Arranging Small Molecules with Subnanometer Precision on DNA Origami Substrates for the Single-Molecule Investigation of Protein–Ligand Interactions

Jingyuan Huang, Antonio Suma, Meiying Cui, Guido Grundmeier, Vincenzo Carnevale, Yixin Zhang,* Charlotte Kielar, and Adrian Keller*

DNA origami nanostructures are versatile substrates for the single-molecule investigation of biomolecular interactions as they enable the display of molecular species in complex arrangements. Herein, the fundamental limitations of this approach are explored by displaying pairs of small-molecule ligands of the protein trypsin on DNA origami substrates and adjusting their ligand–ligand spacing with subnanometer precision. Bidentate binding of trypsin to the ligand pairs is investigated by atomic force microscopy (AFM), microscale thermophoresis (MST), and molecular dynamics simulations. Bidentate trypsin binding is strongly affected by the distance of the ligand pairs and the accessibility of the protein's binding pockets. MST cannot resolve the differences in bidentate trypsin binding because of the nonspecific binding of trypsin to the DNA origami substrates, rendering the AFM-based single-molecule detection of binding events superior to ensemble measurements. Finally, even monodentate binding to a single ligand may be affected by subnanometer variations in its position, highlighting the importance of local microenvironments that vary even over molecular distances. While this single-molecule approach can provide viable information on the effects of ligand arrangements on bidentate protein binding, in-depth investigations into the nature of local microenvironments will be required to exploit its full potential.

1. Introduction

During the last decade, DNA origami nanostructures evolved into versatile platforms for the investigation of numerous biomolecular processes and reactions.[1] This development was initiated by the unique addressability of DNA origami nanostructures, which allows for the site-selective immobilization of multiple biomolecular species with nanometer resolution on the DNA origami surface. Using this approach, a vast variety of biomolecular processes have been investigated so far, including DNA–DNA[2] and DNA–RNA hybridization,[3] protein–DNA binding,[4] conformational transitions in DNA,[5] DNA radiation damage,[6] enzymatic DNA repair,[7] enzyme cascades,[8] antibody–antigen binding,[9–11] and even DNA transcription.[12] The aims of these studies were either to provide fundamental insights into the molecular mechanisms of biological processes or to improve the detection of medically relevant biomolecules. Another class of biomolecular reactions of great importance in a more technological context comprises the binding of proteins to small-molecule ligands, which is at the heart of drug screening and fragment-based drug discovery (FBDD).[13,14]

The fundamental idea behind the FBDD approach is to identify and select few low-molecular mass fragments that efficiently bind to neighboring sites of the same target protein and subsequently merge them to construct a potent small-molecule protein inhibitor.[14] The latter step, however, often presents an enormous challenge as the linking chemistry

J. Huang, Prof. G. Grundmeier, Dr. C. Kielar,[+] Dr. A. Keller
Technical and Macromolecular Chemistry
Paderborn University
Warburger Str. 100, 33098 Paderborn, Germany
E-mail: adrian.keller@uni-paderborn.de

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/sstr.202000038.

[+]Present address: Institute of Resource Ecology, Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstraße 400, 01328 Dresden, Germany

Dr. A. Suma
Dipartimento di Fisica
Università di Bari, and Sezione INFN di Bari
via Amendola 173, 70126 Bari, Italy

Dr. A. Suma, Prof. V. Carnevale
Institute for Computational Molecular Science and Department of Biology
Temple University
Philadelphia, PA 19122, USA

M. Cui, Prof. Y. Zhang
B CUBE - Center for Molecular Bioengineering
Technische Universität Dresden
Tatzberg 41, 01307 Dresden, Germany
E-mail: yixin.zhang1@tu-dresden.de

ADVANCED
SCIENCE NEWS

www.advancedsciencenews.com

**small
structures**

www.small-structures.com

used to merge the discovered fragments usually has a tremendous effect on the affinity of the resulting small-molecule inhibitor toward the target protein and may therefore result in drastically reduced potency.[15] Optimizing fragment linking chemistry and linker properties thus represents an essential yet difficult and time-consuming task in FBDD.

In our previous work,[16] we demonstrated the synthesis of nanoarrays consisting of low-molecular mass fragments immobilized on DNA origami substrates for the single-molecule detection of mono- and bidentate protein binding events. To this end, different ligands were directly conjugated to selected staple strands and incorporated into 2D DNA origami triangles during assembly. Binding of three different proteins to various ligands and ligand pairs was detected by atomic force microscopy (AFM). Several different binding events could be directly observed and quantified in this way, including both symmetric and asymmetric bidentate protein binding. For these bidentate binding events, we could further demonstrate that the lengths of the single-stranded DNA spacers that connected the ligands to the DNA origami substrate had a strong effect on the bidentate binding yields, as they affect the 3D arrangement of the two ligands.

Building on this approach, we here explore the possibility to vary the spacing of chemical modifications on DNA origami substrates with subnanometer precision. While such a precise fine-tuning of ligand spacing will enable the quantitative evaluation of the effect of ligand pair spacing on target protein binding in FBDD, it may also find applications in other fields, such as high-precision DNA origami actuation.[17] To evaluate the viability of this approach, we use the protease trypsin[18] as an established model system that undergoes asymmetric bidentate binding to 4-aminobenzamidine (**B**) paired with 3-iodophenyl isothiocyanate (**I**).[16,19] We have chosen this relatively weak yet well-studied bidentate interaction because such interactions are often the starting point for drug development but difficult for ensemble measurements. We vary the spacing of both ligands in single-bp increments and quantify the trypsin binding yields at a single-molecule level using AFM. We find that the determined

single-molecule bidentate trypsin binding yields are strongly affected by the distance of the ligand pairs on the DNA origami substrates and the accessibility of the protein's binding pockets to the ligands, as determined from molecular dynamics simulations using the oxDNA2 coarse-grained model.[20,21] In particular, all these trends point toward a large sensitivity toward small variations in ligand spacing. Ensemble microscale thermophoresis (MST) measurements, however, cannot resolve this difference in bidentate trypsin binding, as these measurements are dominated by the nonspecific binding of trypsin to the DNA origami substrates. Single-molecule detection of binding events by AFM can thus be considered superior to ensemble measurements. Finally, we also evaluate the effect of local microenvironments[9] at the DNA origami–electrolyte interface on the trypsin–ligand interaction. We find that even subnanometer variations in the position of a single ligand can result in considerable changes in the determined single-molecule binding yield. Despite not being able to yield thermodynamic data such as K_d values, this single-molecule assay nevertheless can provide viable information on the effects of ligand arrangements on bidentate protein binding, in particular when combined with simulations. Our results also highlight the importance of microenvironments, which appear to vary strongly even over molecular distances. Redundant placement of ligands at several positions on a single DNA origami substrate may thus be required to obtain a more realistic measure of the binding yield by averaging over different microenvironments.

