nature nanotechnology

ARTICLES

https://doi.org/10.1038/s41565-020-0720-7

Programming multi-protein assembly by gene-brush patterns and two-dimensional compartment geometry

Ohad Vonshak[1,4], Yiftach Divon[1,4], Stefanie Förste[2], David Garenne[3], Vincent Noireaux[3], Reinhard Lipowsky[2], Sophia Rudorf[2], Shirley S. Daube[1]✉ and Roy H. Bar-Ziv[1]✉

The assembly of protein machines in cells is precise, rapid, and coupled to protein synthesis with regulation in space and time. The assembly of natural and synthetic nanomachines could be similarly controlled by genetic programming outside the cell. Here, we present quasi-two-dimensional (2D) silicon compartments that enable programming of protein assembly lines by local synthesis from surface-immobilized DNA brushes. Using this platform, we studied the autonomous synthesis and assembly of a structural complex from a bacteriophage and a bacterial RNA-synthesizing machine. Local synthesis and surface capture of complexes provided high assembly yield and sensitive detection of spatially resolved assembly intermediates, with the 3D geometry of the compartment and the 2D pattern of brushes dictating the yield and mode of assembly steps. Localized synthesis of proteins in a single gene brush enhances their interactions, and displacement of their genes in separated brushes leads to step-by-step surface assembly. This methodology enables spatial regulation of protein synthesis, and deciphering, reconstruction and design of biological machine assembly lines.

Protein complexes that function as molecular machines are prevalent in nature and offer new opportunities for bottom-up fabrication, and for the development of soft robots made of compliant materials responsive to the environment and compatible with the human body[1–8]. These complexes are inspiring because of their versatile shape, functionality, efficient modes of operation and diverse mechanisms of energy conversion. The synthesis of protein parts in the cell occurs by genetically regulated reactions driven by expenditure of chemical energy. Protein complexes self-assemble at reduced dimensionality, confined to minute volumes in sub-cellular organelles, or localized at internal surfaces where mobility is reduced and interactions are enhanced[9]. Assembly may occur during protein synthesis, impacted by the spatial organization of genes in operons and clusters[10], by ribosome and mRNA localization to specific cellular targets[11] and by the crowded cytoplasm[12]. To implement regulated cell-free synthesis and assembly of functional protein machines we sought to define a compartmentalized system that recapitulates the spatial organizational features of cells.

The cell-free synthesis of proteins from genes in one-pot solution reactions and their assembly into functional machines has recently been demonstrated[13–16], yet these systems lack spatial organization and confinement. As a step beyond bulk solution experiments, the use of surfaces to immobilize DNA templates for protein synthesis adds a spatial component to the reaction[17]. Dense DNA bound to the surface forms a gene brush, which localizes the synthesis of RNA and protein products[18]. Capturing the products on the surface further creates a localized genotype–phenotype linkage[19] and supports multi-protein machine assembly[20], yet products dilute into the large reaction volume in the absence of physical boundaries. Confining gene brushes in quasi-two-dimensional (2D) silicon compartments maintains high concentrations of protein products[21,22] and could thus serve as a platform for confinement of machine assembly reactions.

Here we demonstrate multi-protein machine assembly in 2D compartments with well-defined geometrical parameters. Having the complete synthesis and assembly reaction confined to a volume of a living cell, with all assembly intermediates captured on the entire surface, allows regulation of protein interactions by patterns of gene brushes as well as the geometry of the compartments. As model systems, we reconstitute the assembly of the T4 bacteriophage wedge, a protein complex that is part of the cell-puncturing machine, and *Escherichia coli* RNA polymerase, a central multimeric enzyme, in high-throughput and multiplexed chips. We first show that each compartment holds the capacity for the synthesis and detection of hundreds of potentially different complexes. Then, separating gene brushes in 1D layouts creates self-organized surface-bound assembly intermediates. We provide a means to decipher protein machine assembly lines, and reveal a transition between scaffolded and solution assembly. Finally, we show that assembly yield increases with compartment volume and that spatial partitioning of reaction resources allows to locally silence genes.

Synthesis of protein assemblies in 2D compartments

We etched into a silicon wafer an array of compartments of radius 200 μm and 2 μm height embedded in a relief structure of height 50 μm above the surface. In the centre of each compartment, we immobilized linear DNA polymers, 1–3.5 kilobase pairs (kb) long, packed as a gene brush with a diameter of ~60 μm and a typical gene density of ~1,000 μm^{-2} surrounded by protein traps immobilized on the entire remaining surface (Fig. 1a and Supplementary Fig. 1). Gene brushes localize RNA polymerases (RNAP) and ribosomes to the vicinity of the brush with a concomitant localization of reaction

[1]Department of Chemical and Biological Physics, The Weizmann Institute of Science, Rehovot, Israel. [2]Max Planck Institute of Colloids and Interfaces, Potsdam, Germany. [3]School of Physics and Astronomy, University of Minnesota, Minneapolis, MN, USA. [4]These authors contributed equally: Ohad Vonshak, Yiftach Divon. ✉e-mail: Shirley.Daube@weizmann.ac.il; Roy.Bar-Ziv@weizmann.ac.il

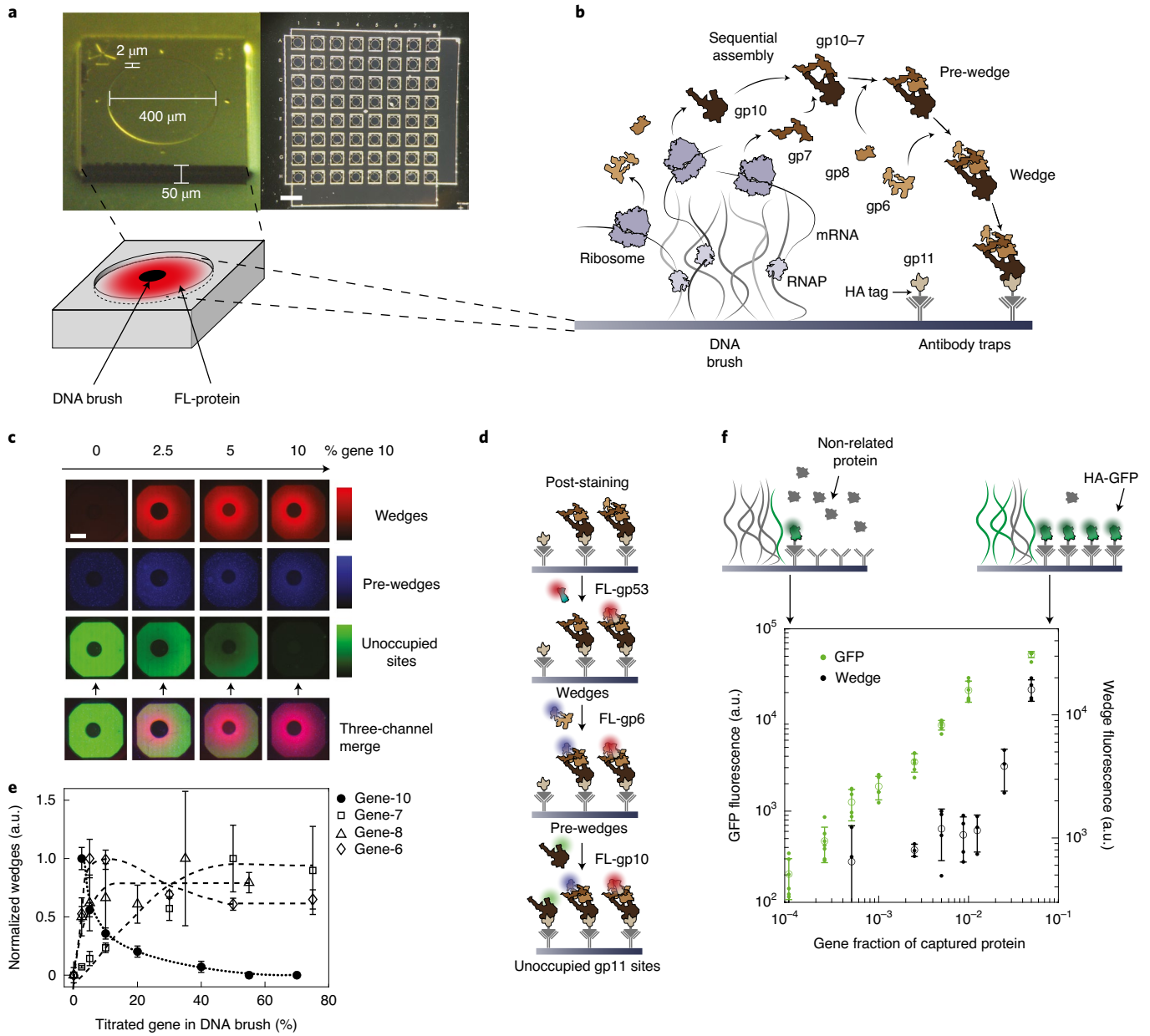


Fig. 1 | Protein synthesis and assembly in 2D compartments. a, Images and scheme of an array of silicon 2D compartments, with a central DNA brush (black circle) surrounded by captured wedges (stained by FL-protein, red). Scale bar in top right image, 1 mm. **b**, T7 RNAPs (light grey) and ribosomes (dark grey) localize to DNA brush and express wedge proteins (brown) that assemble sequentially and bind to gp11 (light brown) on surface antibodies. **c**, Four compartments with variable gene-10 fraction, and fixed gene-6, 7 and 8. Post-staining of wedges, pre-wedges and unoccupied gp11 sites are shown separately and as composite images (bottom row). Scale bar, 100 μm. **d**, Post-staining scheme in reverse order to the assembly pathway. **e**, Normalized dose response curves of wedges formed at different gene fractions, titrated separately (see Methods). Individual data points and similar sets of experiments with similar conditions are presented in Supplementary Fig. 3. Lines are a guide to the eye. **f**, GFP and wedges captured as a function of their gene fraction in the DNA brush, complemented by an expressible non-related gene. Gene compositions (**c**,**e**,**f**) are given in Supplementary Table 3. The number of samples for each data point (**c**,**e**,**f**) is listed in Supplementary Table 4. Data are presented as mean values, error bars represent ± s.d.

products[23], while the protein traps capture one of the brush-encoded proteins, designated as the scaffold. Other proteins could be bound on the surface through complex formation, mediated and dependent on the scaffold protein and the proteins that precede it in the assembly line (see Methods, Fig. 1b and Supplementary Fig. 1a). All the genes were under control of the T7 promoter and optimized for cell-free expression (CFE) in an *E. coli* extract[24,25].

