Nucleic Acids Research, 2024, 1–14
https://doi.org/10.1093/nar/gkae576
Chemical Biology and Nucleic Acid Chemistry

Weak-cooperative binding of a long single-stranded DNA chain on a surface

Giovanni Nava [1,*,†], Thomas Carzaniga [1,†], Luca Casiraghi [1], Erik Bot [1], Giuliano Zanchetta [1], Francesco Damin [2], Marcella Chiari [2], Gerald Weber [3], Tommaso Bellini [1], Luca Mollica [1,‡] and Marco Buscaglia [1,*,‡]

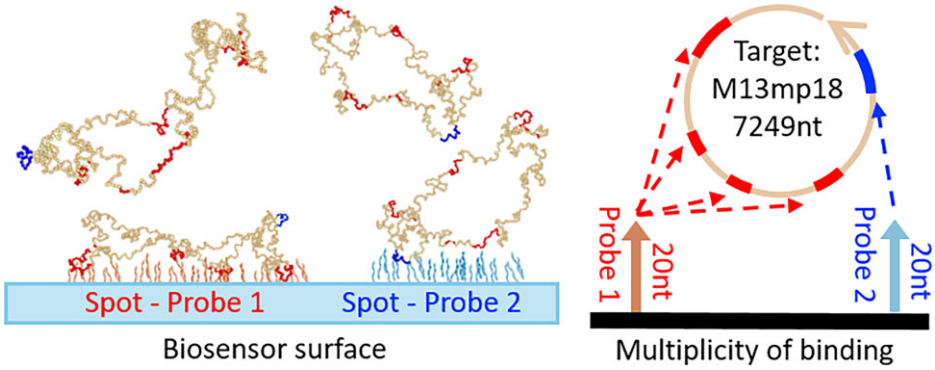
[1]Department of Medical Biotechnology and Translational Medicine, Università degli Studi di Milano, via F.lli Cervi 93, 20054 Segrate (MI), Italy
[2]Istituto di Scienze e Tecnologie Chimiche 'Giulio Natta', National Research Council of Italy (SCITEC-CNR), via Mario Bianco 11, 20131 Milano, Italy
[3]Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil

[*]To whom correspondence should be addressed. Email: giovanni.nava@unimi.it
Correspondence may also be addressed to Marco Buscaglia. Email: marco.buscaglia@unimi.it
[†]The first two authors should be regarded as Joint First Authors.
[‡]The last two authors should be regarded as Joint Last Authors.

Abstract

Binding gene-wide single-stranded nucleic acids to surface-immobilized complementary probes is an important but challenging process for biophysical studies and diagnostic applications. The challenge comes from the conformational dynamics of the long chain that affects its accessibility and weakens its hybridization to the probes. We investigated the binding of bacteriophage genome M13mp18 on several different 20-mer probes immobilized on the surface of a multi-spot, label-free biosensor, and observed that only a few of them display strong binding capability with dissociation constant as low as 10 pM. Comparing experimental data and computational analysis of the M13mp18 chain structural features, we found that the capturing performance of a specific probe is directly related to the multiplicity of binding sites on the genomic strand, and poorly connected with the predicted secondary and tertiary structure. We show that a model of weak cooperativity of transient bonds is compatible with the measured binding kinetics and accounts for the enhancement of probe capturing observed when more than 20 partial pairings with binding free energy lower than -10 kcal mol^{-1} are present. This mechanism provides a specific pattern of response of a genomic strand on a panel of properly selected oligomer probe sequences.

Graphical abstract



Introduction

In the last decades, the research aimed at understanding and controlling nucleic acid (NA) interactions for either analytical or nano-technological applications has been constantly expanding (1,2). Thanks to the simplicity of the nucleotide (nt) pairing code, extremely complex structures self-assembled by the hybridization of several single-stranded (SS) chains can be modelled and fabricated (3,4). Such highly accurate design capability is typically achievable when the equilibrium of the overall hybridization reaction is strongly shifted towards the product structures and kinetic traps are avoided (5). In contrast, the conformational behaviour of long SS NA chains (>1000 nt), dominated by marginally stable and competing pairings, remains elusive. This affects the understanding of

Received: March 22, 2024. **Revised:** June 4, 2024. **Editorial Decision:** June 19, 2024. **Accepted:** June 20, 2024

natural processes, such as the modulation of mRNA translation (6) or the function of non-coding RNAs (7–9). Moreover, the complex behaviour of long SS NA chains strongly impacts the design of analytical tools relying on the hybridization with oligomer probes with complementary sequence (10–12). In such framework, the COVID-19 pandemic has boosted the research on innovative methods for the direct detection of long DNA or RNA strands without extraction and thermal cycles (13–15) or even without amplification (16–18), but the direct multiplex characterization of pools of long NA sequences by methods suitable for mass screening still remains an unmet need (19,20). Despite the complexity of the interactions and the unavoidable presence of secondary structures affecting the hybridization with complementary probe strands (21), recent studies have shown that long NA strands can offer the opportunity of increasing the sensitivity and specificity of detection through multivalent binding (18,22–24).

The complex polymer folding and dynamics of long SS NA plays a major role especially when the binding occurs on a surface rather than in free solution, mainly because of entropic and electrical effects (25,26). The self-assembly of large DNA structures on a surface or the capture of a long SS NA by surface-immobilized probes are non-trivial even when they occur efficiently in solution (27,28). This phenomenon affects surface-based biosensors and microarrays, which typically limit their targets to NA strands with length up to a few hundreds nt (29–31). In contrast to more common fluorescence detection, label-free biosensors and microarrays (32–35) offer access to kinetics and more accurate ratiometric quantification (36). While label-free DNA microarrays have been successfully exploited to quantify and study short oligomers (26,37–39), the mechanism for capturing a long SS NA sequence remains largely unexplored.

Here, we analysed by label-free microarray a wide panel of oligomer probes for isothermal surface capturing of bacteriophage genome M13mp18, a SS DNA chain of 7249 nt commonly used in NA research, as a model for a long SS NA chain. We show that this long DNA genomic strand can provide a characteristic and sequence-specific pattern of response, with only a few selected probes yielding remarkably efficient binding. We find that such binding pattern is primarily encoded in the sequence through the multiplicity of sub-optimal pairings between the long strand and the surface probes, as schematically shown in Figure 1, a behaviour that is consistent with a model of weak cooperativity of multiple transient bonds. Such a model, built on a simple coupling of reaction kinetics, indicates that in a system in which multiple bonds of a single chain is entropically disfavoured, there is still a combinatorial advantage for rapidly alternating weak bonds, yielding an overall enhancement of binding strength and kinetics hence an intrinsic amplification of surface hybridization.

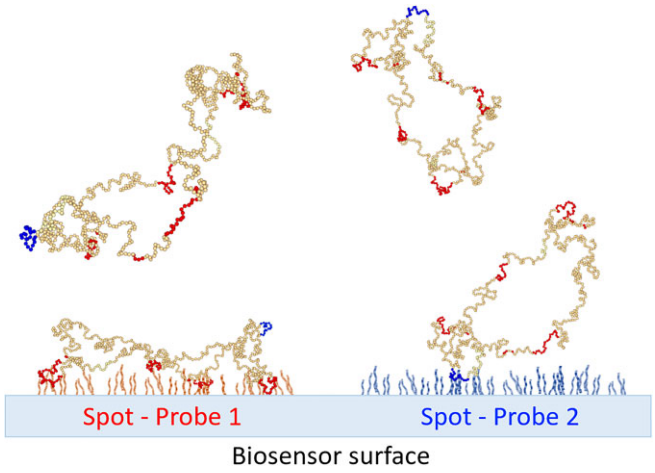


Figure 1. Scheme of the circular single stranded DNA of bacteriophage M13mp18 binding on 20-mer single stranded DNA probes immobilized in different spots on the biosensor surface. Each probe is fully complementary to a specific region of the long strand, but it forms partial pairings also with other regions. The colored fragments on M13mp18 represent regions fully or partially complementary to probe 1 (red) or fully complementary to probe 2 (blue).

Materials and methods

Computational methods

Selection of single stranded regions by base pairing probability matrix—topology prediction

The sequence of the bacteriophage M13mp18, a circular single stranded DNA chain with length $N_{nt} = 7249$ nucleotides (nt) (Supplementary File S1) has been used as input for the RNAFold (40) software as provided by the VIENNA Package. The force field by Mathews (41) has been used for the prediction of the base pairing probability of the circular DNA at 37°C and the centroid structure, i.e., the structure in the entire structure ensemble that has the minimum total base-pair distance to the structures in the set. The centroid secondary structure obtained from RNAFold annotated in .ct2 format (Supplementary File S2) has been processed by in-house scripts in order to extract single strands with length equal or greater than 35 nt resulting in 34 sequence regions. The sequence regions SS_{PM} selected by the analysis of the base pairing probability matrix are reported in Supplementary File S3.

Coarse grained Monte-Carlo simulation—conformation prediction

As schematically shown in Figure 2A, the SS_{PM} regions of M13mp18 have been used as input for performing Virtual Move Monte Carlo (VMMC) (42) as implemented in the code OxDNA (43). A Replica Exchange Monte Carlo (REMC) scheme (44) has been adopted using the canonical NVT ensemble with the temperature scheme $T = 298$, 306, 313, 320 K, and 10 millions of steps, with an attempt of exchange every 5000 steps and a writing rate of 1/2000 frame/steps (i.e. leading to 5000 frames collected for every DNA sequence). For each sequence, the simulations have been performed in a cubic box with the edge length equal to 42.5 nm (i.e. 50 internal units). The trial translational and angular rotational displacements have been respectively set up at ±0.4 Å (delta = 0.1) and ±7 arcdegrees (delta = 0.25). The structures corresponding to the thermal equilibrium condition of REMC simulations (i.e., non-exchanging states) have been extracted and converted to all-atoms models using the TacoxDNA utility (45) for all the temperature explored by means of the adopted REMC scheme. For the sake of homogeneity of sampling for all the inspected sequences, the last 2×10^6 steps (i.e. the last 1000 frames) have been considered for the extraction of atomic structures in every trajectory. The resulting structures have been inspected in the form of a trajectory using the contact maps MDMAT analysis tool of the