2. Results and Discussion

2.1. Tuning Molecular Arrangements on DNA Origami Substrates with Subnanometer Precision

The general approach is schematically shown in **Figure 1**. The two ligands **B** and **I** are attached to the 5′ and 3′ ends of two neighboring staples in the DNA origami triangle, respectively. To allow for some conformational freedom, single-stranded
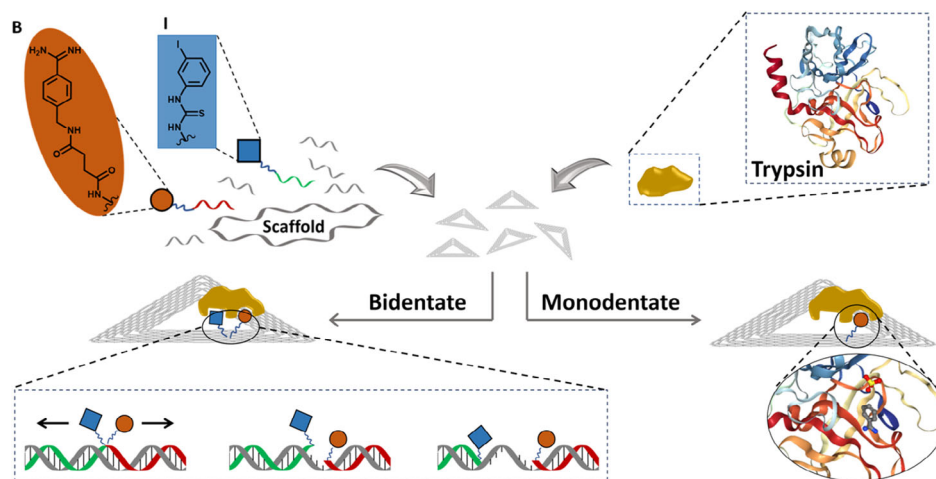

Figure 1. Schematic representation of the experimental approach. **B**- and **I**-modified staples are incorporated into a DNA origami triangle during assembly. Depending on the number and arrangement of the ligands, exposure to trypsin may lead to either monodentate or bidentate binding. By omitting scaffold-binding nucleotides located next to the ligand-modified T_4 spacers from the staple sequences, the spacing of the ligand pairs for bidentate trypsin binding can be varied in single-bp increments. Images of trypsin with and without bound benzamidine from the RCSB PDB (rcsb.org) of PDB ID 1CE5.[22]

DNA spacers consisting of four thymines are introduced between ligand and DNA origami surface. Without any additional modifications of the scaffold-hybridizing domains of the two staples, the spacing of the two ligands is determined only by those DNA spacers. To vary the spacing, a defined number of nucleotides adjacent to the DNA spacers is omitted from one or both of the ligand-carrying staples. This then results in the spacing of the two ligands being increased in single-bp increments, with the T_4 spacers being separated by a defined number of unpaired scaffold nucleotides. In a normal DNA duplex, such a short region of unpaired nucleotides would behave like a fluctuation-prone hinge and thus result in an undefined ligand spacing. As this molecular motif, however, is embedded in the rigid DNA origami framework, the planar duplex arrangement will be stabilized even in the presence of unpaired scaffold nucleotides. In this way, the spacing of the two ligands can be adjusted in a well-defined molecular configuration with a resolution of less than one nanometer.

One further aspect needs additionally to be considered when using this approach. The deletion of staple nucleotides results not only an additional spacing of the two DNA spacers but also causes them to move around the duplex (see Figure 1), which will gradually reduce the distance between ligand and DNA origami, until one or both of the ligands will protrude from the opposite face of the DNA origami substrate. This implies that 1) the single-stranded DNA spacers are becoming more essential for ensuring conformational freedom of the ligands the further they move down into the DNA origami plane, and 2) the total spacing of the DNA spacers that can be achieved is limited to about 6 bp, corresponding to one-quarter rotation around the double helix for each of the staples.

2.2. Effect of Ligand Arrangement on Bidentate Trypsin Binding

Figure 2 shows selected AFM images of triangular DNA origami nanostructures displaying **B** + **I** ligand pairs at different ligand spacings after exposure to trypsin. Trypsin has a diameter of about 3 nm and can thus be easily resolved in AFM images.[16] While a large fraction of the DNA origami triangles appears to be empty in all images, there are always a few DNA origami substrates carrying a single trypsin molecule at the position of the ligand pairs. From such AFM images, the bidentate trypsin binding yields were determined as the percentage of DNA origami triangles with trypsin-occupied ligand pairs.[16] However, some of the DNA origami triangles visible in the AFM images in Figure 2 also carry proteins at positions that have not been decorated with ligands, which indicates that trypsin can also adsorb nonspecifically to the DNA origami surface. To evaluate the importance of nonspecific adsorption versus specific ligand binding, a control experiment has been performed using DNA origami triangles with no displayed ligands. The nonspecific binding yield of trypsin
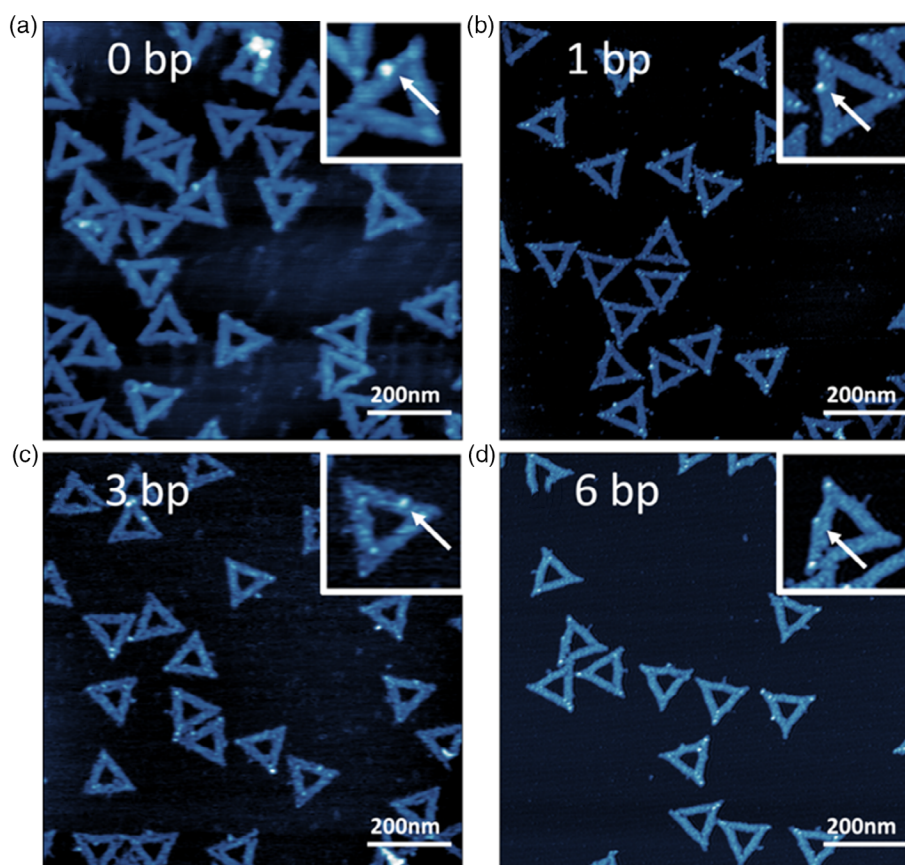

Figure 2. AFM images of DNA origami triangles carrying **B** + **I** ligand pairs with ligand spacings of a) 0, b) 1, c) 3, and d) 6 bp, after exposure to trypsin. Insets show zooms of single DNA origami substrates with ligand-bound trypsin (white arrows).

