JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

An Intrinsically Disordered Peptide-Peptide Stapler for Highly Efficient Protein Ligation Both in vivo and in vitro

Xia-Ling Wu, Yajie Liu, Dong Liu, Fei Sun, and Wen-Bin Zhang

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.8b08250 • Publication Date (Web): 19 Nov 2018

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

An Intrinsically Disordered Peptide-Peptide Stapler for Highly Efficient Protein Ligation Both *in vivo* and *in vitro*

Xia-Ling Wu,[§,1] Yajie Liu,[§,1] Dong Liu,[1] Fei Sun,[2] Wen-Bin Zhang*[1]

[1] Key Laboratory of Polymer Chemistry & Physics of Ministry of Education, Center for Soft Matter Science and Engineering, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P. R. China

[2] Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, P.R. China

KEYWORDS SpyStapler, SpyTag, SpyCatcher, BDTag, Ligase

ABSTRACT: Herein, we report an intrinsically disordered protein SpyStapler that can catalyze the isopeptide bond formation between two peptide tags, i.e. SpyTag and BDTag, both *in vitro* and *in vivo*. SpyStapler and BDTag are developed by splitting SpyCatcher—the cognate protein partner of SpyTag—at the more solvent exposed second loop region. Regardless of their locations in protein constructs, SpyStapler enables efficient covalent coupling of SpyTag and BDTag under a variety of mild conditions *in vitro* (yield ~80%). Co-expression of SpyStapler with telechelic dihydrofolate reductase (DHFR) bearing a SpyTag at N-terminus and a BDTag at C-terminus leads to direct cellular synthesis of a circular DHFR. Mechanistic studies involving circular dichroism (CD) and nuclear magnetic resonance (NMR) spectrometry reveal that SpyStapler alone is disordered in solution and forms a stable folded structure (T_m ~55 °C) in the presence of both SpyTag and BDTag upon isopeptide bonding. No ordered structure can be formed in the absence of either tag. The catalytically inactive SpyStapler-EQ mutant cannot form a stable physical complex with SpyTag and BDTag, but it can fold into ordered structure in the presence of the ligated product (SpyTag-BDTag). It suggests that the isopeptide bond is important in stabilizing the complex. Given its efficiency, resilience, and robustness, SpyStapler provides new opportunities for bioconjugation and creation of complex protein architectures.

Introduction

Peptide/protein tagging offers powerful ways to manipulate proteins in various contexts, such as protein purification, translocation, topology engineering, and bio-imaging.[1-4] The recent emergence of "molecular superglue" techniques provides a new tagging strategy with essentially infinite affinity, as demonstrated by the SpyTag/SpyCatcher,[5] SnoopTag/SnoopCatcher reactive pairs,[6] as well as their variants.[7-12] They are based entirely on natural amino acids and are thus genetically encodable.[13-14] Incorporating latent bio-reactive unnatural amino acids represents another way to achieve covalent tagging or bioconjugation via proximity-enhanced reaction.[15] It can be readily achieved both in vitro and in vivo in various cells, leaving only a tiny residue after ligation.[16] By contrast, genetically encoded peptide-protein reactive pairs often forms a complex after ligation which may complicate the study of structure-property relationship.

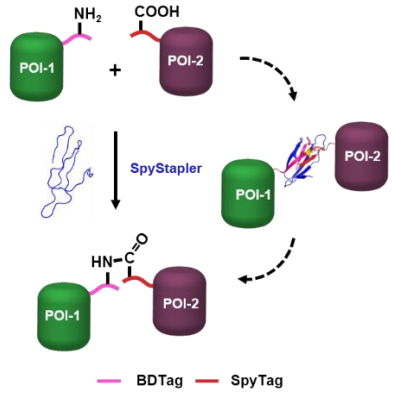
To reduce the sizes of the protein partner (the "Catcher" component), a cleavable variant, SpyCatcher-N^{TEV}, has been successfully developed *via* circular permutation with high reactivity.[11] By splitting SpyCatcher at the first loop, Howarth and co-workers developed a peptide-peptide ligation technique mediated by SpyLigase.[17] With the addition of excess SpyLigase (> 1 *eq*) as well as 1.5 M trimethylamine *N*-oxide (TMAO) as the promoter, it can achieve moderate ligation efficiency (~50%). They further developed a second generation, more effective SnoopLigase/DogTag/SnoopTagJr system via dissecting the RrgA adhesin domain followed by extensive efforts of optimization.[18] Even in the absence of TMAO, the SnoopLigase system can achieve high ligation efficiency (>60% at a 1:1:1 ratio of SnoopTagJr, DogTag and SnoopLigase) with certain amounts of glycerol. SnoopLigase also shows remarkable resilience toward thermal inactivation at 100 °C. It thus represents an important addition to the existing bioconjugation toolbox that includes split intein technology,[19] sortase,[20-21] butelase 1,[22-23] and OaAEP1,[24] etc. Notably, SnoopLigase was developed by improving its conformational rigidity through structure-guided optimization, assuming that a stable folded structure would lead to enhanced catalytic efficiency. Nonetheless, its solution state remains to be elucidated and its performance in cells remains to be demonstrated.

Intrinsically disordered proteins (IDPs) refer to a unique class of proteins without a stable secondary structure under physiological conditions.[25-27] While most proteins assume a pre-existing ordered conformation toward specific functions, IDPs also turn out to be functional and can serve as active enzymes under physiological conditions. It thus prompts scientists to re-assess the structure-property relationship of proteins and ponder on how structure and dynamics modulate function.[28-29] Our previous works show that unlike wild-type SpyCatcher that is folded in solution,[30-31] the supercharged SpyCatcher(-)[8] and the circularly permuted SpyCatcher-N^{TEV} [11] are disordered in solution and remain reactive with SpyTag, giving a folded complex

upon isopeptide bond formation. Hence, we hypothesized that, counterintuitively, ligases may not necessarily have to adopt a well-folded structure in solution to achieve high ligation efficiency.

In this article, we report SpyStapler, an unstructured ligase that recognizes two peptide tags, i.e. SpyTag and BDTag, and catalyzes the isopeptide bond formation between the two both *in vitro* and *in vivo* in high efficiency (Scheme 1, Figure 1). SpyStapler is extremely stable *in vitro* and withstands various inactivation treatments (such as denaturation by urea or heat inactivation for extended times). It also enables nearly quantitative protein cyclization inside cells when co-expressed with the telechelic protein construct SpyTag-DHFR-BDTag. The CD and NMR experiments reveal that the SpyStapler alone is disordered, exemplifying the potential for engineering a functional intrinsically disordered ligase.

Scheme 1. SpyStapler-mediated ligation[a]



[a] Both SpyTag and BDTag can be fused to the proteins of interest (POIs) at the N-terminus, C-terminus, or in the middle of the chain for ligation reaction.

Results and Discussion

Design of SpyStapler. Previously, Howarth and co-workers have demonstrated SpyLigase-mediated ligation between SpyTag and KTag, though with relatively low efficiency.[17] Structure-based computational design has subsequently proven useful in improving the efficiency of SnoopLigase based on the premise that a stable folded structure is essential for high catalytic efficiency.[18] Inspired by the intrinsically disordered SpyCatcher variants that remain active toward SpyTag, we envisioned that a pre-existing folded structure might not be essential for such ligases. A visual inspection over the crystal structure of the SpyTag-SpyCatcher complex (PDB code: 4MLI)[32] revealed the second loop distal from the catalytic triad as the most malleable and plastic region (Figure 1a). We have previously performed circular permutation on SpyCatcher and converted this region to the new N- and C-termini of the SpyCatcher-N^{TEV} variant. Given the high efficiency of the SpyTag/SpyCatcher-N^{TEV} reaction with a second-order rate constant of ~10 $M^{-1}\cdot s^{-1}$,[11] we speculated that the catalytic domain, corresponding to the segment from the third strand to the last strand of SpyCatcher (residue 55-116, named SpyStapler), could be designed as a standalone catalyst to covalently stitch together SpyTag and the remaining reactive fragment of SpyCatcher (residue 27-54, named BDTag) (Figure 1b, c). The name "SpyStapler" comes from the resemblance of its folded shape within the SpyTag-SpyCatcher complex to a real stapler (Figure 1b).