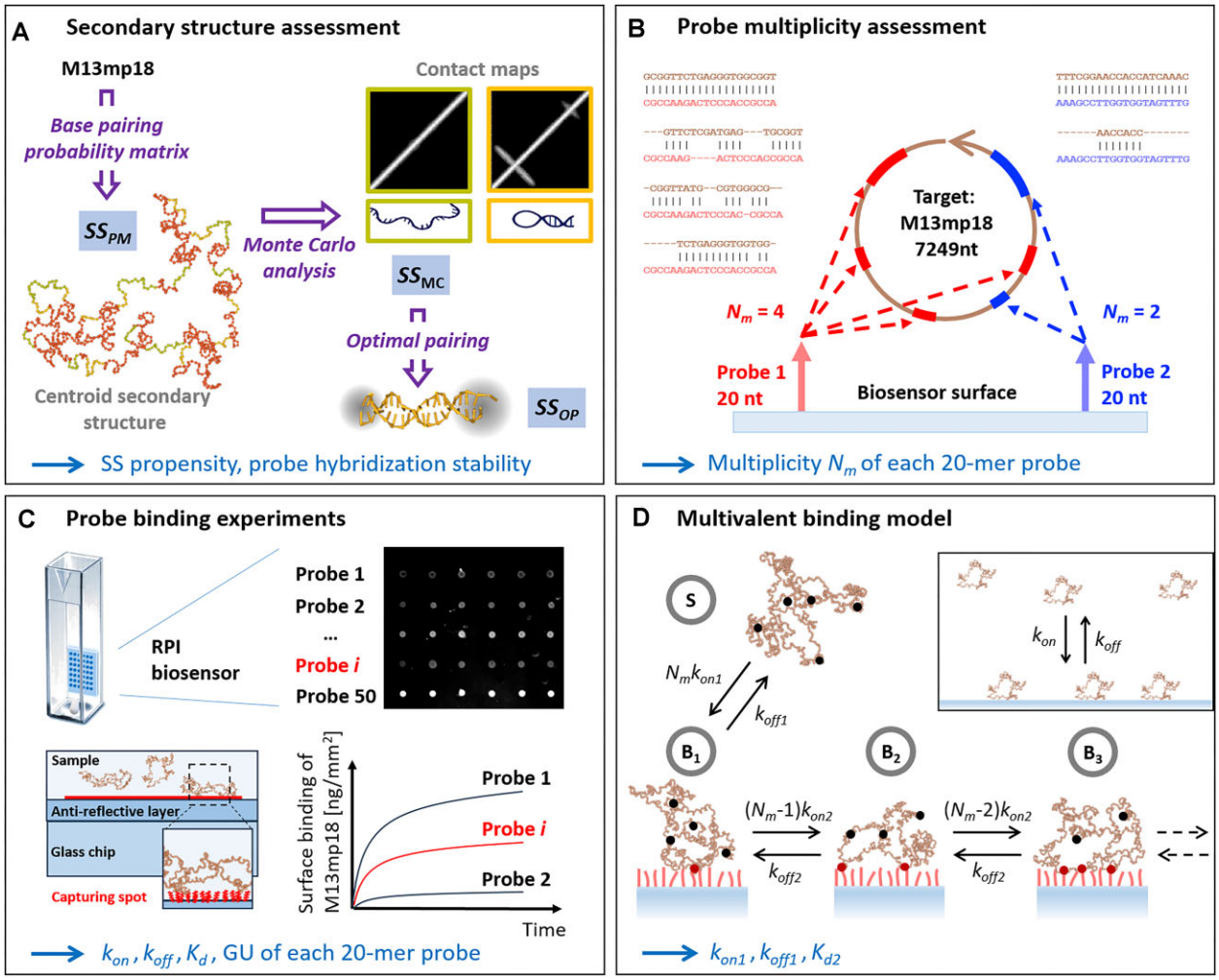


Figure 2. Schematic of computational and experimental assessment of binding between M13mp18 and 20-mer DNA probes. (**A**) The regions of the genomic DNA with the largest SS propensity and hence accessibility to probe binding are computed by RNAFold (40) (SS_{PM}) and OxDNA (43) (SS_{MC}). 20-mer probes with optimal pairing with their complementary fragments on M13mp18 (SS_{OP}) are selected from the SS_{MC} regions. (**B**) For all the 20-mer sequences fully complementary to a fragment of M13mp18, the multiplicity N_m of binding sites forming full or partial pairings on the M13mp18 strand is computed by a local alignment algorithm (49) using a threshold $\Delta G_{th} = -10$ kcal mol^{-1} for the hybridization free energy computed by Nupack (51). (**C**) The equilibrium and kinetic parameters of the binding between the genomic strand and the probes are measured by RPI biosensor. Top left: the probes are immobilized in spots on a glass sensor glued on the inner wall of a cuvette and mixing is provided by a magnetic stirrer. Top right: the spot surface density of molecules is obtained from the images of light reflected by the glass sensor surface. Bottom left: the RPI sensor is formed by a glass substrate with anti-reflective coating for water. Bottom right: the binding parameters (*GU*, K_d, k_{on}, k_{off}) are obtained from the fit of the kinetic binding curves measured on each spot type for different concentrations of genomic strand in solution. (**D**) Multivalent kinetic model for the binding of a long SS NA on surface immobilized oligomer probes. The long strand has N_m possible pairing sites for the oligomers (black and red dots). Each state is characterized by the number of bindings forming with the surface, from 0 (long strand in solution, state *S*), up to N_m. Each of the surface bound sub-states B_i forms *i* bindings with the surface probes. The association kinetic rate from state *S* to B_1 is $N_m k_{on1}$, whereas the dissociation rate is k_{off1}. The kinetic rate of forming an additional paring from a surface bound state is given by k_{on2} multiplied by the number of unbound pairing sites on the long NA strand, whereas the rate of breaking an elementary surface paring is k_{off2} for all sub-states. Inset: the solution of the multi-state system provides the overall kinetic rate k_{on} and k_{off} for surface binding of M13mp18.

GROMACS 2020.1 suite (46) (Supplementary File S3 and S4). The SS_{PM} sequences have been assessed according to the following criteria: (a) completely unpaired regions at 298 K (SST25); (b) completely unpaired regions according to the contacts averaged over all the four temperatures (SSTvar); (c) paired regions according to the contacts averaged over all the four temperatures that display a less permanent secondary structure (i.e. higher averaged nucleotide-nucleotide distances) at higher temperatures (DSTvar). Sequence regions SS_{MC} (reported in Supplementary File S3) satisfying criteria a or b but not c were selected as candidate targets for probe strands with complementary sequence (Figure 2A).

Oligomer probe selection—hybridization prediction

The SS_{MC} regions of M13mp18 selected as single stranded by coarse grained Monte-Carlo analysis were further analysed for their hybridization efficiency with their candidate 20-mer complementary probes. The criteria were: (i) the probe should have little or no probability to form self-complementary hairpins; (ii) it should have the highest possible melting temperature; (iii) it should have little end fraying. For condition (i) we used RNAfold with built-in DNA parameters (40) and retained only those having a predicted folding Gibbs free energy $\Delta G = 0$, meaning no predicted capability of forming hairpins. To evaluate the melting temperature for condition (ii),

we used up-to-date nearest-neighbour parameters from (47) at 119 mM Na^+. For the case where there were multiple probes with the same melting temperature, we evaluated the possibility of end fraying with a mesoscopic model from (48), also calculated at 119 mM Na^+. Condition (iii) was met by selecting the probe with the lowest 5′ and 3′ average opening amplitude (48). The genome sequence regions SS_{OP} providing optimal pairing with their fully complementary 20-mer probes according to the criteria above and the probes' sequences are reported in Supplementary File S3. These selected probes were used for the experimental characterization by optical biosensor.

Probe selection by probe multiplicity analysis

A collection of all the possible 7249 20-mer single strands reverse complementary to M13mp18 was created by scanning the genomic strand with step of 1 nt. For each of these candidate probes we computed all the pairing N_m with M13mp18, including partial pairing with mismatches, bulges and internal loops (Figure 2B), providing a free energy for hybridization below the threshold $\Delta G_{th} = -10$ kcal mol^{-1} for 1 M NaCl at 25°C. This threshold approximately corresponds to the free energy of hybridization of a fully complementary duplex with 6 base pairs and average nt sequence. The choice of ΔG_{th} is meant to cut out shorter random pairings that are numerous in a sequence of length N_{nt} and therefore to save only the ones that characterize the sequence of the long strand. The specificity comes from the fact that a random probe sequence is unlikely to provide many full or partial pairings with free-energy of hybridization lower than ΔG_{th} with a strand of length N_{nt}. To compute the sub-optimal matches, we exploited the Smith-Waterman alignment algorithm (49) to find local alignments between the whole genome sequence and each of its 20-mer fragments. We used a custom scoring matrix approximating the average free energy of pairing for the reverse complementary sequence of the fragment obtained by nearest neighbour model (50) with gap opening penalty corresponding to the free energy cost of a bulge and a cut-off score corresponding to $\Delta G < -6$ kcal/mol. Then, for each optimal and suboptimal pairing with M13mp18, the free energy of hybridization was computed by Nupack 4.0 (51), and the pairings with free energy larger than ΔG_{th} were discarded.

Probe binding measurements and analysis

DNA sequences

M13mp18 purified ssDNA was purchased from Bayou Biolabs (Metairie, LA, USA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA) with Ultramer synthesis. The selected 20-mer DNA probe sequences were elongated at the 3′ side with a 10-mer poly-T strand modified at 3′ terminal with amine for covalent immobilization on the biosensor surface.

Reagents

All buffers and reagents were purchased from Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO, USA) and prepared according to common protocols using Milli-Q pure water.

Biosensor cartridge preparation

DNA probe strands were covalently immobilized on the surface of a Reflective Phantom Interface (RPI) sensing substrate forming a matrix of spots (52). Wedge-like glass chips (F2 optical glass, Schott) with 5° angle and a size of 8 mm × 12 mm were coated with SiO_2 to form an anti-reflective layer of 80 nm. After plasma cleaning, the chips were dip-coated with MCP2 (Lucidant Polymers Inc.; Sunnyvale, CA USA), a copolymer of dimethylacrylamide (DMA), *N*-acryloyloxysuccinimide (NAS) and 3-(trimethoxysilyl) propyl methacrylate (MAPS)—copoly (DMA–NAS–MAPS) (39). Droplets of spotting buffer (Na_2HPO_4, pH 8.5, 150 mM and sucrose monolaurate 0.01%w/v) containing amine-terminated DNA probes at concentrations of 10 μM were deposited on the chip surface by an automated, non-contact dispensing system (sciFLEXARRAYER S3; Scienion AG, Berlin, Germany). Following an overnight incubation, the chip surface was rinsed with blocking buffer (Tris–HCl, pH 8, 10 mM, NaCl 150 mM, ethanolamine 50 mM) and distilled water before being dried. The diameter of the resulting spots of DNA probes was between 150 and 200 μm. The sensor cartridges were assembled by affixing the glass chips onto the inner wall of 1 cm plastic cuvettes (Figure 2C). Prior to use, these cartridges were stored at a temperature of 4°C. All target DNA were suspended in measuring buffer (PBS 1×, pH 7.4, with final concentration of NaCl 1.137 M) before use.

Label-free measurement and analysis

The RPI measurements were performed using the apparatus and the analysis described in (52). The sensor cartridges were filled with 1.3 mL of measuring buffer and maintained at 25°C during the measurement through a thermalized holder. A magnetic stirrer ensured quick mixing of the solution. Spikes of single stranded DNA target, M13mp18 or oligomers, were performed by adding 20 μl of measuring buffer containing different amounts of target molecules obtaining concentrations from 1.5 pM up to 16 nM.

