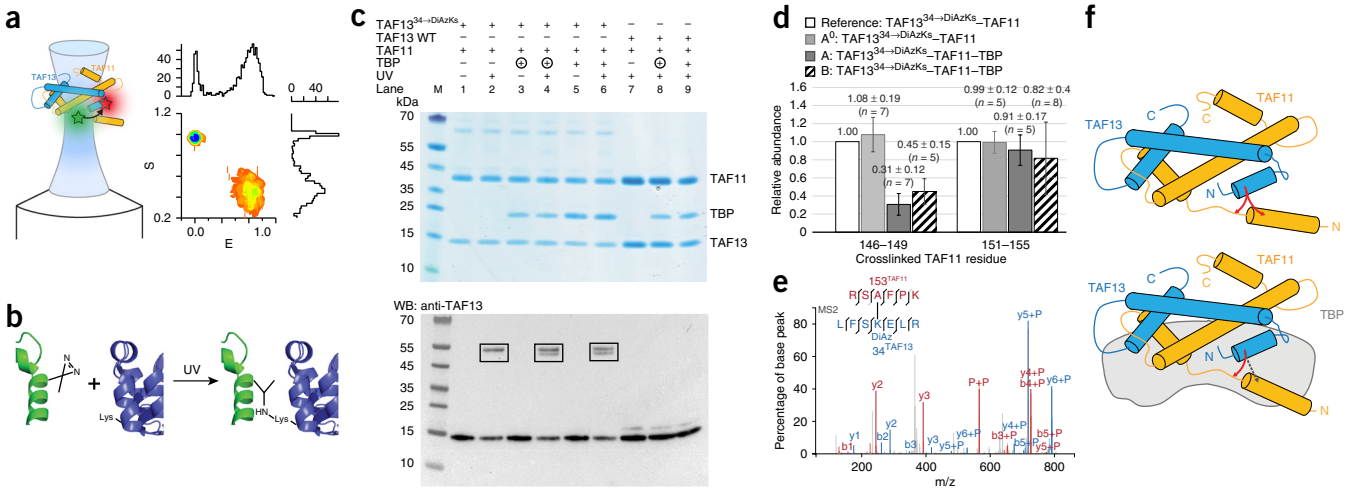

Figure 3 | Crosslinking of TAF11–TAF13–TBP complex. (**a**) A cartoon of the TAF11–TAF13 complex is shown with labeling sites indicated by a green and a red star (donor and acceptor position), as well as FRET efficiency (*E*) versus stoichiometry (*S*) plot revealing a population at $E = 0.8$ (the population around $E = 0$ exists on account of dye photophysics or limited labeling efficiencies). (**b**) Crosslinking between two proteins using DiAzKs and UV light. (**c**) Coomassie-stained SDS–PAGE (top) and the corresponding anti-TAF13 WB of the crosslinked TAF11–TAF13 complex with increasing TBP (1:1:0 (–), 1:1:0.625 (⊕), 1:1:1.25 (+)). (**d**) MS analysis of gel crosslinked products from **c** (analyzed bands boxed schematically in black), revealing two crosslink regions in TAF11 with TAF13$^{34\rightarrow DiAzKs}$. Sample A and B (both TAF11–TAF13$^{34\rightarrow DiAzKs}$–TBP) are biological replica each with their own reference of TAF11–TAF13$^{34\rightarrow DiAzKs}$ without TBP. Relative abundance of crosslinks in presence of TBP were calculated against a reference of TAF11–TAF13$^{34\rightarrow DiAzKs}$ in absence of TBP. To show the variance in the measurements, the reference was also replicated (sample A^0). Center values are the median, error bars show s.d. based on multiple crosslinked peptides; and *n* indicates the number of quantified crosslinked peptides (see **Supplementary Fig. 14** and **Supplementary Table 4** for additional details). (**e**) Annotated high-resolution fragmentation mass spectrum of crosslinked peptide RSAFPK–FLSK$_{DiAz}$ELR, revealing a crosslink between TAF11^{153} and TAF13^{34}. The fragment ion (indicated with b or y) annotated with "+P" contains the crosslinked partner peptide. "P + P" refers to the intact precursor ion. (**f**) Cartoon illustrating that TAF13 (blue) and TAF11 (yellow) form a tight complex (top) yielding two crosslinks (red). Binding to TBP (shown in gray) results in a trimeric complex (bottom) displaying an altered crosslinking pattern (gray dashed arrow).