We chose the T4 wedge protein complex, with much information on its structure and assembly pathway[26,27]. The wedge proteins bind sequentially in a stepwise contingent mode, with gene product 8 (gp8) binding only to gp10–7 complexes (forming pre-wedges), and gp6 binding only to pre-wedges (forming wedges)[26,27] (Fig. 1b and Supplementary Fig. 2a). We verified that each of the wedge proteins could be expressed in bulk off-chip CFE reactions and assemble to wedges when co-expressed (Supplementary Fig. 2b–e). We then asked whether they could assemble on-chip from gene brushes in the compartments. We chose gp11 to be the capture protein for on-chip assembly, as it is the only non-sequential step in wedge assembly

and can bind to all sub-complexes[27] (Supplementary Fig. 2a). We pre-synthesized it off-chip with a haemagglutinin (HA)-affinity tag (gp11-HA) and bound it to anti-HA surface antibodies surrounding a central mixed DNA brush, composed of genes-6, 7, 8 and 10 (see Methods).

In each compartment within the array we immobilized a DNA brush containing different compositions of wedge genes, from zero to maximal value, with all other genes kept constant. We initiated on-chip synthesis by filling all the compartments with the *E. coli* cell-free extract followed by sealing, incubation and washing. The compartments were stained by a three-step post-staining procedure, in reverse order to the assembly line (see Methods), starting with a pre-synthesized off-chip fluorescently-labelled (FL) gp53 protein, known to bind sequentially to gp6 (ref. [27]), thus detecting only complete wedges. The next two staining steps were with FL-gp6, followed by FL-gp10, to quantitatively detect pre-wedges, and unoccupied gp11 sites that remained unbound, respectively (Fig. 1c–e and Supplementary Fig. 3). The complement of the FL-gp10 signal represents all gp11 sites that were bound during the CFE reaction (occupied sites). Compartments with a deletion of any of the genes resulted in background signal levels, confirming that the staining signals were reporting on wedge assembly. The dose response curves exhibited increased wedge formation up to saturation values, except for gp10, which exhibited a narrow peak. We calculated that ~90% of all pre-wedges converted to wedges (see Methods and Supplementary Fig. 4).

We assessed the detection sensitivity of our chips by diluting the wedge genes in the brush by active-gene coding for the unrelated protein dihydrofolate reductase (see Methods), thereby maintaining a constant load on the CFE machinery in every compartment within the array. We detected wedge formation by FL-gp53 post-staining, diluting the wedge genes in the brush at fixed ratios, down to a fraction of 10^{-2} for the capture protein gene (gene-10), before obtaining background levels (Fig. 1f). This implies a capacity for coding, expressing and detecting a protein complex with ~100 different subunits. In addition, as a model for a single-subunit protein, we obtained a linear curve of captured HA-tagged green fluorescent protein (HA-GFP) diluted from 5×10^{-2} to 10^{-4} gene fraction, approaching saturating values above 10^{-2}. Hence, below this value all HA-GFP was captured on the surface, suggesting that ~10,000 different proteins could be synthesized from a single brush and ~20% of those could be captured.

Protein assembly line in a 1D geometry

We interpreted the narrow peak in the gp10 titration (Fig. 1e) as a consequence of the non-sequential nature of the gp11–10 interaction (Supplementary Fig. 2a). Since gp7 can bind either gp10 or gp11–10 (ref. [27]), excess gp10 could saturate gp11 on the surface and sequester wedges from binding to the surface (Extended Data Fig. 1a). This sequestration, known as the 'prozone' or high-dose hook effect[28,29], was not observed in titrations of genes that interact sequentially (Fig. 1e). Sequential steps are robust to variations in stoichiometry because advancing from one step to the next occurs by only one route. In contrast, non-sequential steps may proceed by competing routes, becoming sensitive to variations in stoichiometry[30], and hence decrease in assembly yield. Interestingly, gene-10 and 11 are organized on a single operon in the T4 genome[31], suggesting that in vivo stoichiometry is closely regulated at the level of gene expression.

The above analysis suggested that at least some of the wedge assembly occurred in the brush vicinity before surface binding, rather than by step-by-step surface scaffolding. To test this notion, we created a scenario that generated kinetic competition between these two assembly modes (Extended Data Fig. 1a,b). We pre-bound gp10 on the surface surrounding mixed brushes of gene-10 and gene-7, such that nascent gp7 can bind either to gp10 on the surface or to nascent gp10 in the solution. By post-staining with FL-gp8, we found that nascent gp10 reduced formation of gp10–7 complexes on the surface, suggesting that the rate of coupled synthesis and assembly in the brush vicinity outcompeted the rate of assembly by scaffolding.

We speculated that displacing genes in separated brushes could provide an experimental knob to dial between solution and scaffolded assembly. We designed rectangular compartments of dimensions $200\times1{,}000\,\mu m^2$, which enable positioning of genes along a line (Fig. 2a). In each compartment we pre-bound gp11-HA on the entire surface and patterned brush doublets (see Methods) coding for gp10 and gp7 at variable distances, from close packing to 500 µm separation. Post-staining revealed profiles of gp10–7 complexes and unoccupied gp11 sites (Fig. 2c and Supplementary Fig. 5a). The profiles were consistent with a scenario of localized protein synthesis in the DNA brush followed by diffusion to the next available surface sites. In compartments with mixed brushes, both profiles coincided and displayed slow decay. Separating the two genes by 100 µm resulted in sharper decaying profiles (Fig. 2d). As the two brushes were separated farther apart, the profile of the occupied sites remained unchanged, whereas the gp10–7 profile gradually became narrower concomitantly to a 3-fold reduction in gp10–7 total complexes formed (Fig. 2e).

To explain the data, we simulated synthesis, interaction, diffusion and surface capture of gp10 and gp7 in a 1D layout (see Methods and Supplementary Fig. 11). We reproduced the main features of the observations using a set of experimentally relevant parameters for all configurations, including an adjustment that binding of gp10–7 to pre-bound gp11 is 10-fold weaker than binding of gp10 to pre-bound gp11, possibly due to steric hindrance. Fitting the high-yield and shallow profile of gp10–7 in the mixed brush required increasing the probability of binding of gp10 to gp7 synthesized in the same brush by at least 10-fold higher than by diffusion-limited interaction for separate brushes. The simulation showed that in the mixed brush, 62% of the gp10–7 captured complexes were formed in solution and the rest by scaffolding. Near the brush, the gp10–7 profile was dominated by diffusion and surface binding of free gp10 proteins that prevented binding of gp10–7 complexes, whereas far from the brush, the profile was dominated by solution assembly and diffusion of gp10–7 to the next available site. In contrast, separation of brushes led to reduced overall yields due to lower local protein concentrations, but eliminated the competition between the two modes (Fig. 2d and Supplementary Fig. 5b). This notion was further corroborated by simulating four situations of two genes (gene-10 and gene-7) at two different gene ratios (1:1 and 3:1, respectively) and at two distances (mixed brush and 500 µm) (Fig. 2f). At short times, up to 30 min, there was a substantially higher assembly yield (see Methods) in mixed brushes compared with separate brushes. At longer times, assembly yield from mixed brushes at the imbalanced stoichiometry decreased with time due to sequestration. In contrast, scaffolded assembly at 500 µm separation was enhanced by excess gp10 and surpassed the solution assembly.

We tested the effect of gene separation on wedge assembly by immobilizing either two, three or four of the wedge genes in separate brushes along the compartment axis according to their binding order. Three layouts were compared: mixed in a single brush, separated by 100 µm, and spread out by 250 µm (Fig. 3a,b and Supplementary Fig. 6a,b). After post-staining we obtained 27 reaction profiles depicting the spatial distribution of assembly intermediates. In the mixed layout, all profiles were centred around the single brush and overlapped with the gp10–7 complex profile, while profiles of the separated and spread out layouts had an apparent peak shift away from the gene-10 brush. That is, the relative position of genes-7, 8 and 6 to gene-10 seemed to have dictated a shift from solution to scaffolded assembly as the distance between the brushes increased.