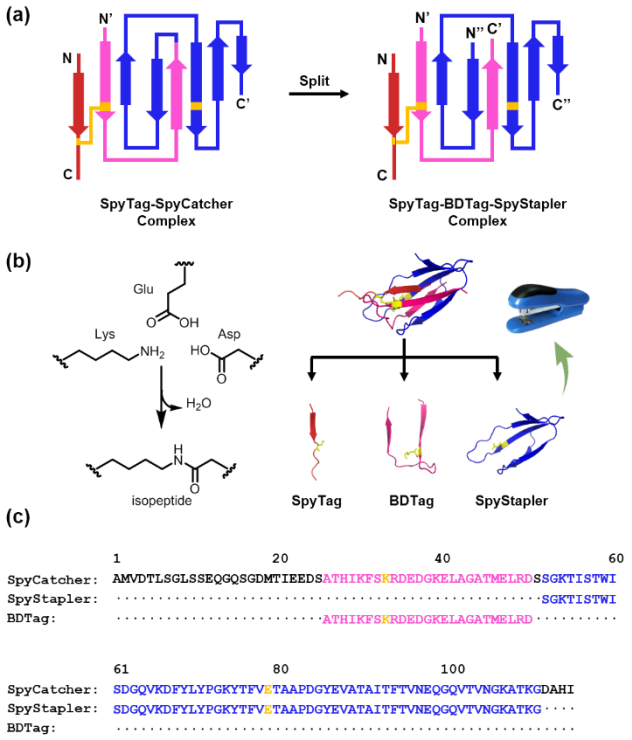


Figure 1. (a) Topology diagram of the SpyTag-SpyCatcher complex, showing the splitting into a SpyTag-BDTag-SpyStapler complex. (b) Reaction between the aspartic acid residue of SpyTag and the lysine residue of BDTag under the catalysis of the glutamic acid residue of SpyStapler. The 3D structure of SpyTag, BDTag and SpyStapler are drawn based on the crystal structure of the SpyTag-SpyCatcher complex (PDB code: 4MLI). (c) Amino acid sequence alignment of SpyCatcher, SpyStapler and BDTag. The SpyTag is colored red, BDTag magenta and SpyStapler blue. The aspartic acid, the lysine and the glutamic acid are colored yellow.

Synthesis and Characterization of SpyStapler. The gene of SpyStapler was cloned from SpyCatcher into pET28a expression vector (for sequences, see Figure S1). The plasmid was transformed into *Escherichia coli* (*E. coli*) BL21 (DE3) strain for expression in LB broth. The product could be purified under either native or denaturing conditions (~5 mg per liter of culture). The SDS-PAGE shows a single band at an apparent molecular weight of 8 kDa. LC-MS further confirms the identity of the SpyStapler with the correct observed mass of 7864 Da versus the calculated value of 7863 Da (Figures 2a). The +178 Da modification was caused by the gluconoylation of N-terminal amino group. The SEC traces for the samples purified under native or denaturing conditions are different. The former exhibits a bimodal elution profile and the latter shows only one monomodal peak (Figure 2b). Even for the former, both peaks are actually identical in SDS-PAGE, indicating that the peak at lower retention volume may be due to aggregation. Denaturing treatment eliminates the aggregation between SpyStapler and gives only one single peak in the elution profile.

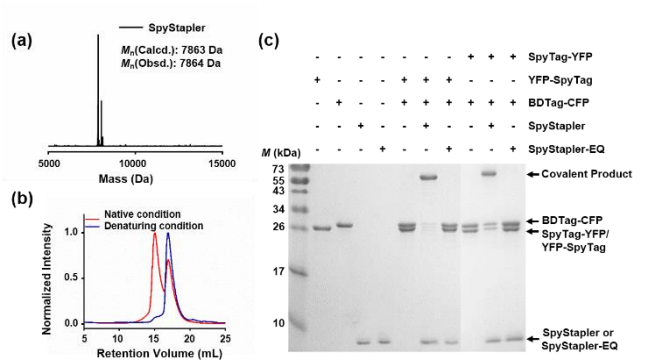


Figure 2. (a) LC-MS spectrum of SpyStapler. (b) SEC overlay of SpyStapler purified under native (red) or denaturing conditions (blue). (c) SDS-PAGE analysis of the reactions between BDTag-CFP and YFP-SpyTag (or SpyTag-YFP) mediated by SpyStapler (or SpyStapler-EQ as a negative control).

SpyStapler Mediates SpyTag-BDTag Ligation. To test the efficiency of the reaction, we designed four telechelic proteins based on cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) by fusing the tags at either N- or C- terminus. They are BDTag-CFP, CFP-BDTag, SpyTag-YFP, and YFP-SpyTag, respectively (see Figure S1 for sequences). Their identities were confirmed by LC-MS (Figure S2). The reactions were run in different combinations with equimolar amounts of reactants in the presence of two equivalents of SpyStapler for 8 hours at 4 °C. The SDS-PAGE clearly shows a new band at much higher position and the densitometry analysis reveals a typical yield of ~80% (Figures 2c and S3). Remarkably, it was achieved without TMAO or glycerol. The control experiments without SpyStapler or with the catalytically inactive EQ mutant of SpyStapler (SpyStapler-EQ, see Figure S1 for sequence) show no product formation (Figure 2c), confirming the catalytic role of SpyStapler. Although the yields seem to vary slightly depending on the location of the reactive sequences, the effect is relatively minor. We proceeded with BDTag-CFP and YFP-SpyTag as the model reactants to examine the effects of various conditions on the reaction, including temperature, pH, TMAO concentration, and the SpyStapler stoichiometry (Figure 3).

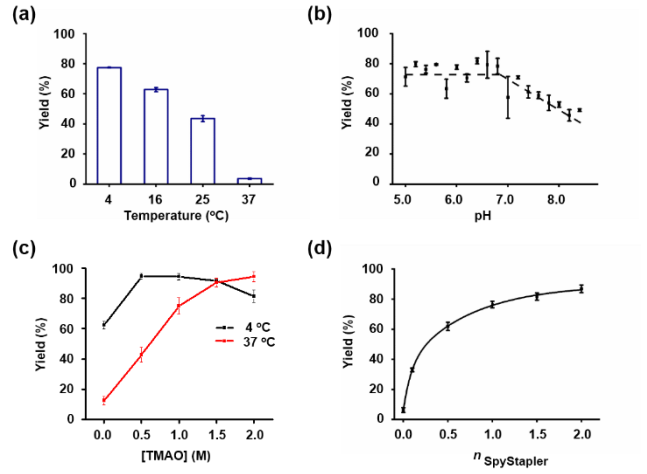


Figure 3. Effects of various conditions on the SpyStapler-mediated ligation between BDTag-CFP and YFP-SpyTag (10μM each): (a) temperature (2 eq. SpyStapler, 8 hours); (b) solution pH (2 eq. SpyStapler, 3 hours, 4 °C); (c) TMAO concentration (2 eq. SpyStapler, 3 hours) at 4 °C and 37 °C respectively; (d) stoichiometry of SpyStapler (4 °C, 8 hours). The experiments were run in triplicates.

SpyStapler is Robust under Various Conditions. Unless otherwise noted, the reactions were run at concentrations of 10 μM each with 2 equivalents of SpyStapler in PBS (pH 7.4) at 4 °C for 8 hours. It turned out that the reactions were efficient at lower temperatures (4 and 16°C), but were discouraged above 25 °C (Figure 3a, Figure S4). The solution pH value has only moderate influence on the ligation efficiency. The yields fluctuates around 80% at pH < 7.0, but drop at higher pH values (Figure 3b, Figure S5). With TMAO, the reactions were so fast that they already reached equilibrium at 8 hours. To differentiate the effects of different TMAO concentration, the reactions were run for as short as 3 hours at 4 and 37 °C, respectively. The results show that the addition of TMAO can promote the reactivity even further and enhance the performance of SpyStapler at higher temperature, which would be extremely useful for manipulating the reactivity *in vitro* (Figure 3c, Figure S6). It is worth noting that within 3 hours, the reaction at 4 °C already achieved a decent yield of 62%. Finally, Figures 3d and S7 show that when the stoichiometry of SpyStapler is lower than 0.6, the yields are generally higher than the stoichiometry of SpyStapler, indicating a limited turnover number per catalyst; at higher stoichiometry, the yields level off at ~80%, representing only minor improvement and suggesting potential product inhibition similar to that in the SnoopLigase system.

