Published in final edited form as:

Anal Chem. 2013 June 18; 85(12): 6080–6088. doi:10.1021/ac400949t.

Oriented covalent immobilization of antibodies for measurement of intermolecular binding forces between zipper-like contact surfaces of split inteins

Mirco Sorci¹, Bareket Dassa², Hongwei Liu¹, Gaurav Anand¹, Amit K. Dutta¹, Shmuel Pietrokovski², Marlene Belfort³, and Georges Belfort^{1,*}

¹Howard P. Isermann Department of Chemical and Biological Engineering and Center for Biotechnology and Interdisciplinary Studies Rensselaer Polytechnic Institute, Troy, NY 12180

²Molecular Genetics Department, Weizmann Institute of Science, Rehovot 76100, Israel

³Department of Biological Sciences, University at Albany, SUNY, Albany, NY 12222

Abstract

In order to measure the intermolecular binding forces between two halves (or partners) of naturally split protein splicing elements called inteins, a novel thiol-hydrazide linker was designed and used to orient immobilized antibodies specific for each partner. Activation of the surfaces was achieved in one step allowing direct force measurements of the formation of a peptide bond catalyzed by the binding of the two partners of the split intein (called protein *trans*-splicing). Through this binding process, a whole functional intein is formed resulting in subsequent splicing. Atomic force microscopy (AFM) was used to directly measure the split intein partner binding at 1 μm/s between native (wild-type) and mixed pairs of C- and N-terminal partners of naturally occurring split inteins from three cyanobacteria. Native and mixed pairs exhibit similar binding forces within the error of the measurement technique (~52 pN). Bioinformatic sequence analysis and computational structural analysis discovered a zipper-like contact between the two partners with electrostatic and non-polar attraction between multiple aligned ion pairs and hydrophobic residues. Also, we tested the Jarzynski's equality and demonstrated, as expected, that non-equilibrium dissipative measurements obtained here gave larger energies of interaction as compared with those for equilibrium. Hence, AFM coupled with our immobilization strategy and computational studies provides a useful analytical tool for the direct measurement of intermolecular association of split inteins and could be extended to any interacting protein pair.

Keywords

Hydrazide; MBP; CBD; Split-inteins; Binding forces; Protein Association

*To whom correspondence should be addressed: Georges Belfort, Howard P. Isermann Department of Chemical and Biological Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA, Tel.: (518) 276-6948; Fax: (518) 276-4030; belfog@rpi.edu.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Author Contributions

Conceived and designed the experiments: MS, MB and GB. Performed all experiments and analyzed the data: MS. Provided I_C and I_N clones: BD and SP. Synthesized the linker: HL. Helped setting up AFM runs: GA and AKD. Wrote the paper: MS, MB and GB.

The authors declare no competing financial interest.

INTRODUCTION

Atomic Force Microscopy (AFM) has been extensively used to investigate structure and function of biological molecules¹. AFM is well known for its high resolution imaging capability and it is also a powerful technique for intermolecular force measurements². Furthermore, the possibility to coat the cantilever tips and substrates with different materials, allows one to study the interaction between different combinations of bound and oriented molecules. Here we have used AFM in force-mode to analyze the interaction between the two binding partners of three naturally split inteins, which are autocatalytic protein splicing elements from diverse cyanobacteria³. Inteins are of both fundamental and biotechnology importance. Some occur in nature in split form and the halves need to associate to function (i.e splice). Here we express three split-intein partners with fused extein tags, which facilitate the immobilization of the different constructs using monoclonal antibodies.

Three main strategies for oriented antibody immobilization were developed over the past 30 years⁴: (i) Antibodies were bound to Fc receptors on solid supports (e.g. protein A, protein G); (ii) Chemical or enzymatic oxidation of the IgG carbohydrate moiety (mainly in the Fc fragment) into aldehyde groups were reacted with hydrazide activated supports to form covalent hydrazone bonds; (iii) Native thiol-groups, after chemical reduction of the disulfide bonds, were attached to gold or maleimide functionalized surfaces⁵. The first strategy, while highly specific⁶, is characterized by non-covalent attachment of the antibody; the third approach, does involve a covalent bound, but could result in inactive antibody fragments due to unintentional reduction of the disulfide bonds⁷. The second method was selected here, since it also results in covalent antibody attachment but without redox chemistry and it has previously been successfully applied^{4b, 8}. Depending on the substrate, this approach could require several steps in order to activate the surface with hydrazide groups⁹. Using gold-coated substrates is convenient since the activation can be implemented in one step¹⁰. To accomplish this and to orient the molecules correctly, we synthesized a novel thiol-linker with a protected hydrazide moiety.

Since their discovery in 1990, more than 600 inteins have been identified in all three domains of life^{3b}. Splicing (and self-cleaving) of protein elements occurs when an intein catalyzes its self-excision from 2 segments of a protein (exteins) located at its N- and C-termini and simultaneously ligates or reunites the remaining host protein components or exteins. The splicing mechanism has been reviewed¹¹ and extensive engineering of inteins has resulted in many new technologies¹². Recent strategies for protein engineering include protein *trans*-splicing, which is performed by split inteins. Split inteins have not yet been found in eukaryotes but have been identified in diverse cyanobacteria (e.g. C-type DNA polymerase III α subunits, DnaE proteins) and archaea^{3a, 13}. To date, there are around 50 unique sequences of split inteins reported, and the number is rapidly growing as more microbial genomes are sequenced¹⁴. Split-intein partners (intein fragment plus the neighboring extein sequence) are generally inactive until they associate and initiate splicing activity¹⁵. The association between two split-intein partners is not a ligand-protein interaction but the reconstruction of a protein fold, which restores intein splicing activity. Here we study three split inteins from diverse cyanobacteria: *Oscillatoria limnetica* (Oli), *Thermosynechococcus vulcanus* (Tvu) and *Nostoc species PCC7120* (Nsp).