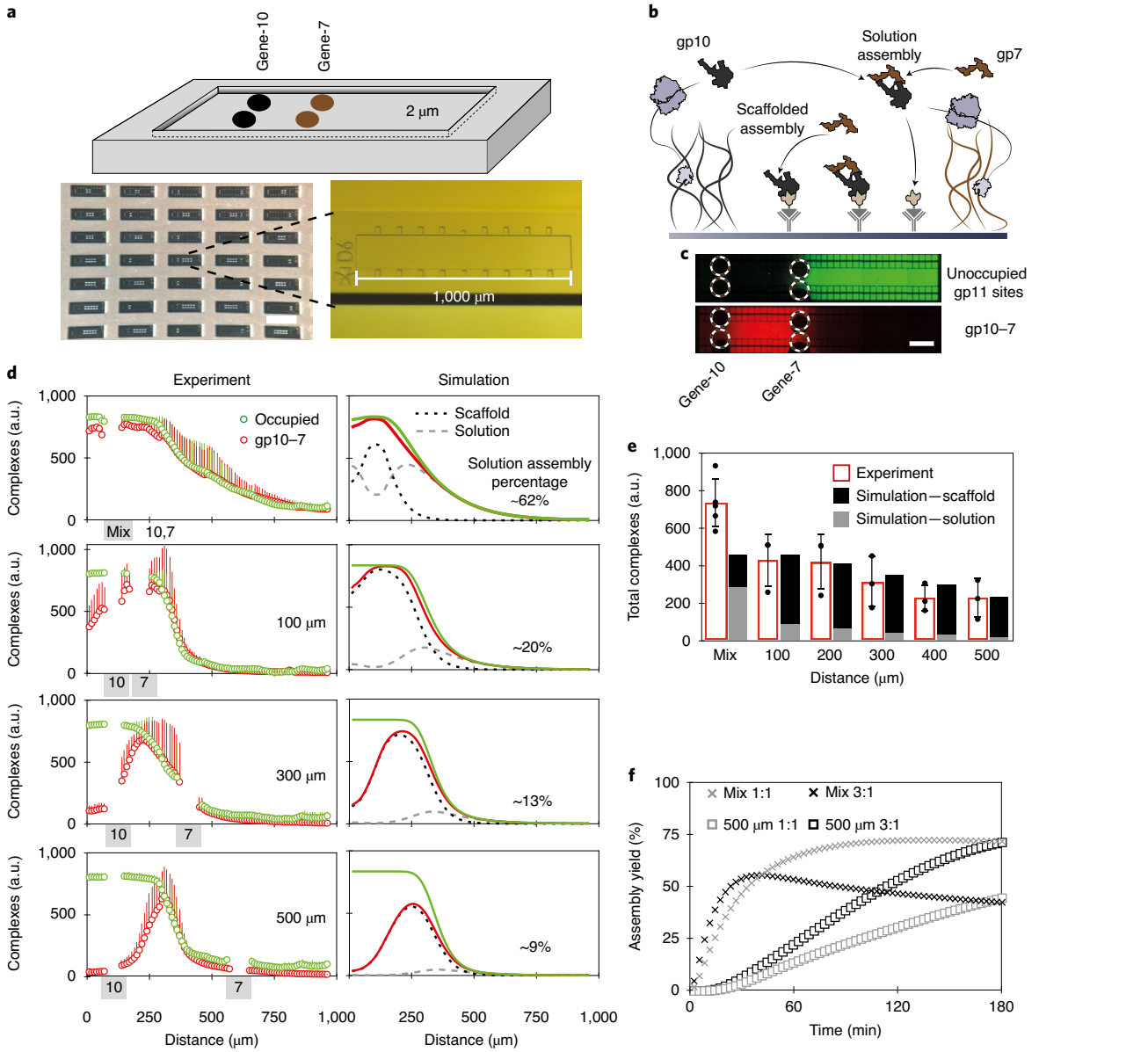


Fig. 2 | Solution assembly and surface scaffolding in 1D layouts. a, Image and scheme of rectangular compartments (200 × 1,000 × 2 μm^3) with brush doublets of gene-10 and 7 patterned on the long axis. Scale bar, 1 mm. **b**, Scheme of solution and scaffolded assembly of gp10 and gp7. **c**, Image of a compartment with gene-10 and gene-7 brushes, 300 μm apart, with surface-captured gp10–7 complexes post-stained with FL-gp8 (red), and unoccupied sites with FL-gp10 (green). Two additional independent similar experiments are shown in Supplementary Fig. 5a. Scale bar, 100 μm. **d**, Profiles of gp10–7 (red) and occupied sites (green) as in **c**, with increasing distance between gene-10 and gene-7 brushes. Computer simulation (see Methods) fitted to 2 h of CFE reveals profiles of occupied sites (green) and total gp10–7 complexes (red), formed by solution (grey) or scaffolded (black) assembly. Relative solution assembly of the total captured gp10–7 (red curve) is noted for each graph. Individual curves are presented in Supplementary Fig. 5b. Similar sets of experiments with different conditions are presented in Supplementary Fig. 5c. Data are presented as mean values, error bars are + s.d. **e**, Total number of gp10–7 complexes in the experiment and simulation in **c** (solution and scaffolded assembly, as denoted). **f**, Computer simulation of gp10–7 assembly yield at two gene ratios (1:1 and 3:1), and in mixed brushes or separated by 500 μm. Data are presented as mean values, error bars represent ± s.d. Detailed gene compositions of all experiments are given in Supplementary Table 3. The number of samples for each data point (**d**,**e**) is listed in Supplementary Table 4.

In addition, we observed that the brush separation dramatically reduced the assembly yield of wedges (~75, 50, 0%) but not of pre-wedges (~90, 70, 75%) in the mixed, separated and spread out layout, respectively (see Methods and Supplementary Fig. 7a,b). To investigate this, we displaced only one brush, of either gene-6 or gene-8, from the rest of the wedge genes at increasing distances up to 750 μm (Fig. 3c and Supplementary Fig. 7). Wedge assembly was not affected by gene-8 displacement, but was substantially reduced for the farthest gene-6 brush (Fig. 3d), suggesting that gp6 binds more weakly to pre-wedges than gp8 does to gp10–7 complexes. Therefore, weak interactions seem to be compensated by high concentrations from a mixed brush or brushes in close proximity, and not in spread out configurations.

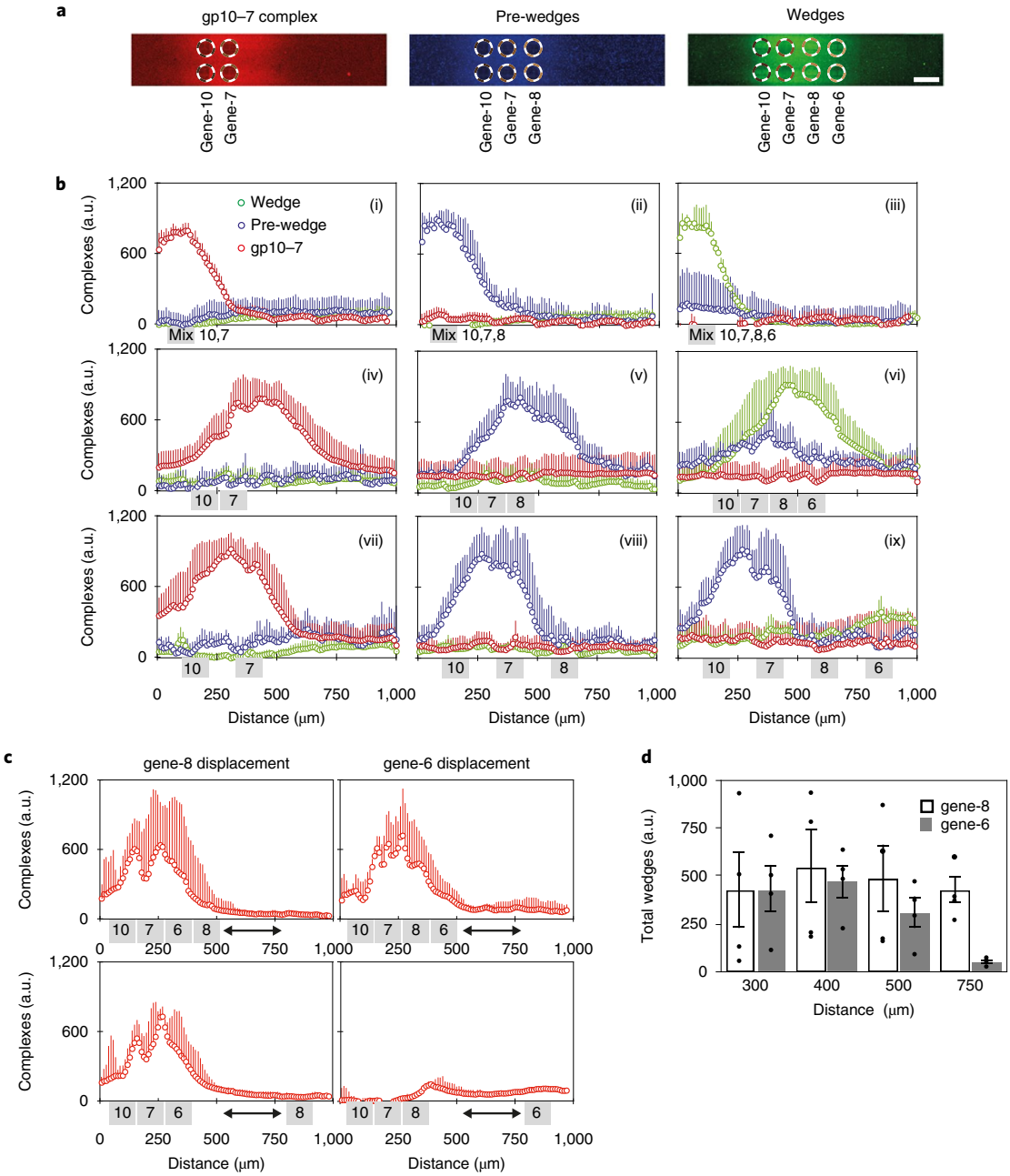


Fig. 3 | The effect of gene brush layout on wedge assembly. a, Images of post-stained compartments of gp10–7 complexes (left), pre-wedges (centre) and wedges (right) from a separated brush layout, post-stained with FL-gp8, FL-gp6 and FL-gp53, respectively. Similar results for four additional independent experiments are shown in Supplementary Fig. 6a. Scale bar, 100 μm. **b**, Twenty-seven profiles, three for each compartment, of wedges and intermediates stained as in **a**, from mixed ((i), (ii), (iii)), separated ((iv), (v), (vi)) and spread-out ((vii), (viii), (ix)) gene brushes. Gene number and position is denoted below each graph. Individual curves are presented in Supplementary Fig. 6b. Data are presented as mean values, error bars are + s.d. **c**, Profiles of wedges, with gene-8 (left) or gene-6 (right) displaced from the rest of wedge genes, post-stained with FL-gp53. Data are presented as mean values, error bars are + s.d. **d**, Total wedges integrated across the entire compartments in **c** as a function of the inter-brush distance between gene-10 and gene-8 or gene-6. Data are presented as mean values, error bars represent ± s.d. Detailed gene compositions of all experiments are given in Supplementary Table 3. The number of samples for each data point (**b**–**d**) is listed in Supplementary Table 4.