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. European Nucleotide Archive, PRJEB12116.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all members of our laboratories for helpful discussions. E.A.L., C.K., P.F.S., M.W., and S.B. acknowledge funding from the BW Stiftung. E.A.L. acknowledges additional support from the Emmy Noether program. E.A.L. and C.S. are grateful for funding by SPP1623 of the Deutsche Forschungsgemeinschaft. I.B. is funded by the European Commission Framework Programme 7 (FP7) ComplexINC project (contract no. 279039). P.S.-B. acknowledges funding from the Laura Bassi Centres of Expertise initiative for the Centre of Optimized Structural Studies, project 253275. M.W. thanks the KSOP for financial support. P.S.T. is supported by the EMBL Interdisciplinary Postdoc Programme (EIPOD) under Marie Curie Actions COFUND. The Wellcome Trust generously funded this work through a Senior Research Fellowship to J.R. (103139), a Centre core grant (092076), and an instrument grant (108504). We also thank the members of the EMBL Genomics Core Facility for sample processing and sequencing, as well as the EMBL FACS facility for technical support.

AUTHOR CONTRIBUTIONS

C.K. planned and performed experiments. P.F.S., M.W., M.B.B., S.B., G.E.G., J.J.M.L., M.H.-Y.F., B.G., S.J., P.S.-B., G.P., A.G., H.B., V.B., J.O.K., K.G., I.B., K.R., M.J., J.-E.H., C.S., Z.A.C., J.Z., J.R., and P.S.T. provided critical instrumental and analytical expertise or reagents. C.K., I.B., and E.A.L. cowrote the manuscript with input from all authors. E.A.L. planned experiments and conceived the project.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. *Trends Biochem. Sci.* **37**, 49–57 (2012).
2. Crépin, T. *et al. Curr. Opin. Struct. Biol.* **32**, 139–146 (2015).
3. Fitzgerald, D.J. *et al. Nat. Methods* **3**, 1021–1032 (2006).
4. Lemke, E.A. *ChemBioChem* **15**, 1691–1694 (2014).
5. Liu, C.C. & Schultz, P.G. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
6. Chin, J.W. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
7. Chatterjee, A., Xiao, H., Bollong, M., Ai, H.W. & Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **110**, 11803–11808 (2013).
8. Chen, P.R. *et al. Angew. Chem. Int. Ed. Engl.* **48**, 4052–4055 (2009).
9. Mukai, T. *et al. Biochem. Biophys. Res. Commun.* **371**, 818–822 (2008).
10. Bianco, A., Townsley, F.M., Greiss, S., Lang, K. & Chin, J.W. *Nat. Chem. Biol.* **8**, 748–750 (2012).
11. Mukai, T., Wakiyama, M., Sakamoto, K. & Yokoyama, S. *Protein Sci.* **19**, 440–448 (2010).
12. Hernandez, G. Jr., Valafar, F. & Stumph, W.E. *Nucleic Acids Res.* **35**, 21–34 (2007).
13. Hancock, S.M., Uprety, R., Deiters, A. & Chin, J.W. *J. Am. Chem. Soc.* **132**, 14819–14824 (2010).
14. Nikić, I. *et al. Angew. Chem. Int. Ed. Engl.* **53**, 2245–2249 (2014).
15. Plass, T. *et al. Angew. Chem. Int. Ed. Engl.* **51**, 4166–4170 (2012).
16. Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E.A. *Angew. Chem. Int. Ed. Engl.* **50**, 3878–3881 (2011).
17. Axup, J.Y. *et al. Proc. Natl. Acad. Sci. USA* **109**, 16101–16106 (2012).
18. Kraemer, S.M., Ranallo, R.T., Ogg, R.C. & Stargell, L.A. *Mol. Cell. Biol.* **21**, 1737–1746 (2001).
19. Robinson, M.M. *et al. Mol. Cell. Biol.* **25**, 945–957 (2005).
20. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P.J. & Seidel, C.A.M. *Methods Enzymol.* **475**, 455–514 (2010).
21. Zhang, M. *et al. Nat. Chem. Biol.* **7**, 671–677 (2011).

ONLINE METHODS

Reagents. Unless further indicated, chemicals were purchased from Sigma. Noncanonical amino acids were prepared inhouse, in the case of DiAzKs; and PrK, SCO, TCO*, and BCN were purchased from Sirius Fine Chemicals (SiChem, Bremen; note, now DiAzKs can also be purchased from SiChem). BOC was purchased from IRIS Biotech (Marktredwitz).

