

# Integrating gene synthesis and microfluidic protein analysis for rapid protein engineering

Matthew C. Blackburn, Ekaterina Petrova, Bruno E. Correia and Sebastian J. Maerkl\*

Institute of Bioengineering, School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Received August 28, 2015; Revised October 15, 2015; Accepted December 09, 2015

## ABSTRACT

**The capability to rapidly design proteins with novel functions will have a significant impact on medicine, biotechnology and synthetic biology. Synthetic genes are becoming a commodity, but integrated approaches have yet to be developed that take full advantage of gene synthesis. We developed a solid-phase gene synthesis method based on asymmetric primer extension (APE) and coupled this process directly to high-throughput, on-chip protein expression, purification and characterization (via mechanically induced trapping of molecular interactions, MITOMI). By completely circumventing molecular cloning and cell-based steps, APE-MITOMI reduces the time between protein design and quantitative characterization to 3–4 days. With APE-MITOMI we synthesized and characterized over 400 zinc-finger (ZF) transcription factors (TF), showing that although ZF TFs can be readily engineered to recognize a particular DNA sequence, engineering the precise binding energy landscape remains challenging. We also found that it is possible to engineer ZF–DNA affinity precisely and independently of sequence specificity and that *in silico* modeling can explain some of the observed affinity differences. APE-MITOMI is a generic approach that should facilitate fundamental studies in protein biophysics, and protein design/engineering.**

## INTRODUCTION

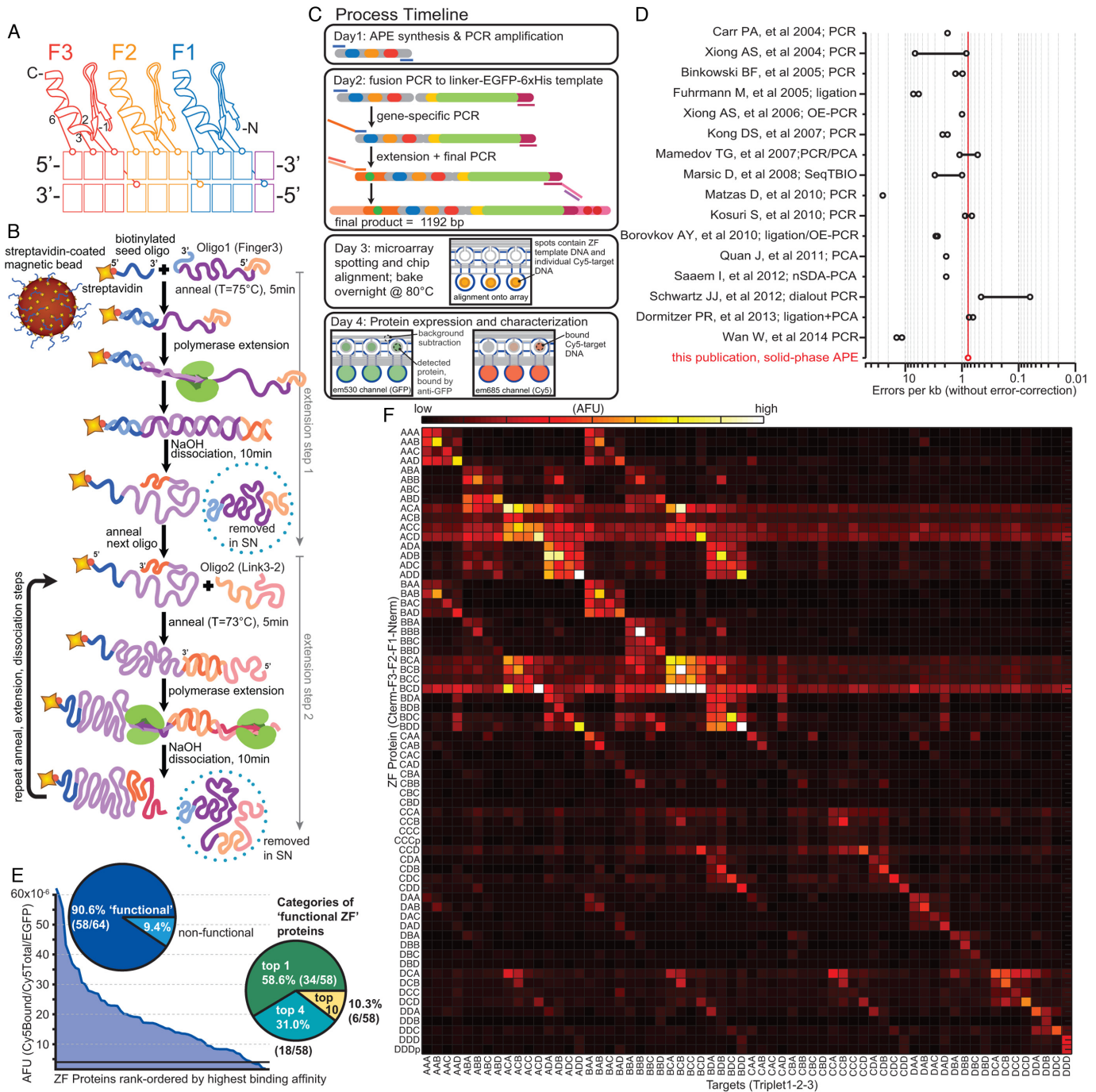
Engineering proteins with novel functions remains a challenging task. Experimental methods that rely on generating large numbers of random protein variants followed by screening are currently the most successful approaches to protein engineering (1). Rational design of protein function, on the other hand, remains a major unsolved problem in biochemistry. Nonetheless, computational approaches are becoming adept at informing protein design and predicting