Despite the many applications and studies on splicing kinetics^{11h, 15a, 16}, very little is known about the molecular recognition events preceding the protein *trans*-splicing process. Contreras Martinez and co-workers¹⁷ identified the hydrophobic core amino acids between the two protein fragments that were crucial for split-protein reassembly. Shi and Muir¹⁸ presented the first thermodynamic and kinetic analysis of only the binding event between

split-intein partners, independently of the splicing reaction. Using fluorescence resonance energy transfer (FRET), they showed that the naturally split intein, *Synechocytis* sp. (*Ssp*) DnaE partners bound with low nanomolar affinity ($K_d = 43 \pm 10$ nM), resulting from very fast association rates ($k_{on} = 2.8 \pm 0.5 \cdot 10^7 \text{ M}^{-1} \text{ s}^{-1}$) and moderately slow dissociation rates ($k_{off} = 1.2 \pm 0.2 \text{ s}^{-1}$). Also, the kinetic constants depended on ionic strength of the buffer, suggesting a contribution by electrostatic interactions between split-intein partners. Later Ludwig and co-workers monitored the association of artificially split-intein partners from a mutant *Ssp* DnaB intein, which blocked protein splicing, by native PAGE and fluorescence anisotropy¹⁹. The kinetic parameters for intein complex formation were $K_d = 1.1 \text{ } \mu\text{M}$, $k_{on} = 16.8 \text{ M}^{-1} \text{ s}^{-1}$, and $k_{off} = 1.8 \cdot 10^{-5} \text{ s}^{-1}$. The authors suggest that the very low k_{on} value for this artificial intein fragment association likely reflects required major structural changes (the split fragments are largely unfolded), compared with the *Ssp* DnaE fragments which have undergone natural evolution for efficient protein *trans*-splicing. Consideration of association of split inteins would clearly benefit from direct experimental measurements.

For direct measurement of the interactions (forces and energies) between split intein partners, a novel thiol-hydrazide linker was designed and used to orient immobilized antibodies specific for each intein partner. The binding force measurements performed here shows that native (from same species) and mixed (from different species) split-intein pairs exhibit similar binding energies within the error of the measurement technique. This finding is in concert with Dassa and co-workers results^{3a}, which showed that all nine combinations of the same split-intein partners can induce protein-splicing and have suggested that split inteins, sharing 50–70% of sequence identity, derive from a common progenitor^{13a}. Molecular details from our bioinformatic sequence analysis of the binding surface demonstrate interdigitated electrostatic and non-polar attraction of aligned ionic and hydrophobic amino acid pairs, explaining how the intein binding partners are guided toward each other for subsequent binding. Thus, our novel immobilization procedure, AFM-force spectroscopy and bioinformatic sequence analysis with computational structural prediction provide a direct analytical tool to investigate the association of split splicing-proteins. These approaches give a fundamental understanding of molecular recognition in *trans*-splicing, while they will facilitate the design of artificial split inteins for practical uses, and they offer a direct way of measuring binding between any proteins.

EXPERIMENTAL SECTION

Linker

A linker was synthesized for direct and oriented immobilization of antibodies on gold surfaces. The linker had two reactive end groups: thiol and hydrazide. Thiol groups react directly with gold; the hydrazide group can be covalently coupled to carbohydrate residues in glycoproteins and other glycoconjugates after oxidation to produce aldehydes²⁰. The disulfide structure and protective hydrazide were important for the synthesis of the linker, rather than the immobilization strategy. Details of the linker synthesis are presented in Supporting Information (Figure S1).

Antibodies

Anti-MBP and anti-CBD monoclonal antibodies were used for protein immobilization (New England Biolabs, Ipswich, MA). AffiniPure goat anti-mouse IgG (H+L) (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA) and HRP-linked anti-mouse IgG (Cell Signaling Technology, Inc., Danvers, MA) were utilized in an ELISA assay, as coating and secondary antibody, respectively. HRP-conjugated anti-mouse antibodies were used for luminol-based detection in Western blots (GE Healthcare, Piscataway, NJ).

Proteins

Three naturally occurring split inteins were chosen from diverse cyanobacteria: *Oscillatoria limnetica*, *Thermosynechococcus vulcanus* and *Nostoc species PCC7120*. Each N-terminal intein domain (I^N) was cloned downstream to a maltose-binding domain (MBP) and each C-terminal intein domain (I^C) was cloned between an upstream MBP to enhance its solubility, and a downstream chitin-binding domain (CBD). The fusion proteins were identical to those studied by Dassa et al.^{3a}. MBP-CBD protein was also expressed as a negative control. I^N , I^C and MBP-CBD were purified using their affinity MBP tags. Details are in Supplementary Information (Figures S2, S3 and S4). Bovine serum albumin (BSA) was certified ACS reagent grade (Sigma-Aldrich, St. Louis, MO).

Chemicals and solutions

The oxidant used to obtain aldehydes groups in anti-MBP and anti-CBD antibodies before immobilization was NaIO_4 (98–100%) (Thermo Scientific - Pierce Biotechnology, Rockford, IL). NaCl was certified ACS (FisherScientific, Pittsburgh, PA). All the other reagents were obtained from Sigma-Aldrich, St. Louis, MO: EDTA (98.5%, cell culture tested), Na azide (99.5%), NaH_2PO_4 (Reagent plus, 99%), Na_2HPO_4 (Reagent plus, 99%), TFA (99% solution), tetrahydrofuran (THF) (99.9% solution), Tris-HCl (minimum 99% titration, performance certified), HCl (37% solution) and ethanol (99.5% solution) were ACS reagent grade. Buffers were filtered prior to use through a 0.22 μm poly(ether sulfone) membrane filter (Millipore Corp., Bedford, MA).

Chemical modification of the substrates

The immobilization of split-intein partners onto gold-coated cover slips and AFM cantilevers (Novascan, Ames, IA) followed a 4-step protocol, which exploited the high affinity of the antibodies for the proteins fused domains (Scheme 1).