The intensity of light reflected by each spot as a function of time t was converted into the total mass surface density of molecules $\sigma(t)$ according to (52). The mass surface density of target single stranded DNA binding the surface-immobilized 20-mer probes was computed as $\Delta\sigma(t) = \sigma(t) - \sigma_0$, where σ_0 is the surface density of oligomer probes measured before the addition of the target molecules. Each binding curve $\Delta\sigma(t)$, obtained by averaging the signal from at least five spots with identical composition, was fitted with the exponential growth function:

$$\Delta\sigma(t) = \Delta\sigma_{eq}(1 - e^{-kt}) \quad (1)$$

where $\Delta\sigma_{eq}$ is the asymptotic amplitude and k is the observed binding rate. The initial slope σ' of each binding curve extrapolated at $\Delta\sigma = 0$ was obtained as $\sigma' = \Delta\sigma_{eq}k$, and the growth unit was computed as $GU = \sigma'/\sigma_0$. The dissociation equilibrium constant K_d and the kinetic constants for association, k_{on}, and dissociation, k_{off}, were obtained from the analysis of the amplitude $\Delta\sigma_{eq}(c)$ and the initial slope $\sigma'(c)$ of each binding curve as functions of the target concentration c in solution, assuming a pseudo-first order behaviour at small concentrations and a deviation from the ideal scaling only when $\Delta\sigma$ approaches the surface saturation value $\Delta\sigma_\infty$. The equations used to fit the concentration dependence were proposed in (26) for modeling the DNA hybridization on a surface. For the analysis of the amplitudes $\Delta\sigma_{eq}(c)$, here we adopted a linear approximation of the exponential coefficient $exp(-\omega\phi)$ proposed in (53), (54) and (26) to account for weaker binding with the increase of the fraction ϕ of bound sites on the surface, where the coefficient ω is proportional to the repulsive

potential that the target has to overcome to form a contact with the surface probes. For small values of ω, the exponential coefficient can be approximated as $1 - \omega\phi$, with $0 < \omega < 1$, and the equilibrium values ϕ_{eq} of $\phi(c)$ can be obtained from the solution of (53,54)

$$\frac{\phi_{eq}}{1-\phi_{eq}} = \frac{c}{K_d}(1-\omega\phi_{eq}) \quad (2)$$

which, considering that $\Delta\sigma_{eq} = \Delta\sigma_\infty\phi_{eq}$, yields

$$\Delta\sigma_{eq}(c) = \Delta\sigma_\infty\frac{\psi - \sqrt{\psi^2 - 4\omega}}{2\omega} \quad (3)$$

where $\psi = 1 + \omega + \frac{K_d}{c}$. Eq. (3) becomes similar to a simple Langmuir isotherm for either $\omega = 0$ or $c \ll K_d$. From the fit of the measured $\Delta\sigma_{eq}(c)$ with Eq. (3), we obtained K_d and $\Delta\sigma_\infty$ for each probe spot. The value of k_{on} was obtained from the fit of the initial slope $\sigma'(c)$ with the function (26)

$$\sigma'(c) = \Delta\sigma_\infty k_{on} c\frac{c^*}{c^* + c} \quad (4)$$

and the dissociation rate constant was given by $k_{off} = K_d k_{on}$. Eq. (4) accounts for a linear dependence of $\sigma'(c)$ for small c and a progressive reduction of the observed binding kinetics when the concentration approaches the characteristic value c^*. In the fit with Eq. (4), c^* was left as free parameter, and its value typically corresponded to large surface coverage (i.e. $\Delta\sigma$ close to $\Delta\sigma_\infty$). The linear dependence of $\sigma'(c)$ at small c and the value of $\Delta\sigma_\infty > 0.2$ ng mm^{-2} were taken as quality check of the fit and the extracted parameters. Examples of fits of the measured binding curves and of their amplitudes and initial slopes as functions of c are reported in Supplementary Note S1. The use of Eqs. (3) and (4) enables to extrapolate the binding parameters of the pseudo-first order behaviour observed for $c < K_d$ by fitting all data points, including those for larger concentrations.

Kinetic model for multivalent binding

The multivalent binding of M13mp18 with the surface oligomer probes was modeled by the multi-state kinetic system schematically represented in Figure 2D. The model comprises a single state for M13mp18 in solution (S), and multiple states for the surface bound condition (B_i), each characterized by the number i of elementary bindings, which the long strand forms with the surface probes, with $1 \le i \le N_m$. Each state B_i is kinetically connected by a first-order reaction only with states B_{i-1} (if $i > 1$) and B_{i+1} (if $i < N_m$). The rate constant from the unbound state S to the first bound state B_1 is given by $N_m k_{on1}$, which accounts for the fact that the probability of binding with an oligomer probe upon contact with the surface linearly scales with the total number of available binding sites on the long strand. Analogously, the rate constant for adding one more elementary binding after the first, i.e. moving from state B_i to B_{i+1}, is $(N_m - i)k_{on2}$, where $(N_m - i)$ is the number of unpaired sites on the long strand. The rate constant k_{on2} is assumed to be the same for all bound states. Similarly, the rate constant k_{off2} for removing one binding is the same for all states with $i > 1$ and can be different from the rate constant k_{off1} for leaving the surface from the state with only one binding (i.e. from B_1 to S), because k_{off1} also accounts for the time required to diffuse away from the surface far enough so that rebinding becomes unlikely. With these assumptions, the fraction ϕ of occupied surface binding sites that can host the long strand ($0 \le \phi \le 1$) is obtained as $\phi = \sum\phi_i$, where ϕ_i is the fraction of surface sites in which the long strand is bound with i pairings (i.e. ϕ_i is the fraction of the surface bound population in state B_i). According to the model, the values of ϕ_i are given by a system of first-order differential equations (see Supplementary Note S2)

By summing all ϕ_i to obtain the total fraction ϕ of occupied surface sites, most of the terms of the system cancel out and it simply gives:

$$\frac{d\phi}{dt} = N_m k_{on1} c(1-\phi) - k_{off1}F_1\phi \quad (5)$$

where $F_1 = \phi_1/\phi$ is the fraction of the bound population in state B_1. In agreement with Eq. 5, the unbinding from the surface only depends on the population of ϕ_1, which in turn depends on the solution of the full kinetic system. Assuming a rapid equilibrium between the surface bound states, much faster than the dissociation rate k_{off1} (i.e. $k_{on2} + k_{off2} \gg k_{off1}$), the value of ϕ_1 is obtained from the equilibrium solution $d\phi_i/dt = 0$ for a constant value of ϕ. In this case, F_1 is given by

$$F_1 = \frac{1}{\sum_{i=1}^{N_m}\frac{(N_m-1)!}{(N_m-i)!}\left(\frac{1}{K_{d2}}\right)^{i-1}} \quad (6)$$

where $K_{d2} = k_{off2}/k_{on2}$ is the adimensional equilibrium constant for the first-order binding process between B_i states. Eq. (5) with F_1 given by Eq. (6) represents a pseudo-first order behaviour with association and dissociation kinetic rates $k_{on} = N_m k_{on1}$ and $k_{off} = k_{off1}F_1$, respectively, and thus with equilibrium dissociation constant $K_d = (k_{off1}/k_{on1})(F_1/N_m)$.

Results and discussion

Computational selection of oligomer probes for surface capture

To search for oligomer probes capable of capturing the long SS strand M13mp18 on a surface, we compared two different approaches for computational selection, analyzing either the internal structure of M13mp18 or the strength and degeneracy of M13mp18-probe binding (Figure 2A, B).

The secondary structure of the whole M13mp18 sequence was investigated by means of base pairing probability matrix (PM) to extract the centroid structure (Supplementary File S2). With the aim of maximizing the accessibility of the M13mp18 to the complementary probes, we selected all the centroid regions corresponding to SS strands of length equal or larger than 35 nt (SS_{PM}). These regions are displayed by green and yellow shades in Figure 3. All the other regions of M13mp18 (orange in Figure 3A and not shaded in Figure 3B) include either partially DS regions or SS regions shorter than 35, therefore we address them as non-SS by PM. The internal dynamics of the SS_{PM} regions have been further investigated by coarse-grained OxDNA Monte Carlo (MC) simulation (55) to distinguish between those with good (SS_{MC}) versus reduced (non-SS by MC) accessibility as obtained through the analysis of the contact maps (Supplementary File S4) and displayed in Figure 3 as green and yellow shading, respectively. Thus, green shadings in Figure 3 mark regions emerging as SS by both PM and MC analysis. We performed a further selection to obtain the 20-mer complementary probes that can bind such SS_{MC} regions with the largest stability based on a combination of minimum free-energy of hybridization and maximum stability of the terminal of the resulting duplex, as estimated by a meso-

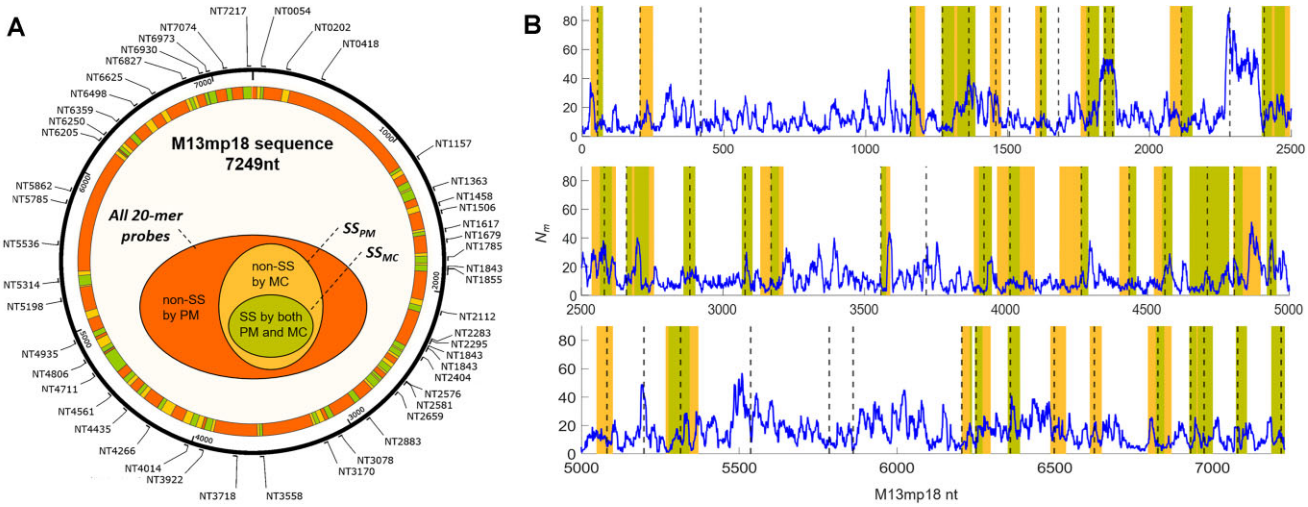


Figure 3. Map of M13mp18 sequence. (**A**) Positions of the regions complementary to the selected 20-mer probes along the M13mp18 sequence, and regions evaluated as SS by base pairing probability matrix (SS_{PM}) and Monte-Carlo simulations (SS_{MC}). The color code is defined in the graphical representation of the sets: regions classified as SS by both PM and MC (SS_{MC}, green); regions classified as SS by PM (SS_{PM}, yellow and green), but not by MC (yellow); the remaining regions were classified as non-SS by PM (orange). (**B**) Multiplicity of binding of all the 20-mer sequences fully complementary to a fragment of M13mp18. For each fragment of 20 nt starting with the position in abscissa, the blue line represents the number of partial pairings N_m with free energy of hybridization lower then $\Delta G_{th} = -10$ kcal mol^{-1} as computed by Nupack 4.0 (51). Green and yellow shadow colors indicate the same classification of the regions according to PM and MC analysis of panel a. The starting nt of the sequence fragments complementary to the probe sequences selected for the experimental assessment by PRI biosensor are indicated as vertical dashed line.

scopic model for hybridization (48). Through the combination of these criteria, among the SS_{MC} regions of M13mp18, we selected 33 20-base-long sequences with optimal probe pairing (SS_{OP}) (Supplementary File S5).