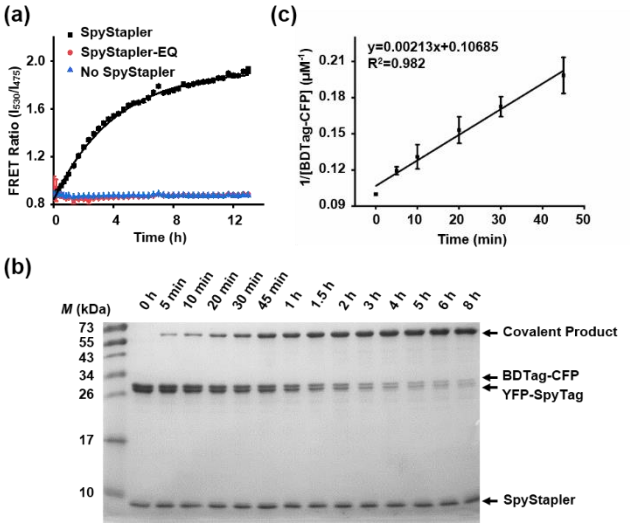


Figure 4. (a) Time course of the FRET signal for the reaction between BDTag-CFP and SpyTag-YFP in PBS (pH=7.4) at 4 °C over a period of 13 h. (b) SDS-PAGE analysis following the time course of the reaction between BDTag-CFP and YFP-SpyTag. (c) Plot of 1/[BDTag-CFP] vs time.

Kinetics of SpyStapler-Mediated Ligation. Once conjugated, the two model fluorescent proteins can undergo Förster resonance energy transfer (FRET), which offers a convenient way to monitor the progress of reaction. To shed light into the kinetics of the SpyStapler reaction, we followed the FRET signal for the

reaction between BDTag-CFP and SpyTag-YFP with SpyStapler and compared it to that without SpyStapler or with SpyStapler-EQ (Figure 4a). No increase in FRET signal was observed in the latter cases as expected. In particular, it suggests that SpyStapler-EQ could not even form a stable physical complex to bring the two components closer. The steady increase of FRET signal thus only represents the isopeptide bond formation.

For a quantitative assessment, we proceeded to determine the rate constant of the reaction by SDS-PAGE analysis (Figure 4b, c). Since BDTag-CFP and YFP-SpyTag are better resolved in SDS-PAGE than the other pairs, quantification by densitometry analysis is supposed to be more accurate in this case. They were thus chosen as the pair for determining the kinetics. The apparent second order rate constant was measured at a given concentration of reactants (10 μM each) and fixed amount of SpyStapler (20 μM). The calculated second-order rate constant is (36 ± 2) $M^{-1} \cdot s^{-1}$. This value is comparable to that of SpyCatcher-N^{TEV} (~10 $M^{-1} \cdot s^{-1}$),[11] a little slower than supercharged SpyCatcher(-) (~110 $M^{-1} \cdot s^{-1}$),[8] and around two orders of magnitude lower than the wild-type SpyCatcher (~10^3 $M^{-1} \cdot s^{-1}$).[5] By saying so, one should keep in mind that SpyStapler does not turn over to a significant level and this is a three-component reaction. Therefore, this value is only relevant to this specific reaction condition at the very early stage of reaction. Nonetheless, the results still suggest that SpyStapler is a fast ligase, with efficiency comparable to most "click" reactions.[33] The results are encouraging for pursuing more challenging targets, such as the "in-chain" reactions and the applications in cells.

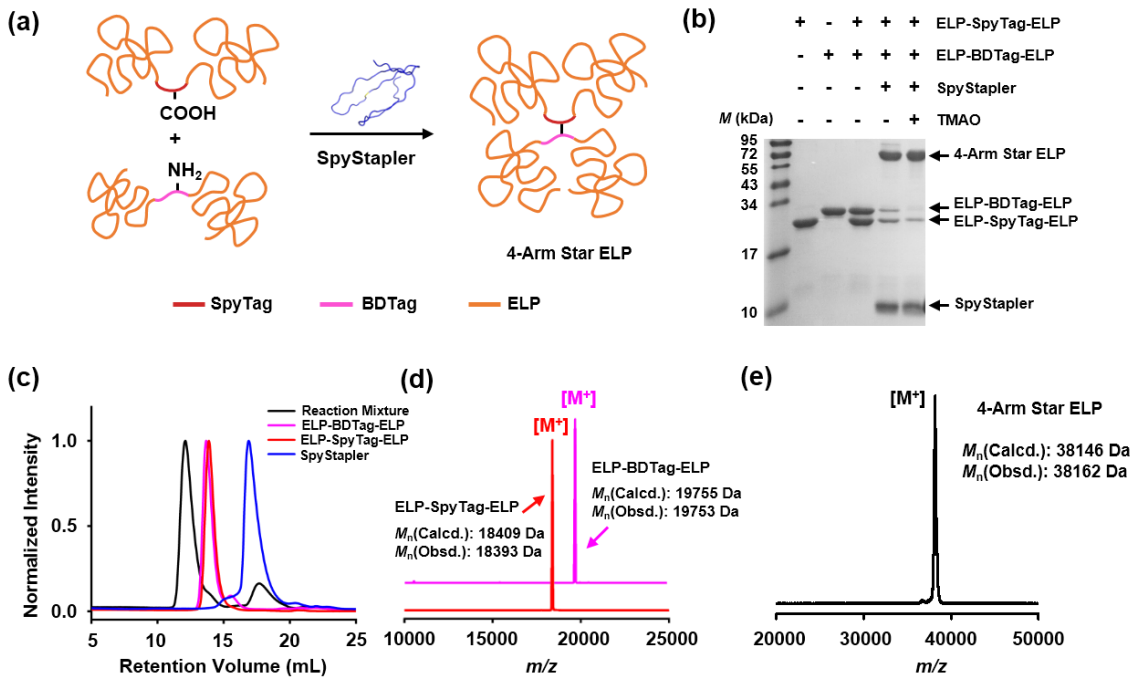


Figure 5. SpyStapler-mediated "in-chain" ligation toward four-arm star ELP. (a) Scheme for the reaction between ELP-SpyTag-ELP and ELP-BDTag-ELP *in vitro*. (b) SDS-PAGE analysis of the reaction with or without TMAO. (c) SEC overlay of SpyStapler (blue), ELP-SpyTag-ELP (red), ELP-BDTag-ELP (magenta), and the reaction mixture containing mainly the 4-arm star ELP (black). (d) MALDI-TOF mass spectra of ELP-SpyTag-ELP (red) and ELP-BDTag-ELP (magenta). (e) MALDI-TOF mass spectrum of the ligation product, 4-arm star ELP.

SpyStapler Enables Ligation between Internal Reactive Sites. When the reactive sequences (i.e. SpyTag and BDTag) are placed in the middle of the chain, the neighboring protein domains may exert considerable steric hindrance for binding and reaction, especially when the neighboring domains have random-coil-like conformation with a large hydrodynamic volume. Hence, few ligases could work on internal reactive sequences to date. Transglutaminase is an exception. It catalyzes the isopeptide bond formation between free amine groups and acyl groups at any sites with no sequence specificity and leads to extensive crosslinking.[34] Other natural enzymes such as sortase A and butelase 1 only recognize specific amino acid sequences locating at the termini of a protein, but not the internal ones. Both SpyLigase and SnoopLigase exposed their limitation when the ligation reactions involved internal sites between or within protein domains. It is therefore highly desired to have a protein ligase that can robustly stitch together the peptides even when they are inserted into the middle of a protein construct, or even in the loop region of a protein domain.[35] Hence, we challenged the catalytic capability of SpyStapler for its activity on two telechelic proteins based on elastin-like polypeptide (ELP) with reactive sequence in the middle of the chain, namely ELP-SpyTag-ELP and ELP-BDTag-ELP (see Figure S8 for sequences) (Figure 5a).