- i. *Preparation and cleaning of the substrates.* Gold-coated surfaces were prepared by deposition onto glass cover slips of a first layer of 15 nm of titanium (99.999% International Advanced Materials) and a second layer of 50 nm of gold (99.999% International Advanced Materials), using an electron beam evaporator²¹. The gold coated surfaces and the AFM probes (tips), supplied with an immobilized 10 μm diameter gold-coated borosilicate glass sphere, were washed with ethanol before chemical modification.
- ii. *Linker immobilization.* Cover slips and AFM probes were immersed in a 2 mM protected disulfide-hydrazide linker solution in ethanol, at 20°C for 24 h, and then thoroughly washed with THF. Before the antibody immobilization, the linker had to be de-protected. To do that, the cover-slips were immersed in trifluoroacetic acid (TFA), at 20°C for 1 h, and then thoroughly washed with 0.1 M phosphate buffer (PB) pH 7.0.
- iii. *Anti-MBP and anti-CBD antibodies immobilization.* In order to functionalize the antibodies with the linker, the glycosylated groups at the Fc region of the antibodies were oxidized, so that the resultant cis-diols were transformed into reactive aldehyde moieties. These aldehydes then were combined with the exposed hydrazide groups of the linker to form stable, leak-resistant hydrazone linkages, so both the antigen-binding sites were free to interact with the antigen in the final step. The oxidation was performed in an amber vial, as the reaction was light sensitive, dissolving 1.5 mg NaIO_4 with 400 μl of phosphate buffer (PB). Then 2 μl of anti-MBP or anti-CBD were added, gently mixed and incubated for 30 min at 20°C. Purification was unnecessary, as only oxidized antibodies could bind and any non-oxidized antibodies were washed away during the binding to the substrates. The

oxidized antibodies were then directly added to the reactive substrates in order to bind the available primary amine groups, and incubated at 20°C for 1 h. Afterwards, the substrates were thoroughly washed with 20 mM Tris HCl, 200 mM NaCl, 1 mM EDTA, 1 mM Na azide, pH 7.4 (intein buffer, IB).

- iv. *Split-intein immobilization.* The reactive substrates were finally incubated with a split-intein solution, obtained from the purification step, at 20°C for 1 h. The samples were then thoroughly washed with IB.

Detection of immobilized antibodies

An ELISA assay was designed to check the anti-MBP and anti-CBD antibodies on gold coated substrates, before intein partner immobilization. The ELISA assay was first optimized to detect the antibodies in solution. Details are given in the Supporting Information (Figure S5). The assay was then adjusted to detect the antibodies immobilized on the gold-coated surfaces. The modified substrate was placed into an empty well of a 96-well plate and the ELISA assay was performed on the samples adding the HRP-linked anti-mouse IgG, the substrate 3,3',5,5'-tetramethylbenzidine and finally the stop reagent. The solution was then transferred into a new well for absorbance reading at 450 nm.

Force measurements

Atomic force microscopy in force mode was used to measure the interactions between native and mixed pairs of split inteins. The adhesion force (F_{ad}) was defined as the maximum adhesive force observed on retraction of the sample away from the tip²¹. Technically, F_{ad} measures an “unbinding” event. Binding between split-intein partners and subsequent splicing (unbinding) of the whole intein are the two steps of biological interest. We focus here on the first step and thus we refer to F_{ad} as a “binding force” rather than an “unbinding force” which would be misleading due to the second step. The average F_{ad} was calculated from a set of 400 data-points for each run. All measurements were taken at an approach speed of 0.6 $\mu\text{m/s}$ and with a dwell time toward the surface of 0.99 s. For imaging with AFM, high resolution topographic images are usually achieved using sharp tips, but for force measurements the mechanical properties of the AFM probe or tip become critical. Also tips with spherical (used here) or paraboloid shapes generate conditions of lowest stresses and strains compared with other profiles, including the pyramidal profile of the most common AFM silicon nitride tips²². For each experiment the spring constant was estimated using a standard 3-steps procedure: (i) determine the virtual deflection in optical path; (ii) determine the slope of the contact region from a deflection-distance curve to determine the “inverse optical lever sensitivity” of the cantilever (in nm/V); (iii) withdraw tip and perform a thermal tune to determine the cantilever’s resonant frequency. The spring constants were 65.88, 51.38 and 38.67 pN/nm for *Nsp*, *Oli* and *Tvu* modified tips, respectively (the cantilever spring constant supplied by the manufacturer was 60 pN/nm). A typical force-distance curve between split-intein partners is shown in Scheme 1. The presence of multiple-peaks during retraction is a fingerprint of the interaction between multiple domains, while the first peak on the left is due to nonspecific interaction between tip and substrate and the farthest peak to the right is the rupture force between the last single molecule interacting across the gap²³.

Equipment

An electron beam evaporator was used to deposit gold onto the substrate (Temescal BJD-1800, BOC Edwards, UK). The force measurements were performed using an atomic force microscope (MFP-3D™ system, Asylum Research, Santa Barbara, CA). The data were collected and analyzed using commercial software (IGOR Pro 5, WaveMetrics, Inc., Lake Oswego, OR).

THEORETICAL SECTION

Analysis of non-equilibrium and equilibrium binding energies

In order to explicitly compare our non-equilibrium measured binding energies to the equilibrium binding energies of the natural Ssp DnaE intein partners ($K_d = 43 \pm 10$ nM) of Shi and Muir¹⁸, we invoke and use the Jarzynski equality relating irreversible work, Δw_i , to the equilibrium free energy difference, ΔG ²⁴. Repeating each force (binding) measurement, $F_i(x)$, $i = N$ times, we obtain the mean work needed to pull the force from position, x_1 , to position x_2 , as:

$$\langle \Delta w \rangle = \frac{\sum_{i=0}^N \Delta w_i}{N} = \frac{\sum_{i=0}^N \int_{x_1}^{x_2} F_i(x) dx}{N} \quad (1)$$

In general $\Delta G \neq \Delta w_i$ and their difference is the dissipative work. The Jarzynski equality states that when the average taken over an *infinite* number of trials (far from the equilibrium) is Boltzmann (K_B) and temperature (T) - weighted according to the work required for each repetition, the result will always be the free energy difference, ΔG :

$$\Delta G = -K_B T \ln \left\langle \exp \left(\frac{-\Delta w_i}{K_B T} \right) \right\rangle \quad (2)$$

Here, we compare equilibrium free energy difference measurements, ΔG , estimated from the literature with our dissipative work, Δw_i , from direct force measurements.