As alternative approach for selecting the oligomer probes, we counted the number of binding sites available on M13mp18 to each of the 20 base-long sequences complementary to the 7249 20 base-long fragments present on the circular genome. In the counting, we included all matches with binding free energy ΔG computed by nearest-neighbour model lower than $\Delta G_{th} = -10$ kcal mol^{-1}. Such threshold was chosen to discriminate between the very large number of short pairings statistically unavoidable, hence not meaningful, and the few sequence-specific matches less likely to be found by chance (see Materials and Methods and Supplementary Note S3). The result is shown in Figure 3B, where we plot the number of matches N_m versus the starting position of each complementary 20-mer (blue line). As apparent, N_m can vary from 1—no stable bond other than the perfect complementary region—up to about 80. Most of the sequence positions yield $N_m < 10$, whereas only a few give $N_m > 30$ (see Supplementary Note S3). The adoption of the threshold ΔG_{th} enables evaluating the multivalency of the interaction between M13mp18 and the surface oligomer probes.

Considering all this information, we made an overall selection of 50 M13mp18-capturing probes to be experimentally tested. Of each, the starting positions of the complementary sequence on M13mp18 are marked in Figure 3A and B by name labels and vertical dashed lines, respectively. The total number of hits is given by the cross point between the dashed line and the blue line (N_m). The selection includes: at least one probe per SS_{PM} region (33 probes); probes complementary to M13mp18 fragments inside non-SS by MC (6 probes) or non-SS by PM (10 probes) regions (Figure 3A); probes corresponding to maxima in N_m (11 probes with $N_m > 20$) and probes providing low values of N_m (23 probes with $N_m < 10$). An additional negative control (DEB-NEG) was selected to have minimum complementarity with the genome. The full list of the 50 probes is reported in Supplementary File S5, where, with the exception of DEB-NEG, each sequence is named after the initial position of the corresponding fully complementary strand on M13mp18 (e.g. NT2283 is the probe with sequence reverse complementary to the portion 2283–2303 of M13mp18).

Binding of M13mp18 to surface probes

The selected probe sequences were experimentally tested by measuring their binding with M13mp18 by RPI microarray biosensor. The probe strands were immobilized in different spots on the surface of an RPI sensor and M13mp18 was added with stepwise increments of concentration c in the measuring cartridge. The optical signal from each spot as a function of time t was converted into surface density $\Delta\sigma(t)$, expressed in ng mm^{-2}, providing a quantification of the amount of M13mp18 binding to the spot. The obtained binding curves for a concentration c of 0.5 nM are reported in Figure 4A–E.

We find that only about 10% of the probes displayed a strong binding capability characterized by large amplitude and fast response, whereas most of the probes provided a very weak binding. Given the large diversity of responses, in terms of both amplitudes and kinetics, we quantified the binding on the different probes by the initial slope of the binding curves obtained as the growth unit GU (see Materials and methods). The values of GU for all probes are reported in Figure 4F, in which the color of the bars indicates the class of probes according to the PM and MC calculation (see Figure 3A). Remarkably, two of the probes with the strongest binding are targeting complementary sequences on M13mp18 predicted to be non-SS by PM assessment (orange). In contrast, none of the probes targeting strands resulting as non-SS by MC simulations (yellow) displayed a large binding. Figure 4F also reports the multiplicity N_m of pairing sites with $\Delta G < \Delta G_{th}$ for

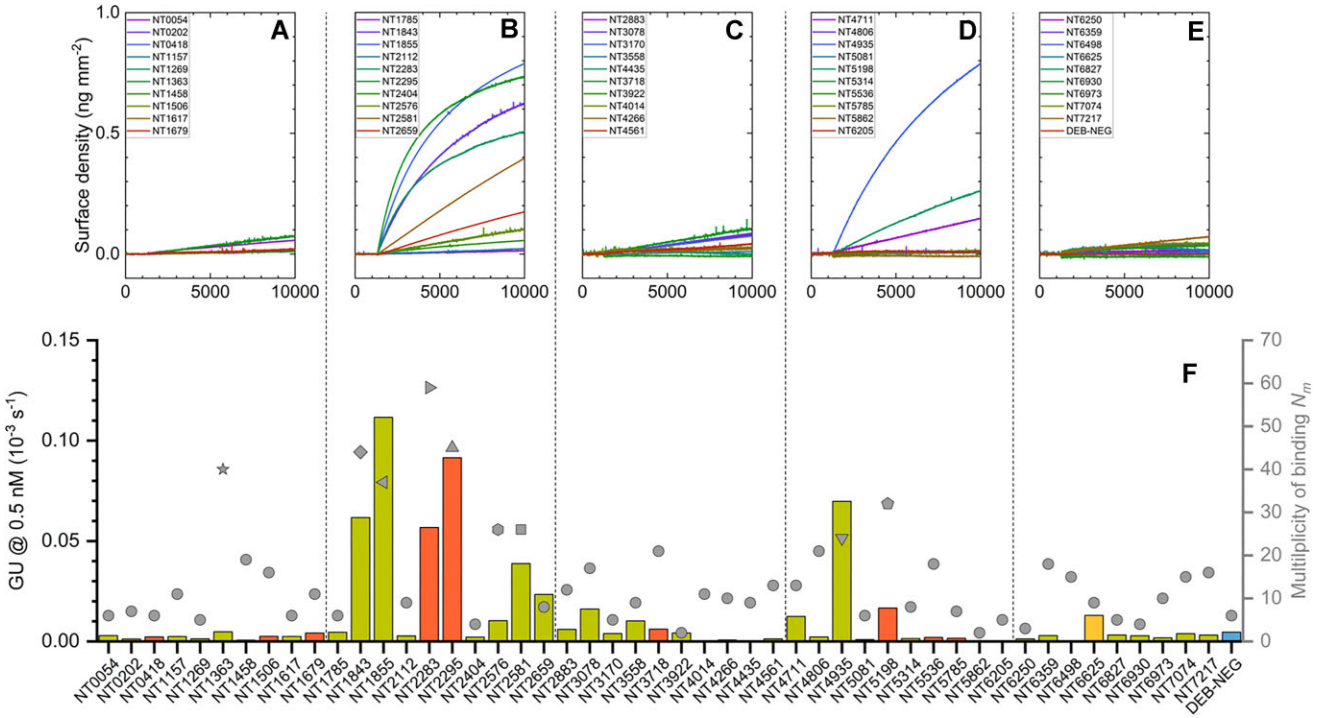


Figure 4. Binding pattern of M13mp18 on the surface-immobilized 20-mer probes. (**A–E**) Binding curves of M13mp18 measured by RPI as surface density as function of time. After the acquisition of a baseline, M13mp18 was added into the RPI cartridge at a concentration of 0.5 nM at 1200 s. The binding curves are divided into five graphs for clarity. The line colors indicate the probe spots reported in the legend of each panel. (**F**) Pattern of binding kinetics of M13mp18 expressed by *GU* (bar plot, left axis) and pattern of multiplicity of binding expressed as N_m (points, right axis) for different 20-mer probes immobilized on the RPI sensor. The *GU* values are obtained from the fit of the binding curves of panels A–E. The color of the bars is the same of Figure 3A and indicates probes predicted as SS by both PM and MC calculation (green), as SS by PM but not by MC (yellow), and as not fully SS by PM (orange), or indicates a control sequence without a fully complementary region in M13mp18 (blue). The data points corresponding to the largest values of N_m ($N_m > 22$) are represented by different symbols.

each probe on the M13mp18 sequence. All probes yielding a large *GU* also display a large value of N_m, hence indicating a fundamental role of this parameter for efficient binding of M13mp18 to the surface probes.

The binding pattern of Figure 4F is specific for the sequence of M13mp18 as confirmed by control experiments performed with a different genomic strand on the same panel of probes, which overall provided different and much smaller responses (Supplementary Note S4). However, neither the full complementarity of the probe strand nor the secondary structure prediction provide sufficient support to identify strong binding probes. In contrast, a large multiplicity of binding sites N_m appears as a requirement to obtain rapid capture of M13mp18 by the surface probes.

Effect of the multiplicity of binding sites

To asses the role of the multiplicity of binding on the capacity of each probe to capture the genomic strand, we studied the effect of N_m on binding efficiency. Figure 5A compares the measured *GU* with the number of binding sites N_m. As anticipated in Figure 4, none of the 40 probes with the lowest value of N_m ($N_m < 22$) displayed a large *GU*. In contrast, all the 6 probes with the largest *GU* have also a large N_m. The color of the data points in Figure 5A indicates the sequence accessibility predicted by PM and MC calculation of the secondary structure (Figure 3A). The data confirm that in the presence of a large N_m the SS propensity of a M13mp18 region becomes a secondary fact.

The figure also demonstrates that a large value of N_m is a necessary but not sufficient condition to achieve a rapid surface capture of the M13mp18 strand. More generally, different sequence features were found to contribute to the complex hybridization kinetics of oligomers (56,57). In the case of this study, among the secondary factors to consider, even a weak probe-probe interaction on the surface can contribute to slightly decrease the binding kinetics of M13mp18 (37,58), but only for the probes with large N_m (Supplementary Note S5). Therefore, the distribution of data points in Figure 5A suggests an upper limit of the *GU* that increases with N_m, as represented by the dashed line separating a region of the plot with $GU < \alpha N_m$ with $\alpha = 3.3 \times 10^{-6}$ s^{-1} (green shaded area), and a not accessible region (pink shaded area) for $GU > \alpha N_m$. Since the parameter *GU* is obtained as the initial slope σ′ of the binding curves divided by the surface density of probes, and σ′ is given by the kinetic rate for association k_{on} times the saturation surface density of M13mp18 (see Materials and methods), the observed effect of N_m on *GU* suggests a direct contribution of N_m on k_{on}.