Deciphering the order of an assembly line

We noted that our design tools may in principle be used to decipher the assembly order of an unknown set of proteins. However, a modification was required, since the choice of the HA-protein and the order of the post-staining procedure in the above wedge analysis relied on available information. We next demonstrate an in situ labelling methodology enabling us to identify the optimal scaffold protein and to verify the wedge assembly order without prior knowledge. The main change is the coding of the HA- and FL-proteins in the gene brushes and their expression on-chip rather than by adding them through the pre-binding and post-staining steps, respectively (see Methods). Different combinations of HA and FL gene brushes patterned in each compartment could reveal the optimal HA-protein for capturing the maximal amount of

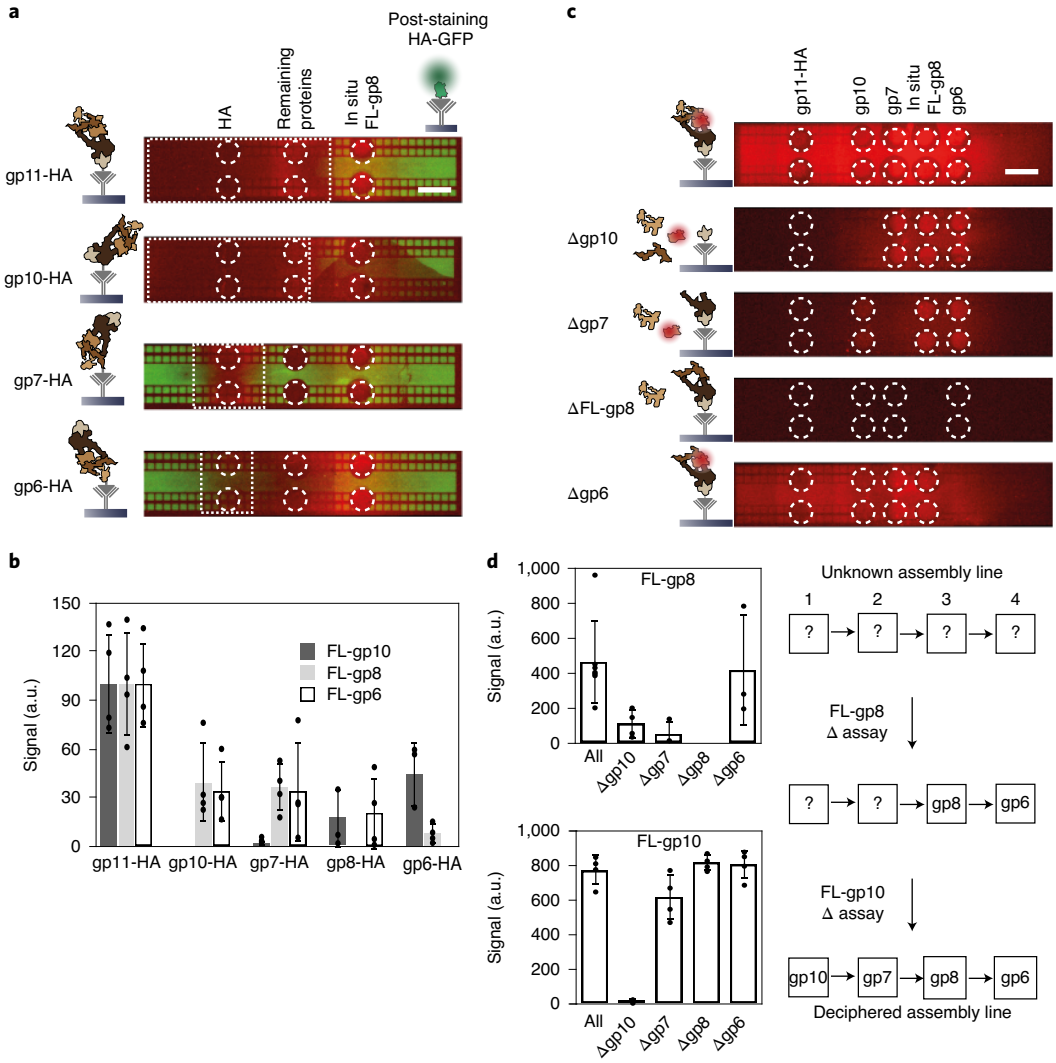


Fig. 4 | Deciphering assembly line order. a, Images of compartments, in situ labelled with FL-gp8 during the on-chip reaction (red). In situ expressed HA-protein revealed by post-staining with HA-GFP (green). The overlap of the red and green areas is marked with a dotted white box. Each compartment had a different HA-protein, which is indicated on the left along with a scheme of the possible conformation of the captured wedge. The position of the brushes coding for the HA-protein (left), FL-gp8 (right) and the remaining genes (centre) is marked with white circles. **b**, Integral of the red signal in the dotted white box in **a** (see Methods) for all combinations of three different FL-proteins and five different HA-proteins, as in **a**. **c**, Images of wedge profiles, in situ labelled with FL-gp8 with HA-gp11 chosen in **b**, Separated brushes of wedge genes are marked with white circles, in each compartment a different brush is deleted. Schematics on the left depict interactions of wedge genes due to the deletion. **d**, Integral of wedge profiles (in **c**), in situ labelled with FL-gp8 (top) and with FL-gp10 (bottom). The flow-chart on the right depicts how the assembly order is deciphered based on the deletion assays. Signal is an integral of the entire compartment. For all experiments, gene compositions are given in Supplementary Table 3. Individual data points and mean values are presented, error bars represent ± s.d. Scale bars, 100 μm. The number of samples for each data point (**a–d)** is listed in Supplementary Table 4.

complexes. Then, combinations of deleted genes and FL-proteins could reveal the assembly order. For example, if no signal appeared on the surface, the corresponding deleted protein preceded the FL-protein in the assembly line.

In every 1D compartment we immobilized three brushes, one of an HA-gene, one of a FL-gene and one with the rest of genes (Fig. 4a). The profile of the HA-protein was revealed by post-staining with HA-GFP that bound to all unoccupied sites. If assembly occurred, the profile of the FL-protein and of the HA-protein (complement of the HA-GFP profile) would coincide. We found that gp11 had the highest overlap region between the two profiles for all FL-proteins, suggesting that it was indeed the optimal surface capture of wedges (Fig. 4a,b). To decipher the assembly order using gp11-HA as the capture protein, we then tested combinations of gene deletions and FL-proteins (Fig. 4c). The integrals of all FL-protein profiles were compared, revealing that FL-gp8 preceded gp6 as its deletion did not affect the profile but follows that of gp10 and 7 as their deletions diminished the profile. FL-gp10 was found to precede gp6, 7 and 8 as its profile was not affected by their deletion (Fig. 4d). For a five-protein complex, this analysis was sufficient to determine order. For larger complexes, a more rigorous scan may be conducted.

Machine assembly and gene silencing impacted by geometry

To expand the scope of our approach, we studied the on-chip synthesis and assembly of *E. coli* RNAP, a five-protein molecular machine responsible for the transcription of every gene in *E. coli*, and present in the *E. coli* extract that we have used so far[32]. We immobilized

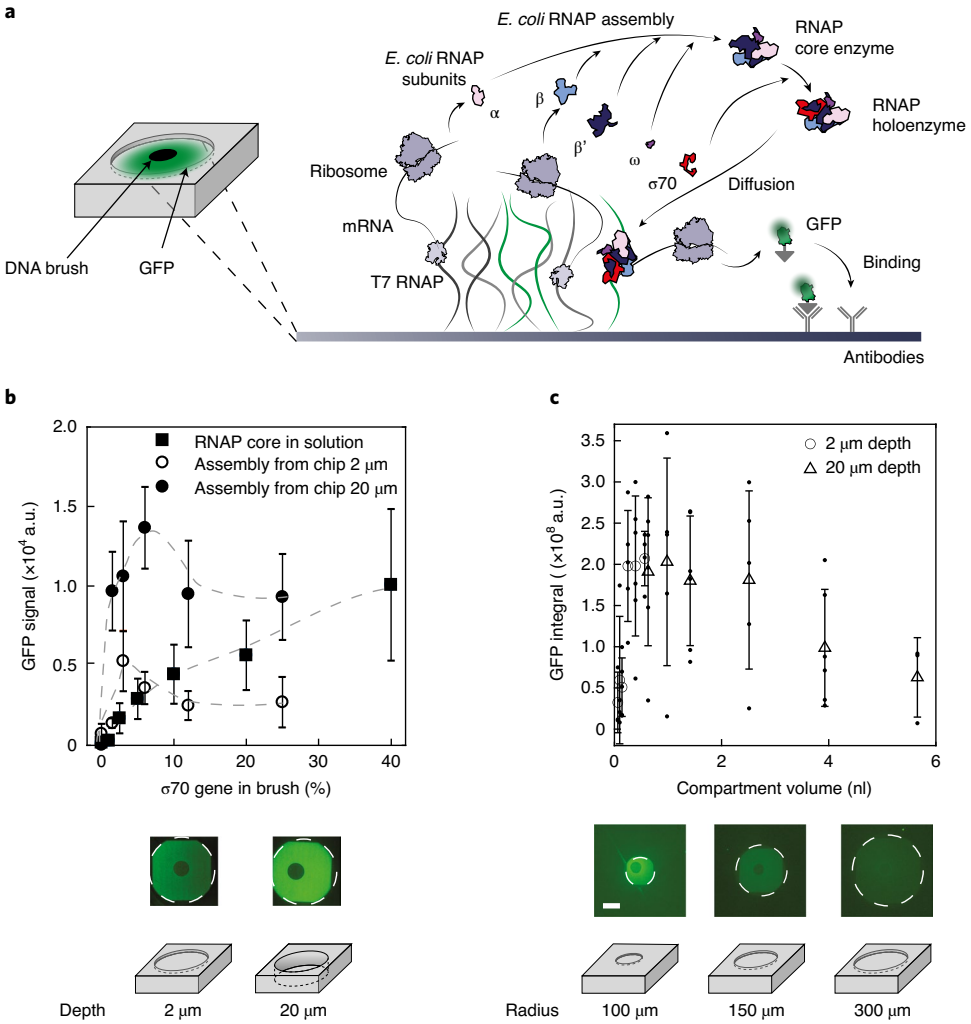


Fig. 5 | Gene composition and compartment geometry impact *E. coli* RNAP assembly. a, Schematic of a cascaded reaction of synthesis, assembly and activity of *E. coli* RNAP core enzyme and σ70 subunits (under control of T7 promoter) from a mixed brush. Nascent assembled *E. coli* RNAP enzyme binds to GFP genes (green strands) under the control of the *E. coli* P70 promoter. Colours and object scheme as in Fig. 1a. **b**, GFP signal as a function of σ70 gene fraction in the DNA brush, in 2-μm-deep (empty circles) and 20-μm-deep (full circles) compartments. All core genes were present in the brush. Purified core enzyme was added to 2-μm-deep compartments with only σ70 gene brushes (square symbols). Lines are a guide to the eye. Representative images are of compartments with RNAP core genes and the σ70 gene at the peak of each graph. Individual data points are presented in Supplementary Fig. 9a–c. Additional independent similar experiments are shown in Supplementary Fig. 9d. **c**, Integral of HA-GFP signal over entire compartment surface, at variable compartment radii (100–300 μm) and depth (2 μm and 20 μm). DNA brushes contained an identical mix of RNAP and P70-GFP genes. Images are of 2-μm-deep compartments with the indicated radii. Detailed gene compositions of all experiments are given in Supplementary Table 3. Data are presented as mean, error bars represent ± s.d. All images are in the same scale. Scale bar, 100 μm. The number of samples for each data point (**b,c**) is listed in Supplementary Table 4.