RESULTS AND DISCUSSION

First, evidence is provided that the antibodies are oriented and immobilized covalently on the gold surface correctly. Next, results of direct force measurements between (i) antibodies with fused tags (70–90 pN), (ii) split-intein partners (50–52 pN) and (iii) negative controls (< 30 pN) are presented below. Then, analyses of a comparison of non-equilibrium with equilibrium binding energies and of the molecular structures and sequences at the binding interfaces of the split inteins are presented. Finally, we provide a molecular rationale for similar binding forces among split-intein partners with a bioinformatic analysis that points to a zipper-like interface involving charged and apolar residues.

Evidence of covalent immobilization of antibody via ELISA

The AFM sample preparation requires the covalent immobilization of antibodies on gold coated substrates. In order to verify this arrangement, these surfaces were placed in a 96-well plate for an ELISA assay that was designed for this task. Two controls were included: (i) an empty well, and (ii) a gold coated substrate, after linker immobilization. The empty well was necessary to estimate the nonspecific binding of the HRP-linked antibody, while the gold-coated substrate gave the interaction between the HRP-linked antibody with the substrate (gold and linker) before anti-MBP or anti-CBD antibody immobilization. Results, confirming the detection of immobilized antibodies, are summarized in Table 1. As expected, the background signals, both from the empty well and the substrate (controls), were small compared with those for the modified substrates. Using the calibration curves established for antibodies in solution (Figure S5), it was possible to estimate “equivalent concentrations” of bound antibodies of ~30 ng/ml and ~20 ng/ml from the A450 values of 1.222 and 0.633 for anti-MBP and anti-CBD, respectively. Although we could not translate these values to a surface density, the aim here was to confirm the presence of antibodies on

the gold coated substrates (cover slip and AFM cantilever) and, hence, validate part of the modification protocol.

Direct Force measurements

Antibody-fused tag interactions—In order to measure the molecular interactions between native and mixed pairs of split-intein partners, we first confirmed that the molecular constructs for each surface were present and that the weakest link among these connecting supports was between the pairs of split inteins and not elsewhere (Scheme 1). The disulfide organic links, attached to gold, and the hydrazide links, attached to the oxidized carbohydrate groups on the Fc region of the antibodies, were covalent in nature and therefore could not have released prior to the non-covalent pair interactions of the split inteins²⁵. The only other possible weak link was between the variable region of the monoclonal antibodies and the tagged domains (MBP and CBD) attached as pendants to the split-intein pairs. To test these interactions, we measured their binding energies directly using intein C-domain, I^C, from all three species, which were cloned with both MBP and CBD. Hence, to measure the interaction between CBD and the anti-CBD antibody, we immobilized the anti-CBD antibody on the cantilever probe and the I^C construct on a substrate using MBP and anti-MBP antibody; we swapped the antibodies to measure the interaction between MBP and anti-MBP (Scheme 1). The analysis of the multiple peaks is summarized in Figure 1, where the distribution of the interaction forces is reported for the three inteins and the two antibodies. The resulting distributions exhibited lognormal behavior and were very similar among the three proteins, for the same antibody. Hence, it was possible to stack each of the three distributions for the different proteins and the same antibody, resulting in a “global” distribution with an expected value of 70 [4.39±0.38]** pN and 90 [4.59±0.29] pN for anti-MBP and anti-CBD antibodies, respectively. These values were in good agreement with those reported in the literature for similar experiments between antigens and IgG antibodies, ranging from 60 to 244 pN²⁶. While ELISA only proved the covalent attachment of the antibodies, but not their orientation (e.g. some antibodies may lie down on the surface), the reproducibility of these results with force values in the expected range confirms the covalent and oriented immobilization of antibodies. This clearly would not be observed with non-specific binding. This confirms that the antibody variable regions are exposed and interact with their antigens, i.e. MBP and CBD which are attached as pendants to the split-inteins pairs and hence guarantee their immobilization in the last step of the modification protocol.

Split-inteins interactions—The substrates and cantilever tips were modified with the I^C and I^N partners, respectively (Scheme 1). In particular, the MBP domain that was originally in the I^C fragment together with a CBD tag, was removed during protein purification to avoid cross-interaction between the MBP tag and the anti-MBP antibody immobilized on the substrate and holding the I^N fragment. The cleaved CBD-I^C (11 kDa) was immobilized on the cantilever tip, while the MBP-I^N (55 kDa) was immobilized on the substrate (gold-coated cover slips and AFM cantilevers). Split inteins from three cyanobacteria, *Tvu*, *Nsp* and *Oli* gave 9 possible combinations: 3 native (partners from the same species) and 6 mixed (partners from different species). Single force-distance curves were similar in profile to the antibody/fused tag experiments, showing multiple peaks for each run (Figure 2 including Table 2). Statistical analysis for all the data showed an indistinguishable difference in intermolecular forces between native and mixed pairs. A force of ~52 [4.17±0.47] pN was measured for all 9 combinations. The lognormal profiles, for both the antibody/fused tag and the split-intein experiments, could be interpreted as the result of multiple pair interactions.

**We report the *mode*, point of global maximum of the probability density function of log-normal distributed values, defined as $\exp(\mu - \sigma^2)$, and in parenthesis the mean, μ , $\pm$ the standard deviation, σ , of the variable's natural logarithm.

Thus, the data can be fitted with multiple Gaussian curves, with multiple mean values: 1 μ , 2 μ , 3 μ and so on. A similar force value (50.8 ± 15.8 pN, for the first peak) to the lognormal analysis (~ 52 pN) was obtained by fitting the data with three Gaussian curves, i.e. deconvoluting larger interaction forces into components that represent multiple simultaneously interacting split-intein pairs. Details of this alternative analysis are also presented in Figure 2.

The antibody-fused tag interaction forces of 70 pN for MBP and 90 pN for CBD, were statistically higher (35 and 73%, respectively) than 50–52 pN, measured for split-intein partners. Thus, the possibility of rupturing the antibody-fused tag bond, which was the same for all experiments, was eliminated. Other forces in the constructs were covalent in nature.