function (2,3), and techniques that permit the rapid generation and quantitative characterization of designed protein variants would greatly aid protein engineering and further improve the accuracy of computational approaches. Numerous gene synthesis methods have been developed (4,5) and synthetic genes are becoming a commodity, but little to no attention has been given to streamlining the downstream processing steps required to quantitatively characterize the large number of proteins that can potentially be generated using gene synthesis based approaches.

Here we present a pipeline for the rapid synthesis and characterization of rationally designed synthetic proteins. We developed a bench-top solid-phase gene synthesis method based on APE, and demonstrate that expression ready linear templates generated by APE can be used directly for on-chip high-throughput protein expression, purification and quantitative characterization by MITOMI (6,7). APE synthesizes genes with fidelity comparable to the best currently available methods and can synthesize genes with internal sequence redundancies. APE-MITOMI completely circumvents any requirement for molecular cloning and cell-based protein expression, and can synthesize and characterize hundreds of novel protein variants per week. As a proof-of-concept we applied APE-MITOMI to the engineering and characterization of Cys<sub>2</sub>His<sub>2</sub> ZF TFs. We found that ZF TF affinity can be precisely tuned independently of specificity and although it is possible to engineer specificity, the precise binding energy landscape is more difficult to rationally engineer.

Individual ZF domains provide a convenient structure for refactoring due to their relatively small size and composability (8,9) (Figure 1A). ZFs fused to nucleases (ZFN) are one of the main tools currently used for clinical genome editing (10,11). The versatility of ZFs can be further expanded by fusing them to other effector domains allowing them to perform a variety of site-specific genetic modifications beyond DNA cleavage (12,13). Synthetic ZFs have also been used to construct artificial transcriptional regulatory circuits in yeast (14,15). ZF TFs are thus ideal targets for exploring the biophysics of transcription factor DNA specificity, and the ability to engineer ZF TFs makes them useful tools in biotechnology and synthetic biology. But

\*To whom correspondence should be addressed. Tel: +41 21 69 37835; Email: sebastian.maerkl@epfl.ch