Sequencing and analysis of the Sf21 genome. The Sf21 genome was sequenced by Illumina sequencing technology using three types of libraries. Two short-insert paired-end libraries (2 × 104 bp of ~288 bp insert size and 2 × 36 bp of ~590 bp insert size), two long-insert mate-pair libraries (2 × 94 bp and 2 × 101 bp of ~4,500 bp insert size), and one TruSeq Synthetic Long-Read library were generated and sequenced. The data obtained with the last library was assembled into long synthetic reads using the TruSeq Long-Read Assembly app version 1.1 available on BaseSpace (Illumina Inc.). At first the paired-end reads were corrected and filtered with SGA (version 0.9.43; ref. 22). The resulting ~87.2×10^6 read pairs were used as input to perform contig assembly, scaffolding, and gap closing using SOAPdenovo2 (version 2.4; ref. 23). Second, mate-pair reads were processed with FLASH[24] (version 1.2.6), and all overlapping read pairs were discarded. The resulting ~32.4×10^6 pairs were employed with SOAPdenovo2 for scaffolding and then gap closing of the previous assembly. Third, the 18.3×10^4-long synthetic reads were used to scaffold the assembly obtained with paired-end and mate-pair sequencing data. All data types were then finally used for a final gap closing step (SOAPdenovo2).

Eight U6 snRNA genes were found (U6-1–U6-8) using *Bombyx mori* snRNA U6 isoform E gene as query sequence (RefSeq: AY649381.1), with at least 400 bp upstream (promoter region) and 100 bp downstream sequences (termination signal) (**Supplementary Fig. 4**). We decided to work with the U6 promoter and the 3′ termination signal out of the second scaffold (17011_2962_3036_+) that was found, and we called this U6 promoter U6(Sf21)-2.

MultiBac baculovirus system for transduction of insect cells. *Construction of amber suppressor genomes (MultiBacTAG^{WT}, MultiBacTAG^{AF}).* We generated a baculoviral genome that contained the genes encoding for both the synthetase and the tRNA for amber suppression by using Cre-recombinase-mediated insertion into the LoxP present on the MultiBac viral backbone (**Fig. 1**). Thus, the attachment site for Tn7 transposition (mini-attn7) remained fully accessible to accept multigene constructs of target proteins and their complexes. We inserted the expression cassette U6(Sf21)-2-tRNA^{Pyl}-3′ term into the pUCDM donor plasmid module by using ClaI and XbaI restriction enzymes. Next, by means of NsiI and XhoI digestion and ligation we added the MM PylRS or MM PylRS AF into the p10-driven expression cassette, giving rise to MultiBacTAG^{WT} and MultiBacTAG^{AF} viral genomes, respectively. For all cloning steps of the pUCDM plasmid, BW23474 cells were used to provide the Pir$^+$ background required by the conditional origin present on the donor[3]. The resulting dual-expression plasmid pUCDM-U6(Sf21)-2-tRNA^{Pyl}-3′ term-PylRS was transformed into electrocompetent DH10MultiBac^{Cre} cells following established protocols[3,25]. Tetracycline antibiotic challenge was applied during all transformation steps to ensure maintenance of the pHelper plasmid, which encodes for the Tn7 transposase and is required for inserting multigene constructs encoding for target proteins. Cell stocks were validated by preparing composite baculoviral genomes from eight blue colonies each and by transfection of Sf21 cells. V_0 virus was harvested after 60 h of incubation, and the V_1 generation was started. V_0 virus is the primary virus stock and is used to prepare the V_1 virus, which is utilized for protein expression. Cells were harvested 60 h after proliferation arrest[3]. Cell pellets were resuspended in 4× phosphate-buffered saline (PBS, pH 8), resulting in 1 million cells/ml. Glycerol stocks of cells containing MultiBacTAG^{WT} and MultiBacTAG^{AF} were prepared, and from those stocks electrocompetent cells were prepared following standard protocols, and they were stored at −80 °C.

Plasmids. *Reporter plasmids.* First, a reporter plasmid was constructed. GFP(Y39TAG)–6His and mCherry–GFP(Y39TAG)–6His were separately cloned into acceptor pACEBacDual plasmid under the polyhedrin (polh) promoter using BamHI and PstI restriction enzymes. The resulting pACEBac-Dual–GFP(Y39TAG)–6His and pACEBac-Dual–mCherry–GFP(Y39TAG)–6His acceptors were transformed into cells containing MultiBacTAG^{WT} and MultiBacTAG^{AF}, respectively, for integration into the Tn7 attachment site.

Herceptin. Synthetic genes encoding for the variable and constant regions of the heavy and light chain of the Herceptin Fab fragment (for simplification referred to as Herceptin in the whole manuscript) were codon optimized for insect cell expression and inserted into pACEBac-Dual acceptor into the polh- and p10-driven expression cassettes, respectively. A C-terminal 6His tag was fused to the Herceptin heavy chain. Two individual amber mutations were inserted at positions A121 and A132 of the heavy chain.