the genes coding for the core RNAP (α, β, β', ω subunits) and the promoter-specific σ70 subunit, together forming the holoenzyme, packed as a mixed brush in the centre of compartments with radius of 200 μm, and replaced the *E. coli* extract with a minimal gene expression system devoid of *E. coli* RNAP[33]. The synthetic operon thus coded for a cascaded reaction initiated by T7 transcription of *E. coli* RNAP genes, which once expressed and assembled led to the synthesis of GFP under the control of a σ70-specific promoter[14] (P70-GFP) (Fig. 5a). In contrast to the T4 wedge, RNAP assembly was programmed to assemble in solution since it needs to bind immobilized P70-GFP genes.

In the presence of all RNAP subunit genes, GFP expression originated from the central brush with an ~50 min delay compared with the expression of the *E. coli* RNAP subunits. The GFP profile was consistent with diffusion from the brush source and capture on the next available site (Supplementary Fig. 8a). Deletion of each of the RNAP genes, except for those coding for the ω subunit, a non-essential subunit for RNAP functionality[34], abolished the GFP signal, proving that nascent *E. coli* RNAP machines were assembled in a functional form (Supplementary Fig. 8b). We further showed that σ70 subunits expressed from a gene brush could complement the activity of purified core enzyme added to the minimal gene expression reaction. The GFP signal increased with σ70 genes, unlimited by the excess of core enzyme in solution but peaked when all subunits were expressed from brushes with a fixed amount of core genes (Fig. 5b and Supplementary Fig. 9a–d). Since both steps of the cascaded reaction consume the same resources, which are limited in the minimal gene expression system[35], protein synthesis in the initial step of the reaction may consume resources for the expression of GFP in the second step[14].

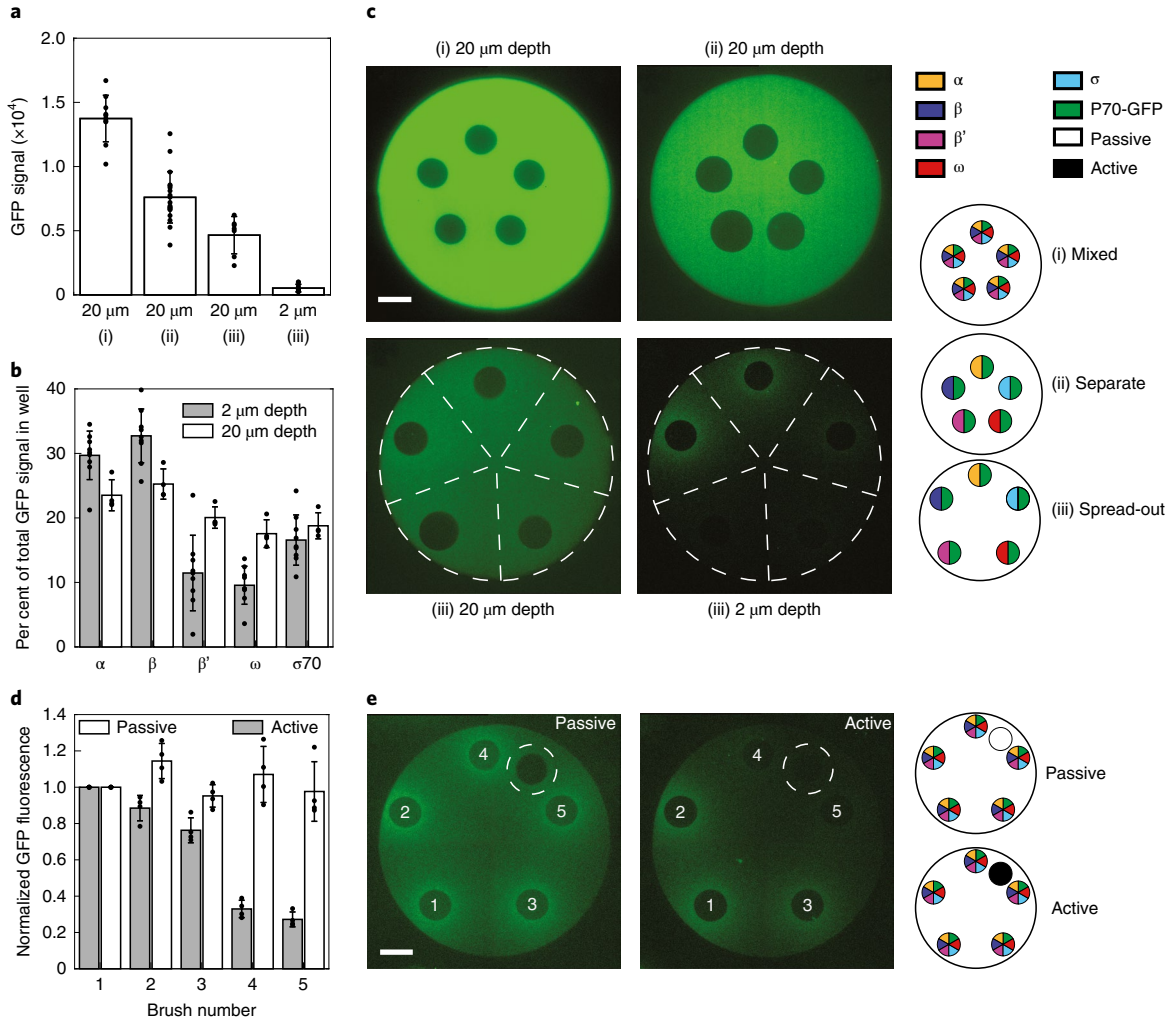


Fig. 6 | Local regulation of gene expression by resource partitioning. a–c, Total GFP signal (**a**) in compartments with brush layouts (mixed (i), separate (ii) and spread-out (iii)) and colour coding of genes as defined in the scheme in **c** (right). Localized GFP signal (**b**) in 2-μm-deep and 20-μm-deep compartments with brush layout (iii). Bar graph displays the percentage of the GFP signal in each slice out of the total GFP signal in the compartment. In **c**, the depth and layout for each representative image are marked above and below. The image of the 2-μm-deep compartment is presented with an enhanced contrast to demonstrate localization. **d**,**e**, GFP signal (**d**), integrated around each of the five identical brushes (1–5 in the respective representative images in **e**) and normalized to the signal around brush 1. Brush layout and gene content as in the schematic on the right. The position of an additional brush (white circle) containing passive or active DNA is marked (see Methods). Detailed gene compositions of all experiments are given in Supplementary Table 3. Data are presented as mean, and error bars represent ± s.d. Scale bars, 100 μm. The number of samples for each data point (**a**,**b**,**d**) is listed in Supplementary Table 4.

We attempted to increase the available resources by expanding the depth of the compartments from 2 to 20 μm (Fig. 5b). Indeed, a threefold higher GFP signal was observed, possibly due to localization of the reactions in the vicinity of the brush creating a sink for resources that are more abundant in the large-volume compartments. Consistently, the addition of increasing amounts of active genes to the DNA brush dramatically reduced GFP expression (Supplementary Fig. 8c). Thus, at maximal GFP production (Fig. 5b), the highest local concentrations of RNAP subunits were created to drive efficient assembly, leaving enough resources for GFP expression. Any further increase in protein synthesis would increase the consumption of resources, but to a lesser extent in the large-volume compartments.

We further varied compartment volume from ~60 pl to ~6 nl by systematically changing the diameter of the 2 μm and 20 μm compartments (Fig. 5c and Supplementary Fig. 9e–g). Indeed, the total amount of GFP produced in each compartment increased with the diameter, and hence compartment volume. The diameter series of both types of compartment merged into a continuous trend, which reached a constant GFP yield, that reduced only at high volumes beyond 3 nl. The reduction may stem from lower local concentrations of resources diffusing over large distances.

The large-diameter compartments provided an opportunity to immobilize several brushes in a circular pattern within a compartment. We immobilized five brushes, each coding for one of the RNAP subunits, with the P70-GFP genes added to each brush (Fig. 6a(ii),(iii)), and compared to GFP expression from the same number of mixed brushes (Fig. 6a(i)). GFP expression levels were higher in compartments with mixed RNAP gene brushes (Fig. 6b), consistent with the T4 wedge assembly in compartments (Figs. 2 and 3). The reduced GFP expression due to brush separation could be overcome by placing the brushes more closely at the compartment centre or by increasing the compartment volume (Fig. 6b). In the 2-μm compartments, GFP expression was localized to the α and β subunit

brushes (Fig. 6c), even when the P70-GFP gene brush was patterned separately (Supplementary Fig. 10), but was almost homogenous in the 20-µm-deep compartment, as in Fig. 5c.

Finally, we demonstrate how resource partitioning could be used to spatially regulate gene expression. We patterned five identical brushes composed of P70-GFP and RNAP subunit genes and immobilized an additional non-related yet active brush (Fig. 6d). GFP expression was suppressed by competing expression from this brush only in nearby brushes. Consistently, a passive DNA brush with no coding sequences had no effect on GFP expression. This mode of gene expression silencing regulated by spatial positioning cannot be realized in any off-chip cell-free reaction.