Figure 5B shows the scaling with M13mp18 concentration of the initial slope σ′ of the binding curves (see Materials and Methods) measured on the spots of the probes with the largest N_m. In general, we find that σ′ linearly scale with c for all probes for surface densities of M13mp18 well below the surface saturation (see Supplementary Note S1). This behaviour is consistent with a pseudo-first order binding process at small concentrations (35,59), indicating that each M13mp18 chain binds the surface independently if not too crowded. The fitting

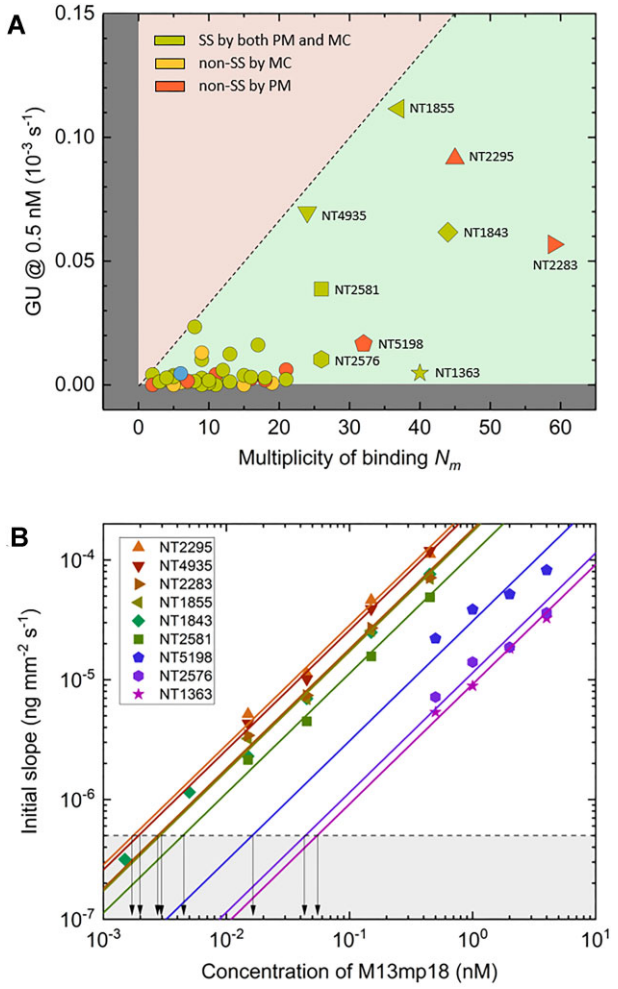


Figure 5. Effect of the multiplicity of binding on *GU* for M13mp18. (**A**) *GU* measured from the binding curves of Figure 4A-E plotted as a function of the multiplicity of binding N_m of each probe. The symbol colors are the same of the corresponding bars of Figure 4F and represent the accessibility estimated by PM and MC, as also indicated in the figure legend. The name of the probes with $N_m > 22$ is reported in the figure and the corresponding data points are represented by the same symbols of Figure 4F. The pink-shaded area represents a region of the *GU* – N_m plane without data points. (**B**) Dependence of the initial slope of the binding curves on M13mp18 concentration for the probes with $N_m > 22$, as indicated in the legend. The continuous lines are linear fit to the data points with the same color. The horizontal dashed line represents the *GU* corresponding to 3σ from the baseline noise and the vertical arrows indicate the limit of detection for the different probes.

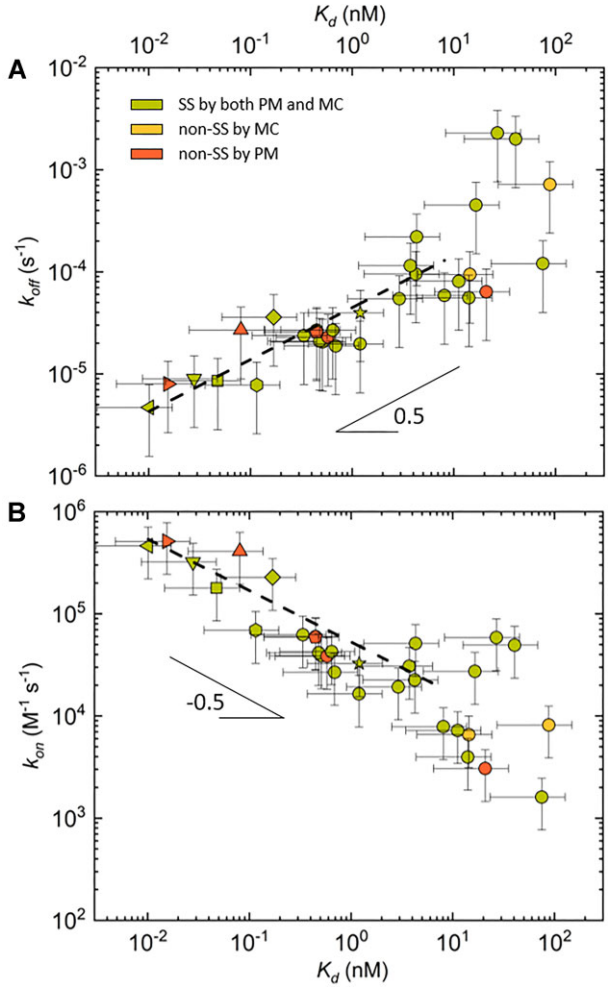


Figure 6. Equilibrium and kinetic parameters measured by RPI for M13mp18 binding on the 20-mer surface probes. (**A**) Kinetic constant for dissociation as a function of the equilibrium constant. (**B**) Kinetic constant for hybridization as a function of the equilibrium constant. In both graphs: vertical and horizontal error bars represent the average standard deviations from multiple measurements for each probe; the dashed lines represent power low fits to the data points yielding exponents 0.509 (A) and –0.504 (B); the symbol color and type is the same of Figure 5A; only data points corresponding to saturation amplitude $\Delta\sigma_\infty > 0.2$ ng mm^{-2} are reported.

of *GU* vs *c* enables to estimate the limit of detection (LOD) of M13mp18, defined as the concentration (indicated by arrows) yielding the lowest detectable signal (horizontal dashed line). The LOD for probes NT1843, NT1855, NT2283, NT2295 and NT4935 is within 3 pM, which is about 100 times lower than most of the probes with low N_m value.

Probe binding strength and kinetics

The large difference in the observed *GU* for the selected oligomer probes demonstrates a significant heterogeneity in their strength of hybridization with M13mp18. To further investigate the origin of this diversity, we analyzed the concentration dependence of the amplitudes and rates of the binding curves to obtain equilibrium and kinetic binding parameters (see Materials and methods and Supplementary Note S1).

Figure 6 reports the equilibrium and kinetic constants for all the probes that provide a signal large enough to enable an accurate analysis ($\Delta\sigma_\infty > 0.2$ ng mm^{-2}). Remarkably, the measured K_d spans 4 orders of magnitude, down to 10 pM, whereas both k_{off} and k_{on} extend for less than 3 orders of magnitude, reaching extreme values of 5×10^{-6} s^{-1} and 5×10^5 M^{-1}s^{-1}, respectively, for the lowest K_d. The maximum value of k_{on} observed is well within the limit for transport limitation estimated for M13mp18 in our set-up, as shown in Supplementary Note S6. The figure shows that both kinetic rate constants approximately scale with K_d especially for $K_d <$ 10 nM. The observed behaviour indicates a weaker than linear increase of k_{off} and decrease of k_{on}, compatible with $k_{off} \propto K_d^{0.5}$ and $k_{on} \propto K_d^{-0.5}$. We can then conclude that both kinetic rate constants contribute to the large range of K_d values. Data with $K_d > 10$ nM are less accurate because the value of K_d

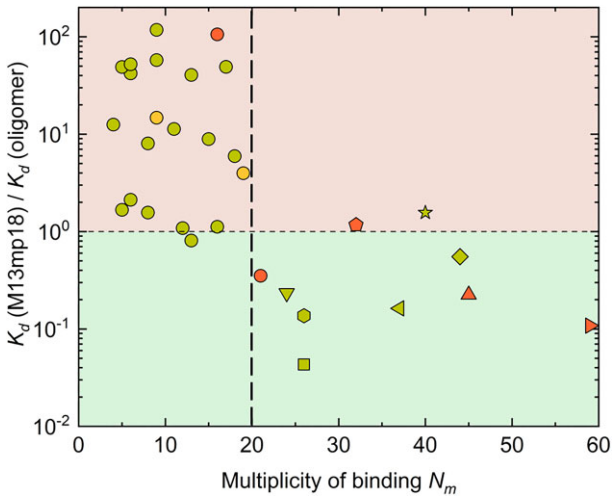


Figure 7. Ratio between the equilibrium dissociation constant for M13mp18 and for oligomer sequences fully complementary to each surface probe as a function of the multiplicity of binding N_m of the probes. The shaded areas represent the regions with K_d of M13mp18 larger (pink) or smaller (green) than that of the oligomer complementary to the corresponding probe sequence. The vertical dashed line is a guide to the eye indicating $N_m = 20$. The symbol color and type is the same of Figure 5A; only data points corresponding to saturation amplitude for M13mp18 $\Delta\sigma_\infty > 0.2$ ng mm^{-2} are reported.

approaches the maximum concentration of M13mp18 used in the experiments.

The saturation of the spot surface by binding of M13mp18 was obtained as the asymptotic value $\Delta\sigma_\infty$ of the surface density at large concentration of M13mp18 (see Materials and methods). The average value of $\Delta\sigma_\infty$ was 0.49 ng mm^{-2} (see Supplementary Note S7), which corresponds to an average distance between surface bound M13mp18 of 86 nm, slightly larger than the hydrodynamic diameter of 70 nm ± 10 nm measured by dynamic light scattering (Supplementary Note S8). The larger size of M13mp18 on the probe spots suggests a spreading on the surface, in agreement with the notion of multivalent binding with the immobilized probes.

Data in Figure 6 are represented with the same symbols and colors of Figure 5A, which enable noticing that the seven probes with the lowest values of K_d are also among those with the largest N_m, hence confirming that the multiplicity of binding sites on M13mp18 strongly affects the binding strength. In particular, four probes yielded extreme values of both k_{off} and k_{on} (NT1855, NT2283, NT4935, NT2581), whereas for one probe the low value of K_d is primarily ascribed to k_{off} (NT2576), and for other two probes to k_{on} (NT1843 and NT2295).

We then compared the binding parameters of M13mp18 to those of oligomer strands fully complementary to the selected probes. In this way, we evaluated the effects of probe sequence composition, which can either directly determine the strength of hybridization, or the propensity to form transient internal hairpins or dimers, hence reducing the binding with the target strand. Figure 7 reports the ratio between the K_d obtained for M13mp18 and that of the corresponding oligomer strands. The plot clearly shows two regimes: for low value of N_m ($N_m < 20$) the ratio of the K_d is larger than 1 (and even reaches 100), indicating a much weaker binding of M13mp18; for the largest N_m, M13mp18 binds stronger than the oligomer complementary to the probe by a factor up to 25.