TAF11–TAF13–TBP complex. pFastBac-Dual–6His–TAF11–TAF13 was constituted from pFastBac-Dual by inserting the genes encoding for human TBP-associated factors 11 (TAF11) and 13 (TAF13) into the polh- and p10-driven expression cassettes. TAF11 contains an N-terminal 6His tag followed by a tobacco etch virus (TEV) NIa protease site. Two Amber stop codons were introduced separately into the TAF13 gene at positions A20 and K34. Human TATA-box binding protein (TBP) core (residues 155–333) was cloned into pET28aHis plasmid, resulting in a 6His tag at the N-terminal domain of TBP (pET28aHis-TBP was a generous gift of T.J. Richmond, ETH, Zurich).

Cell culture. *Sf21.* Following standard protocols[26], Sf21 cells were cultured in Erlenmeyer flask at 27 °C with shaking at 180 r.p.m., using Sf-900 III SFM medium (Thermo Fisher Scientific) at the Protein Expression and Purification core facility (PEPcore) at EMBL, Heidelberg. Cells were split every day to 0.6×10^6 cells/ml or every third day to 0.3×10^6 cells/ml. For bacmid transfection, 3 ml per well of 0.3×10^6 cells/ml were seeded in a six-well multidish (Nunclon Delta Surface, Thermo scientific). Bacmid DNA was prepared and Sf21 cells transfected using FuGENE HD Transfection Reagent (Promega). V_0 virus was harvested after 70 h post -transfection and the V_1 generation started. For small-scale test expression, 100 ml of Sf21 cells at 0.6×10^6 cells/ml was transfected with 0.1 ml of V_1 virus, and the respective ncAA was added to a final concentration of 1 mM. As negative control, a 100 ml culture was set up the same way, but without ncAA. After cell proliferation stopped, the cultures were kept another 48–60 h at 27 °C with shaking at 180 r.p.m. The cells were harvested at 500 r.p.m. for 10 min, and the pellets were stored at −20 °C.

Flow cytometry analyses. Flow cytometry analyses were done on a BD LSRFORTESSA (BD Biosciences). Therefore, Sf21 cells were transduced with the corresponding virus in a six-well multidish. After 3 d of incubation time, the cells were harvested at 500 r.p.m. for 10 min at 4 °C and resuspended in 500 µl sterile 1× PBS. The suspension was filtered through a cell strainer (Falcon, 70 µm, Fisher Scientific) and kept on ice until measurements. Data from 500,000 cells for each sample were acquired and analyzed with FlowJo X software (FlowJo Enterprise).

Protein expression and purification. *GFP(Y39TAG) & mCherry–GFP(Y39TAG).* The plasmids pACEBac-Dual–GFP(Y39TAG)–6His and pACEBac-Dual–mCherry–GFP(Y39TAG)–6His were transformed into cells containing MultiBacTAG (WT and AF variants) and plated on agar plates containing X-Gal and IPTG (for blue–white selection) as well as ampicillin (100 µg/ml), kanamycin (30 µg/ml), tetracycline (10 µg/ml), and gentamycin (10 µg/ml). Four white colonies of each construct were picked and composite baculoviral DNA prepared. After transfecting Sf21 cells, the four V_0 virus preparations were harvested after 60 h. V_1 virus was produced using all four V_0 viruses in parallel, and 0.1 ml of each virus was added to 100 ml of fresh Sf21 cells. Five cultures were set up in the same way, one for each of the four V_1 viruses, in which ncAA at a final concentration of 1 mM was added; and one culture was set up without ncAA as a negative control. After cell propagation stopped, the cells were harvested after an additional 48–60 h.

For purification, cell pellets were resuspended in 4× PBS (5 mM imidazol, 0.2 mM TCEP, 1mM PMSF) and centrifuged at 40,000 r.p.m. at 4 °C using a Beckman ultracentrifuge (SW Ti60 rotor) after sonication. The cleared lysate was incubated on Ni beads for 1-2 h at 4 °C. Immobilized metal ion affinity chromatography (IMAC) was carried out by washing with 10 mM imidazol in 4× PBS (0.2 mM TCEP and 1 mM PMSF), followed by an elution step using 500 mM imidazol in the same buffer. Finally the elution fraction was analyzed by SDS–PAGE and stored at –20 °C.

Herceptin. For the expression of Herceptin, the plasmid pACEBac-Dual–Herceptin–6His was transformed in both MultiBacTAG^{WT}- and DH10MultiBacTAG^{AF}-containing cells. Expression and purification was carried out following the same steps as described above for GFP(Y39TAG).