Conclusions

Our results signify a paradigm shift from bulk cell-free reactions to confined, quasi-2D, surface-localized reactions with a capacity for genome-scale synthesis, highly efficient and tunable assembly, capture of protein clusters, and resource partitioning, facilitating high yields of cascaded reactions. The genetic layout and the geometry of the compartment on a sub-millimetre scale tune nanoscale protein assembly yield. We find that coupled synthesis and assembly in the brush vicinity enhances weak interactions of sequential assembly steps, while surface-scaffolded assembly enhances non-sequential assembly steps by eliminating competing interactions. The genetic pattern could discriminately silence otherwise identical genetic regulatory units and the in situ labelling methodology could be used to decipher assembly order. Surface profiles may be used in the future to characterize assembly intermediates by high-resolution imaging approaches[19].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41565-020-0720-7.

Received: 1 September 2019; Accepted: 27 May 2020;
Published online: 20 July 2020

References

1. Gu, H., Chao, J., Xiao, S.-J. & Seeman, N. C. A proximity-based programmable DNA nanoscale assembly line. *Nature* **465**, 202–205 (2010).
2. Douglas, S. M. et al. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
3. Bath, J. & Turberfield, A. J. DNA nanomachines. *Nat. Nanotechnol.* **2**, 275–284 (2007).
4. Leung, K., Chakraborty, K., Saminathan, A. & Krishnan, Y. A DNA nanomachine chemically resolves lysosomes in live cells. *Nat. Nanotechnol.* **14**, 176–183 (2019).
5. Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012).
6. Schwarz-Schilling, M. et al. Optimized assembly of a multifunctional RNA-protein nanostructure in a cell-free gene expression system. *Nano Lett.* **18**, 2650–2657 (2018).
7. Freeman, R. et al. Reversible self-assembly of superstructured networks. *Science* **362**, 808–813 (2018).
8. Strackharn, M., Pippig, D. A., Meyer, P., Stahl, S. W. & Gaub, H. E. Nanoscale arrangement of proteins by single-molecule cut-and-paste. *J. Am. Chem. Soc.* **134**, 15193–15196 (2012).
9. Daube, S. S. & Bar-Ziv, R. H. Protein nanomachines assembly modes: cell-free expression and biochip perspectives. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **5**, 613–628 (2013).
10. Shieh, Y.-W. et al. Operon structure and cotranslational subunit association direct protein assembly in bacteria. *Science* **350**, 678–680 (2015).
11. Holt, C. E., Martin, K. C. & Schuman, E. M. Local translation in neurons: visualization and function. *Nat. Struct. Mol. Biol.* **26**, 557–566 (2019).
12. Minton, A. P. Implications of macromolecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* **10**, 34–39 (2000).
13. Daube, S. S., Arad, T. & Bar-Ziv, R. Cell-free co-synthesis of protein nanoassemblies: tubes, rings, and doughnuts. *Nano Lett.* **7**, 638–641 (2007).
14. Asahara, H. & Chong, S. In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. *Nucleic Acids Res.* **38**, e141 (2010).
15. Matthies, D. et al. Cell-free expression and assembly of ATP synthase. *J. Mol. Biol.* **413**, 593–603 (2011).
16. Rustad, M., Eastlund, A., Jardine, P. & Noireaux, V. Cell-free TXTL synthesis of infectious bacteriophage T4 in a single test tube reaction. *Synth. Biol.* **3**, ysy002 (2018).
17. Bracha, D., Karzbrun, E., Daube, S. S. & Bar-Ziv, R. H. Emergent properties of dense DNA phases toward artificial biosystems on a surface. *Acc. Chem. Res.* **47**, 1912–1921 (2014).
18. Daube, S., Bracha, D., Buxboim, A. & Bar-Ziv, R. H. Compartmentalization by directional gene expression. *Proc. Natl Acad. Sci. USA* **107**, 2836–2841 (2010).
19. Heyman, Y., Buxboim, A., Wolf, S. G., Daube, S. S. & Bar-Ziv, R. H. Cell-free protein synthesis and assembly on a biochip. *Nat. Nanotechnol.* **7**, 374–378 (2012).
20. Levy, M., Falkovich, R., Daube, S. S. & Bar-Ziv, R. H. Autonomous synthesis and assembly of a ribosomal subunit on a chip. *Sci. Adv.* **6**, eaaz6020 (2020).
21. Karzbrun, E., Tayar, A. M., Noireaux, V. & Bar-Ziv, R. H. Programmable on-chip DNA compartments as artificial cells. *Science* **345**, 829–832 (2014).
22. Tayar, A. M., Karzbrun, E., Noireaux, V. & Bar-Ziv, R. H. Synchrony and pattern formation of coupled genetic oscillators on a chip of artificial cells. *Proc. Natl. Acad. Sci. USA* **114**, 11609–11614 (2017).
23. Efrat, Y., Tayar, A. M., Daube, S. S., Levy, M. & Bar-Ziv, R. H. Electric-field manipulation of a compartmentalized cell-free gene expression reaction. *ACS Synth. Biol.* **7**, 1829–1833 (2018).
24. Shin, J., Jardine, P. & Noireaux, V. Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. *ACS Synth. Biol.* **1**, 408–413 (2012).
25. Garamella, J., Marshall, R., Rustad, M. & Noireaux, V. The all *E. coli* TX-TL toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth. Biol.* **5**, 344–355 (2016).
26. Yap, M. L. et al. Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proc. Natl Acad. Sci. USA* **113**, 2654–2659 (2016).
27. Yap, M. L. et al. Sequential assembly of the wedge of the baseplate of phage T4 in the presence and absence of gp11 as monitored by analytical ultracentrifugation. *Macromol. Biosci.* **10**, 808–813 (2010).
28. Bray, D. & Lay, S. Computer-based analysis of the binding steps in protein complex formation. *Proc. Natl Acad. Sci. USA* **94**, 13493–13498 (1997).
29. Roy, R. D., Rosenmund, C. & Stefan, M. I. Cooperative binding mitigates the high-dose hook effect. *BMC Syst. Biol.* **11**, 74 (2017).
30. Murugan, A. et al. Undesired usage and the robust self-assembly of heterogeneous structures. *Nat. Commun.* **6**, 6203 (2015).
31. Luke, K. et al. Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* **299**, 182–191 (2002).
32. Marshall, R. & Noireaux, V. Quantitative modeling of transcription and translation of an all-*E. coli* cell-free system. *Sci. Rep.* **9**, 11980 (2019).
33. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
34. Mathew, R. & Chatterji, D. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol.* **14**, 450–455 (2006).
35. Hillebrecht, J. R. & Chong, S. A comparative study of protein synthesis in in vitro systems: from the prokaryotic reconstituted to the eukaryotic extract-based. *BMC Biotechnol.* **8**, 58 (2008).

Methods

DNA preparations. *Cloning of genes.* Bacteriophage T4 wedge genes were amplified from the T4 GT7 genome (Nippon Gene, Japan) using appropriate primers (Integrated DNA technologies; Supplementary Table 1) and standard polymerase chain reaction (PCR) with KAPA HotStart ready mix (Kapa Biosystems). Primers were designed using SnapGene software (GSL Biotech LLC). The enhanced GFP (eGFP) F64L/S65T mutant sequence was used in all experiments containing eGFP. The genes were cloned into plasmids pIVEX 2.6 and pIVEX 2.5 (5′) under control of the T7 promoter for N or C terminus HA-tagging, respectively, using published restriction-free cloning protocols[36] (adapted from the reference and personal communications with Y. Peleg, Proteomics unit, Weizmann Institute of Science). For FL-proteins, a UAG amber stop codon was introduced into all T4 wedge genes as the second codon after the initiation AUG, using PCR amplification with mutated primers.

RNA polymerase genes were amplified from *E. coli* DH5α using the appropriate primers (IDT; Supplementary Table 1). Genes were inserted into pIVEX 2.5 under control of the T7 promoter using NEBuilder HiFi DNA Assembly (New England Biolabs). The eGFP gene was inserted by inverted PCR into pIVEX 2.6 in frame with N terminus HA-tagging, under the *E. coli* P_R(T-30A) promoter[37] replacing the T7 promoter.

Plasmids were transformed into *E. coli* DH5α and purified using Wizard SV-Gel (Promega) either at the miniprep or midiprep scale. Plasmid concentrations were determined using a NanoDrop ND-1000 (NanoDrop Technologies, Inc./ThermoFisher Scientific).

Preparation of linear DNA fragments for DNA brush formation. Linear double-stranded DNA fragments were amplified by PCR, as above, with 0.1 ng μl^{-1} plasmid template and 300 nM of 5′ modified forward and reverse primers (IDT), the forward primer conjugated to a fluorescent marker (either ATTO 488, ATTO 647 or Alexa Fluor 647) and the reverse primer conjugated to biotin, following a previously published protocol[38]. Non-expressible DNA fragments (termed passive DNA) were prepared by amplifying gene-7 without promotor sequences. DNA fragments coding for the dihydrofolate reductase gene (termed active DNA), serving as an expressible unrelated protein to wedges or RNAP, was amplified from a commercial plasmid (Cosmo Bio, Japan). Fragments of passive and active DNA were not conjugated to a fluorescent marker. Amplified fragments were purified by the Wizard SV-Gel and PCR Clean-Up System (Promega) and mixed with streptavidin (Sigma-Aldrich) at a 1.5:1 ratio in 1× phosphate-buffered saline (PBS), forming a DNA–streptavidin conjugate. DNA concentration was adjusted to 150–300 nM, and glycerol (J.T. Baker) was added to 5% final concentration to minimize evaporation during surface deposition. DNA–streptavidin conjugates were evaluated using 1% agarose gel electrophoresis.

Off-chip cell-free gene expression. All wedge proteins were expressed in an *E. coli* extract prepared according to published protocols[39]. Each reaction was supplemented with 5 μM His_6-GamS protein and 0.2 μM His_6-T7 RNA polymerase, which had been purified using published protocols[26,40]. Template DNA was added either as plasmids at a final concentration of 0.01–5 nM or as linear fragments at a final concentration of 0.5–5 nM. Reactions were incubated at 30 °C and were stopped by the addition of sodium dodecyl sulfate (SDS) loading buffer before resolution on denaturing protein gels.