The ELP is an unstructured protein.[36-38] The large excluded volume effect is expected to impose considerable steric hindrance. Considering the tremendous entropy penalty, we anticipate that it would be difficult for SpyStapler to recognize two individual tags buried in the middle of the random coils, bring

them close in space, and staple them together with isopeptide bond. Indeed, the reaction under normal conditions (PBS, pH = 7.4, 4 °C, SpyTag: BDTag: SpyStapler = 10 μM:10 μM:20 μM) only gives a compromised yield of 27%. To enhance the reaction, we used 3 equivalents of SpyStapler (30 μM), the yield could be improved to ~60% (Figure 5b). The addition of 1.5 M TMAO further improved the ligation efficiency to 83% (Figure 5b). The MALDI-TOF mass spectra of the starting materials and the final product demonstrate the ligation between two components (Figure 5d and 5e). The SEC overlay clearly shows that there is only minor residual starting material as well as the remaining catalyst SpyStapler (Figure 5c). This is consistent with the results by SDS-PAGE analysis, which suggests that SpyStapler is indeed a robust ligase, even for "in-chain" ligation.

To demonstrate the broad applicability of SpyStapler-mediated ligation, we also explored the catalytic capability of SpyStapler for its activity on AffiHER2-SpyTag-AffiHER2 (see Figure S8 for sequence) and ELP-BDTag-ELP or BDTag-CFP. AffiHER2 is one kind of affibody that binds with high affinity to HER2 (an important tyrosine kinase cancer antigen).[39] AffiHER2 is folded and comprised of a bundle of three helixes. Again, reaction under normal conditions is low yielding even after 20 hours (Figure S9). However, with the addition of 1.5 M TMAO and slight excess of the SpyTag component, the reaction could be driven almost to completion, thus promising a broad scope for SpyStapler-mediated ligation.

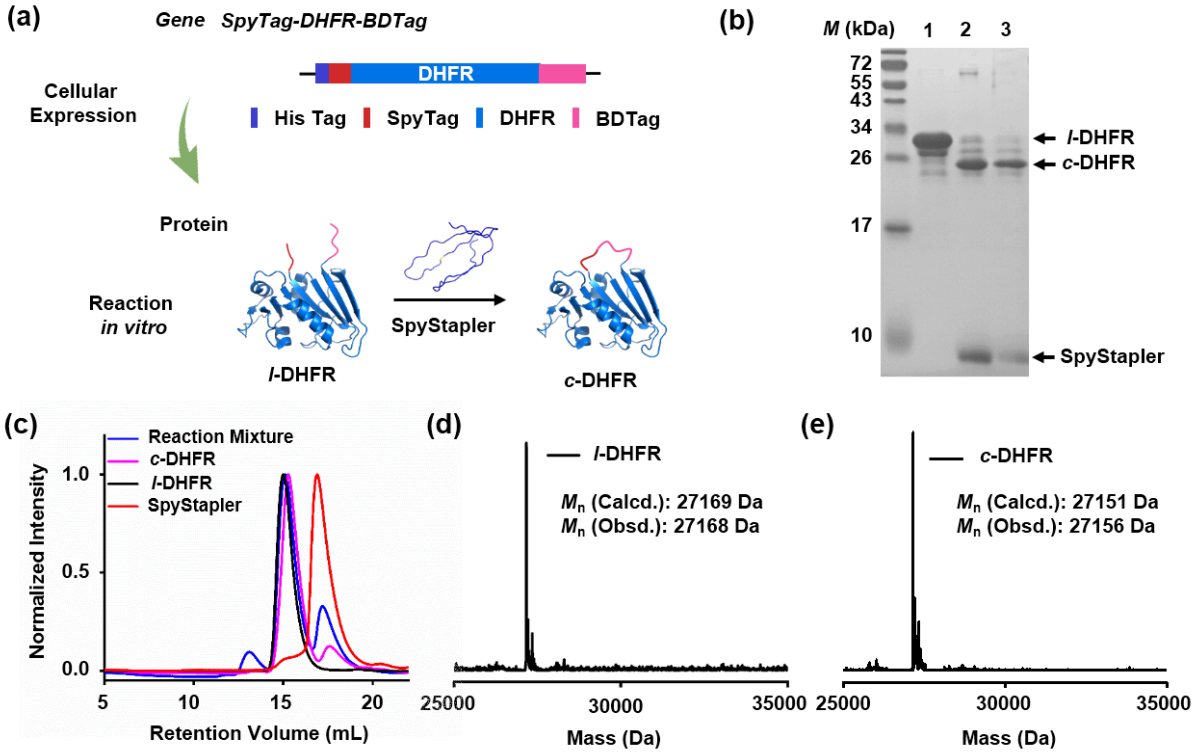


Figure 6. SpyStapler-mediated macrocyclization of SpyTag-DHFR-BDTag *in vitro*. (a) Scheme for the cyclization reaction. (b) SDS-PAGE analysis of *l*-DHFR (lane 1), reaction mixture (lane 2), and the purified *c*-DHFR (lane 3). (c) SEC overlay of SpyStapler (red), *l*-DHFR (black), the reaction mixture (blue), and the purified *c*-DHFR (magenta). (d) LC-MS spectrum of *l*-DHFR. (e) LC-MS spectrum of *c*-DHFR. PDB code: 4NX6 for DHFR.

SpyStapler Possesses Cyclase Activity Both *in vitro* and *in vivo*. Macrocyclization has been commonly used to evaluate the efficiency of bioconjugation.[23-24, 30, 40] We designed a telechelic DHFR bearing SpyTag at N-terminus and BDTag at C-terminus (named SpyTag-DHFR-BDTag or *l*-DHFR, see Figure S10 for sequence) to test the cyclization efficiency of SpyStapler (Figure 6a). After reaction under optimized condition for 12 hours, the reaction mixture was analyzed by SEC and SDS-PAGE. As shown in Figure 6c, the SEC elution profile of the reaction mixture (blue) contains mainly three fractions: the fraction F1 at retention volume of ~13 mL (which is supposedly chain-extended products), the fraction F2 with similar retention volume (~15 mL) to the starting material (which is probably the cyclized SpyTag-DHFR-BDTag, or *c*-DHFR), and the fraction F3 at the retention volume of ~ 17 mL (which is the excess SpyStapler). The assignment is corroborated by the SDS-PAGE analysis (Figure 6b) where lane 1 is the starting material *l*-SpyTag-DHFR-BDTag, lane 2 is the reaction mixture and lane 3 is fraction F2 in SEC, the purified *c*-DHFR. It is impressive that the monomers almost exist exclusively in the cyclic form. In addition, *c*-DHFR appears in a slightly lower position than *l*- DHFR, which is consistent with the compact conformation and fast mobility associated with the cyclic topology. There are also certain amounts of dimers in the reaction mixture, which is not surprising considering the competition between chain extension and

cyclization. The reaction concentration was set at a relatively low value (10 μM) so as to favor cyclization. This is further helped by the close proximity of the N- and C-termini of DHFR in space.[41] The LC-MS spectra of *l*-DHFR and *c*-DHFR are shown in Figure 6d and 6e, respectively. Within range of error, their mass difference is consistent with the loss of one H_2O molecule upon cyclization. Hence, we concluded that the SpyStapler-mediated macrocyclization of DHFR is very effective with almost quantitative conversion of the monomers *in vitro*.

While *in vitro* cyclization is routinely performed using a variety of enzymes like sortase, butelase 1, and OaAEP1, their applications in cells *in vivo* are often plagued by the difficulties associated with the heterologous expression of catalytically active forms of ligase, the relatively poor catalytic activity, or the reversible nature of certain ligases. The complex cellular environment is characteristic of a very crowded, heterogeneous mixture where millions of biological processes occur simultaneously. It would be very challenging for a ligase to find each component efficiently, bind them selectively, and perform the ligation effectively in cells. To test the activity of SpyStapler in cellular environments, we designed the co-expression of SpyTag-DHFR-BDTag and SpyStapler based on the pACYCDuet-1 vector (Figure 7a). The genes for two proteins are arranged in tandem sequence and express sequentially in cells under the control of T7 promoter. The expression temperature was set to be 16 °C at which SpyStapler exhibits excellent *in vitro* activity and the rate for cell growth and protein production is reasonably fast to ensure proper yield. The expression time was set to be 20 hours so as to ensure complete reaction.