TAF11–TAF13 complex. For producing TAF11–TAF13 complex, MultiBacTAG^{AF} was used for both wild-type TAF11–TAF13 complex as well as for the amber mutants (see above). Again, the same protocol was followed as described above for GFP(Y39TAG).

The cell pellet was resuspended in 150 ml Tris buffer (25 mM Tris, 150 mM NaCl, 5 mM imidazol, 1 mM PMSF, pH 8) per 1 l expression culture. After sonication, the insoluble fraction was spun down at 40,000 r.p.m. at 4 °C (Beckman SWTi60 rotor). The supernatant was incubated on nickel beads for 1–2 h, and the protein was eluted after several washing steps with increasing imidazol concentrations. To finalize the IMAC purification procedure, the protein was further purified by size exclusion chromatography (SEC) using a Superdex column, equilibrated beforehand with Superdex running buffer (25 mM Tris, 300 mM NaCl, 1 mM EDTA, 1mM DTT, pH 8) and analyzed by SDS–PAGE.

TATA-box binding protein (TBP), residues 155–333. pET28aHis–TBP was transformed into BL21(DE3) Rosetta cells and expressed in LB medium at 18 °C overnight. Cells were harvested by centrifugation (4,500 r.p.m. for 20 min at 4 °C) and stored at –20 °C.

The cells of 1 l expression culture were lysed in 20 ml TBP lysis buffer (25 mM Tris, 1 M NaCl, 10 mM imidazol, 1 mM PMSF, pH 8) using a sonicator. After spinning down the insoluble fraction, the cleared supernatant was purified by IMAC. Washing was done with increasing concentration of imidazol, and the protein was finally eluted. After loading the protein on a Superdex column, which was equilibrated with Superdex running buffer, the purity was checked by SDS–PAGE analysis.

Single-molecule FRET experiments. Dual-labeled TAF11–TAF13$^{20\rightarrow A488,\ 37\rightarrow A594}$ complexes were diluted to ~100 pM and subjected to multiparameter single-molecule FRET (smFRET) spectroscopy on a custom-built confocal detection setup as detailed previously[27]. In brief, the sample was excited through a 1.2 numerical aperture (NA) 63× Olympus objective with alternating LASER pulses from a 485 LDH diode Laser and an 570 nm filtered while light LASER (Koheras). Emission signal was split into green and orange color channels and detected on photon-counting diodes (MPD and APD), directed to Hydraharp (Picoquant) counting electronics, and analyzed further using IgorPro (Wavemetrics) as previously detailed[27]. The signal intensities were analyzed according to the following equations, with I_A and I_D being the recorded photon counts during donor Laser excitation and I_A^{dir} the intensity of the acceptor during acceptor LASER excitation. The plot shown in **Figure 4a** shows a 2D E_{FRET} versus S plot. At $E = 0$ and $S = 1$ sites, the so-called 'Zero' peak arises from inactive acceptor and is not of relevance in this analysis. From the known γ (a correction factor for the apparent brightness of our dye pair) and the known distance at which FRET from donor dye to acceptor dye is 50% efficient (R_0) for our dye pair[28], we can estimate that the measured FRET intensity corresponds to an approximate distance (r) of around 30 Å.

$$E_{FRET} = \frac{I_A}{\gamma I_D + I_A} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}; S = \frac{I_A + I_D}{I_D + I_A + I_A^{dir}}$$

Crosslinking experiments. *Western blot analysis of crosslinked samples.* The crosslinking reactions contained 40 µM of TAF11–TAF13 complex. TBP was added in two different molar ratios to the reaction. The first ratio was 1:1:0.625, corresponding to TAF11:TAF13:TBP. For this ratio, we used 12.5 µM of TBP. The second ratio was 1:1:1.25, which resulted in 25 µM of TBP per reaction. For each crosslinking experiment, we set up 20 µl reactions containing the respective proteins in Superdex running buffer and incubated the reactions on ice for 2 h. These reactions were then split into 2 × 10 µl, and one of the 10 µl reactions was exposed to UV light. UV irradiation was performed for 15 min on ice using a 345 nm filter with an approximately 40 cm distance to the 1,000 W lamp. The crosslinking experiments were performed with a TAF13$^{34\rightarrow DiAzKs}$ mutant.