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**small
structures**

www.small-structures.com

molecules adsorbed nonspecifically at the potential modification site was about 10% (see Figure S3, Supporting Information), which can thus be considered as background noise.

As shown in **Figure 3**a, there is a pronounced and complex dependence of the single-molecule binding yield on the ligand spacing. At a spacing of 0 bp, a binding yield of 22% is obtained. Increasing the spacing by only 1 bp, however, results in a drastically increased binding yield of almost 60%. Such drastic increase in the binding yield is indicative of a change from monodentate binding to the weak trypsin binder **B** to strong bidentate binding to the ligand pair **B + I**.[19] This suggests that at a spacing of 0 bp, the two ligands are too close to each other to facilitate efficient bidentate binding to the protein. At a larger spacing of 2 bp, the binding yield drops again to around 30%, suggesting that a spacing of 1 bp is indeed the optimum spacing, which allows both ligands to bind to their respective pockets simultaneously.

For ligand spacings larger than 2 bp, a continuous decrease in the binding yield with spacing distance is observed (Figure 3a). As the recorded binding yield is averaged over several hundred

DNA origami substrates, it may include both mono- and bidentate binding events, which, unfortunately, cannot be distinguished by AFM. The continuous decrease in the binding yield in Figure 3a thus can be explained by the continuous decrease in the fraction of bidentate binding events with increasing ligand spacing until the recorded binding yield reaches the value of the pure monodentate binding yield. This indicates that with increasing ligand spacing, bidentate binding is increasingly disfavored. This can be rationalized in the following way. With a further increase in distance, both ligands attached to the flexible spacers may still reach their respective pockets. However, the restriction of a small molecule's motion upon binding to a protein causes a loss of configurational entropy and thus a penalty in binding affinity. Increasing the distance between two ligands on two flexible spacers will increase the configurational entropy and thus result in a higher entropic penalty upon binding to the protein. Nevertheless, as long as the two ligands on the flexible T_4 spacers can overlap with each other, bidentate binding events may still occur but with lower probability. Note that this
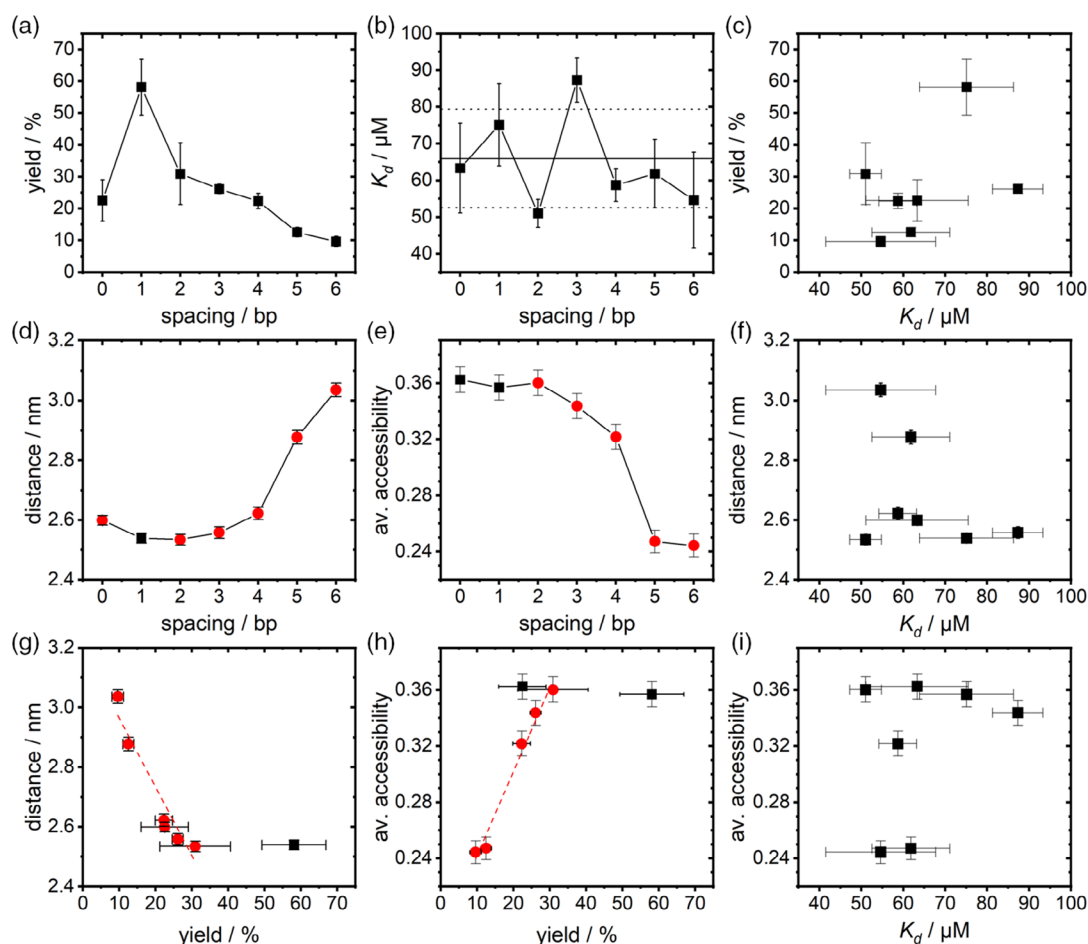

Figure 3. Analysis of bidentate trypsin binding to **B + I** ligand pairs as a function of ligand spacing. a) Single-molecule binding yields determined by AFM. b) Dissociation constants K_d determined by MST. The solid horizontal line indicates the K_d determined for a blank DNA origami substrate without any ligands with the error range given by the dotted horizontal lines. c) Cross-correlation between K_d and binding yield. d) Distance of the ligands determined from the simulations. e) Average accessibility of the ligand pairs determined from the simulations. f) Cross-correlation between K_d and ligand distance. g) Cross-correlation between binding yield and ligand distance. h) Cross-correlation between binding yield and average accessibility. i) Cross-correlation between K_d and average accessibility. For the red data points in parts (d), (e), (g), and (h), the linear correlations represented by the straight broken lines in parts (g) and (h) were obtained.

situation is markedly different from the recent experiments of Zhang et al., where the authors investigated bidentate binding of IgG antibodies to differently spaced epitope pairs immobilized on DNA origami substrates.[11] Here, the epitopes were directly attached to the ends of the staple strands without any additional single-stranded DNA spacers. Therefore, only the two arms of the IgG antibodies could provide for some flexibility, which, however, are more rigid than DNA single strands.

Optimizing the linkage between two ligands to achieve optimal bidentate binding to a target protein represents a long-lasting challenge in biomolecular design. For example, various linkages have recently been incorporated into the design of a DNA-encoded chemical library and it was observed that not only the length but also the structure of the linkages can have a remarkable effect on target binding.[23] While our approach could provide a potential platform for the detailed study of such effects, the immobilization of the DNA-bound ligand pairs on a DNA origami substrate may come at the cost of additional negative effects on protein binding. For instance, increasing the spacing distance will cause the ligands to move closer to the surface of the DNA origami (see earlier), which may result in additional steric hindrance and unfavorable conformational changes in the DNA spacers. Furthermore, as discussed later, the presence of the DNA origami surface in the close vicinity of the ligand may modulate its affinity for the protein in nontrivial ways.