**Figure 1.** APE-MITOMI applied to ZF TF module combinatorics. **(A)** Cartoon model of canonical Cys<sub>2</sub>His<sub>2</sub> ZF TF binding to DNA with residues -1, 2, 3 and 6 of the recognition helix primarily encoding DNA specificity. Residue 2 makes a cross-strand contact, which creates 'context dependent' effects. **(B)** Schematic of the APE solid-phase gene assembly technique, showing assembly through the first two extension steps. **(C)** Process timeline from gene assembly to protein characterization. **(D)** Comparison of APE error rate with values from previously published gene assembly techniques. A line between two points indicates a range of error rates from different experimental conditions. **(E)** Overview of experimental results obtained from combinatoric assembly of ZF TFs demonstrating protein expression and functional DNA binding success rates. **(F)** Heatmap of relative binding affinities for each assembled ZF TF (y-axis) to 64 predicted consensus DNA targets (x-axis). Protein naming convention indicates ZF domain from C-to-N (F3 to F1), where AAA (A<sub>F3</sub>A<sub>F2</sub>A<sub>F1</sub>) = Zif268, BBB = 37-12, CCC = 92-1, DDD = 158-2 (14); for example, protein ABC = F3 from Zif268, F2 from 37-12, F1 from 92-1; target ABC = Zif268 F3 binding consensus triplet GCG, 37-12 F2 binding consensus GAC, 92-1 F1 binding consensus triplet GCC (5'-GCG GAC GCC). The values represent averages of multiple measurements and the precise number of technical repeats and a histogram thereof are shown in Supplementary Figures S7 and S8, respectively. Oligomer assembly and target sequences are given in Supplementary Tables S7 and S8.

even with the most recent data sets and models (16–24), engineering novel ZF TFs with precise sequence specificities remains a challenging problem. The Zinc Finger Consortium's online database of ZFs (25) is in principle a useful resource of recognition helices (RHs) that supposedly bind a particular DNA triplet, but characterization assays vary and are relatively incomparable in terms of measuring DNA target specificity and affinity, leading to incongruities in ZF studies (26,27).

## MATERIALS AND METHODS

### Synthetic DNA design for gene assembly

All ZF modules used in this study were either taken from the Zinc Finger Consortium's online Zinc Finger Database (25), or from references (14) and (22), which contain natural and designed single ZF modules or concatenated, multi-finger arrays. Other ZF modules were designed using online DNA binding predictors (20–23). For our purposes, we use only the 7 amino acid residues in positions –1 to 6 of the  $\alpha$ -helix. Other parts of the protein that are indirectly involved in DNA sequence recognition, called the framework, are taken directly from the three-finger murine transcription factor Zif268 (RCSB PDB 1AAY; Supplementary Figure S1). The Zif268 coding sequence (90 aa) was converted to an *Escherichia coli* (strain K12) codon optimized nucleotide sequence using JCat (Java Codon Adaptation Tool (28), [www.jcat.de](http://www.jcat.de)). The resulting 270 nt sequence was then partitioned into five oligomers: three 'finger' oligomers containing the sequence coding for  $\alpha$ -helix residues with flanking sequence, and two 'linking' oligomers containing sequences bridging the three 'finger' oligomers (see Supplementary Figure S2 and Table S1 for sequences). Each oligomer has a 25 or 28 nt overlap with the oligomer preceding it for annealing. This way, all oligomers with the appropriate flanking sequence can be interchanged with each other since they contain the necessary complementary sequence for annealing. This allows single oligomers to be used in multiple zinc finger assemblies, but limits them to the same position in the assembly process. Linking oligomers 'Link3–2' and 'Link2–1' are used for all assemblies, whereas libraries of O1F3 (Oligo1 Finger3), O3F2 (Oligo3 Finger2) and O5F1 (Oligo5 Finger1) oligos are used to generate different three-finger assemblies (21 nt located in the colored regions of Supplementary Figure S2). Oligomers were ordered from IDT with standard desalting only and were rehydrated to 500  $\mu$ M in 1x Tris-EDTA (TE) buffer for stock solutions, and diluted to 50  $\mu$ M with PCR grade water for working solutions.