For preparing the samples for SDS–PAGE, 5 µl of each reaction was mixed with 35 µl Superdex running buffer and 10 µl 5× SDS loading dye, then the samples were heated up for 1 min at 95 °C. 15 µl of these samples was loaded in a well of a ten-well SDS–PAGE (NuPAGE 4–12% Bis-Tris, Thermofisher). After running the gels using MES buffer, the gels were plotted using the Trans-Blot

doi:10.1038/nmeth.4032

Turbo Transfer system (Bio-Rad). With the Trans-Blot Turbo Mini Nitrocellulose Transfer Packs (Bio-Rad) the transfer was done in 7 min, and the membranes were blocked for 1 h at room temperature with 5% milk in 1× PBS. The primary antibodies (anti-TAF13 (Abcam, ab201090), anti-TBP (kind gift from L. Tora, IGBMC, France) and anti-Flag (Monoclonal Antibodies Core Facility, EMBL)) were diluted 1:1,000 (for anti-TAF13) and 1:2,000 (for anti-TBP and anti-Flag) in 5% milk, 1× PBS; and the membrane was incubated over night at 4 °C. After three washes with 1× PBS, 0.2% Tween 20, the secondary antibody was incubated for 1 h at room temperature. For the anti-TAF13, an anti-rabbit secondary antibody (Peroxidase AffiniPure goat anti-rabbit IgG (H + L), Jackson ImmunoResearch, 111-035-003) was used in a 1:5,000 dilution in 1× PBS, 0.2% Tween 20 and for the anti-TBP and anti-Flag antibodies an anti-mouse secondary antibody was diluted 1:10,000 in 1× PBS, 0.2% Tween 20 (Amersham ECL Mouse IgG, HRP-linked whole Ab (from sheep), GE Healthcare, NA931-1ML). After three more washes with 1× PBS, 0.2% Tween 20, a chemiluminescence kit (ECL Western Blot reagent, GE Healthcare) in combination with a Chemidoc Touch system (Biorad) was used to visualize the western blot signal.

Sample preparation for mass spectrometric analysis. For mass spectrometric analysis, the crosslinking reaction was set up in the ratio 1:1:1.25, corresponding to TAF11:TAF13:TBP. The $TAF13^{34\rightarrow DiAzKs}$ crosslinked samples were prepared in replicates as given in the text (**Fig. 3**). For each reaction 40 µM of TAF11–TAF13 complex was mixed with 25 µM of TBP in a 30 µl reaction volume, incubated on ice, crosslinked by UV light (15 min, 345 nm filter, 1,000 W lamp) and loaded on an SDS–PAGE gel. 1.5 µl of each reaction was loaded on the same gel in a separate well and was used to identify the crosslinked species by western blot. The gel bands of crosslinked TAF11–TAF13 complexes were excised, in-gel reduced and alkylated, then digested using trypsin following a standard protocol[29]. The peptide mixture was then desalted using C18-Stage-Tips[30] for mass spectrometric analysis.

Mass spectrometric analysis. LC–MS/MS analysis was performed using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific), applying a 'high-high' acquisition strategy. Peptides were separated on a 75 µm × 50 cm PepMap EASY-Spray column (Thermo Scientific) fitted into an EASY-Spray source (Thermo Scientific), operated at 50 °C column temperature. Mobile phase A consisted of water and 0.1% v/v formic acid. Mobile phase B consisted of 80% v/v acetonitrile and 0.1% v/v formic acid. Peptides were loaded at a flow rate of 0.3 µl/min and eluted at 0.2 µl/min using a linear gradient going from 2% mobile phase B to 4% mobile phase B over 139 min, followed by a linear increase from 45% to 95% mobile phase B in 11 min. The eluted peptides were directly introduced into the mass spectrometer. MS data were acquired in the data-dependent mode with the top-speed option. For each 3 s acquisition cycle, the survey level spectrum was recorded in the Orbitrap with a resolution of 120,000. The ions with a precursor charge state between 3^+ and 8^+ were isolated and fragmented using high-energy collision dissociation (HCD). Precursor priority for fragmentation was set to 'highest charge state' then 'most intense'. The fragmentation spectra were recorded in the Orbitrap with a resolution of 15,000. Dynamic exclusion was enabled with single repeat count and 60 s exclusion duration.