To evaluate the effect of omitting staple nucleotides on the conformation of the DNA spacers and the geometric arrangement of the ligands in detail, we turned to molecular dynamics simulations, using the oxDNA2 coarse-grained model.[20,21] More specifically, we studied only the trapezoid portion of the Rothemund triangle[24] containing the ligand pairs (**Figure 4**a), considering all seven cases of ligand spacing (0–6 bp). The oxDNA2 model parametrizes DNA at the level of a single nucleotide, and reproduces typical experimental properties of DNA

probed in solution. These include base pairing, torsional rigidity, persistence length, and melting temperature. We are therefore confident that the fluctuations we sampled for small strand distances are consistent with those of real experiments of DNA origami in solution. The quantities determined from these simulations were the ligand distance (Figure 3d and 4d) and the protein's accessibility (Figure 3e). The former quantity is tightly related to the actual orientation and conformation of the two single-stranded T_4 spacers. As the oxDNA2 model takes spiral directionality automatically into account when reproducing the helical form and has single nucleotide resolution as its core element, it will accurately reproduce the 3D structure of the protruding DNA spacers for the various spacings (see Figure S4, Supporting Information). The latter quantity measures, for each trajectory's frame, the steric overlap of trypsin, approximated as a sphere with the same radius of gyration R_g (Figure 4b,c), with the DNA origami substrate (see Section 4). This rather crude approximation was chosen for three reasons. First, the protein itself has a globular fold (see Figure 4b), which makes a sphere a rather obvious choice. Second, the binding site of fragment **I** is not known yet[19] and its identification would require atomistic modeling, which is out of scope for this work. Third, we are performing coarse-grained simulations for the DNA origami, and higher precision would not be beneficial in this context. Thus, further assumptions regarding the relevant geometrical shape beyond that of a simple sphere are currently not justified in the context of bidentate binding to **B** + **I**.

As shown in Figure 3d, there is a rather large discrepancy between the designed spacing of the staple modifications and the actual distance of the ligands. Approximating the spacing increment of one base pair with a distance of 0.34 nm, the former should range from 0 nm at a spacing of 0 bp to about 2 nm at 6 bp. In contrast, the average ligand distances shown in Figure 3d range from about 2.5 to more than 3 nm.
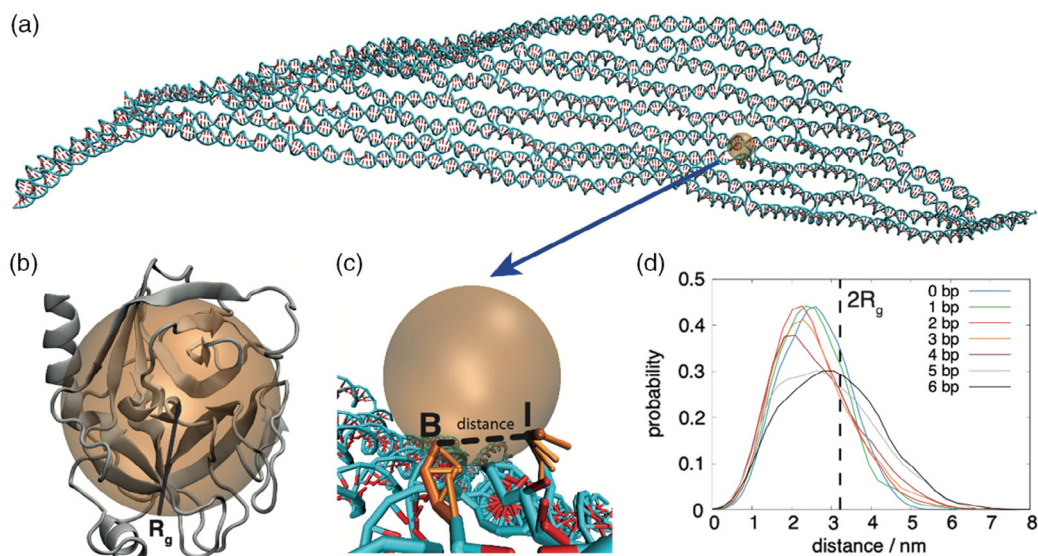

Figure 4. a) oxDNA conformation of the trapezoid of the Rothemund triangle carrying DNA spacers at a 4 bp spacing. b) Approximating the volume of trypsin by a sphere with the same radius as the radius of gyration $R_g = 1.6$ nm of the protein. Image from the RCSB PDB (rcsb.org) of PDB ID 1CE5.[22] c) Same sphere docked onto the edges of the two DNA spacers (highlighted in orange) corresponding to **B** and **I**. d) Probability distributions of the edges of the DNA spacers for spacings of 0–6 bp.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**small
structures**

www.small-structures.com

Furthermore, for the spacings between 0 and 4 bp, there is only a very weak dependence of the measured ligand distance on the designed spacing. These rather unexpected observations can be attributed to the important role of the T_4 DNA spacers. In our previous work, we observed a strong influence of the length of the single-stranded spacers on the bidentate binding yield, which we explained by the two negatively charged single strands behaving as neighboring entropic springs that repel each other.[16] This interpretation is now supported by our simulations, as is evident from the snapshots shown in Figure S4, Supporting Information, which clearly show that small variations in the designed spacing result in rather pronounced changes in the actual conformation and orientation of the two DNA spacers. Therefore, we can conclude that while the designed spacing of the modifications can indeed be used to fine-tune the ligand–ligand distance at a subnanometer level, the actual distances may deviate quite drastically from the designed ones because of the strong contributions of the used DNA spacers. This underlines the great importance of combining experimental measurements of protein binding to simulations of the designed DNA origami structures to determine the actual ligand pair distance.

Qualitatively, the calculated average distance of the ligand pair has a similar (inverted) trend as the bidentate binding yield determined by AFM (see Figure 3a,d). While the latter has a maximum at 1 bp spacing, the former has a minimum at around 1–2 bp. Indeed, as shown in Figure 3g, we find an inverse linear correlation between the single-molecule binding yield and the ligand distance for most data points. Only the highest binding yield obtained for the 1 bp spacing cannot be described by this correlation. At the same time, the average accessibility (Figure 3e) points to an effective reduction in the steric space available for docking onto the ligands: small spacings of 0–2 bp yield about identical accessibilities, while for larger spacings, accessibility is decreasing continuously with spacing distance. For this range of spacings ≥ 2 bp, we again obtain an almost perfect linear correlation with the single-molecule binding yields determined from AFM, as shown in Figure 3h. These linear correlations between binding yield and distance and between binding yield and accessibility indicate that within this spacing range, there is a superposition of mono- and bidentate binding: the larger the ligand distance, the fewer DNA origami will show bidentate binding of trypsin to the **B + I** ligand pair. In the absence of other effects (see later), monodentate binding to **B**, however, should not be affected by ligand spacing.