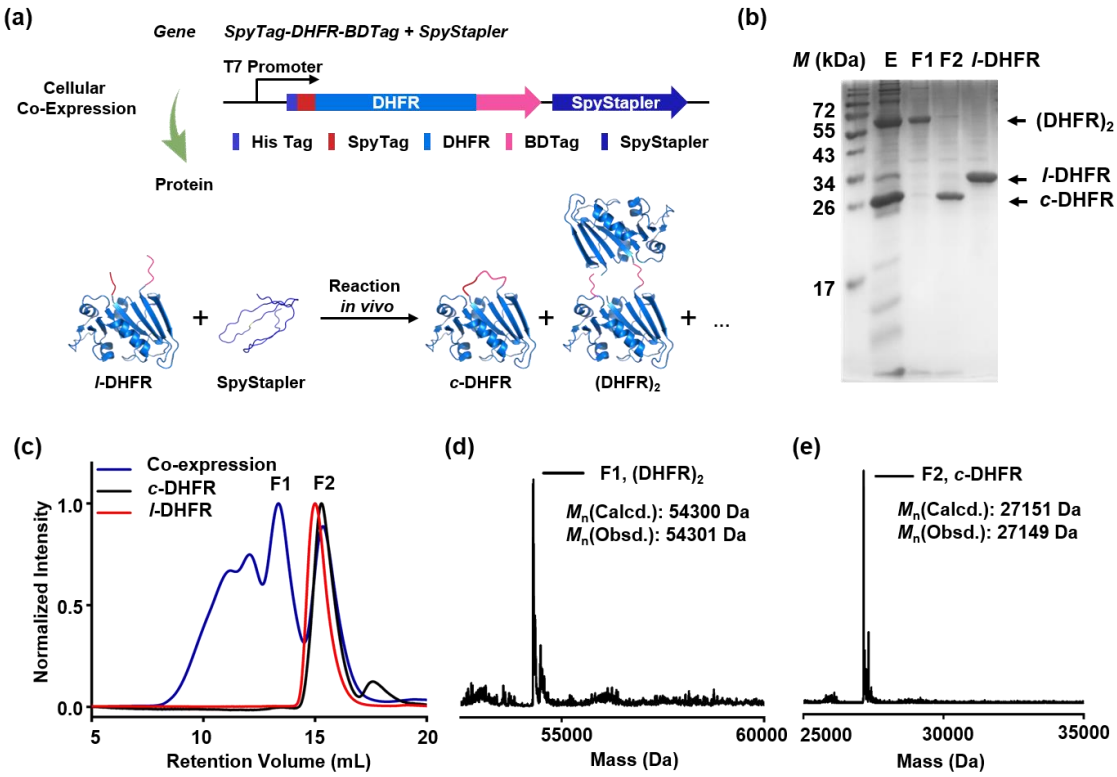


Figure 7. SpyStapler-mediated cyclization of SpyTag-DHFR-BDTag in cells. (a) Gene construct on pACYCDuet-1 for co-expression of SpyTag-DHFR-BDTag and SpyStapler in *E. coli*. (b) SDS-PAGE analysis of the crude product from co-expression (lane 2, E), the fractions F1 (lane 3) and F2 (lane 4) from SEC, and the linear SpyTag-DHFR-BDTag (*l*-DHFR) as a control (lane 5). (c) SEC overlay of the crude products from co-expression after affinity purification (blue), the purified fraction F2 (*c*-DHFR, black), and the control *l*-DHFR (red). (d) LC-MS spectrum of fraction F1 which is the dimer $(DHFR)_2$. (e) LC-MS spectrum of fraction F2 which is *c*-DHFR.

From the SDS-PAGE analysis of crude lysate and noninduced samples (Figure S11), there is only low level of target protein overexpression, which is probably probably due to the low plasmid copy number of the pACYCDuet-1.[42] Since only the SpyTag-DHFR-BDTag construct was designed with a His-Tag at N-terminus, affinity purification with Ni-NTA resin would presumably only pick up the components containing DHFR. After affinity purification, the crude product was analyzed by SEC chromatography and SDS-PAGE. In SEC, there are considerable amounts of high molecular weight products and two major peaks at larger retention volumes referred to as fractions F1 and F2, respectively (Figure 7c). The SDS-PAGE analysis confirm the presence of oligomers and verify that fraction F1 is mainly the dimer and F2 is the monomer (Figure 7b). It suggests that the reaction proceeds effectively in cells, but with much less control than the *in vitro* experiment. The lack of control could probably be explained by the relatively slow reaction rate as compared to the protein expression rate, which leads to the accumulation of reactive proteins over time. At later stages, the high intracellular telechelic protein concentration favors chain-extension over cyclization. Fraction F1 is the pure dimer $(DHFR)_2$. The LC-MS spectrum shows the loss of two water

molecules, confirming that both tags were reacted ($M_{calcd.}$ = 54300 Da vs $M_{obsd.}$ = 54301 Da) (Figure 7d). Considering the geometry of the tags, it is probably a cyclic dimer, rather than protein catenanes.[43-44] Fraction F2 is the pure cyclic monomer, *c*-DHFR. It appears at a lower position in SDS-PAGE than the linear counterpart, which is consistent with the previous macrocyclization experiments *in vitro*. The LC-MS spectrum further confirmed the structure (Figure 7e). It was quite remarkable that the monomer exists almost exclusively in cyclic form, indicating high cyclization efficiency *in vivo*. The high reactivity is also evidenced by the significant amounts of oligomers due to intermolecular reactions. In fact, the performance of SpyStapler in cells is almost comparable to that of the SpyRing system.[30, 40]

One interesting observation is that, even though SpyStapler does not have a His-Tag, it was also observed in the crude products (Figure S11). It suggests that SpyStapler may form a complex with the ligated product and small amounts of SpyStapler may co-elute with the product. While residual SpyStapler may not necessarily be a problem, it could be removed, if desired, by means of extensive wash on resin with competing peptide ligands or denaturing buffers, as previously demonstrated by Howarth and coworkers in similar SnoopLigase system.[18]