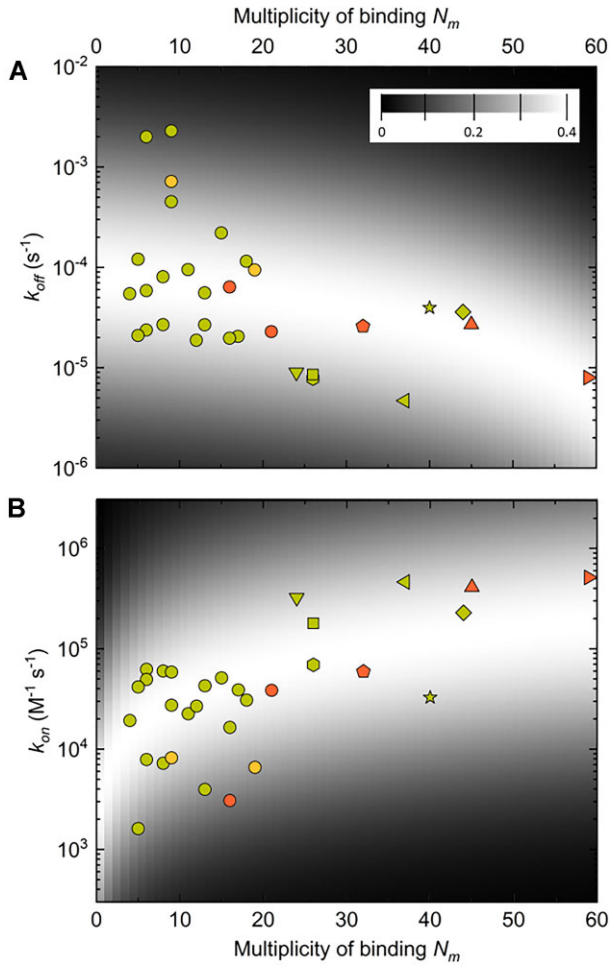


Figure 8. Dependence of M13mp18 kinetic constants on the multiplicity of binding. Kinetic rate constants for (**A**) dissociation and (**B**) association as a function of N_m of the probes. The symbol color and type is the same of Figure 5A. Only data points corresponding to saturation amplitude for M13mp18 $\Delta\sigma_\infty > 0.2$ ng mm^{-2} are reported. The background shading represents, for each value of N_m, the probability density function of finding the corresponding value of kinetic constant computed by Eqs. (5) and (6) with $K_{d2} = 55.6$, and assuming a Gaussian distribution for the values of log(k_{on1}), log(k_{off1}) with mean log(3 × 10^3 M^{-1}s^{-1}) and log(8 × 10^{-5} s^{-1}), respectively, and a standard deviation of one decade. The inset in panel a reports the shading gray scale of the probability density function.

As expected, the kinetic rate constants also show a strong dependence on N_m. This is apparent in Figure 8 where k_{off} tends to decrease with N_m whereas k_{on} displays an opposite behaviour. This suggests that the genomic strand can find more rapidly a conformation suitable to binding if the number of sequence regions partially complementary to the surface probe is larger. Analogously, the lifetime of binding of the genomic strand increases with this number.

Multivalent binding model

The observed effect of N_m on equilibrium and kinetic binding parameters of M13mp18, made reliable by the large ensemble of tested probe sequences, raises questions that we addressed by developing a kinetic model for multivalent binding on the surface-immobilized oligomer probes. Specifically, we aimed

to understand why both kinetic rates seem to scale with N_m in a similar way. A particular challenge is the rather limited range of values of k_{off}. Indeed, one would expect that, with N_m as large as 60, the binding lifetime of M13mp18 were many orders of magnitude larger than for smaller N_m values. The narrow range of k_{off} in the experimental observation suggests that, despite the large N_m, only a limited number of elementary bindings takes place at the same time.

The dependence of the surface binding of M13mp18 on N_m is described by the kinetic model schematically represented in Figure 2D. As detailed in the Materials and Methods section, the parameters of the model are: the kinetic rate k_{on1} for association with the first elementary pairing that M13mp18 forms with the surface probes, and the rate k_{off1} for dissociation from the last elementary pairing, before diffusing into the bulk solution; the kinetic rates for adding or removing one elementary pairing with the surface probes after the first, k_{on2} and k_{off2}, respectively, where $k_{off2}/k_{on2} = K_{d2}$ is the equilibrium constant for such first-order binding process; the concentration c of M13mp18 in solution; the fraction ϕ of surface sites occupied by M13mp18; the multiplicity N_m of available binding sites on M13mp18 for a given oligomer probe.

According to the model, the observed k_{on} rate for surface binding of M13mp18 linearly increases with N_m to account for the increased probability of finding at least one pairing upon contact with the surface, hence $k_{on} = N_m k_{on1}$. At a certain time, the binding with the surface can involve more than one pairing site, but we assumed that the unbinding of the whole M13mp18 strand from the surface is always preceded by a single binding. Therefore, the observed overall rate of unbinding is $k_{off} = F_1 k_{off1}$, where the factor F_1 represents the fraction of bound population forming only one transient pairing with the surface probes, hence it accounts for the combinatorial probability of observing a single binding, which decreases with the multiplicity N_m.

To account for the fact that the experimentally observed binding curves of M13mp18 are well fitted by simple exponential growths (see Supplementary Note S1), we hypotized that the kinetics of adding or removing a single elementary pairing with the surface probes is much faster than the rate k_{off1} for dissociation of the whole strand from the surface. Indeed, k_{off1} can be longer than k_{off2} because it also accounts for the diffusion of M13mp18 far enough from the surface so that the rebinding probability becomes unlikely. In this framework, the binding curves of M13mp18 are described by the pseudo-first order model given by Eqs. (5) and (6).

The expression of F_1 given by Eq. (6) accounts both for the cooperativity of parallel bindings through the power law dependence on K_{d2} and for the combinatorial effect of multiple independent binding sites (i.e. combinatorial entropy). Similar equations have been proposed by others to describe the thermodynamics of systems composed by multivalent flexible ligands interacting with receptors on a surface (60,61). Moreover, the consistency of a pseudo-first order binding behaviour for this class of systems in presence of weak and independent multivalent interactions has been postulated and demonstrated in previous work (62,63). Here, the model represented in Figure 2D enables a kinetic description in addition to the equilibrium solution based on the same concepts. Eq. (5) does not account for the non-ideal scaling of the amplitudes and the rates of the binding curves experimentally observed approaching the surface saturation, which is ascribed to surface crowding of M13mp18 (see Materials and methods and Supplementary Note S1). Therefore, Eq. 5 provides the basis for the interpretation of the observed kinetics of binding for surface coverage not too close to saturation (e.g. $\phi \leq 0.5$).

In the limit of rapid exchange of elementary pairings when M13mp18 is on the surface, the coefficient F_1 depends only on N_m and K_{d2} (Eq. 6), and it enables to account for two different limiting regimes (see Supplementary Note S2): (i) if the binding of each single elementary pairing is strong (i.e. the equilibrium constant K_{d2} of the pairings following the first one is small), all M13mp18 strands bind to the surface probes with a number of pairings close to N_m, and the observed k_{off} can be several orders of magnitude smaller than k_{off1} (*strong cooperativity*); (ii) in contrast, for a weak binding of each single pairing (i.e. K_{d2} is large), the genomic strand only binds the surface probes with one or a few sequence regions at the same time undergoing a dynamic equilibrium, and the observed dissociation rate scales with N_m as $k_{off} \approx k_{off1}/[1 + (N_m - 1)/K_{d2}]$ (*weak cooperativity*). The small range of variation of k_{off} reported in Figure 8, which does not decrease exponentially with N_m, is consistent with the case ii of weak cooperativity.

Despite the overall dependence of the surface binding of M13mp18 on N_m, as shown in Figures 7 and 8, for similar values of N_m different oligomer probe sequences can yield binding parameters as different as 2 orders of magnitude. Accordingly, assuming a range of values for k_{on1} and k_{off1} distributed on a Gaussian with standard deviation of a factor of 10, Eq. (5) was used to predict the range of observed kinetic constants as a function of N_m. Such behaviour is shown in Figure 8, where the grey shading represents the probability of observing values of k_{on} and k_{off} given the mean values of the Gaussian distributions of k_{on1} and k_{off1} for $N_m = 1$ and the value of K_{d2}. These parameters were varied to maximize the overlapping between the computed probability of k_{on} and k_{off} and the measured values (see caption to Figure 8). The fit of the shaded region to the measured data yields $K_{d2} = 55.6$, hence confirming the scenario of weak cooperativity emerging as a result of multivalency.

Weak cooperativity

The experimental observation confirms that the number of optimal and suboptimal pairing sites defined by N_m is highly correlated to the probe binding performance of M13mp18. This reveals the complex nature of the binding of a large genomic strand to short probes on a surface, in which the relevant factor is not the strength of the best DNA–DNA hybridization, but the availability of multiple interaction sites. An insight of the mechanism of such cooperative effect is given by the observation that the plurality of binding sites is necessary to provide strong binding but does not yield an exponential growth of the binding strength. The multi-state kinetic model accounts for this behaviour (Eqs. 5 and 6) suggesting a model of weak cooperativity, in which each interaction site of M13mp18 undergoes a rapid binding and unbinding to the surface bed of probes, with a small number of parallel bonds present at each time (see Supplementary Note S2). Therefore, the meaning of a large value of N_m is having a pool of interaction sites on the genomic strand large enough that it is highly probable for each detaching bond to be soon replaced by another one, leading to a serial rather than parallel pattern of cooperativity.

As shown in Figures 5A and 7, probe sequences with $N_m > 20$ provide a large binding signal for M13mp18 and a K_d lower than that of the corresponding complementary

oligomers. This means that for such value of N_m, the equilibrium probability of at least two pairings between the genomic strand and the probe bed is rather large. The equilibrium constant $K_{d2} = k_{off2}/k_{on2}$ represents also the ratio between the probabilities of finding a single probe-specific site on M13mp18 in the unbound and bound state. Therefore, the value of K_{d2} extracted from the fit of the observed kinetics with Eq. 5 (Figure 8) corresponds to a probability $p_b = (1 + K_{d2})^{-1}$ as low as 1.8% that a specific M13mp18 site is bound to the surface probes. Thus, for 20 independent sites this corresponds to a probability of $(20 - 1)p_b = 34\%$ of forming a second binding when the first is present. Figure 7 shows that for $N_m = 20$ the binding strength of M13mp18 becomes similar to that of oligomers with sequence complementary to the surface probes. This value of N_m corresponds to a decrease of the overall K_d of a factor $F_1/N_m = 30$ due to the multiplicity of binding, and in turn represents a quantification of the difference of binding strength for complementary 20-mers and for a long chain with a complementary 20-mer, without multivalent binding.

The reduced affinity of M13mp18 sites to the surface probes—involving each single bond and the weak cooperativity—can be understood as an intrinsic entropic effect necessarily present every time a polymer binds a surface. This effect has been studied and quantified because of its general relevance not only in biology but in any process involving polymer adsorpion/desorption. An estimate of the entropic penalty ΔS_s inherent to the reduced conformational space of a polymer of gyration radius r_g bound in one position to a flat surface is $\Delta S_s = k_B \ln(\mathrm{Erf}(\ell/2r_g))$, where Erf is the error function and ℓ is the distance between the binding site and the surface (64). Assuming $\ell \approx 2$ nm and $r_g \approx 35$ nm, (equal to the hydrodynamic radius r_h measured by dynamic light scattering, see Supplementary Note S8), we expect a weakening of the equilibrium constant of about 30 times, in line with the values of K_d (M13mp18) $/K_d$ (oligomer) measured for low binding multiplicity from Figure 7, as well as with the enhancement factor F_1/N_m obtained from the model for K_d (M13mp18) $=K_d$ (oligomer) (i.e. $N_m = 20$). An even stronger entropy penalty ΔS_2 is associated to the formation of a second bond to the surface when the first is set. We can approximate ΔS_2 as the entropy loss when a second point of the circular M13mp18 genome is pinned on the first bound position. In this case $\Delta S_2 \approx 3/2 k_B \ln[n(n-m)^{-1}m^{-1}]$, where we used the notion that $3/2 k_B \ln(n_L)$ is the entropy of creating a loop in a polymer of length n_L (65). n and m are, respectively, the lengths of the entire M13mp18 genome and of the distance between the two binding positions, both measured in Khun lengths. With this simple approach we obtain $T\Delta S_2 = 4.5$–6.5 kcal mol^{-1} ($T = 25^\circ$C), depending on m. Having considered the entropy of loop with no surface, this result is an underestimate of the real ΔS_2.