Only the two smallest spacings of 0 and 1 bp do not follow the linear correlation between binding yield and accessibility. This is because our approach to determining ligand pair accessibility assumes identical ligand binding affinities for the whole surface of the protein. In reality, however, each ligand is binding to a specific site. Whereas benzamidine binds to the trypsin catalytic pocket[22] (see crystal structure in Figure 1), the binding site of **I** is not known.[19] Based on our AFM measurements (Figure 3a) and molecular dynamics simulations (Figure 3d), we estimate that the **I**-binding site addressed in our experiments has a distance of less than 2.54 nm from the benzamidine binding pocket. When the two ligands can reach their respective binding pockets at this ligand distance, bidentate binding results in a drastically enhanced affinity,[19] which is reflected in the sharp peak in the single-molecule binding yield at 1 bp spacing (Figure 3a).

Protein–ligand interactions, as well as other effects due to explicitly accounting for electrostatic interactions and solvent exposure, however, are not considered in our simulations at this level of coarse graining (see Section 4), which thus cannot predict how geometric changes in the ligand arrangement will affect protein binding affinity.

Aforementioned single-molecule experiments were conducted at very low trypsin concentrations of 80 nM to minimize nonspecific protein adsorption at the mica substrate and thus reduce imaging artifacts.[16] Therefore, we next set out to determine whether the observed ligand spacing dependence of the bidentate trypsin binding yield is also reflected in the corresponding K_d values determined by ensemble measurements. For this, we turned to MST, which can readily be used in bulk solution and is thus not affected by nonspecific protein adsorption at the liquid–solid interface. It allowed us to measure dissociation constant using fluorescently labeled DNA origami nanostructures at a low nM concentration and varied concentrations of trypsin in the range from about 0.7 to 500 μM. As shown in Figure 3b, the obtained K_d values scatter between 50 and 90 μM without any clear trend. Consequently, no correlations with the single-molecule binding yield, the ligand distance, or the average ligand pair accessibility could be established either (see Figure 3c,f,i). The origin of this discrepancy between the single-molecule AFM and the ensemble MST measurements is shown in Figure 3b, where the horizontal lines indicate the K_d and corresponding margin of error of a blank DNA origami triangle without any trypsin-binding ligands. The margin of error covers almost the complete K_d range of the ligand-presenting DNA origami nanostructures. This suggests that the trypsin–ligand interaction is masked by a nonspecific interaction between the protein and the DNA origami substrate. Indeed, bovine trypsin has an isoelectric point >10 and is thus positively charged at pH 7.5.[25] Therefore, we assume that at μM trypsin concentrations, the interaction between trypsin and the negatively charged DNA origami substrate is dominated by nonspecific electrostatic attraction. In contrast, the AFM-based single-molecule assay uses a much lower trypsin concentration, employs sample washing to remove unbound proteins, and specifically detects ligand-bound trypsin (see Figure S3, Supporting Information), whereas nonspecifically adsorbed proteins at nonligand sites, as shown in Figure 2, are not included in the analysis. It thus enables the detailed investigation of protein binding events that are not accessible by standard ensemble measurements.

2.3. The Role of Local Microenvironments

Recently, it was shown by Zhang et al. that different binding yields may be obtained for the same protein–ligand system if the ligand is immobilized at different positions on the same DNA origami surface.[9] These differences were attributed to the presence of locally different microenvironments at the DNA origami–electrolyte interface, which somehow affect the protein–ligand interaction. However, the shortest distance between two ligand sites that was evaluated in this study was about 20 nm. To assess whether such microenvironment-related effects also act over much shorter molecular length scales, we

investigated monodentate trypsin binding to each ligand at the four different positions used for adjusting ligand pair spacing (see Figure 1). The distance between two ligand sites thus ranges from 1 to 3 bp. Despite these small differences in position, an astonishingly strong influence on the monodentate binding yields for both ligands is shown in **Figure 5**. Even more surprising, there is no continuous trend. Rather, most positions exhibit a very similar monodentate binding yield, whereas single positions have a much lower one. The lowest monodentate binding yield close to the nonspecific binding yield is observed for the **B** ligand with three omitted nucleotides, which should thus have the smallest vertical distance to the DNA origami surface and may thus experience the strongest steric hindrance. However, our simulations revealed almost identical ligand accessibilities for the different positions, ranging from 0.87 to 0.95 (see Figure S6, Supporting Information). Therefore, it is rather unlikely that steric hindrance is the origin of this peculiar distance dependence. This interpretation is further supported by the fact that a completely different distance dependence is observed for the **I** ligand, where the lowest binding yield is obtained at only one omitted nucleotide. In prinicple, such differences could also be caused by different incorporation probabilities of the different staples.[26] However, shortening the scaffold-hybridizing region of the staples by three nucleotides reduces their melting temperatures T_m by a mere 3–4 °C (see Table S1, Supporting Information),[27] so that even the shortened staples lie right in the center of the T_m distribution of the complete DNA origami triangle.[28] Therefore, we assume very similar incorporation probabilities of the different staples. Different microenvironments thus remain the most likely explanation for the observed differences in the binding yields. These may include local differences in the DNA origami's hydration shell and electric double layer, both of which play important roles in protein–ligand, protein–protein, and protein–DNA interactions.[29] Furthermore, it is well known that both DNA hydration and ion–DNA interactions depend on the sequence, conformation, and mechanical properties of the DNA duplex.[30] Therefore, it appears rather likely that the hydration shell and the electric double layer of a large and highly strained DNA origami nanostructure exhibit significant local variations, which, when overlapping with the small-molecule ligand, may result in locally different ligand accessibilities and protein binding affinities, and even render the ligand virtually inaccessible for the protein.

The data shown in Figure 5 also exhibit a few other, similarly surprising features. Most importantly, **B** is known as a weak trypsin inhibitor with an IC_{50} value of 100 μM, whereas the trypsin affinity of **I** alone is so low that no trypsin binding could be detected yet.[19] In contrast, our single-molecule assay revealed monodentate **I** binding yields that in most cases are similar to or even higher than those of **B**. While we can rule out that this is caused by a direct interaction between trypsin and the single-stranded DNA spacer (see Figure S3, Supporting Information), we can only speculate about the origin of this observation. For instance, **I** is a trypsin binder discovered from a DNA-encoded chemical library. Conjugation of **I** to DNA could facilitate the interaction. A similar effect has also been observed for an albumin binder discovered from a DNA-encoded chemical library.[31] Aforementioned single-molecule (AFM) and ensemble (MST) experiments indicate the nonspecific, reversible adsorption of trypsin at the DNA origami surface. In the present case, the nonspecific binding of trypsin to the DNA origami surface might thus facilitate its specific binding to the DNA origami-immobilized **I** ligand, for instance, by helping the **I** ligand to overcome a repulsive potential barrier upon entering its binding pocket. In this case, the DNA origami substrate would play a similar role as the **B** ligand in previous studies that observed drastically enhanced trypsin binding for **B** + **I** conjugates.[19] Finally, the presence of the DNA origami substrate itself may have an effect on trypsin–ligand binding through the same mechanisms discussed with respect to the local microenvironments, i.e., hydration shell and electric double layer.