Fluorescent labelling of wedge proteins. An unnatural fluorescent amino acid was incorporated as the second amino acid in each T4 wedge gene by supplementing each 30 μl CFE reaction with 5 nM of plasmid DNA and 0.5 μl CloverDirect tRNA reagent with one of the following fluorophores: HiLyte Fluor 488 AF, TAMRA-C6 AF, ATTO 633 or ATTO 655-X-AF (Cosmo Bio, Japan). Reactions were incubated at 30 °C for 2 h and were used directly for post-staining. For on-chip in situ fluorescent labelling (Fig. 4), 0.5 μl CloverDirect tRNA reagent was added per every 50 μl *E. coli* extract but no plasmid DNA was added.

Purification and pull-down of proteins. gp6-His, gp7, 8, 10, 11 and FL-gp53 were co-expressed in a 100 μl CFE reaction. After 2 h at 30 °C, the reaction was diluted 1:1 v/v with binding buffer (50 mM NaCl, 50 mM Tris and 20 mM imidazole, pH 8) and mixed with Ni^{2+} affinity beads (Ni-NTA His•Bind Resin, Milipore/Merck) pre-equilibrated in binding buffer. After 30 min incubation, beads were separated by centrifugation at 1,200 relative centrifugal force. Excess solution was removed by pipetting and the beads were washed with 30 volumes of binding buffer in repeated centrifugation steps. Wedges were eluted with a 1:1 bead volume of elution buffer (50 mM NaCl, 20 mM Tris and 500 mM imidazole). Washing efficiency and elution were evaluated using the Bradford reagent (Sigma-Aldrich) with 3:1 reagent-to-sample ratio.

Polyacrylamide gel electrophoresis and western blot. Tris-glycine (TG) linear gradient gels were purchased from GeBa at 4–20%. Gels were pre-equilibrated with SDS by a pre-run (160 V for 12 min) with TG-SDS buffer (Bio-Lab, Israel) for denaturation conditions. A western blot was preformed using Mini Trans-Blot (Bio-Rad) and polyvinylidene difluoride membranes (Moshe Stauber Biotec Application). Following transfer, the membrane was blocked with 3% bovine serum albumin (BSA) fraction V (Sigma-Aldrich) in 0.1% Tween 20 (Sigma-Aldrich) in PBS (PBS-T) for 1 h with gentle shaking. 50 ng ml^{-1} Anti-HA-Biotin, High Affinity (3F10) (Roche, Sigma-Aldrich) antibody in 3% PBS-T was applied to the membrane and incubated for 1 h while shaking. The membrane was washed three times in large volumes of PBS-T for 15 min each while shaking. The membrane was incubated with 5 nM ATTO 488 conjugated to streptavidin (Sigma-Aldrich) in 3% PBS-T for 15 min while shaking. Gels containing fluorescent proteins were imaged using a FLA5100\FLA9500 scanner (FUJI Typon/GE Typhoon). Images were analysed using Fiji software[41].

Chip preparations. *Multi-well array fabrication by deep silicon etching.* A two-step silicon etching was performed on a 5″ <100> Si wafer (University Wafer, USA). A pattern created in a dxf format using AutoCAD software (AutoDesk) was photolithographed using a μPG101 laser writer (Heidelberg Instruments) and etched with an inductively coupled plasma (ICP) machine (Multiplex ICP, SPTS Technologies). The 2-μm-deep chambers were formed using a single 25–30 s etching process (30 mtorr, 130 s.c.c.m. SF_6, bias voltage 500 W applied to the 13.56 MHz radio-frequency (RF) coil and 100 W to the platen) applied on a S1805 photoresist (MicroChem) patterned wafer, resulting in a 2–3 μm deep etch. Then, AZ4562 photoresist (MicroChemicals) was spin-coated on the wafer and the separation channels were etched using 50 cycles of an SF_6 etch alternating with C_4F_8 polymer deposition (Bosch process; etching: 12 s, 30 mtorr, 130 s.c.c.m. SF_6, 13 s.c.c.m. O_2, bias voltage 500 W applied to the RF coil and 100 W to the platen; passivation: 10 s, 30 mtorr, 30 s.c.c.m. C_4F_8, bias voltage 500 W applied to RF coil), resulting in a 40–50 μm deep etch. All heights were measured using a stylus profiler (DektakXT, Dektak/Bruker, MA, USA). The 20-μm-deep chambers were formed using a similar two-step etching process, using AZ4562 photoresist for both etching steps. Silicon etching was achieved by the Bosch process in an ICP machine (LPX ICP, SPTS Technologies). Forty cycles were used to achieve compartments with a depth of 20 μm and 100 cycles were used to achieve separation channels with a depth of 50 μm.

SiO_2 deposition. The etched wafer was divided by hand into separate chips. Each chip was coated with a 50-nm SiO_2 layer by plasma-enhanced chemical vapour deposition (PECVD) using a VERSALINE PECVD machine (Plasma-Therm, Saint Petersburg, FL, USA) at the following conditions: 21 s, pressure of 1,200 mtorr, 5% SiH_4 in He, flow rate of 750 s.c.c.m., N_2O flow rate of 1,250 s.c.c.m., N_2 flow rate of 400 s.c.c.m., RF power of 110 W, upper electrode temperature of 200 °C and lower electrode temperature of 300 °C.

Photosensitive biocompatible monolayer coating. The protocol to form a photosensitive and biocompatible monolayer coating on silicon chips has been described elsewhere[42]. Briefly, the chips were coated with a polymer composed of a polyethylene glycol backbone with a protected amine at one end, and a triethoxysilyl group at the other end. The slides were incubated with a toluene solution of the polymer (0.2 mg ml^{-1}) for 20 min, rinsed with toluene and dried.

Photosensitive biocompatible monolayer ultraviolet photolithography. De-protection of surface amines was preformed using the μPG101 laser writer (Heidelberg Instruments) and a pattern created in a dxf format using AutoCAD software (AutoDesk). The coated surfaces were exposed with either write-head 4 mm (35 mW 50%) or with write-head 20 mm (70 mW 100%). Exposed amines were immediately coupled to biotin by incubating the surfaces with 0.5 mg ml^{-1} biotin 3-sulfo-*N*-hydroxysuccinimide ester (EZ-link, Pierce) in 0.2 M borate-buffered solution pH 8.6 (ThermoFisher Scientific) for 15 min, followed by rinsing and drying.

DNA deposition. DNA–streptavidin conjugates at 150 nM concentration, unless stated otherwise in Supplementary Table 3, were deposited on biotin-patterned surfaces using a GIX II microplotter (Sonoplot). A 60-μm-diameter tip apparatus was used for all DNA deposition experiments. The minimum spacing between microdroplets of DNA–streptavidin conjugates was 100 μm. Patterns for GIX II were made by the SonoGuide software (Sonoplot). Microdroplets were incubated for at least 2 h. For 1D layouts (Figs. 2–4 and Supplementary Figs. 5–7), DNA brushes, two for each gene (termed doublets), were deposited along the width of the compartment to create a uniform source of expression along the compartment length.

Immobilization of antibodies and tagged proteins. Biotinylated high-affinity (3F10) anti-HA antibodies (50 μg ml^{-1}, ~500 nM, Roche, Sigma-Aldrich) were mixed at a 2:1 ratio of streptavidin (Sigma-Aldrich) in 1× PBS. After 30 min incubation at 4 °C the mix was diluted to 25–50 nM in 1× PBS supplemented with 0.2 mg ml^{-1} BSA, applied to the surface and incubated for at least 1 h. After washing with 1× PBS, the surface was either directly covered with cell-free extract or first covered with a tagged protein (gp11-HA), freshly synthesized in a CFE reaction. The crude reaction (30–90 μl, depending on chip size) was applied to the biochip without drying the antibodies, and incubated for 1 h followed by washing in 1× PBS. For reactions with pre-adsorbed gp10, the process was identically repeated.

On-chip cell-free protein expression. The chip was rinsed with 1× PBS and excess solution was carefully blotted using paper (WhatmanTM #1) while keeping the chambers wet. Fresh CFE solution (30–90 µl, depending on chip size, prepared as for the off-chip reactions) for T4 assembly experiments or PURE solution (PURExpress 2.0, Cosmo Bio, Japan) for RNAP assembly experiments was applied on the chip and spread uniformly to cover all chambers. After excess solution was removed, the chip was covered either with 1–2-mm-thick polydimethylsiloxane (PDMS) slabs (SYLGARD 184 silicone elastomer kit, Dow Corning), or with PDMS-covered coverslips (24 mm × 50 mm × 0.170 mm, Marienfeld, Germany) for experiments with compartment diameters larger than 400 µm. CFE reactions were carried out at 30 °C for 2 h (*E. coli* extract), and 37 °C for 1.5–3 h (PURE; exact times for each experiment are indicated in Supplementary Table 3) in a PCR machine (Mastercycler gradient (Eppendorf, Hamburg, Germany) or Labcycler 48, SensQuest) fitted with a slide hybridization adapter. After expression, the PDMS slab was removed and the chip was washed with 1× PBS. For visualization of on-chip dynamic expression (Supplementary Fig. 8a), the chip was placed in an incubator chamber (Bold-line stage top incubator, Okolab) installed on the microscope stage.

Post-staining. The crude protein-labelling reaction was applied on the washed chip directly (for FL-gp10, FL-gp8 and GFP), or after 1:4 dilution in PBS (for FL-gp53), or after centrifugation for 5 min at 16 K relative centrifugal force at 4 °C to remove labelled aggregates (for FL-gp6). The chip was incubated for 1 h and washed with 1× PBS. For multiple post-staining steps, the labelled proteins were introduced one by one, with 1× PBS washing between steps, in the reverse assembly order to ensure that a labelled protein would not bind to labelled complexes. FL-gp53 reported on wedge, FL-gp6 on pre-wedge and FL-gp8 on the gp10–7 complex. To deduce the profile of the unoccupied sites of surface gp11-HA or anti-HA antibodies, the surface was post-stained with either FL-gp10 or HA-GFP, respectively, and the complementary signal was calculated by subtracting the signal from control compartments with no trap.