### Asymmetric primer extension (APE) assembly

Prior to gene-assembly, an aliquot (800–1200  $\mu$ l) of MyOne™ Streptavidin T1 beads (Life Technologies) are placed in a 1.5 ml Eppendorf tube, pelleted using a magnetic stand, and resuspended in an equivalent volume of 0.2 M NaOH in water. The beads are preconditioned for at least 1 h at room temperature before use, then stored at 4°C for longer conditioning times. These conditioned beads can be used for up to one month after being suspended in the 0.2 M NaOH solution. It was found that preconditioning of beads

in NaOH releases labile streptavidin monomers (29) and in our hands this translated to a reduction in non-specific PCR bands during intermediate quality control PCR steps and during amplification of the final assembly. Each individual assembly reaction requires 25  $\mu$ l of preconditioned beads. Lower bead quantities may work as well, but to account for losses during washing, buffer exchanges and transfer steps, we have continued using 25  $\mu$ l with consistent success. Larger reactions are also possible by scaling all volumes accordingly. This is particularly useful during the creation of ZF combinatoric array variants. By starting with a large pool of beads, beginning the assembly together with the same Oligo1Fin3 and Link3–2, followed by partitioning the pool into smaller volumes and continuing assembly with different Oligo3Fin2 parts in separated reactions, followed by a final partitioning for Oligo5Fin1 parts, where the final volume in each Oligo5Fin1 assembly is 25  $\mu$ l, many different genes can be assembled within the same workflow.

For a single APE reaction, 25  $\mu$ l of preconditioned beads are pelleted using a magnetic stand (Invitrogen DynaMag™-Spin) for 30–60 s until the solution is clear. The supernatant (0.2M NaOH) is carefully aspirated. The beads are then washed twice with 25  $\mu$ l of 1x binding and washing buffer (B&W; 2x contains 10mM Tris-HCl pH 7.5, 1mM EDTA, 2M NaCl) containing 0.01% (v/v) Tween20 (BW+Tween; to reduce non-specific binding (30)). Each washing step involves adding the wash solution, mixing the solution by aspiration until the beads are resuspended, then pelleting the beads and removing the supernatant. Then the beads are pelleted again and resuspended in 25  $\mu$ l of 2x B&W Buffer (without Tween20), to which 25  $\mu$ l of 'seed' oligomer solution (0.12  $\mu$ M biotinylated seed oligomer in PCR grade water; Supplementary Table S1) is added and mixed. This mixture is incubated at room temperature for at least 15 min on a lab rotisserie. Following incubation, the beads are pelleted against the magnetic stand, washed twice with 25  $\mu$ l of 1x HF Buffer without detergent (Phusion® HF Buffer Detergent-free (5x), New England Biolabs) to prevent bubble formation during resuspension, and finally resuspended in 25  $\mu$ l Oligo1Fin3 extension mix (final concentrations: 1x HF Buffer with detergent, 0.2 mM dNTPs, 5% DMSO, 8  $\mu$ M Oligo1Fin3, 0.3 units Phusion High-Fidelity Polymerase (NEB)). This mixture is then placed on a thermocycler and run through a brief annealing and extension routine (5.5 min at  $T_{\text{anneal}}$ , 2 min at 72°C, then hold at 25°C;  $T_{\text{anneal}}$  for each oligomer is given in Supplementary Figure S2). The tube is removed from the thermocycler, the beads are pelleted, and the supernatant is removed and discarded. The beads are then washed twice with 50  $\mu$ l 1x SSC buffer (saline sodium citrate, Sigma), resuspended in 50  $\mu$ l 0.15 M NaOH and incubated at room temperature on a rotisserie for 10 min to facilitate strand dissociation. The beads are then pelleted, the supernatant is removed, and the beads are washed once with 50  $\mu$ l 0.15M NaOH, once with 50  $\mu$ l 1x BW+Tween, and once with 50  $\mu$ l 1x HF buffer without detergent. The beads are then resuspended in Oligo2 (Link3–2) extension mix (same recipe as for Oligo1, except Link3–2 is used), then placed back on the thermocycler, and run through the annealing and extension routine, where  $T_{\text{anneal}}$  has been adjusted to the temperature required for this annealing reaction. This procedure

of extension, strand dissociation via 0.15 M NaOH, and buffer exchanges is repeated for each oligomer in the assembly. After the final extension reaction (Oligo5Fin1), there is no NaOH dissociation step. Instead, the beads are pelleted, washed twice with 50  $\mu$ l 1x SSC buffer and resuspended in a final volume of 20  $\mu$ l 10 mM Tris-Cl pH 8.5 buffer.