Independent of the origin of the astonishingly high monodentate **I** binding yields, the data shown in Figure 5 raise yet another interesting issue. For a large ligand spacing of 6 bp, the bidentate binding yield in Figure 3a approaches a value of about 10%, similar to the nonspecific binding yield (see Figure S3, Supporting Information). For the same ligand position, however, a monodentate **I** binding yield of almost 60% is determined (see Figure 5b). Similar observations, i.e., monodentate binding yields that are larger than the corresponding bidentate yield, can be made also for several other spacings and positions, including the full-length staples used for the 0 bp spacing. Again, we can
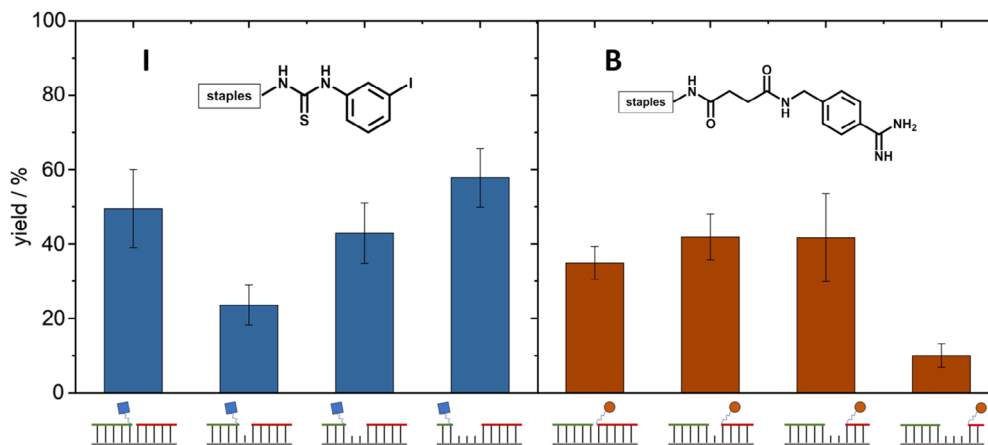

Figure 5. Single-molecule yields of monodentate trypsin binding to either **I** or **B** immobilized at different positions on the DNA origami surface.

only speculate that this counterintuitive behavior results from the local microenvironment surrounding the ligand site, which seems to be affected by the presence of another ligand and/or single-stranded DNA spacer in the close vicinity. In this context, it is not unreasonable to assume that a 2.5 nm long negatively charged polyelectrolyte protruding from a negatively charged surface affects the electric double layer structure of that surface within a 2.5 nm radius and thereby influences the binding behavior of a positively charged protein with 3 nm diameter at a neighboring site within this radius.

3. Conclusion

In summary, we have investigated the possibility to arrange small-molecule ligands with subnanometer precision on DNA origami substrates to study the effect of ligand pair spacing on bidentate protein binding. It is important to note that we have used weakly binding ligands for this study. Such systems are often the starting points of drug design and development, especially for FBDD. By varying the spacing of the two ligands in single-bp increments, we could determine the optimum arrangement of the **B** + **I** ligand pair for efficient bidentate trypsin binding. This was achieved by a combination of single-molecule AFM detection of protein binding and molecular dynamics simulations of the molecular structure of the DNA origami-immobilized ligand pairs, which was found to depend not only on the designed spacing but also critically on the conformation of the used single-stranded DNA spacers. We furthermore found that such single-molecule AFM measurements are superior to ensemble MST measurements, in which the specific protein–ligand interaction is masked by a nonspecific electrostatic interaction of the positively charged trypsin molecules with the negatively charged DNA origami substrates at high trypsin concentrations. Even though this single-molecule assay does not yield absolute protein–ligand affinities, it can nevertheless provide viable information on the effects of ligand arrangements on bidentate protein binding, in particular when combined with simulations.

However, as we could further demonstrate in this work, the anchoring of small-molecule ligands to a DNA origami surface itself may affect protein binding. For instance, surprisingly high monodentate trypsin binding yields have been determined for ligand **I**, which may even overcome those of the much stronger trypsin binder **B**. This may again be caused by the additional electrostatic interaction between protein and DNA origami, or by the properties of the DNA origami–electrolyte interface, that differ from those of the bulk solution. This DNA origami–electrolyte interface furthermore appears to be rather inhomogeneous even over molecular length scales, resulting in different monodentate trypsin binding yields for the same ligand attached to different locations with subnanometer distances. In addition, our results suggest that these local microenvironments are affected by the presence of small molecules and/or DNA sequences protruding from the DNA origami surface, which may result in altered protein binding at a nearby ligand site at few nanometers distance. Furthermore, it can be expected that these microenvironment-related effects will show different magnitudes and dependencies for different protein–ligand systems.

Currently, it appears that only the implementation of high redundancy, i.e., the placement of several identical ligands on multiple positions on the same DNA origami surface to average over different microenvironments, may allow for the determination of realistic binding yields. However, this will significantly affect high-throughput screening campaigns as it will drastically reduce the number of ligands that can be screened on a single DNA origami substrate and simultaneously increase the effort required for data analysis. Circumventing such issues in future implementations of this assay will require further in-depth investigations into the nature of these microenvironments. While elucidating the structure, behavior, and diverse effects of microenvironments represents a great experimental challenge, especially at nanometer or subnanometer resolution, exploring such effects will ultimately enable the engineering of molecular recognition at a precision not possible with current technologies.

4. Experimental Section

DNA Origami Assembly: Rothemund triangles[24] were assembled from the 7249 nt M13mp18 scaffold and 208 staple strands (Metabion) as previously described[16] in 1 × TAE buffer (Calbiochem) supplemented with 10 mM $MgCl_2$ (Sigma-Aldrich). The modified staples were added in tenfold excess over the unmodified staples. After assembly, the samples were purified by spin filtering using a 100 kDa centrifugal filters (Amicon Ultra, Millipore) to remove the excess staple strands. The DNA origami concentration was measured by UV–vis absorption with an IMPLEN Nanophotometer P330 and adjusted to a final concentration of 20 nM with 1 × TAE–Mg^{2+} buffer.

B- and **I**-modified staple strands were synthesized as previously described.[16] For monodentate trypsin binding, only one of the modified staples (see Table S1, Supporting Information) was used in DNA origami assembly, whereas for bidentate trypsin binding, one **B**- and one **I**-modified staple were used in combination (see Table S2, Supporting Information).

Determination of Trypsin Binding Yields by AFM: Trypsin (type I from bovine pancreas; Sigma-Aldrich) was dissolved in 0.1 M HCl (Strockmeier Chemie) to yield a concentration of 125 µM. Right before incubation, the trypsin solution was diluted with the 1 × TAE–Mg^{2+} buffer, which was adjusted to pH 7.5 using 0.1 M HCl (Strockmeier Chemie). Then, trypsin and DNA origami at final concentrations of 80 and 4 nM, respectively, were incubated for 30 min at room temperature.