The hypothesis that the splicing reaction occurred during the force measurements was also rejected. Previous studies of DnaE *trans*-splicing suggest that association of the split-intein fragments is rapid relative to the subsequent splicing reaction, which occurs in minutes to hours^{16c, 18}. In particular, for *Oli*, *Tvu* and *Nsp*, 30 min is the minimal time at room temperature to clearly observe splicing products on SDS-PAGE^{3a}, while each single AFM run was performed over a few seconds with a dwell time of 1 s. Also, if splicing and extein ligation had taken place during the force measurements, forces of the order of 70 pN and 90 pN for antibody-fused tag interactions or forces < 50 –52 pN between antibodies would have been observed. Neither of these forces was recorded. Thus, splicing was not likely during the AFM measurements based on time scales or force observations.

Negative control—To validate the results and the specific interactions between split-intein partners, the MBP-CBD construct was expressed, purified and used as a negative control. Since it is lacking a split-intein domain, the construct is expected to be non-interacting toward both MBP-I^C-CBD and MBP-I^N. MBP-CBD and MBP-I^C-CBD were immobilized on the substrate and on the cantilever tip, respectively, and the AFM force experiment was repeated for all three I^C proteins, *Tvu*, *Nsp* and *Oli* (Scheme 1). Results are summarized in Figure S6A and they show that a majority of interactions (over 90%, for *Nsp* and *Oli*, and 73%, for *Tvu*) was well below 50 pN. In addition, the histogram plots (Figure S6A) highlight events that showed either no interaction at all or forces below 30 pN, which was chosen as the minimum value of measurable force. An example of raw data is also reported in Supporting Information (Figure S6B).

Comparison of non-equilibrium with equilibrium binding energies

We wish to compare equilibrium with non-equilibrium binding energies (Figure S7). The dissociation constant $K_d = 43 \pm 10$ nM calculated by Shi and Muir¹⁸ was used to estimate the free energy, $\Delta G = -RT \ln(K_d) = 21.4 K_B T$. The distribution of Δw_i measurements was used to estimate the average non-equilibrium work, $\langle \Delta w \rangle = 197 \pm 97 K_B T$ (Figure S7). The results indicate that $\Delta G < \langle \Delta w \rangle$ and ΔG and $\langle \Delta w \rangle$ are of the same order of magnitude ($21.4 K_B T$ and $197 \pm 97 K_B T$, respectively). Since only a finite number of work trajectories were measured here ($N \sim 113$) this result is expected²⁷. These results validate their similar binding energies and hence the reasonableness of their values.

Molecular structures and sequences at the binding interfaces

To probe a molecular rationale for similar binding forces among the different interacting split-intein partners, a bioinformatic sequence analysis was performed. A sequence Logo alignment of *Oli*, *Tvu* and *Nsp* (Figure 3) was examined in relationship to opposing residues in the two halves of the split inteins as modeled using the Phyre2 program²⁸. This analysis shows how many of the highly conserved charged amino acids are found to be (i) clustered in the two long strands comprising most of the contact area between the I^N and I^C split-

intein parts and (ii) participating in intermolecular ion pairs (e.g. Glu85-Arg123) and ion triads (e.g. His48-Asp133-His72 and Glu93-Lys119-Glu61). These two long “contact strands” display three additional non-polar patches comprising hydrophobic residues interspersed between each ion pair previously listed: Pro43/Ile44/Ala45/Trp47-Ile134/Leu136, Val55-Val126/Trp128/Val131 and Leu60-Ile120/Val121. The conservation of these interdigitated electrostatic and non-polar interactions provides a molecular explanation for the similar forces measured among the different split-intein pairs described above.

Shi and Muir¹⁸ and Dassa and co-workers^{3a} have postulated the role of ionic interactions in split-intein assembly. In their work, Dassa et al.^{3a} showed that all 9 combinations of the same split-intein partners studied here can induce protein-splicing despite a 30–50% sequence variation between native split inteins, whether the split-intein halves had opposite charges or when both halves were negative. Similar results were found here in terms of binding forces between, for example, Oli^N (−10.0) with Nsp^C (+2.8) (opposite charges; $\Delta\text{Charge} \sim |12.8|$) or with Oli^C (−0.2) (both negative; $\Delta\text{Charge} \sim |9.8|$). While previously postulated, Shah et al.^{11f} have recently investigated the role of specific ionic interactions for fragment assembly and splicing for the *Nostoc punctiforme* split intein, both *in vivo* and *in vitro*. The authors found that charge complementation or repulsion could bias relative split-intein binding affinities, which in turn could affect relative splicing kinetics. However, the magnitude of these energetic effects did not fully explain the splicing selectivity observed *in vivo*. Knowing the importance of electrostatic²⁹ and non-polar³⁰ attractions in binding, the multiple interdigitated highly conserved ion and non-polar pairs seem to form a complementary components of a zipper which could explain the indistinguishable intermolecular forces between native and mixed pairs measured with AFM. We have repeated the same bioinformatics (sequence Logo) analysis described above with the 13 naturally occurring split inteins (including the original three: *Tvu*, *Nsp* and *Oli*) reported by Dassa et al.^{3a}. Both the analysis of the three split inteins and the 13 inteins (ours plus 10 more inteins) for the C- and N-termini are included in the Supporting Information (Figure S8). It is impressive to see how the same conserved amino acids are present in all 13 split inteins for both the charged amino acids and the non-polar amino acids. These results strongly support and generalize our findings (“zipper-like interface”) and are in concert with the splicing results of Dassa and co-workers^{3a}, suggesting that not only splicing activity, but also fragment association could be the result of local (rather than global) charge distributions between split-intein halves. Given similar interactions between split-intein partners, we speculate that, in spite of 30–50% sequence variation between the constructs, they originated from the same progenitor and diverged while maintaining key electrostatic and non-polar interactions required for association.

One could potentially use this same assay to explore additional fundamental questions on split intein recognition, e.g. one could mutate the highly conserved amino acids along the zipper-like interface, and investigate the importance of both electrostatic and hydrophobic interactions. This immobilization method could also be adopted to intein technologies. For example one could design immobilized split inteins arrays and use them to screen for binding partners by AFM.