The beads in Tris-Cl buffer from the final extension step are used directly as template for a PCR amplification of the complete five-step assembly product. PCR primers were designed to amplify the 239 bp product (Supplementary Table S1). For a 20  $\mu$ l PCR reaction, final concentrations are as follows: 1x HF Buffer with detergent (New England Biolabs), 0.2mM dNTPs, 5% DMSO, 0.5  $\mu$ M each primer (assembly check-f and -r), 0.6  $\mu$ l suspension of beads in Tris-Cl (template), 0.3 units Phusion High-Fidelity Polymerase; touchdown PCR: 98°C, 30 s; 74>72°C, 30 s then 17 cycles at 71°C, 30 s; 72°C, 30 s). 4  $\mu$ l of this PCR is then run on a 2% agarose gel with 0.4x GelGreen (Biotium) at 110V for 1 h. Due to non-specific primer interactions with the template and interactions with truncated assembly products, some PCR amplifications can result in the formation of multiple truncated bands, the highest of which is the complete assembly product. To overcome this problem, we have taken advantage of a modified band-stab technique (31) to isolate the band of interest and re-amplifying it to reduce the amount of nonspecific products in downstream steps. Briefly, the 2% gel is imaged using a blue-light transilluminator and the band of interest is captured using a 200  $\mu$ l pipette tip, with the end cut off about 1 cm from the tip, by stabbing into the gel at the location of the band. The pipette tip is then placed into a 1.5 ml Eppendorf tube, and the agarose gel core inside the pipette tip is pushed out using a second sterile pipette tip. 20  $\mu$ l of Qiagen EB buffer (10 mM Tris-Cl, pH 8.5) is added to the agarose gel sample, briefly vortexed, centrifuged and incubated at 80°C for at least 10 min with the tube cap closed. The sample is then vortexed and centrifuged again, before a 0.25  $\mu$ l sample of the buffer is taken as template for a second PCR amplification. This PCR is prepared and thermocycled following the same recipe from the first PCR (assembly check PCR; Supplementary Figure S3).

### Expression-ready linear template preparation

Following the second assembly check PCR, the core region coding for the three linked  $\alpha$ -helices is complete, but the final template will consist of a C-terminal proline-linker and EGFP fusion, a 6x histidine tag, and 5' and 3' UTRs for expression within a cell-free, transcription/translation mixture. All of these parts are added to the ZF assembly via four different PCRs: a fusion PCR (for adding the proline-linker, EGFP domain and 6x-histidine tag; results in a 1019 bp product), a gene-specific PCR (for adding part of the 5' and 3' UTRs; results in a 1084 bp product) and an extension+final two-step PCR (which completes the template construction and amplifies the full-length product of 1192 bp; Supplementary Table S2). The fusion PCR requires two templates: the 239 bp ZF construct and a previously amplified EGFP domain from the pKT127 plasmid, including a 5'-proline linker and 3'-6x histidine tag. Briefly, in a 20  $\mu$ l PCR containing 1x HF Buffer with de-