After incubation, 10 µL of the sample solution was deposited on a freshly cleaved mica surface, together with 20 µL 1 × TAE–Mg^{2+} buffer (pH 7.5). After 1 min, the sample was rinsed with HPLC-grade water (VWR) and dried carefully with a stream of ultrapure air. The dry mica sample was imaged using an Agilent 5100 and an Agilent 5500 AFM in intermittent contact mode with NSC18/AlBS cantilevers (MikroMasch). AFM images were recorded with a scan size of 2.5 × 2.5 µm² and a resolution of 1024 × 1024 px². By manually counting the DNA origami substrates with occupied and unoccupied binding sites, a trypsin binding yield was determined for each AFM image. The average binding yields shown in Figure 3 and 5, and S3, Supporting Information, were obtained by averaging over several AFM images recorded for the same protein–ligand system, with error bars corresponding to the standard deviations. For each protein–ligand system, between 4 and 15 AFM images corresponding to 322–1662 individual DNA origami substrates were analyzed (see Table S3 and S4, Supporting Information, for details).

Simulation Details: We performed Langevin molecular dynamics simulations, for each of the seven structures characterized experimentally (spacing between ligands from 0 to 6 bp). In particular, we focused only on simulating the trapezoid part of the structure containing the ligands. Each trapezoid is composed by about 4700 nucleotides and is modeled using a coarse-grained model of DNA, oxDNA2,[20,21] which is parametrized to accurately reproduce mechanical and thermodynamic properties

of DNA at different monovalent salt concentrations.[32] Note that the effect of the solvent is implicitly accounted for through a friction and a random force in the Langevin equation, and the effect of the electrolytes is also implicitly taken into account through a screening of Coulomb interactions, according to the Debye–Hückel theory.

The initial configurations were obtained starting from the caDNAno file,[33] which was converted to the oxDNA representation using the tacoxDNA package.[34] Simulations were conducted with the LAMMPS software.[35] The temperature was set to $T = 300$ K and the monovalent salt concentration to 1 M. The systems were evolved for about 3×10^6 simulation time units, sampling 3000 configurations. Equilibration was observed after 10^4 simulation time units.

The distance between the ligands was approximated by the distance between the edges of the single-stranded spacers. Trypsin accessibility was measured using a similar protocol as previously described.[36] The protein was approximated by a sphere with the same radius as the radius of gyration of trypsin (Figure 4b), which is about 1.6 nm. For each trajectory frame, the sphere surface was docked to the spacer edges, checking if any nucleotide overlapped with it (Figure 4c). The same operation was repeated while rotating the sphere around the edges (3.6° for each rotation, for a total of 100 states). The accessibility was then defined as the fraction of states with no overlapping nucleotides. If the edges' distance was more than twice the sphere radius, the accessibility was considered to be zero. This definition of accessibility does not take into account a preferential distance between the binding pockets.

K_d Determination by MST: MST measurements were performed with a Monolith NT115 equipped with Cy5 Pico detector (NanoTemper). For this, the DNA origami substrates were additionally labeled using a staple (t5s16e) carrying a Cy5-modification at the 5′ end (Metabion). Each Cy5-modified DNA origami was used at a constant final concentration of 2 nM, while the final concentration of trypsin varied from 500 μM to 77 nM in 1 × TAE buffer (10 mM $MgCl_2$, 0.05% Tween 20). After 10 min incubation of trypsin with the labeled DNA origami in darkness at room temperature, the solution was loaded into a MST standard capillary. Measurements were performed with 10% light-emitting diode power and 40% MST power. All experiments were performed in triplicates and the error bars indicate the standard deviation. The relative fluorescence measured 1.5 s after MST-on was plotted against the concentration of trypsin. All data sets were evaluated by Origin 8.5 (OriginLab).

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

Keywords

[1] a) A. Rajendran, M. Endo, H. Sugiyama, *Angew. Chem. Int. Ed.* **2012**, *51*, 874; b) I. Bald, A. Keller, *Molecules* **2014**, *19*, 13803; c) A. J. Lee, C. Wälti, *Comput. Struct. Biotechnol. J.* **2019**, *17*, 832.

[2] a) G. P. Acuna, F. M. Möller, P. Holzmeister, S. Beater, B. Lalkens, P. Tinnefeld, *Science* **2012**, *338*, 506; b) H. K. K. Subramanian, B. Chakraborty, R. Sha, N. C. Seeman, *Nano Lett.* **2011**, *11*, 910; c) M. B. Scheible, G. Pardatscher, A. Kuzyk, F. C. Simmel, *Nano Lett.* **2014**, *14*, 1627.

[3] a) Y. Ke, S. Lindsay, Y. Chang, Y. Liu, H. Yan, *Science* **2008**, *319*, 180; b) D. Wang, Y. Fu, J. Yan, B. Zhao, B. Dai, J. Chao, H. Liu, D. He, Y. Zhang, C. Fan, S. Song, *Anal. Chem.* **2014**, *86*, 1932; c) S. E. Ochmann, C. Vietz, K. Trofymchuk, G. P. Acuna, B. Lalkens, P. Tinnefeld, *Anal. Chem.* **2017**, *89*, 13000.

[4] a) G. Raghavan, K. Hidaka, H. Sugiyama, M. Endo, *Angew. Chem. Int. Ed.* **2019**, *58*, 7626; b) S. Rinker, Y. Ke, Y. Liu, R. Chhabra, H. Yan, *Nat. Nanotechnol.* **2008**, *3*, 418; c) R. Subramani, S. Juul, A. Rotaru, F. F. Andersen, K. V. Gothelf, W. Mamdouh, F. Besenbacher, M. Dong, B. R. Knudsen, *ACS Nano* **2010**, *4*, 5969.

[5] a) Y. Sannohe, M. Endo, Y. Katsuda, K. Hidaka, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 16311; b) A. Rajendran, M. Endo, K. Hidaka, H. Sugiyama, *J. Am. Chem. Soc.* **2013**, *135*, 1117; c) P. C. Nickels, B. Wünsch, P. Holzmeister, W. Bae, L. M. Kneer, D. Grohmann, P. Tinnefeld, T. Liedl, *Science* **2016**, *354*, 305.

[6] a) J. Rackwitz, I. Bald, *Chem. Eur. J.* **2018**, *24*, 4680; b) J. Rackwitz, J. Kopyra, I. Dąbkowska, K. Ebel, M. L. Ranković, A. R. Milosavljević, I. Bald, *Angew. Chem. Int. Ed.* **2016**, *55*, 10248; c) A. Keller, J. Rackwitz, E. Cauët, J. Liévin, T. Körzdörfer, A. Rotaru, K. V. Gothelf, F. Besenbacher, I. Bald, *Sci. Rep.* **2014**, *4*, 7391.

[7] a) M. Endo, Y. Katsuda, K. Hidaka, H. Sugiyama, *Angew. Chem. Int. Ed.* **2010**, *49*, 9412; b) M. Tintoré, I. Gállego, B. Manning, R. Eritja, C. Fàbrega, *Angew. Chem. Int. Ed.* **2013**, *52*, 7747; c) Y. Suzuki, M. Endo, C. Cañas, S. Ayora, J. C. Alonso, H. Sugiyama, K. Takeyasu, *Nucleic Acids Res.* **2014**, *42*, 7421.