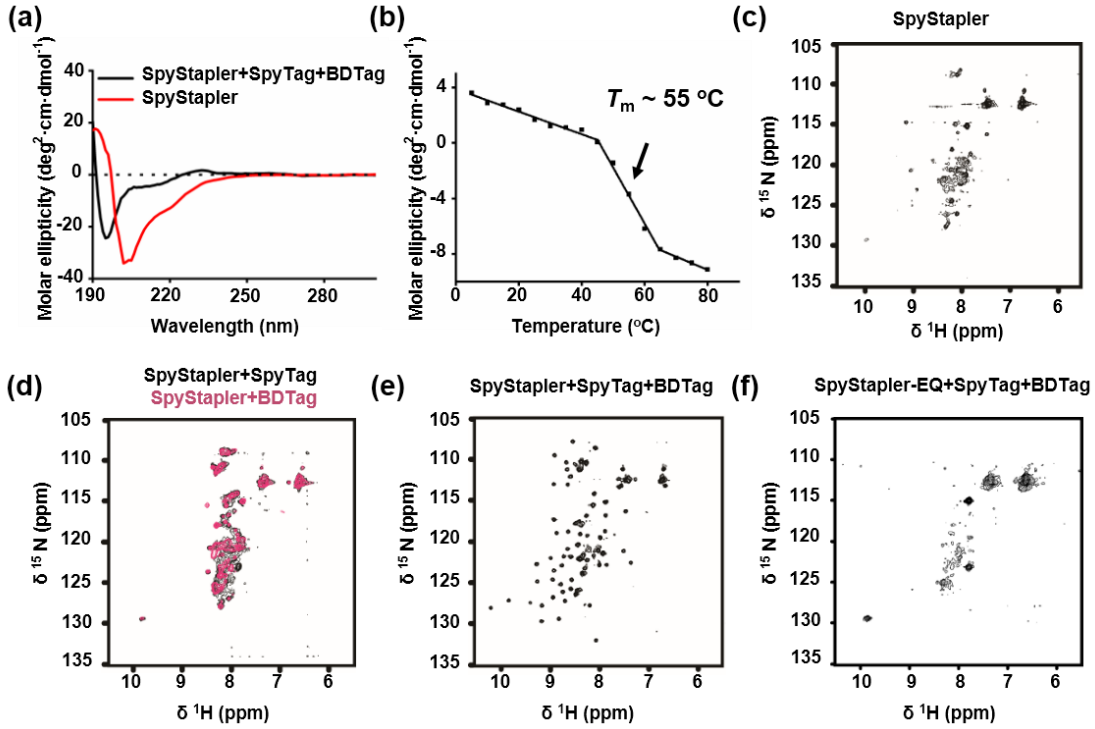


Figure 8. (a) CD spectra of SpyStapler (red) and the complex of SpyStapler+SpyTag+BDTag (black). (b) Temperature dependence of molar ellipticity of the complex of SpyStapler+SpyTag+BDTag at the wavelength of 230 nm. (c) HSQC NMR spectrum of SpyStapler. (d) Overlay of HSQC NMR spectra of SpyStapler+SpyTag (black) and SpyStapler+BDTag (magenta). (e) HSQC NMR spectrum of the complex of SpyStapler+BDTag+SpyTag. (f) HSQC NMR spectrum of the mixture of SpyStapler-EQ+SpyTag+BDTag.

SpyStapler is Disordered in Solution and Forms an Ordered Complex with the Ligated Product. Previously, engineering ligases often assumes a well-folded structure for activity optimization.[45-46] The crystal structure of transglutaminase, sortase A, and OaAEP1 have been elucidated, which facilitates structure-based design and engineering.[24, 47] Along this line, Howarth and co-workers used computer simulation to stabilize the SnoopLigase and optimize its activity.[18] In our case, however, the CD spectrum of SpyStapler revealed an unstructured polypeptide (Figure 8a, red). It was further verified by the 2D ^{1}H-^{15}N HSQC NMR spectrometry. Indeed, the HSQC spectrum of SpyStapler obtained in PBS buffer (pH=7.4) at 4 °C (Figure 8c) shows clustered peaks at the chemical shift of 7~9 ppm for ^{1}H and 120~125 ppm for ^{15}N NMR, which is typical for disordered proteins. The addition of either SpyTag or BDTag alone does not induce ordered structure formation even after incubation for 8 hours at 4 °C (Figure 8d). Only in the presence of both tags, a well-dispersed spectrum could be observed (Figure 8e, S12), indicating that SpyStapler is folded upon reaction. The spectrum is similar to that of the CnaB2 domain.[48] The NH signal of the isopeptide bond was not observed because the bond is formed between the unlabeled SpyTag and BDTag. The formation of a complex is also evidenced by the CD spectrum of the mixture of SpyTag, BDTag and SpyStapler (Figure 8a, black). It differs from that of SpyStapler in that the minimum shifts from 200 nm to 195 nm and there appears a bump at ~230 nm characteristic of the β-turn structure.[48] The spectrum resembles that of the CnaB2 domain or the SpyTag-SpyCatcher complex.[48]

To determine the stability of the complex, we followed the temperature dependence of the signal at 230 nm. With increasing temperature, the signal decreases and there is a clear transi-

tion (Figure 8b). The T_m was thus determined to be ~55 °C, suggesting that the complex is relatively stable. By comparison, the ancestor CnaB2 domain actually resists unfolding by boiling at 100 °C[48] and the parent SpyTag-SpyCatcher complex has a T_m of ~85 °C.[49] It is not surprising to see further reduced stability in the SpyTag-BDTag-SpyStapler complex. Hence, it is also reasonable to assume that the complex is dynamic and may exist as an equilibrium between the associated and unassociated forms depending on temperature (Figure 9). It also explains the fact that cellularly synthesized *c*-DHFR contains a small amount of SpyStapler.

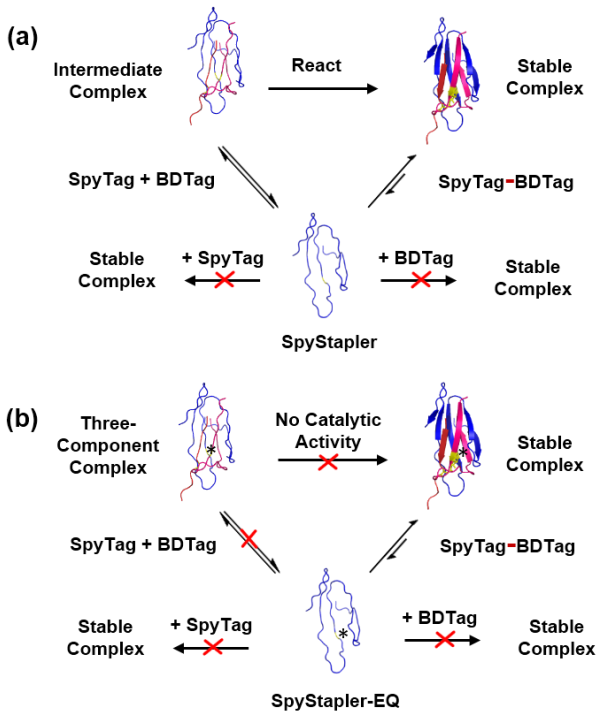


Figure 9. (a) SpyStapler could not form stable complex with either SpyTag or BDTag but could form a physical complex in the presence of both tags and transform into a stable complex upon isopeptide bond formation. Dissociation into free SpyStapler and the covalent product, SpyTag-BDTag, may occur. (b) SpyStapler-EQ could not form stable complex with SpyTag, BDTag, or both, but could form a stable physical complex with the ligated product SpyTag-BDTag.

To probe the role of covalent bond in the complex, we furthered looked into the interaction between the catalytically inactive SpyStapler-EQ and the tags, as well as the ligated product of SpyTag-BDTag. The SpyTag-BDTag conjugate was prepared by SpyStapler-mediated ligation and purified by high performance liquid chromatography (HPLC). As shown by the HSQC NMR spectra, SpyStapler-EQ is unfolded in solution and could not become structured upon interaction with SpyTag, or BDTag, or both tags (Figure 8f, S13). However, it does form a stable complex with the SpyTag-BDTag conjugate (Figure S14). The overlay of the NMR spectra of SpyStapler and SpyStapler-EQ in the complex with the SpyTag-BDTag shows very minor differences that are probably due to the residues at or near the mutation site. It indicates that mutation in SpyStapler-EQ does not weaken the binding to the product. Therefore, it is clear that covalent bonding plays a critical role in stabilizing the complex (Figure 9).

The mechanistic study also improves our understanding on the stoichiometry-dependence of the reaction (Figure 3d). We envisioned that when SpyTag and BDTag are fused to other proteins, the resulting complex in the ligated product may experience more steric hindrance from the neighboring proteins and it is more likely to undergo dissociation after covalent bond formation. Indeed, when SpyStapler is present in less than 0.5 equivalence, it shows considerable turnover and the yields were generally higher than the stoichiometry. Increasing SpyStapler stoichiometry could push the reaction further toward completion. However, significant product inhibition may arise with increasing amounts of the ligated product, leading to the plateau of yield at around 80%. After all, the reaction system is an equilibrium among the three components, the ligated product, and the complex. For SpyStapler-EQ, the absence of covalent bonding fails to push the equilibrium to the product side and thus it could not form any order complex with SpyTag and BDTag, even though it retains the capability to bind the ligated product SpyTag-BDTag and fold.