tergent, 0.2 mM dNTPs, 5% DMSO, 0.5  $\mu$ M each primer (Proline-linker-EGFP-f and EGFP-His6-r; Supplementary Table S2), 1 ng of pKT127, and 0.3 units Phusion High-Fidelity Polymerase are thermocycled for 25 cycles (98°C, 30 s; 61.7°C, 30 s; 72°C, 1 min). This product (Prolink-EGFP-6His, 805nt) is used without purification in the fusion PCR. The fusion PCR is carried out in two steps, the first reaction contains all the necessary ingredients for PCR except the nucleotide mix and polymerase. A 15  $\mu$ l reaction is prepared containing 1x HF Buffer with detergent, 5% DMSO, 0.5  $\mu$ M of each primer (assembly check-f and EGFP-His6-r), and 0.25  $\mu$ l each of the assembly PCR from band-stab and Prolink-EGFP-6His PCR. This mixture is placed on a thermocycler and heated to 98°C for 4 min, then cooled down (10% ramp) to 25°C for annealing. Then 5  $\mu$ l of an extension mixture (1x HF Buffer with detergent, 0.2 mM dNTPs, 0.3 units Phusion) is spiked into the annealing mixture (20  $\mu$ l final volume) and cycled 20 times (98°C, 30 s; 72°C, 30 s; 72°C, 1 min). The gene-specific PCR uses the product generated in the fusion PCR as template. A 20  $\mu$ l reaction is prepared: 1x HF Buffer with detergent, 0.2 mM dNTPs, 5% DMSO, 0.5  $\mu$ M each primer (genespecific-f and EGFP-His6-r), 0.25  $\mu$ l of the fusion PCR and 0.3 units Phusion polymerase. This reaction is cycled using a short touchdown PCR (98°C, 30 s; 75>72°C, 30 s; 72°C, 1 min), followed by 16 cycles (98°C, 30 s; 72°C, 30 s; 72°C, 1 min). The two-step extension+final PCR uses the product generated in the gene-specific PCR as template. In this reaction, it is very important to use the HF Buffer without detergent, since this product will be used directly for microarray spotting and the presence of detergent will result in large spots. A 20  $\mu$ l reaction is prepared: 1x HF Buffer without detergent, 0.2 mM dNTPs, 5% DMSO, 2.5 nM each primer (extension-f and -r), 0.25  $\mu$ l of 1:10 diluted gene-specific PCR (in Tris-Cl or water) and 0.3 units Phusion polymerase. This mixture is thermocycled 10 times (98°C, 30 s; 61°C, 30 s; 72°C, 1 min). Then the reaction is kept at 72°C for 2 min, and cooled to 25°C. At this point, 0.1  $\mu$ l of each final-highTm primer (50  $\mu$ M stock; final 0.25  $\mu$ M in 20.2  $\mu$ l) are spiked into the mixture, and it is thermocycled again via a short touchdown PCR, 98°C, 30 s; 75>72°C, 30 s; 72°C, 1 min, then 20 cycles 98°C, 30 s; 71°C, 30 s; 72°C, 1 min. Successful amplification is determined by running 1.5  $\mu$ l of the product on a 1% agarose gel, and checking for the 1192 bp product (see Supplementary Figure S3).

### Error rate analysis

Gene synthesis reaction on beads was carried out using either DNA Polymerase I, Large (Klenow) Fragment (NEB) for APE assembly reactions entirely carried out at room temperature (no specific annealing temperatures used), or Phusion High-Fidelity Polymerase with brief annealing and extension steps on a thermal cycler. Both approaches used unpurified oligomers for the construction of ZF array 92–1 (14) within a Zif268 backbone. For Klenow assembly error analysis, following 20 cycles of PCR using Phusion polymerase for amplification of template detached from beads using an SDS-boiling and reannealing protocol, the reaction was run on an agarose gel and the product band was gel-stabbed, and a second PCR (20 cycles with Phusion

polymerase) was run. For Phusion assembly error analysis, following 20 cycles of PCR using Phusion polymerase for amplification of the template attached to beads in 1x SSC buffer, the reaction was run on an agarose gel and the product band was gel-stabbed, and a second PCR (20 cycles) was run. The PCR product (239 bp) from each of the band-stab PCRs was purified and cloned via Gibson assembly into the pUC19 plasmid with assembly-check overhangs. In addition, the PCR from the Phusion assembly was used without the band-stab procedure, to determine whether the band-stab has an effect on error rate.