Identification of crosslinked peptides. The raw mass spectrometric data files were processed into peak lists using MaxQuant version 1.5.3.30 (ref. 31) with default parameters, except that 'FTMS top peaks per 100 Da' was set to 100, and 'FTMS de-isotoping' was disabled. The peak lists were searched against the sequences as well as the reversed sequences (decoy) of TAF11 and $TAF13^{34\rightarrow DiAzKs}$ using Xi software (ERI, Edinburgh) for identification of crosslinked peptides and noncrosslinked linear peptides. In the protein sequences, DiAzKs was represented as 'Xd'. Search parameters were as follows: MS accuracy, 6 p.p.m.; MS2 accuracy, 20 p.p.m.; enzyme, trypsin; specificity, fully tryptic; allowed number of missed cleavages, four; fixed modifications, carbamidomethylation on cysteine; variable modifications, oxidation on methionine. The crosslinking reactivity of DiAzKs is toward any other amino acid residues. All fragmentation spectra of all identified crosslinked residue pairs were validated manually. In addition, we identified linear peptides from TAF11 and TAF13. Linear peptides with Xi score above 7 were used for quantitation to estimate the relative protein abundance in each sample.

Quantitation of crosslinking data using Pinpoint software. Identified crosslinked peptides and selected linear peptides were quantified based on their MS1 signals. The quantitative proteomics software tool Pinpoint (Thermo Fisher Scientific) was used to retrieve intensities for each crosslinked and linear peptide[32]. To construct the input library of Pinpoint, the sequence of every crosslinked peptide was converted into a linear version with identical mass[33]. The five most abundant signals in the isotope envelope were used for quantitation. The error tolerance for precursor m/z was set to 6 p.p.m. Signals were only accepted within a window of retention time (defined in the spectral library) ± 10 min. Manual inspection was carried out to ensure the correct isolation of elution peaks. 'Match between runs'[34] was carried out for all crosslinked peptides in Pinpoint interface manually, based on high mass accuracy and reproducible LC retention time.

The signal intensities of crosslinked peptides were normalized against abundance of TAF13, which was calculated as summed signal intensities of seven linear peptides. The relative abundance of crosslinks in samples with and without TBP was compared.

Statistics. QCLMS analysis was repeated in two separate experiments. In experiment I, three samples were analyzed: two $TAF13^{34\rightarrow DiAzKs}$ + TAF11 samples (reference and A^0) and one $TAF13^{34\rightarrow DiAzKs}$ + TAF11 + TBP sample (A). In experiment II, two samples were analyzed: one $TAF13^{34\rightarrow DiAzKs}$ + TAF11 sample (reference) and one $TAF13^{34\rightarrow DiAzKs}$ + TAF11 + TBP sample (B).

The TAF11 residues that were crosslinked to DiAzKs fall into five regions. For each sample, the relative intensity of crosslinks to each region was calculated as the median of all their supporting crosslinked peptides. The numbers of supporting crosslinked peptides (n) for crosslinkages to each TAF11 region are listed in **Figure 3d** and **Supplementary Figure 14**.

Click reactions. *Copper-catalyzed alkyne-azide cycloaddition (CuAAC).* Purified protein, which contains an ncAA (Propargyllysine, PrK) with an alkyne group incorporated at the amber stop codon side, was exchanged to 1× PBS buffer, pH 7.5 (0.2 mM TCEP), and 5 nmol were used for the click reaction, following the protocol as described in ref. 35. Cycloaddition reactions were followed up by SDS–PAGE.

Strain-promoted alkyne-azide cycloaddition (SPAAC). Protein, expressed in the presence of 1 mM of BCN (Sichem), was

purified and exchanged into 1× PBS buffer, pH 8. For the labeling reaction, 2 nmol of protein mixed with 100 nmol of glycan-azide (PSZ170) were incubated over night at room temperature[17]. Labeling reactions were loaded on a Superdex column and analyzed by SDS–PAGE.

Strain-promoted Diels–Alder cycloaddtion (SPDAC). Protein, expressed in the presence of 1 mM of SCO (Sichem) or TCO* (Sichem), was purified and exchanged into 1× PBS buffer, pH 8. For the labeling reaction, 1 nmol of protein containing SCO mixed with 5 nmol of TAMRA-Tetrazine (Jena Bioscience) was incubated for 1 h at room temperature[16]. In the case of protein harboring TCO*, 5 nmol of protein was used in a reaction with 50 nmol of Tetrazine-5-TAMRA. Labeling reactions were loaded on a Superdex column and analyzed by SDS–PAGE.

Immunofluorescence analysis. Tissue sections were processed for immunofluorescence staining and incubated with Herceptin$^{121\rightarrow TCO^*}$-TAMRA-labeled antibody (diluted 1:100) overnight at 4 °C, washed in PBS, and mounted in ProLong Gold antifade with DAPI (Invitrogen). Images were obtained on a Leica TCS SP5, LAS AF version 2.7.3.9723 (Leica Microsystems CMS GmbH). Objective: HCX PL APO lambda blue 63.0×/1.40 OIL UV.