Proposed Free Energy Landscape of SpyStapler-Mediated Ligation. The above analysis reveals the important role of covalent bonding and its interplay with folding. We did attempt to characterize the thermodynamics of SpyStapler-mediated ligation by isothermal titration calorimetry. However, little exothermic or endothermic effect were observed when mixing the SpyTag or BDTag with SpyStapler. Increasing concentration of the peptide tags led to aggregation and precipitation that further plagued the characterization. Combined with CD and NMR results, we deduced that SpyStapler probably exists as a molten globule in solution and its interaction with either SpyTag or BDTag has little enthalpic effect. Even in the presence of both SpyTag and BDTag, the three components can barely form a stable ordered complex before the covalent bond formation. However, this transient complex is catalytically active and the isopeptide bond formation could dramatically change the free energy landscape for folding and create a thermodynamic sink to stabilize the complex. The lifetime of the folded complex is long enough to be detected by NMR techniques (Figure 8e). Nonetheless, disassociation of the folded complex could still occur occasionally upon thermal fluctuation, thus regenerating SpyStapler and releasing the ligated product. Based on these arguments, we tentatively proposed a reaction energy profile for the SpyStapler-mediated ligation (Figure 10).

This process is unique in that it couples assembly with reaction. The catalyst, SpyStapler, is unstructured by itself, yet could assemble with SpyTag and BDTag to reconstitute the structural domain for the catalytic formation of isopeptide bond. Moreover, after ligation, SpyStapler could disassociate from the complex to regenerate the active catalyst, though product inhibition is inevitable at high SpyStapler stoichiometry (Figure 3d). This is an excellent example of "the synergy between molecular assembly and chemical reaction".[50] It is reasonable to speculate that in nature, many intrinsically disordered proteins may adopt similar modes of action for enzymatic activities. It also makes SpyStapler quite resilient toward various denaturing treatments. For example, SpyStapler can also be purified under denaturing conditions in the presence of 8 M urea without losing activity

once buffer-exchanged to the normal buffer. Even after SpyStapler is boiled at 100 °C for extended times (for example, 20 min), there is no aggregation or precipitation and SpyStapler remains catalytically active (Figure S15).

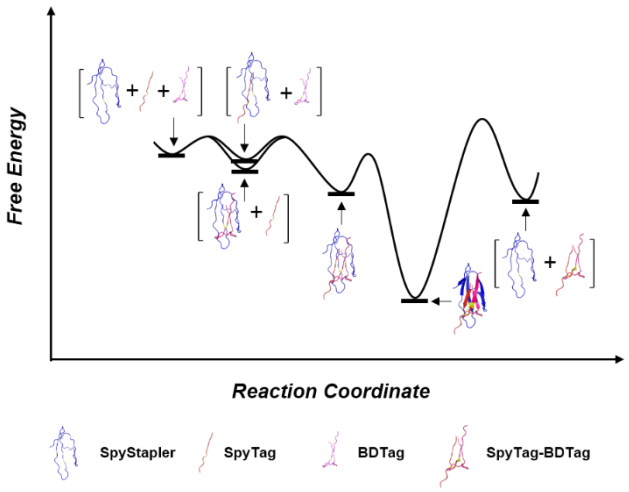


Figure 10. Postulated free energy profile for the SpyStapler-mediated ligation.

Conclusions

To sum up, we report a powerful and unstructured ligase SpyStapler that catalyzes the isopeptide bond formation between SpyTag and BDTag both *in vitro* and *in vivo*. We show that, although disordered in solution, SpyStapler enables efficient ligation between SpyTag and BDTag regardless of their relative positions within protein molecules (N-, C-terminus, or in the middle of the chain), even in the absence of promoters like TMAO or glycerol. The typical second order rate constant of SpyStapler-mediated ligation ($\sim 36 \pm 2\ M^{-1} \cdot s^{-1}$) is comparable to most bioorthogonal "click" reactions. High coupling efficiency further facilitates the applications in cells, such as the macrocyclization of DHFR during expression. Detailed study by CD and NMR spectrometry further shed light into the possible mechanisms by which SpyStapler binds SpyTag and BDTag for isopeptide bond formation. The disordered nature of SpyStapler makes it highly resilient toward various inactivating treatments such as denaturation in 8 M urea or boiling at 100 °C for extended time without losing activity. We believe that this tool would be of broad interest to researchers in the field of protein engineering, biomaterials, chemical biology, and synthetic biology. The interplay between covalent bonding and protein folding also suggests that chemical reaction can be programmed and coupled with protein-protein interactions for mutual reinforcement to enhance protein functions and impart resilience.

ASSOCIATED CONTENT

Supporting Information.

Full amino acid sequence of the telechelic proteins, SDS-PAGE, LC-MS spectra of the telechelic protein and the NMR spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* wenbin@pku.edu.cn

Author Contributions

§ X.-L. Wu and Y.–J. Liu contributed equally to the work.

Funding Sources

National Natural Science Foundation of China (Grants 21474003, 91427304) and "1000 Plan (Youth)".

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We are grateful for the financial support from the National Natural Science Foundation of China (Grants 21474003, 91427304), the 863 Program (2015AA020941), and "1000 Plan (Youth)". We also thank Beijing NMR Centre and the NMR Facility at the National Centre for Protein Sciences at Peking University for assistance with the solution NMR experiments.

REFERENCES

[1] Lotze, J.; Reinhardt, U.; Seitz, O.; Beck-Sickinger, A. G., Peptide-tags for site-specific protein labelling in vitro and in vivo. *Mol. BioSyst.* 2016, *12* (6), 1731-1745.

[2] Kimple, M. E.; Brill, A. L.; Pasker, R. L., Overview of affinity tags for protein purification. *Curr. Protoc. Protein Sci.* 2013, *73* (1), 9.9.1-9.9.23.

[3] Xu, L.; Zhang, W.-B., Topology: a unique dimension in protein engineering. *Sci. Chi. Chem.* 2018, *61* (1), 3-16.

[4] Wang, X. W.; Zhang, W. B., Chemical Topology and Complexity of Protein Architectures. *Trends Biochem. Sci.* 2018, *43* (10), 806-817.

[5] Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. USA* 2012, *109* (12), E690-E697.

[6] Veggiani, G.; Nakamura, T.; Brenner, M. D.; Gayet, R. V.; Yan, J.; Robinson, C. V.; Howarth, M., Programmable polyproteams built using twin peptide superglues. *Proc. Natl. Acad. Sci. USA* 2016, *113* (5), 1202-1207.

[7] Liu, Y.; Liu, D.; Yang, W.; Wu, X.-L.; Lai, L.; Zhang, W.-B., Tuning SpyTag-SpyCatcher mutant pairs toward orthogonal reactivity encryption. *Chem. Sci.* 2017, *8* (9), 6577-6582.

[8] Cao, Y.; Liu, D.; Zhang, W.-B., Supercharging SpyCatcher toward an intrinsically disordered protein with stimuli-responsive chemical reactivity. *Chem. Commun.* 2017, *53* (63), 8830-8833.

[9] Tan, L. L.; Hoon, S. S.; Wong, F. T., Kinetic Controlled Tag-Catcher Interactions for Directed Covalent Protein Assembly. *PloS one* 2016, *11* (10), e0165074.

[10] Proschel, M.; Kraner, M. E.; Horn, A. H. C.; Schafer, L.; Sonnewald, U.; Sticht, H., Probing the potential of CnaB-type domains for the design of tag/catcher systems. *PLoS One* 2017, *12* (6), e0179740.

[11] Zhang, X.-J.; Wu, X.-L.; Wang, X.-W.; Liu, D.; Yang, S.; Zhang, W.-B., SpyCatcher-N(TEV): A circularly permuted, disordered SpyCatcher variant for less trace ligation. *Bioconjugate Chem.* 2018, *29* (5), 1622-1629.

[12] Fang, J.; Zhang, W.-B., Genetically Encoded Peptide-Protein Reactive Pairs. *Acta Polym. Sin.* 2018, (4), 429-444.

[13] Veggiani, G.; Zakeri, B.; Howarth, M., Superglue from bacteria: unbreakable bridges for protein nanotechnology. *Trends Biotechnol.* 2014, *32* (10), 506-512.

[14] Sun, F.; Zhang, W.-B., Unleashing chemical power from protein sequence space toward genetically encoded click chemistry. *Chin. Chem. Lett.* 2017, *28* (11), 2078-2084.