Chemically competent DH5 $\alpha$  *E. coli* cell aliquots (30  $\mu$ l) were transformed with 1.5  $\mu$ l of each Gibson assembly product via heat shock (30 s at 42°C), recovered in 300  $\mu$ l SOC medium for 1 h at 37°C and plated on ampicillin plates for overnight growth at 37°C. Colonies from each plate (Klenow+gel stab, Phusion+gel stab, Phusion no gel stab) were picked with a sterile 200  $\mu$ l pipette tip, briefly stirred in 20  $\mu$ l PCR-grade water, boiled for 15 min and centrifuged. Water from the colony boils was used as template for an insert-check PCR using the primers pUC19-f and pUC19-r with Phusion polymerase. All of these PCR reactions were run on 1.5% agarose gels with GelRed to determine which colonies had the correct sized insert. Colonies were picked and analyzed in this way until 32 colony PCRs for each assembly method were identified with a single band corresponding to the correct sized insert. The insert-check PCRs were submitted for Sanger sequencing in a 96-well plate without PCR cleanup (Microsynth AG), and the resulting sequencing reads were aligned with the expected sequence to analyze the error-rate and identify which types of errors were prevalent.

The highest error rate (Supplementary Table S3) was observed in the colonies cloned with the Klenow APE assembly method, however the majority of these products would have resulted in nonfunctional ZF TFs since they coded for early stop codons or caused frame-shifts which would prevent the EGFP tag from folding and being captured or detected in a MITOMI experiment. The Phusion APE assembly products yielded significantly lower erroneous sequences, with no effect seen from the band-stab PCR. The high error rate seen in the Klenow reaction product can be partly explained by comparing the relative error rates documented for Klenow and Phusion polymerase.

### dsDNA target synthesis

Double stranded DNA targets (dsDNA) for ZF array binding were prepared via isothermal Klenow extension as previously described (32). Oligomers were designed such that the DNA contains the 9 nt target sequence with single nucleotide flanks (11 nt total). At the 3' end of the DNA is the complementary sequence for a Cy5-labeled primer (5' CompCy5; Supplementary Table S4). The reaction consists of two steps, one annealing step, and one extension step. Each 20  $\mu$ l annealing reaction contains 1x NEB Buffer 2, 10  $\mu$ M 5' CompCy5 and 15  $\mu$ M Target oligomer. This mixture was placed on a thermocycler, heated to 94°C for 5 min, then cooled to 37°C (10% ramp) for 5 min, then held at 20°C. Following the annealing program, 10  $\mu$ l of extension mix (1x NEB Buffer 2, 3mM dNTPs, and 2.5 units Klenow exo-)

are spiked into the annealing mix, and the reaction is thermocycled according to the following routine: 37°C, 90 min; 75°C heat kill, 20 min; 10% ramp down to 30°C, 30 s; hold at 4°C.

### Microarraying

*Preparation of epoxy-silane glass slides.* For microarray printing, epoxy-silane functionalized glass slides were prepared, following an adapted protocol (33). A MilliQ water and ammonia solution (NH<sub>4</sub>OH, 25%) mixture was prepared in a 5:1 ratio, respectively, and heated to 80°C. Then, 150 ml of hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>, 30%) was added to the mixture, and glass microscope slides were placed in the cleaning bath for 30 min. The glass slides were then removed and rinsed in fresh MilliQ water, dried with N<sub>2</sub>, and placed in a second bath containing 1% 3-glycidypropyltrimethoxymethylsilane (97% purity) in toluene, and incubated for at least 20 min at ambient temperature. Then, the glass slides were removed, rinsed in fresh toluene, dried with N<sub>2</sub> and baked at 80°C for 30 min. The glass slides were removed from the oven, allowed to cool and stored under vacuum at room temperature in opaque storage boxes until used. Immediately prior to microarray printing, both sides of each glass slide is briefly rinsed with fresh isopropanol, dried with N<sub>2</sub>, rinsed with fresh toluene and dried with N<sub>2</sub> again.