CONCLUSIONS

Since the molecular details of binding between split-intein partners before *trans*-splicing are lacking, the goal of this research was to directly measure the intermolecular interactions between the various partners of three split intein species, *Oli*, *Tvu* and *Nsp*, using AFM in force mode. To accomplish this, we synthesized a novel thiol-hydrazide-linker that permitted single-step covalent and oriented immobilization of fusion-tag antibodies (anti-MBP and anti-CBD) onto opposing gold surfaces. The split-intein fragments, I^N and I^C,

were then immobilized via their fusion tags (MBP and CBD), leaving them free to interact with their partner. All 9 pairs – native and mixed – exhibited similar binding forces (~50–52 pN), which were statistically higher than the negative controls of 30 pN, and below 70–90 pN, for the antibody-fused tag interactions.

The main conclusions are that: (i) Split inteins from different species of Cyanobacter had similar binding energies, (ii) both electrostatic and non-polar attraction of aligned ionic and hydrophobic pairs guided partner binding between native and mixed split-intein pairs for all three inteins in a zipper-like fashion (10 other split inteins exhibit the same zipper-like structure at their binding surface), (iii) a new linker was synthesized that is generally useful for oriented covalent immobilization of antibodies on gold for measuring intermolecular forces between proteins or other molecular pairs, and (iv) as required, non-equilibrium dissipative measurements gave larger energies of interaction as compared with those for equilibrium but values of similar order.

Although previous researchers^{3a, 11f, 18} have proposed that electrostatic interactions play a role in split-intein binding, none have suggested that *both* electrostatic and hydrophobic amino acids are important in such binding and that these interactions are interspersed and aligned along a zipper-like binding interface. That these interspersed polar and non-polar amino acids on both strands have very high homology across the three species of Cyanobacteria provides a molecular explanation why split inteins from different species have similar binding energy. The similar avidity is evolutionarily important and relevant to biotechnology and biomedicine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank John T Dansereau, (Wadsworth Center, Albany, NY) Carol L Piazza, Dorie Smith, Brian P Callahan and Natalia I Topilina (SUNY, Albany, NY) for help with cell growth, polyacrylamide gels, Western blots and bioinformatic sequence analysis; Bob Linhardt (Rensselaer Polytechnic Institute, Troy, NY) for suggesting the use of antibodies for the oriented immobilization of the split inteins; Gil Amitai (Weizmann Institute of Science, Rehovot, Israel) for discussions on split-intein results; Amish Patel and Shekhar Garde (Rensselaer Polytechnic Institute, Troy, NY) for discussion on the Jarzynski equality; and Ryan Fuierer (Asylum Research, Santa Barbara, CA) for help in analyzing the AFM force runs.

Funding Sources

This work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, via DOE grants DE-FG02-90ER14114 and DE-FG02-05ER46249 to GB, the US National Science Foundation via NSF-NIRT grants CTS-0304055 and CBET-1122780 to GB, and the National Institutes of Health via NIH grant GM44844 to MB.

ABBREVIATIONS

Oli	<i>Oscillatoria limnetica</i>
Tvu	<i>Thermosynechococcus vulcanus</i>
Nsp	<i>Nostoc species PCC7120</i>
MBP	maltose binding domain
CBD	chitin binding domain
I^N	N-terminal intein domain

I^C	C-terminal intein domain
TFA	trifluoroacetic acid
THF	tetrahydrofuran
PB	phosphate buffer
IB	intein buffer
AFM	atomic force microscopy
<i>F_{ad}</i>	adhesion force

References

1. (a) Bowen WR, Lovitt RW, Wright CJ. *Biotechnol Lett.* 2000; 22:893–903.(b) Horber JK, Miles MJ. *Science.* 2003; 302:1002–1005. [PubMed: 14605360] (c) Hugel T, Seitz M. *Macromol Rapid Comm.* 2001; 22:989–1016.(d) Yang J. *Cell Biochem Biophys.* 2004; 41:435–449. [PubMed: 15509891]
2. Butt HJ, Cappella B, Kappl M. *Surf Sci Rep.* 2005; 59:1–152.
3. (a) Dassa B, Amitai G, Caspi J, Schueler-Furman O, Pietrokovski S. *Biochemistry.* 2007; 46:322–330. [PubMed: 17198403] (b) Perler FB. *Nucleic acids research.* 2002; 30:383–384. [PubMed: 11752343]
4. (a) Gersten DM, Marchalonis JJ. *J Immunol Methods.* 1978; 24:305–309. [PubMed: 82592] (b) Hoffman WL, Oshannessy DJ. *J Immunol Methods.* 1988; 112:113–120. [PubMed: 2457052] (c) O'Shannessy DJ, Hoffman WL. *Biotechnology and applied biochemistry.* 1987; 9:488–496. [PubMed: 2449896] (d) O'Shannessy DJ, Quarles RH. *Journal of applied biochemistry.* 1985; 7:347–355. [PubMed: 3912370] (e) Werner S, Machleidt W. *Eur J Biochem.* 1978; 90:99–105. [PubMed: 213273]
5. Lu B, Smyth MR, O'Kennedy R. *The Analyst.* 1996; 121:29R–32R.
6. Song HY, Zhou X, Hobley J, Su X. *Langmuir: the ACS journal of surfaces and colloids.* 2012; 28:997–1004. [PubMed: 22126088]
7. Alves NJ, Kiziltepe T, Bilgicer B. *Langmuir: the ACS journal of surfaces and colloids.* 2012; 28:9640–9648. [PubMed: 22612330]
8. (a) Brne P, Lim YP, Podgornik A, Barut M, Pihlar B, Strancar A. *Journal of chromatography. A.* 2009; 1216:2658–2663. [PubMed: 19203754] (b) Han HJ, Kannan RM, Wang SX, Mao GZ, Kusanovic JP, Romero R. *Adv Funct Mater.* 2010; 20:409–421.(c) Hong SR, Choi SJ, Jeong HD, Hong S. *Biosensors & bioelectronics.* 2009; 24:1635–1640. [PubMed: 18838265] (d) Qian WP, Xu B, Wu L, Wang CX, Yao DF, Yu F, Yuan CW, Wei Y. *Journal of colloid and interface science.* 1999; 214:16–19. [PubMed: 10328891] (e) Shmanai VV, Nikolayeva TA, Vinokurova LG, Litoshka AA. *BMC biotechnology.* 2001; 1:4. [PubMed: 11545680] (f) Zara JJ, Wood RD, Boon P, Kim CH, Pomato N, Bredehorst R, Vogel CW. *Analytical biochemistry.* 1991; 194:156–162. [PubMed: 1867379]
9. Gering JP, Quaroni L, Chumanov G. *Journal of colloid and interface science.* 2002; 252:50–56. [PubMed: 16290761]
10. (a) Kang JH, Choi HJ, Hwang SY, Han SH, Jeon JY, Lee EK. *Journal of chromatography. A.* 2007; 1161:9–14. [PubMed: 17543979] (b) Mallidi S, Yantsen E, Larson T, Sokolov K, Emelianov S. *P Soc Photo-Opt Ins.* 2008; 6856:85616–85616.
11. (a) Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. *P Natl Acad Sci USA.* 2009; 106:11005–11010.(b) Evans TC Jr, Xu MQ, Pradhan S. *Annual review of plant biology.* 2005; 56:375–392.(c) Noren CJ, Wang JM, Perler FB. *Angew Chem Int Edit.* 2000; 39:450–466.(d) Paulus H. *Annu Rev Biochem.* 2000; 69:447–496. [PubMed: 10966466] (e) Pereira B, Shemella PT, Amitai G, Belfort G, Nayak SK, Belfort M. *Journal of molecular biology.* 2011; 406:430–442. [PubMed: 21185311] (f) Shah NH, Vila-Perello M, Muir TW. *Angewandte Chemie.* 2011; 50:6511–6515. [PubMed: 21656885] (g) Shemella PT, Topilina NI, Soga I, Pereira B, Belfort G,