[15] Wang, N.; Yang, B.; Fu, C.; Zhu, H.; Zheng, F.; Kobayashi, T.; Liu, J.; Li, S.; Ma, C.; Wang, P. G.; Wang, Q.; Wang, L., Genetically Encoding Fluorosulfate-l-tyrosine To React with Lysine, Histidine,

and Tyrosine via SuFEx in Proteins in Vivo. *J. Am. Chem. Soc.* 2018, *140* (15), 4995-4999.
[16] Xiang, Z.; Ren, H.; Hu, Y. S.; Coin, I.; Wei, J.; Cang, H.; Wang, L., Adding an unnatural covalent bond to proteins through proximity-enhanced bioreactivity. *Nat. Methods* 2013, *10* (9), 885-8.
[17] Fierer, J. O.; Veggiani, G.; Howarth, M., SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture. *Proc. Natl. Acad. Sci. USA* 2014, *111* (13), E1176-E1181.
[18] Buldun, C. M.; Jean, J. X.; Bedford, M. R.; Howarth, M., SnoopLigase catalyzes peptide-peptide locking and enables solid-phase conjugate isolation. *J. Am. Chem. Soc.* 2018, *140* (8), 3008-3018.
[19] Evans, T. C.; Benner, J.; Xu, M.-Q., The in vitro ligation of bacterially expressed proteins using an intein from Methanobacterium thermoautotrophicum. *J. Bio. Chem.* 1999, *274* (7), 3923-3926.
[20] van't Hof, W.; Hansenova Manaskova, S.; Veerman, E. C. I.; Bolscher, J. G. M., Sortase-mediated backbone cyclization of proteins and peptides. *Biol. Chem.* 2015, *396* (4), 283-293.
[21] Hou, Y.; Yuan, J.; Zhou, Y.; Yu, J.; Lu, H., A Concise Approach to Site-Specific Topological Protein-Poly(amino acid) Conjugates Enabled by in Situ-Generated Functionalities. *J. Am. Chem. Soc.* 2016, *138* (34), 10995-11000.
[22] Nguyen, G. K.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P., Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* 2014, *10* (9), 732-738.
[23] Nguyen, G. K. T.; Kam, A.; Loo, S.; Jansson, A. E.; Pan, L. X.; Tam, J. P., Butelase 1: A versatile ligase for peptide and protein macrocyclization. *J. Am. Chem. Soc.* 2015, *137* (49), 15398-15401.
[24] Yang, R.; Wong, Y. H.; Nguyen, G. K. T.; Tam, J. P.; Lescar, J.; Wu, B., Engineering a catalytically efficient recombinant protein ligase. *J. Am. Chem. Soc.* 2017, *139* (15), 5351-5358.
[25] Wright, P. E.; Dyson, H. J., Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* 1999, *293*, 321-331.
[26] Dyson, H. J.; Wright, P. E., Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 2005, *6* (3), 197-208.
[27] Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N., Introducing protein intrinsic disorder. *Chem. Rev.* 2014, *114* (13), 6561-6588.
[28] Uversky, V. N., *Intrinsically disordered proteins*. Springer: New York, 2014; p 51.
[29] Sugase, K.; Dyson, H. J.; Wright, P. E., Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 2007, *447* (7147), 1021-5.
[30] Schoene, C.; Fierer, J. O.; Bennett, S. P.; Howarth, M., SpyTag/SpyCatcher cyclization confers resilience to boiling on a mesophilic enzyme. *Angew. Chem. Int. Ed.* 2014, *53* (24), 6101-6104.
[31] Hinrichsen, M.; Lenz, M.; Edwards, J. M.; Miller, O. K.; Mochrie, S. G. J.; Swain, P. S.; Schwarz-Linek, U.; Regan, L., A new method for post-translationally labeling proteins in live cells for fluorescence imaging and tracking. *Protein Eng. Des. Sel.* 2017, *30* (12), 771-780.
[32] Li, L.; Fierer, J. O.; Rapoport, T. A.; Howarth, M., Structural analysis and optimization of the covalent association between SpyCatcher and a peptide Tag. *J. Mol. Biol.* 2014, *426* (2), 309-17.
[33] McKay, C. S.; Finn, M. G., Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem. Biol.* 2014, *21* (9), 1075-101.
[34] Griffin, M.; Casadio, R.; Bergamini, C. M., Transglutaminases: nature' s biological glues. *Biochem. J.* 2002, *368*, 377-396.
[35] Liu, M.; Ji, Z.; Zhang, M.; Xia, J., Versatile site-selective protein reaction guided by WW domain-peptide motif interaction. *Bioconjugate Chem.* 2017, *28* (8), 2199-2205.
[36] Chilkoti, A.; Christensen, T.; MacKay, J. A., Stimulus responsive elastin biopolymers: Applications in medicine and biotechnology. *Curr. Opin. Chem. Biol.* 2006, *10* (6), 652-7.
[37] Chen, Y.; Guan, Z., Bioinspired Modular Synthesis of Elastin-Mimic Polymers To Probe the Mechanism of Elastin Elasticity. *J. Am. Chem. Soc.* 2010, *132*, 4577-4579.
[38] Urry, D. W.; Hugel, T.; Seitz, M.; Gaub, H. E.; Sheiba, L.; Dea, J.; Xu, J.; Parker, T., Elastin: a representative ideal protein elastomer. *Philos. Trans. Royal Soc. B* 2002, *357* (1418), 169-184.
[39] Eigenbrot, C.; Ultsch, M.; Dubnovitsky, A.; Abrahmsen, L.; Hard, T., Structural basis for high-affinity HER2 receptor binding by an engineered protein. *Proc. Natl. Acad. Sci. USA* 2010, *107* (34), 15039-44.
[40] Zhang, W.-B.; Sun, F.; Tirrell, D. A.; Arnold, F. H., Controlling macromolecular topology with genetically encoded SpyTag-SpyCatcher chemistry. *J. Am. Chem. Soc.* 2013, *135* (37), 13988-97.
[41] Fenwick, R. B.; van den Bedem, H.; Fraser, J. S.; Wright, P. E., Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc. Natl. Acad. Sci. USA* 2014, *111* (4), E445-E454.
[42] Held, D.; Yaeger, K.; Novy, R., New coexpression vectors for expanded compatibilities in E. coli. *inNovations* 2003, *18*, 4.
[43] Wang, X.-W.; Zhang, W.-B., Cellular Synthesis of Protein Catenanes. *Angew. Chem. Int. Ed.* 2016, *55* (10), 3442-3446.
[44] Wang, X.-W.; Zhang, W.-B., Protein Catenation Enhances Both Stability and Activity of Folded Structural Domains. *Angew. Chem. Int. Ed.* 2017, *56* (45), 13985-13989.
[45] Dorr, B. M.; Ham, H. O.; An, C.; Chaikof, E. L.; Liu, D. R., Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci. USA* 2014, *111* (37), 13343-13348.
[46] Bentley, M. L.; Gaweska, H.; Kielec, J. M.; McCafferty, D. G., Engineering the substrate specificity of Staphylococcus aureus Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J. Biol. Chem.* 2007, *282* (9), 6571-6581.
[47] Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R. T., Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus. *Proc. Natl. Acad. Sci. USA* 2001, *98* (11), 6056-6061.
[48] Hagan, R. M.; Bjornsson, R.; McMahon, S. A.; Schomburg, B.; Braithwaite, V.; Buhl, M.; Naismith, J. H.; Schwarz-Linek, U., NMR spectroscopic and theoretical analysis of a spontaneously formed Lys-Asp isopeptide bond. *Angew. Chem. Int. Ed.* 2010, *49* (45), 8421-8425.
[49] Liu, D.; Wu, W.-H.; Liu, Y.-J.; Wu, X.-L.; Cao, Y.; Song, B.; Li, X.; Zhang, W.-B., Topology engineering of proteins in vivo using genetically encoded, mechanically interlocking SpyX modules for enhanced stability. *ACS Cent. Sci.* 2017, *3* (5), 473-481.
[50] Wang, Y.; Lin, H. X.; Chen, L.; Ding, S. Y.; Lei, Z. C.; Liu, D. Y.; Cao, X. Y.; Liang, H. J.; Jiang, Y. B.; Tian, Z. Q., What molecular assembly can learn from catalytic chemistry. *Chem. Soc. Rev.* 2014, *43* (1), 399-411.

SYNOPSIS TOC