Fluorescent microscopy imaging. Fluorescent images were obtained using an AxioObserver Z1 inverted microscope with a motorized stage (Zeiss) and Plan-Apochromat 20×/0.8 M27, EC Plan-Neofluar 40×/0.9 Pol M27 (Zeiss) and 10×/0.3 MPlanFL N (Olympus) objectives. Illumination was performed using a Colibri2 LED illumination system equipped with 470-nm and 625-nm LED module (Zeiss) and filter sets 38 HE (Zeiss; excitation 470/40 nm, dichroic mirror 495 nm, emission 525/50) or filter set 50 (Zeiss; excitation 640/30 nm, dichroic mirror 660 nm, emission 690/50). Images were captured using an iXon Ultra CCD camera (Andor Technology, Belfast, UK). Chip alignment and multi-image acquisition was performed using the Zeiss ZEN 2012 software.

Data analysis. Circular compartments with diameter of 500 µm and larger (Figs. 5c and 6, and Supplementary Figs. 9e–g and 10) and 1D compartments (Figs. 2–4 and Supplementary Figs. 5–7) required the acquisition of multiple images to cover the entire compartment (4 images for circular compartments and 5–6 for 1D compartments, with 5–10% overlap). First, inhomogeneous lighting was corrected by normalizing to reference images, and then images were stitched together. Background signal from control compartments was subtracted from all images of the same set of experiments. For circular 200–700-µm-diameter compartments (Figs. 1, 5 and 6, Extended Data Fig. 1, and Supplementary Figs. 3 and 8–10), the total fluorescent signal was measured in the entire patterned area of the compartment excluding the brush area. For 1D compartments (Figs. 2–4 and Supplementary Fig. 7), integrated total fluorescence was taken in the entire compartment.

The 1D profile along the *x* axis (Figs. 2 and 3 and Supplementary Figs. 5–7) was generated from an average along an 80-µm strip in the centre of the *y* axis. The profile was smoothed using a moving average of a 15–20 µm window. Radial profiles (Supplementary Fig. 8a) were calculated using Fiji software[41].

When comparing fluorescent signals in different sets of experiments, each signal of every set was normalized to its maximal value. To compensate for differences in expression levels between compartments, the amount of occupied sites was used as a normalization factor, except for the data points of gene-10 = 0% (Fig. 1e and Supplementary Fig. 3a) to avoid division by 0, and these points were therefore omitted.

The yield of a given assembly step, facilitated by gene *i*, was calculated by:

$$\mathrm{yield}(i) = (N_{i=0} - N)/N_{i=0}$$

with $N_{i=0}$ is the mean signal in compartments without gene-*i*, stained with FL-gp-*i*, and N is the total mean signal in a compartment with all genes up to gene-*i* in the assembly line. The error in the yield is calculated by:

$$\Delta\mathrm{yield}(i) = \sqrt{\left(\frac{N}{N_{i=0}^2} \times \Delta N_{i=0}\right)^2 + \left(\frac{1}{N_{i=0}} \times \Delta N\right)^2}$$

where ΔN and $\Delta N_{i=0}$ are the standard deviation of the mean.

The in situ fluorescent labelling (Fig. 4) resulted in high fluorescent background localized to the area of the brush coding for the FL-gene, even if the FL-protein was not assembled on the surface (Fig. 4c). This prevented using in situ fluorescent labelling in a mixed brush. The background signal was taken from compartments containing the FL gene but lacking a scaffold gene. To account for differences in the profile area of various HA-proteins (Fig. 4a,b), the integrated signal of the in situ FL-protein was measured only in the area occupied by the HA-tagged protein. Different FL -proteins were compared by normalizing to the maximum signal received for each one.

Images presented in the figures were contrast-adjusted and coloured using Fiji software[41] only for presentation purposes. In Fig. 6a, configuration (i) was measured after 120 min incubation, configuration (ii) was measured in two independent sets, after 90 min and after 120 min, and configuration (iii), 2 µm and 20 µm, was measured after 90 min. Configuration (ii), which was measured both at 90 and 120 min was used for normalization.

Computer simulations. We computationally modelled the synthesis, diffusion, binding and assembly of two types of proteins, protein A (gp10) and protein B (gp7) in a confined 2D geometry considering all interactions shown in Fig. 2b. Modelling was simplified by neglecting steric exclusion of the proteins in the solution, which allows a reduction of the model by one spatial dimension via projection of the short axis ('*y*') onto the long axis ('*x*'). The 1D model was spatially and temporally discretized and solved numerically using the Euler method, using a bin size of $\Delta x = 10$ µm and a time step of $\Delta t = 0.18$ s. The underlying first-order differential equations describing the time evolution of the position-dependent concentrations of all seven distinct molecular species involved in the process are given in Supplementary Fig. 11, and a list of all values used to parameterize the model are given in Supplementary Table 2.

The assembly yield (Fig. 2f) was calculated at each time point as the fraction of complexes captured on the surface out of the total possible complexes that could have formed. The expression rate of gp7 was identical in all four scenarios and was the limiting factor for the yield calculation.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data necessary to interpret, replicate and build upon this work appears in the article and its Supplementary information. Additional data can be made available from the corresponding authors upon reasonable request.

Code availability

The information needed for the computer simulation in Fig. 2 appears in the article and its Supplementary information.

References

36. Erijman, A., Dantes, A., Bernheim, R., Shifman, J. M. & Peleg, Y. Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. *J. Struct. Biol.* **175**, 171–177 (2011).
37. Kincade, J. M. & deHaseth, P. L. Bacteriophage lambda promoters pL and pR sequence determinants of in vivo activity and of sensitivity to the DNA gyrase inhibitor, coumermycin. *Gene* **97**, 7–12 (1991).
38. Buxboim, A., Daube, S. S. & Bar-Ziv, R. Ultradense synthetic gene brushes on a chip. *Nano Lett.* **9**, 909–913 (2009).
39. Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription–translation system. *Biochimie* **99**, 162–168 (2014).
40. He, B. et al. Rapid mutagenesis and purification of phage RNA polymerases. *Protein Express. Purif.* **9**, 142–151 (1997).
41. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
42. Buxboim, A. et al. A single-step photolithographic interface for cell-free gene expression and active biochips. *Small* **3**, 500–510 (2007).

Acknowledgements

We acknowledge funding from the Israel Science Foundation (grant no. 1870/15), the United States–Israel Binational Science Foundation (grant no. 2014400), and the Minerva Foundation (grant no. 712274) for the work on the T4 wedges. We thank the Office of Naval Research (award no. N62909-18-1-2094) for funding the work on RNAP assembly. We thank M. Levy for discussions.

Author contributions

O.V., S.S.D., R.H.B.-Z. and V.N. conceived the cell-free assembly of phages. O.V., Y.D., S.S.D. and R.H.B.-Z. designed the experiments and analysed the data. O.V. and Y.D.

performed the experiments. D.G. and V.N. provided the cell-free *E. coli* extract. O.V., S.F., S.R. and R.L. did the modelling and computational work. O.V., Y.D., S.F., S.R., S.S.D. and R.H.B.-Z. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

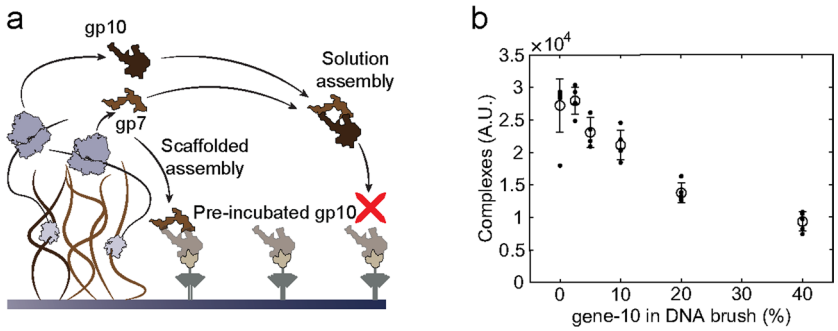
The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41565-020-0720-7.

Supplementary information is available for this paper at https://doi.org/10.1038/s41565-020-0720-7.

Correspondence and requests for materials should be addressed to S.S.D. or R.H.B.-Z.



Extended Data Fig. 1 | Competition between on-chip solution and scaffolded assembly. a, Scheme, co-expression and interaction of gp10 and gp7 in solution prior to surface binding sequester the scaffolding of gp7 onto gp10 pre-bound to the surface. **b**, Dose response of gene-10 fraction in a mixed DNA brush with fixed amount of gene-7. Surface gp11 traps were pre-bound with gp10. Surface bound gp10-7 complexes revealed by post-staining with FL-gp8. Detailed gene composition of all experiments appears in Supplementary Table 3. Individual data points and mean values are shown, error bars represent ± s.d. The number of samples for each data point (**b**) is listed in Supplementary Table 4.

nature research

Corresponding author(s): Roy H. Bar-Ziv
Shirley S. Daube

Last updated by author(s): May 28, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Image acquisition was preformed using the Zeiss ZEN 2012 software
Data analysis	All data were analyzed using standard Matlab code

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data necessary to interpret, replicate and build upon this work appears in the article and its supporting information. Raw data of Figs. 1, 2, 3 and 5 appear in Supplementary Figs. 3, 5, 6, 7 and 9. Additional data can be made available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Exact sample size for each set of experiments is listed in 'Statistics and reproducibility' section. Each chip contained 64-144 simultaneous experiments. Muliple chips were used when a larger sample size was needed. An internal positive control was used in every chip to normalize between chips.
Data exclusions	No compartments were excluded except for physically damaged ones.
Replication	For every data point 3-12 replica compartments were used.
Randomization	The position of different replicas on the chip was randomized.
Blinding	All reactions in the compartments were preformed simultanously. All the analysis was preformed automatically on the entire data set.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	We used commercially available antibodies (anti-HA-biotin (3F10), sigma aldrich)
Validation	These antibodies were only used to anchor proteins to the surface.