Human tissue samples. The European Institute of Oncology (IEO) Division of Biostatistics selected from its institutional database consecutive breast cancer (BC) patients fulfilling the following criteria: (i) histologically proven invasive BC treated by neoadjuvant therapy; (ii) any age (premenopausal or postmenopausal status allowed); (iii) any intrinsic subtype (Luminal A/B-like, Her-2 positive, Triple Negative subtypes allowed). All the patients prospectively entered the IEO BC database and were discussed at the weekly multidisciplinary meeting. Data on patients' medical history, concurrent diseases, surgery, pathological evaluation, radiotherapy, neoadjuvant systemic treatments, and clinicopathological results of preneoadjuvant and postneoadjuvant treatment staging procedures were retrieved. All the biopsies were fixed in 4% buffered formalin for less than 24 h immediately after the core biopsy procedure. All the surgical samples were fresh sampled in accordance to the criteria issued by Provenzano *et al.*[36] and fixed in 4% buffered formalin for less than 24 h. All the biopsies and surgical samples were routinely processed and embedded in paraffin. Detailed information regarding tumor type and grade, ER/PgR and Her-2 status, and K_i-67 labeling index were available in all the cases. ER/PgR and HER2 immunoreactivity were assessed in line with the clinical practice procedures applicable at diagnosis. Her-2 immunoreactivity was assessed using the monoclonal antibody CB11 (Novocastra, 1:800) from 1995 to 2005 and using the HercepTest (Dako) thereafter. Cases classified as Her-2 2^+ by immunohistochemistry were tested by FISH analysis with Vysis probes, in accordance with the ASCO/CAP guidelines[37]. K_i-67 labeling index was assessed by the Mib-1 monoclonal antibody (Dako, 1:200) by counting at least 500 invasive tumor cells, independent of their staining intensity and without focusing on hotspots[38]. Tumors were classified as Luminal A-like (ER and PgR positive, absence of Her-2 overexpression and K_i-67 < 20%), Luminal B-like (ER positive, Her-2 negative and at least one of K_i-67 ≥ 20% and PgR < 20%), Luminal B-like/Her-2 positive (ER and Her-2 positive, any PgR and K_i-67), Her-2 positive (Her-2 3^+ and/or amplified by FISH, ER/PgR negative), and Triple Negative (ER, PgR and Her-2 negative) in accordance with St. Gallen recommendations[39]. For tumor-specific information please refer to **Supplementary Table 3**. All the patients included gave an informed consent for using their clinicopathological data and samples for research purposes at the time of admission to the hospital, and the study was approved by the IEO Review Board.

22. Simpson, J.T. & Durbin, R. *Genome Res.* **22**, 549–556 (2012).
23. Luo, R. *et al. Gigascience* **1**, 18 (2012).
24. Magoč, T. & Salzberg, S.L. *Bioinformatics* **27**, 2957–2963 (2011).
25. Berger, I., Fitzgerald, D.J. & Richmond, T.J. *Nat. Biotechnol.* **22**, 1583–1587 (2004).
26. Nie, Y., Bieniossek, C. & Berger, I. *ACEMBL Expression System User Manual* version 09.11 (EMBL, 2009).
27. Milles, S. & Lemke, E.A. *Biophys. J.* **101**, 1710–1719 (2011).
28. Milles, S. *et al. J. Am. Chem. Soc.* **134**, 5187–5195 (2012).
29. Maiolica, A. *et al. Mol. Cell. Proteomics* **6**, 2200–2211 (2007).
30. Rappsilber, J., Ishihama, Y. & Mann, M. *Anal. Chem.* **75**, 663–670 (2003).
31. Cox, J. & Mann, M. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
32. Tomko, R.J. Jr. *et al. Cell* **163**, 432–444 (2015).
33. Chen, Z.A., Fischer, L., Cox, J. & Rappsilber, J. *Mol. Cell. Proteomics* **15**, 2769–2778 (2016).
34. Thakur, S.S. *et al. Mol. Cell. Proteomics* **10**, M110.003699 (2011).
35. Tyagi, S. & Lemke, E.A. *Methods Cell Biol.* **113**, 169–187 (2013).
36. Provenzano, E. *et al. Mod. Pathol.* **28**, 1185–1201 (2015).
37. Wolff, A.C. *et al. J. Clin. Oncol.* **31**, 3997–4013 (2013).
38. Polley, M.Y. *et al. J. Natl. Cancer Inst.* **105**, 1897–1906 (2013).
39. Goldhirsch, A. *et al. Ann. Oncol.* **24**, 2206–2223 (2013).

doi:10.1038/nmeth.4032