The free energy penalty for the formation of an additional surface binding after the first can also be approximately determined from K_{d2} extracted from the experiments. Once the first bond is made, the position of the second probe-specific binding site on M13mp18 is confined to a volume of order r_b^3 in contact with the surface, corresponding to an effective concentration of $c_2 = 16.5$ μM. We can thus convert the first order equilibrium constant K_{d2} into a corresponding bimolecular dissociation constant as $K_{d2B} = c_2 K_{d2} \approx 1$ mM, which should be compared to the binding of the same probe-specific sequence when disconnected from the rest of the polymer. As a reference we take the equilibrium constant for the binding of freely diffusing 12-mers to their complementary probes on a surface, whose equilibrium constant for hybridization was found to be of the order of 10 nM (26,37), i.e. about 5 orders of magnitude stronger, in turn corresponding to a free energy penalty ascribed to the decrease of polymer entropy upon binding of $\Delta G \approx 6.8$ kcal mol^{-1} at 25°C. This value is in line with the upper extreme of the evaluation above and provides a further indication of the key role played by entropy in the phenomenology here described.

Despite the weak cooperativity of multiple binding of M13mp18, such effect can provide a large amplification factor of probe capturing capability. The kinetic model yielding to Eqs. (5) and (6) enables a quantification of the enhancement due to the multiplicity of binding (Supplementary Note S9). Most of the probes provide enhancement factor F_1/N_m of about 10 for K_d, and those with $N_m > 35$ display more than 100-fold decrease of the equilibrium constant (i.e. larger binding strength). Only one probe (NT2283) displays an enhancement factor larger than 1000 due to $N_m = 59$. Such nonlinear dependence of the enhancement factor on N_m is due to its combined effects on association and dissociation kinetics. Specifically, for large N_m the decrease of k_{off} through the F_1 factor becomes the major contribution.

Conclusion

We analyzed 50 surface-immobilized 20-mer probes complementary to different regions of the bacteriophage genome M13mp18 and observed very different capturing capabilities. Overall, the equilibrium constants of the whole set of probes spanned four orders of magnitude. Most of the probes showed a very slow binding as quantified by the the growth unit GU, whereas a few sequences provided enhanced capturing performance, with K_d as low as 10 pM, and k_{on} of 5×10^5 M^{-1} s^{-1}. The observed response pattern represents a fingerprint specific for the target genomic strand and we investigated how such response pattern is encoded in the sequence of the target strand. We estimated the accessibility of each region of the genomic strand by pairing probability matrix and coarse grained Monte Carlo, and the pairing stability with the probes by a mesoscale model for hybridization. Unexpectedly, we found that these features play a minor role in the binding strength and kinetics between the long genomic strand and the surface oligomer probes. Instead, we observed that the capturing capability largely depends on the number of binding sites N_m with small enough free-energy of hybridization of the long strand for the surface probes ($\Delta G < -10$ kcal mol^{-1}). Probe sequences with $N_m > 20$ displayed a large enhancement of both kinetic rates for association and dissociation, yielding K_d smaller than those for hybridization of oligomers with the same sequence. This represents a fundamental criterion of design for oligomer probes targeting long DNA or RNA strands with high affinity and specificity. Following the principle of the multiplicity of binding sites, an optimal set of probes can be selected for the specific capture and discrimination of long target strands, like viral or cellular genomic DNA, mRNA or long non-coding RNA. For the experimental conditions and sequence lengths used in this study, $N_m > 20$ appears as an effective threshold for the selection of probes with enhanced capturing strengths and kinetics.

The multiplicity of partial pairings with the surface probes represents a form of multivalent binding yielding a strong en-

hancement of the capturing capability for a few probes, corresponding to a LOD as small as 2 pM for direct, label-free detection. We found that the observed experimental dependence of the surface binding parameters with the multiplicity N_m is consistent with a model of weak cooperativity, in which the binding of the long strand to the surface by more than one site at the same time is strongly disfavoured by the entropic penalty associated to the constrained chain conformation. In such scenario, the cooperativity does not arise from the multiple parallel binding of many sites, which would lead to an exponential decrease of k_{off} with N_m, but from the combinatorial probability of forming more than one binding among several transient bindings, which leads to a smooth dependence on N_m, as shown in Figure 8. Such weak cooperativity of binding modulates in a predictable fashion the kinetics and strength of each probe sequence. The discovery of the proposed rational criterion to control the direct capture of M13mp18 highlights the importance of the multiplicity of weak interactions as a possibly general mechanism of binding of long NA strands. This finding is expected to be foundational towards its exploitation in diverse contexts, including DNA microarrays, surface-based sequencing technologies, and rapid methods for the direct detection of genomic strands (22–24), as well as for the design of DNA nanostructures on a surface (66). In all these cases, the sensitivity and specificity (61) of the capture of long NA strands can be finely modulated by taking into account the multiplicity of binding with oligomers that is inherent in the sequences.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding authors.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

Author contributions: G.N., G.Z., L.M. and M.B. conceived the study; T.C. and L.C. designed and performed the experiments; E.B., G.W. and L.M. performed the computational analysis for probe selection according to secondary structure assessment; G.N. developed the algorithm for the assessment of probe binding multiplicity; M.B. conceived the kinetic model for multivalent binding; T.B. conceived the thermodynamic modelling; G.N. and M.B. interpreted the experimental results and prepared the figures; F.D. and M.C. provided biosensor surfaces; G.N., G.Z., T.B., L.M. and M.B. wrote the manuscript with input from all of the authors.

Funding

This work has received funding from Ministero dell'Università e della Ricerca through grants: CN00000041 'National Center for Gene Therapy and Drugs based on RNA Technology', Spoke 5 and Spoke 9 (CUP G43C22001320007, PNRR MUR-M4C2-Investimento 1.4, funded by European Union—NextGenerationEU); PRIN 2022, 'MuSIx' project No. 2022H7MH23 (CUP G53D23000810006); PRIN 2022, 'CUBO' project No. 2022FJJA8W (CUP PRIN202223LMOLL_01). Luca Casiraghi was supported by young scientist fellowships from the UNIMI GSA-IDEA project. Gerald Weber was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brazil). Giovanni Nava and Luca Mollica were supported by 'Piano Sostegno alla Ricerca - anno 2020 Azione A—Linea A' (CUP G49C20001270001). The authors acknowledge support from the University of Milan through the APC initiative.

Conflict of interest statement

None declared.

References

1. Seeman,N.C. and Sleiman,H.F. (2017) DNA nanotechnology. *Nat. Rev. Mater.*, **3**, 17068.
2. Del Grosso,E., Franco,E., Prins,L.J. and Ricci,F. (2022) Dissipative DNA nanotechnology. *Nat. Chem.*, **14**, 600–613.
3. Engelhardt,F.A.S., Praetorius,F., Wachauf,C.H., Brüggenthies,G., Kohler,F., Kick,B., Kadletz,K.L., Pham,P.N., Behler,K.L., Gerling,T., *et al.* (2019) Custom-size, functional, and durable DNA origami with design-specific scaffolds. *ACS Nano*, **13**, 5015–5027.
4. Dey,S., Fan,C., Gothelf,K.V., Li,J., Lin,C., Liu,L., Liu,N., Nijenhuis,M.A.D., Saccà,B., Simmel,F.C., *et al.* (2021) DNA origami. *Nat. Rev. Meth. Primers*, **1**, 13.
5. Wei,X., Nangreave,J. and Liu,Y. (2014) Uncovering the self-assembly of DNA nanostructures by thermodynamics and kinetics. *Accounts Chem. Res.*, **47**, 1861–1870.
6. Leppek,K., Byeon,G.W., Kladwang,W., Wayment-Steele,H.K., Kerr,C.H., Xu,A.F., Kim,D.S., Topkar,V.V., Choe,C., Rothschild,D., *et al.* (2022) Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat. Commun.*, **13**, 1536.
7. Mattick,J.S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays*, **25**, 930–939.
8. Chillón,I. and Marcia,M. (2020) The molecular structure of long non-coding RNAs: emerging patterns and functional implications. *Crit. Rev. Biochem. Mol. Biol.*, **55**, 662–690.
9. Ken,M.L., Roy,R., Geng,A., Ganser,L.R., Manghrani,A., Cullen,B.R., Schulze-Gahmen,U., Herschlag,D. and Al-Hashimi,H.M. (2023) RNA conformational propensities determine cellular activity. *Nature*, **617**, 835–841.
10. Leonardi,A.A., Lo Faro,M.J., Petralia,S., Fazio,B., Musumeci,P., Conoci,S., Irrera,A. and Priolo,F. (2018) Ultrasensitive label- and PCR-Free genome detection based on cooperative hybridization of silicon nanowires optical biosensors. *ACS Sensors*, **3**, 1690–1697.
11. Kong,D., Wang,X., Gu,C., Guo,M., Wang,Y., Ai,Z., Zhang,S., Chen,Y., Liu,W., Wu,Y., *et al.* (2021) Direct SARS-CoV-2 nucleic acid detection by Y-shaped DNA dual-probe transistor assay. *J. Am. Chem. Soc.*, **143**, 17004–17014.
12. Hu,J., Safir,F., Chang,K., Dagli,S., Balch,H.B., Abendroth,J.M., Dixon,J., Moradifar,P., Dolia,V., Sahoo,M.K., *et al.* (2023) Rapid genetic screening with high quality factor metasurfaces. *Nat. Commun.*, **14**, 4486.
13. Alafeef,M., Moitra,P., Dighe,K. and Pan,D. (2021) RNA-extraction-free nano-amplified colorimetric test for point-of-care clinical diagnosis of COVID-19. *Nat. Protoc.*, **16**, 3141–3162.
14. Dighe,K., Moitra,P., Alafeef,M., Gunaseelan,N. and Pan,D. (2022) A rapid RNA extraction-free lateral flow assay for molecular point-of-care detection of SARS-CoV-2 augmented by chemical probes. *Biosens. Bioelectron.*, **200**, 113900.
15. Calorenni,P., Leonardi,A.A., Sciuto,E.L., Rizzo,M.G., Faro,M. J.L., Fazio,B., Irrera,A. and Conoci,S. (2023) PCR-free innovative strategies for SARS-CoV-2 detection. *Adv. Healthc. Mater.*, **12**, 2300512.