[8] a) L. Sun, Y. Gao, Y. Xu, J. Chao, H. Liu, L. Wang, D. Li, C. Fan, *J. Am. Chem. Soc.* **2017**, *139*, 17525; b) V. Linko, M. Eerikäinen, M. A. Kostiainen, *Chem. Commun.* **2015**, *51*, 5351; c) W. P. Klein, R. P. Thomsen, K. B. Turner, S. A. Walper, J. Vranish, J. Kjems, M. G. Ancona, I. L. Medintz, *ACS Nano* **2019**, *13*, 13677.

[9] P. Zhang, F. Wang, W. Liu, X. Mao, C. Hao, Y. Zhang, C. Fan, J. Hu, L. Wang, B. Li, *ACS Appl. Mater. Interfaces* **2019**, *11*, 21973.

[10] A. Shaw, I. T. Hoffecker, I. Smyrlaki, J. Rosa, A. Grevys, D. Bratlie, I. Sandlie, T. E. Michaelsen, J. T. Andersen, B. Högberg, *Nat. Nanotechnol.* **2019**, *14*, 184.

[11] P. Zhang, X. Liu, P. Liu, F. Wang, H. Ariyama, T. Ando, J. Lin, L. Wang, J. Hu, B. Li, C. Fan, *Nat. Commun.* **2020**, *11*, 3114.

[12] M. Endo, K. Tatsumi, K. Terushima, Y. Katsuda, K. Hidaka, Y. Harada, H. Sugiyama, *Angew. Chem. Int. Ed.* **2012**, *51*, 8778.

[13] a) D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, *Nat. Rev. Drug Discov.* **2016**, *15*, 605; b) D. C. Rees, M. Congreve, C. W. Murray, R. Carr, *Nat. Rev. Drug Discov.* **2004**, *3*, 660.

[14] D. E. Scott, A. G. Coyne, S. A. Hudson, C. Abell, *Biochemistry* **2012**, *51*, 4990.

[15] a) G. E. de Kloe, D. Bailey, R. Leurs, I. J. P. de Esch, *Drug Discov. Today* **2009**, *14*, 630; b) S. Chung, J. B. Parker, M. Bianchet, L. M. Amzel, J. T. Stivers, *Nat. Chem. Biol.* **2009**, *5*, 407.

[16] C. Kielar, F. V. Reddavide, S. Tubbenhauer, M. Cui, X. Xu, G. Grundmeier, Y. Zhang, A. Keller, *Angew. Chem. Int. Ed.* **2018**, *57*, 14873.

[17] a) B. Saccà, Y. Ishitsuka, R. Meyer, A. Sprengel, E.-C. Schöneweiß, G. U. Nienhaus, C. M. Niemeyer, *Angew. Chem. Int. Ed.* **2015**, *54*, 3592; b) M. Marini, L. Piantanida, R. Musetti, A. Bek, M. Dong, F. Besenbacher, M. Lazzarino, G. Firrao, *Nano Lett.* **2011**, *11*, 5449; c) M. A. Goetzfried, K. Vogele, A. Mückl, M. Kaiser, N. B. Holland, F. C. Simmel, T. Pirzer, *Small* **2019**, *15*, e1903541.

[18] a) M. Hirota, M. Ohmuraya, H. Baba, *J. Gastroenterol.* **2006**, *41*, 832; b) E. Vandermarliere, M. Mueller, L. Martens, *Mass Spectrom. Rev.* **2013**, *32*, 453.

[19] S. Melkko, Y. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, *Angew. Chem. Int. Ed.* **2007**, *46*, 4671.

[20] T. E. Ouldridge, A. A. Louis, J. P. K. Doye, *J. Chem. Phys.* **2011**, *134*, 85101.

[21] B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, J. P. K. Doye, *J. Chem. Phys.* **2015**, *142*, 234901.

[22] N. Ota, C. Stroupe, J. M. S. Ferreira-da-Silva, S. A. Shah, M. Mares-Guia, A. T. Brunger, *Proteins* **1999**, *37*, 641.

[23] Y. Deng, J. Peng, F. Xiong, Y. Song, Y. Zhou, J. Zhang, F. S. Lam, C. Xie, W. Shen, Y. Huang, L. Meng, X. Li, *Angew. Chem. Int. Ed.* **2020**, *59*, 14965.

[24] P. W. K. Rothemund, *Nature* **2006**, *440*, 297.

[25] K. A. Walsh, *Methods Enzymol.* **1970**, *19*, 41.

[26] M. T. Strauss, F. Schueder, D. Haas, P. C. Nickels, R. Jungmann, *Nat. Commun.* **2018**, *9*, 1600.

[27] W. A. Kibbe, *Nucleic Acids Res.* **2007**, *35*, W43

[28] S. Ramakrishnan, G. Krainer, G. Grundmeier, M. Schlierf, A. Keller, *Nanoscale* **2016**, *8*, 10398.

[29] a) B. Jayaram, T. Jain, *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 343; b) D. Leckband, S. Sivasankar, *Colloids Surf. B* **1999**, *14*, 83.

[30] a) V. P. Chuprina, U. Heinemann, A. A. Nurislamov, P. Zielenkiewicz, R. E. Dickerson, W. Saenger, *Proc. Natl. Acad. Sci.* **1991**, *88*, 593; b) V. A. Buckin, B. I. Kankiya, D. Rentzeperis, L. A. Marky, *J. Am. Chem. Soc.* **1994**, *116*, 9423; c) Y. Yonetani, H. Kono, *Biophys. J.* **2009**, *97*, 1138.

[31] C. E. Dumelin, S. Trüssel, F. Buller, E. Trachsel, F. Bootz, Y. Zhang, L. Mannocci, S. C. Beck, M. Drumea-Mirancea, M. W. Seeliger, C. Baltes, T. Müggler, F. Kranz, M. Rudin, S. Melkko, J. Scheuermann, D. Neri, *Angew. Chem. Int. Ed.* **2008**, *47*, 3196.

[32] a) M. C. Engel, D. M. Smith, M. A. Jobst, M. Sajfutdinow, T. Liedl, F. Romano, L. Rovigatti, A. A. Louis, J. P. K. Doye, *ACS Nano* **2018**, *12*, 6734; b) B. E. K. Snodin, F. Romano, L. Rovigatti, T. E. Ouldridge, A. A. Louis, J. P. K. Doye, *ACS Nano* **2016**, *10*, 1724; c) A. Suma, C. Micheletti, *Proc. Natl. Acad. Sci.* **2017**, *114*, E2991; d) L. Zhou, A. E. Marras, C.-M. Huang, C. E. Castro, H.-J. Su, *Small* **2018**, *14*, e1802580.

[33] S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, W. M. Shih, *Nucleic Acids Res.* **2009**, *37*, 5001.

[34] A. Suma, E. Poppleton, M. Matthies, P. Šulc, F. Romano, A. A. Louis, J. P. K. Doye, C. Micheletti, L. Rovigatti, *J. Comput. Chem.* **2019**, *40*, 2586.

[35] O. Henrich, Y. A. Gutiérrez Fosado, T. Curk, T. E. Ouldridge, *Eur. Phys. J. E* **2018**, *41*, 57.

[36] A. Suma, A. Stopar, A. W. Nicholson, M. Castronovo, V. Carnevale, *Nucleic Acids Res.* **2020**, *48*, 4672.