- Belfort M, Nayak SK. *Biophys J*. 2011; 100:2217–2225. [PubMed: 21539790] (h) Xu MQ, Perler FB. *Embo J*. 1996; 15:5146–5153. [PubMed: 8895558]
12. (a) Cheriyan M, Perler FB. *Advanced drug delivery reviews*. 2009; 61:899–907. [PubMed: 19442693] (b) Chong SR, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ. *Gene*. 1997; 192:271–281. [PubMed: 9224900] (c) Mootz HD. *Chembiochem: a European journal of chemical biology*. 2009; 10:2579–2589. [PubMed: 19708049]
13. (a) Caspi J, Amitai G, Belenkiy O, Pietrokovski S. *Molecular Microbiology*. 2003; 50:1569–1577. [PubMed: 14651639] (b) Choi JJ, Nam KH, Min B, Kim SJ, Soll D, Kwon ST. *Journal of molecular biology*. 2006; 356:1093–1106. [PubMed: 16412462] (c) Iwai H, Zuger S, Jin J, Tam PH. *FEBS letters*. 2006; 580:1853–1858. [PubMed: 16516207] (d) Liu XQ, Yang J. *Journal of Biological Chemistry*. 2003; 278:26315–26318. [PubMed: 12766150] (e) Wu H, Hu ZM, Liu XQ. *P Natl Acad Sci USA*. 1998; 95:9226–9231.
14. Shah NH, Muir TW. *Israel Journal of Chemistry*. 2011; 51:854–861. [PubMed: 23620603]
15. (a) Brenzel S, Kurpiers T, Mootz HD. *Biochemistry*. 2006; 45:1571–1578. [PubMed: 16460004] (b) Mills KV, Lew BM, Jiang SQ, Paulus H. *P Natl Acad Sci USA*. 1998; 95:3543–3548. (c) Volkmann G, Sun WC, Liu XQ. *Protein Sci*. 2009; 18:2393–2402. [PubMed: 19768808] (d) Wu H, Xu MQ, Liu XQ. *Biochimica et biophysica acta*. 1998; 1387:422–432. [PubMed: 9748659]
16. (a) Al-Ali H, Ragan TJ, Gao XX, Harris TK. *Bioconjugate Chem*. 2007; 18:1294–1302. (b) Lew BM, Mills KV, Paulus H. *Biopolymers*. 1999; 51:355–362. [PubMed: 10685046] (c) Martin DD, Xu MQ, Evans TC. *Biochemistry*. 2001; 40:1393–1402. [PubMed: 11170467] (d) Mills KV, Manning JS, Garcia AM, Wuerdeman LA. *Journal of Biological Chemistry*. 2004; 279:20685–20691. [PubMed: 15024006] (e) Nichols NM, Benner JS, Martin DD, Evans TC. *Biochemistry*. 2003; 42:5301–5311. [PubMed: 12731871] (f) Zettler J, Schutz V, Mootz HD. *FEBS letters*. 2009; 583:909–914. [PubMed: 19302791]
17. Martinez LMC, Quintana EEB, Escobedo FA, DeLisa MP. *Biophys J*. 2008; 94:1575–1588. [PubMed: 17993485]
18. Shi J, Muir TW. *J Am Chem Soc*. 2005; 127:6198–6206. [PubMed: 15853324]
19. Ludwig C, Schwarzer D, Mootz HD. *The Journal of biological chemistry*. 2008; 283:25264–25272. [PubMed: 18625708]
20. O'Shannessy DJ, Wilchek M. *Analytical biochemistry*. 1990; 191:1–8. [PubMed: 2127657]
21. Sethuraman A, Han M, Kane RS, Belfort G. *Langmuir: the ACS journal of surfaces and colloids*. 2004; 20:7779–7788. [PubMed: 15323531]
22. Dimitriadis EK, Horkay F, Maresca J, Kachar B, Chadwick RS. *Biophys J*. 2002; 82:2798–2810. [PubMed: 11964265]
23. (a) Fisher TE, Marszalek PE, Fernandez JM. *Nat Struct Biol*. 2000; 7:719–724. [PubMed: 10966637] (b) Lee G, Abdi K, Jiang Y, Michaely P, Bennett V, Marszalek PE. *Nature*. 2006; 440:246–249. [PubMed: 16415852]
24. (a) Crooks GE. *J Stat Phys*. 1998; 90:1481–1487. (b) Crooks GE. *Phys Rev E*. 1999; 60:2721–2726. (c) Jarzynski C. *Phys Rev E*. 1997; 56:5018–5035. (d) Jarzynski C. *Phys Rev Lett*. 1997; 78:2690–2693.
25. Grandbois M, Beyer M, Rief M, Clausen-Schaumann H, Gaub HE. *Science*. 1999; 283:1727–1730. [PubMed: 10073936]
26. (a) Dammer U, Hegner M, Anselmetti D, Wagner P, Dreier M, Huber W, Guntherodt HJ. *Biophys J*. 1996; 70:2437–2441. [PubMed: 9172770] (b) Hinterdorfer P, Baumgartner W, Gruber HJ, Schilcher K, Schindler H. *P Natl Acad Sci USA*. 1996; 93:3477–3481. (c) Stuart JK, Hlady V. *Langmuir: the ACS journal of surfaces and colloids*. 1995; 11:1368–1374.
27. Liphardt J, Dumont S, Smith SB, Tinoco I, Bustamante C. *Science*. 2002; 296:1832–1835. [PubMed: 12052949]
28. Kelley LA, Sternberg MJE. *Nature protocols*. 2009; 4:363–371.
29. (a) Schreiber G, Fersht AR. *Nat Struct Biol*. 1996; 3:427–431. [PubMed: 8612072] (b) Selzer T, Albeck S, Schreiber G. *Nat Struct Biol*. 2000; 7:537–541. [PubMed: 10876236]
30. Sigalov AB. *Prog Biophys Mol Bio*. 2011; 106:525–536. [PubMed: 21867726]