16. Ji,D., Guo,M., Wu,Y., Liu,W., Luo,S., Wang,X., Kang,H., Chen,Y., Dai,C., Kong,D., *et al.* (2022) Electrochemical detection of a few copies of unamplified SARS-CoV-2 nucleic acids by a self-actuated molecular system. *J. Am. Chem. Soc.*, **144**, 13526–13537.
17. Nikolaou,P., Sciuto,E.L., Zanut,A., Petralia,S., Valenti,G., Paolucci,F., Prodi,L. and Conoci,S. (2022) Ultrasensitive PCR-Free detection of whole virus genome by electrochemiluminescence. *Biosens. Bioelectron.*, **209**, 114165.
18. Roychoudhury,A., Allen,R.J., Curk,T., Farrell,J., McAllister,G., Templeton,K. and Bachmann,T.T. (2022) Amplification free detection of SARS-CoV-2 using multi-valent binding. *ACS Sensors*, **7**, 3692–3699.
19. Mercer,T.R. and Salit,M. (2021) Testing at scale during the COVID-19 pandemic. *Nat. Rev. Genet.*, **22**, 415–426.
20. Nava,G., Zanchetta,G., Giavazzi,F. and Buscaglia,M. (2022) Label-free optical biosensors in the pandemic era. *Nanophotonics*, **11**, 4159–4181.
21. Gao,Y. (2006) Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Res.*, **34**, 3370–3377.
22. Curk,T., Brackley,C.A., Farrell,J.D., Xing,Z., Joshi,D., Direito,S., Bren,U., Angioletti-Uberti,S., Dobnikar,J., Eiser,E., *et al.* (2020) Computational design of probes to detect bacterial genomes by multivalent binding. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 8719–8726.
23. Deal,B.R., Ma,R., Narum,S., Ogasawara,H., Duan,Y., Kindt,J.T. and Salaita,K. (2023) Heteromultivalency enables enhanced detection of nucleic acid mutations. *Nat. Chem.*, **16**, 229–238.
24. Xu,P., Cao,T., Fan,Q., Wang,X., Ye,F. and Eiser,E. (2023) Whole-genome detection using multivalent DNA-coated colloids. *Proc. Natl. Acad. Sci.*, **120**, 37.
25. Halperin,A., Buhot,A. and Zhulina,E.B. (2006) On the hybridization isotherms of DNA microarrays: the Langmuir model and its extensions. *J. Phys.: Cond. Matter*, **18**, S463–S490.
26. Vanjur,L., Carzaniga,T., Casiraghi,L., Chiari,M., Zanchetta,G. and Buscaglia,M. (2020) Non-Langmuir kinetics of DNA surface hybridization. *Biophys. J.*, **119**, 989–1001.
27. Qiao,W., Chiang,H.-C., Xie,H. and Levicky,R. (2015) Surface vs. solution hybridization: effects of salt, temperature, and probe type. *Chem. Commun.*, **51**, 17245–17248.
28. Oliviero,G., Federici,S., Colombi,P. and Bergese,P. (2011) On the difference of equilibrium constants of DNA hybridization in bulk solution and at the solid-solution interface. *J. Mol. Recogni.*, **24**, 182–187.
29. Chou,C.-C. (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res.*, **32**, e99.
30. Kreil,D.P., Russell,R.R. and Russell,S. (2006) In: *Microarray Oligonucleotide Probes*. Elsevier, pp. 73–98.
31. Bumgarner,R. (2013) Overview of DNA microarrays: types, applications, and their future. *Curr. Protoc. Mol. Biol.*, **Chapter 22**, Unit 22.1.
32. Šípová,H. and Homola,J. (2013) Surface plasmon resonance sensing of nucleic acids: a review. *Anal. Chimica Acta*, **773**, 9–23.
33. Jebelli,A., Oroojalian,F., Fathi,F., Mokhtarzadeh,A., Guardia and,M.d.l. (2020) Recent advances in surface plasmon resonance biosensors for microRNAs detection. *Biosens. Bioelectron.*, **169**, 112599.
34. Li,N., Canady,T.D., Huang,Q., Wang,X., Fried,G.A. and Cunningham,B.T. (2021) Photonic resonator interferometric scattering microscopy. *Nat. Commun.*, **12**, 1733.
35. Giavazzi,F., Salina,M., Cerbino,R., Bassi,M., Prosperi,D., Ceccarello,E., Damin,F., Sola,L., Rusnati,M., Chiari,M., *et al.* (2013) Multispot, label-free biodetection at a phantom plastic–water interface. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 9350–9355.
36. Carzaniga,T., Casiraghi,L., Nava,G., Zanchetta,G., Inzani,T., Chiari,M., Bollati,V., Epis,S., Bandi,C., Lai,A., *et al.* (2024) Serum antibody fingerprinting of SARS-CoV-2 variants in infected and vaccinated subjects by label-free microarray biosensor. *Front. Immunol.*, **15**, 1323406.
37. Nava,G., Ceccarello,E., Giavazzi,F., Salina,M., Damin,F., Chiari,M., Buscaglia,M., Bellini,T. and Zanchetta,G. (2016) Label-free detection of DNA single-base mismatches using a simple reflectance-based optical technique. *Phys. Chem. Chem. Phys.*, **18**, 13395–13402.
38. Chiodi,E., Damin,F., Sola,L., Ferraro,L., Brambilla,D., Ünlü,M.S. and Chiari,M. (2021) A reliable, label free quality control method for the production of DNA microarrays with clinical applications. *Polymers*, **13**, 340.
39. Vanjur,L., Carzaniga,T., Casiraghi,L., Zanchetta,G., Damin,F., Sola,L., Chiari,M. and Buscaglia,M. (2021) Copolymer coatings for DNA biosensors: effect of charges and immobilization chemistries on yield, strength and kinetics of hybridization. *Polymers*, **13**, 3897.
40. Lorenz,R., Bernhart,S.H., Höner zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithm. Mol. Biol.*, **6**, 26.
41. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci.*, **101**, 7287–7292.
42. Whitelam,S. and Geissler,P.L. (2007) Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles. *J. Chem. Phys.*, **127**, 154101.
43. Šulc,P., Romano,F., Ouldridge,T.E., Rovigatti,L., Doye,J.P. and Louis,A.A. (2012) Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.*, **137**, 135101.
44. Okabe,T., Kawata,M., Okamoto,Y. and Mikami,M. (2001) Replica-exchange Monte Carlo method for the isobaric–isothermal ensemble. *Chem. Phys. Lett.*, **335**, 435–439.
45. Suma,A., Poppleton,E., Matthies,M., Šulc,P., Romano,F., Louis,A.A., Doye,J.P.K., Micheletti,C. and Rovigatti,L. (2019) TacoxDNA: A user-friendly web server for simulations of complex DNA structures, from single strands to origami. *J. Comput. Chem.*, **40**, 2586–2595.
46. Abraham,M.J., Murtola,T., Schulz,R., Páll,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
47. Weber,G. (2015) Optimization method for obtaining nearest-neighbour DNA entropies and enthalpies directly from melting temperatures. *Bioinformatics*, **31**, 871–877.
48. Weber,G. (2013) Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures. *Nucleic Acids Res.*, **41**, e30.
49. Smith,T. and Waterman,M. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
50. SantaLucia,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
51. Zadeh,J.N., Steenberg,C.D., Bois,J.S., Wolfe,B.R., Pierce,M.B., Khan,A.R., Dirks,R.M. and Pierce,N.A. (2010) NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170–173.
52. Salina,M., Giavazzi,F., Lanfranco,R., Ceccarello,E., Sola,L., Chiari,M., Chini,B., Cerbino,R., Bellini,T. and Buscaglia,M. (2015) Multi-spot, label-free immunoassay on reflectionless glass. *Biosens. Bioelectron.*, **74**, 539–545.
53. Vainrub,A. and Pettitt,B.M. (2002) Coulomb blockage of hybridization in two-dimensional DNA arrays. *Phys. Rev. E*, **66**, 041905.
54. Halperin,A., Buhot,A. and Zhulina,E. (2004) Sensitivity, Specificity, and the Hybridization Isotherms of DNA Chips. *Biophys. J.*, **86**, 718–730.

55. Mambretti,F., Pedrani,N., Casiraghi,L., Paraboschi,E.M., Bellini,T. and Suweis,S. (2022) OxDNA to study species interactions. *Entropy*, **24**, 458.
56. Ouldridge,T.E., Šulc,P., Romano,F., Doye,J.P.K. and Louis,A.A. (2013) DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.*, **41**, 8886–8895.
57. Zhang,J.X., Fang,J.Z., Duan,W., Wu,L.R., Zhang,A.W., Dalchau,N., Yordanov,B., Petersen,R., Phillips,A. and Zhang,D.Y. (2017) Predicting DNA hybridization kinetics from sequence. *Nat. Chem.*, **10**, 91–98.
58. Zanchetta,G., Carzaniga,T., Vanjur,L., Casiraghi,L., Tagliabue,G., Morasso,C., Bellini,T. and Buscaglia,M. (2021) Design of a rapid, multiplex, one-pot miRNA assay optimized by label-free analysis. *Biosens. Bioelectron.*, **172**, 112751.
59. Lanfranco,R., Giavazzi,F., Salina,M., Tagliabue,G., Di Nicolò,E., Bellini,T. and Buscaglia,M. (2016) Selective adsorption on fluorinated plastic enables the optical detection of molecular pollutants in water. *Phys. Rev. Appl.*, **5**, 054012.
60. Kitov,P.I. and Bundle,D.R. (2003) On the nature of the multivalency effect: a thermodynamic model. *J. Am. Chem. Soc.*, **125**, 16271–16284.
61. Dubacheva,G.V., Curk,T. and Richter,R.P. (2023) Determinants of superselectivity-practical concepts for application in biology and medicine. *Accounts Chem. Res.*, **56**, 729–739.
62. Huskens,J., Mulder,A., Auletta,T., Nijhuis,C.A., Ludden,M. J.W. and Reinhoudt,D.N. (2004) A model for describing the thermodynamics of multivalent host-guest interactions at interfaces. *J. Am. Chem. Soc.*, **126**, 6784–6797.
63. Martinez-Veracoechea,F.J. and Frenkel,D. (2011) Designing super selectivity in multivalent nano-particle binding. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10963–10968.
64. Lee,C.-T. and Terentjev,E.M. (2019) Hard-wall entropic effect accelerates detachment of adsorbed polymer chains. *Phys. Rev. E*, **100**, 032501.
65. Schubert,U., Winter,A. and Newkome,G. (2021) In: *Supramolecular Polymers and Assemblies chapter Supramolecular Polymers: General Considerations*. Wiley, pp. 1–28.
66. Cao,H.H., Abel,G.R., Gu,Q., Gueorguieva,G.-A.V., Zhang,Y., Nanney,W.A., Provencio,E.T. and Ye,T. (2020) Seeding the self-assembly of DNA origamis at surfaces. *ACS Nano*, **14**, 5203–5212.

Received: March 22, 2024. **Revised:** June 4, 2024. **Editorial Decision:** June 19, 2024. **Accepted:** June 20, 2024