Figure 1. Antibody/fused-tag experiments. Measurements of AFM binding forces for: (A) CBD versus anti-CBD antibody and (B) MBP versus anti-MBP antibody. These experiments were performed using I^C (with both fused domains CBD and MBP, according to Scheme 1) with all 3 species: *Nsp*, *Oli* and *Tvu* (top histograms). The results are stacked together (bottom histogram) and fitted with a lognormal distribution (insert): For CBD $\mu = 4.59$ and $\sigma = 0.29$, while for MBP $\mu = 4.39$ and $\sigma = 0.38$. Each run was repeated 400 times for all combinations.

Figure 2.

Intein/intein experiments. Measurements of AFM binding forces for all 9 combinations of split-intein pairs: I^N from all 3 split inteins *Tvu*, *Oli* and *Nsp* were combined with all I^C according to Scheme 1. (Left) Lognormal analysis for the 3 native (grey background, diagonal) and 6 mixed (white background, off diagonal) pairs. The inserts represent the normal distributions of the variable's natural logarithm (blue square) and their fitting (red line). (Right) Multiple Gaussian peaks analysis. The table summarizes the interaction forces for the individual pair in terms of mean, μ , $\pm$ the standard deviation, σ . The histogram represents the combined AFM intermolecular forces for all 9 split-intein pairs, fitted (red line) using multiple (3) Gaussian curves (white lines). Each run was repeated 400 times for all combinations.

Figure 3.

Analysis of the molecular interfaces of the split-inteins pairs. (Top) Sequence Logo based on the segment alignment of *Nsp*, *Oli* and *Tvu*, for both I^N (above) and I^C (below), highlighting the highly conserved charged (left, basic in blue and acidic in red) and non-polar (right, black) residues. (Bottom) The structure of the complete molecule *Tvu* (middle) was determined using Phyre2²⁸: I^N , residues 1–118 (orange), and I^C , residues 119–152 (green). Selected portions of the molecule highlight the intermolecular electrostatic (left) and non-polar (right) interactions between I^N , residues 43–62 (orange) and I^C , residues 119–136 (green). The red and blue outlines (left) identify acidic and basic patches, respectively. Interacting residues are shown in the different boxes.

Scheme 1.

Experimental design. (Top) Chemistry cartoon of the immobilization of split inteins on gold-coated substrates: (A) linker synthesis; (B) immobilization of the disulfide hydrazide linker and its deprotection using trifluoroacetic acid (TFA); (C) immobilization of the antibodies, after oxidation of glycosylated groups at the Fc region; (D) immobilization of split-intein partners. (Bottom) Cartoon of the AFM experimental set up: Control (antibody/fused-tag) experiments, for measuring the interaction between the antibodies and the tagged domains of the split-intein partners, CBD and MBP; negative control, using a non-interacting protein, MBP-CBD in this case; and intein/intein interaction measurement, for measuring the association between the two split proteins. In the dashed circle, a typical AFM force-distance curve between split-intein partners, where the approach (into contact) and the separation (out of contact) curves are shown. The rupture force F_R is shown as the last force during separation. I^C and I^N are the C- and N-terminal intein domain, respectively; MBP, maltose-binding domain; CBD, chitin-binding domain; anti-MBP and anti-CBD monoclonal antibodies against the fused domains MBP and CBD, respectively.

Table 1

ELISA assay of substrates: A_{450}^a of controls and samples after antibody immobilization.

Sample	A_{450}^a
Empty well ^b (control)	0.062
Gold + linker ^c (control)	0.061
Gold + linker + anti-MBP antibody	1.222
Gold + linker + anti-CBD antibody	0.633

^a Absorbance reading at 450 nm.

^b Polystyrene microtiter plate.

^c Disulfide hydrazide.

Table 2

Interaction forces between immobilized split-intein partners.

F_{ad} (pN)	Tvu^C	Oli^C	Nsp^C
Tvu^N	53 [4.19±0.47]	52 [4.17±0.48]	52 [4.22±0.51]
Oli^N	52 [4.18±0.48]	52 [4.17±0.48]	60 [4.33±0.49]
Nsp^N	49 [4.16±0.51]	52 [4.14±0.44]	52 [4.19±0.50]

Tvu, *Oli* and *Nsp* denote the 3 cyanobacteria, *O. limnetica*, *T. vulcanus* and *Nostoc species PCC7120*, respectively. C- and N- denote the I^C and I^N components.