*Sample microarraying.* All samples to be printed onto a microarray were prepared in a 384-well microtiter plate. Each zinc-finger array assembly (1192 bp PCR product) was co-spotted with a target DNA (Klenow extension products) in duplicate onto epoxy-silane coated glass slides using a microarray robot (QArray2) with a 946MP4 microspotting pin (Arrayit). Up to four slides were prepared in a single printing session. Each spot on the array was generated by four consecutive printing programs. Immediately following completion of printing from a given sample plate, the sample wells were covered with adhesive PCR foil seals (ThermoScientific) and stored at -20°C until needed for future printing runs. In general, Klenow target plates were re-used for up to five printing runs before the volume of each well became too low to be used. After the final printing program, the microspotting pin is cleaned by sonication for 15 min in a 15 ml Falcon tube containing one drop of dish detergent and 10 ml of deionized water, then sonicated for 30 min in a 15 ml Falcon tube containing 10 ml of 70% ethanol in water. Before use, the pin tip is rinsed under deionized water, dried using a high pressure (100 psi) compressed air gun, then inspected under a microscope to verify the tip is clean and free of debris. Finally, the shaft of the pin is briefly polished with a dry Kimwipe to prevent sticking when placed into the robotic printer head.

*Microarray printing routine.* The first BSA (bovine serum albumin; Sigma) printing routine is to block the surface of the glass, to limit the amount of template and target DNA that is irreversibly attached to the epoxy-silane surface. The second BSA printing routine is implemented to deposit a thin blocking layer on top of the printed template spots, to reduce the risk of cross-contamination due to template

carry-over during target printing (since target printing does not involve same-sample washing steps). The full printing routine and operational details can be found in Supplementary Table S5 and Figure S3.

### MITOMI chip fabrication and operation

**Mold fabrication.** The MITOMI microfluidic device (6,7) (Supplementary Figure S4) consists of two superimposed layers, the flow layer and the control layer. Each layer is fabricated in polydimethylsiloxane (PDMS) using 4" silicon wafer molds fabricated using standard lithography techniques (34). Each wafer (control and flow) contained three pattern replicates for a 1024-chamber (16 rows by 64 columns) MITOMI device.

Mask fabrication was carried out using a Heidelberg DWL200 laser lithography system with 10mm writing head and solid state wavelength stabilized laser diode (max. 110 mW at 405 nm). Each layer of the MITOMI device was reproduced as a chrome mask. After laser writing, the chrome mask is cycled twice for 15 s in developer mixture (1:5 MP351 and deionized water, respectively), 45 s of agitation, then rinsed and dried. The developed mask is then chrome etched for 110 s, rinsed, cleaned twice for 15 min in 1165-remover bath, rinsed and air dried.

The flow layer mold is first cleaned for 7 min in a Tepla300 plasma stripper with 400 ml/min O<sub>2</sub> at 500W and 2.45 GHz. The wafer is then treated with hexamethyldisilazane (HMDS) using an ATMsse hotplate at 125°C for 12 min. Positive photoresist AZ9260 is spin-coated on the cleaned wafer for 10 s at 800 rpm, then 40 s at 1800 rpm (ramp 1000 rpm/s) to produce a substrate height of 14 μm. The wafer is then baked on a 115°C hotplate for 6 min. The soft-baked positive resist is then allowed to rehydrate for 1 h. The wafer is then exposed during three intervals of 18 s with a 10 s pauses between each exposure on a MA6 mask aligner (power 360 mJ/cm<sup>2</sup>, intensity 10 mW/cm<sup>2</sup>, broad-spectrum lamp, hard contact exposure mode). After a 1 h relaxation time, the wafer is developed in a DV10 chamber via multiple, automated cycles of rinsing/agitation with development mixture (1:4 ratio of AZ400K and deionized water, respectively) until the features are visible. Finally, the wafer is heated to 160°C for 20 min to anneal and round the features of the flow wafer to create a profile that allows complete valve closure.

The control layer mold is first cleaned following the same plasma treatment protocol as the flow layer mold. Negative photoresist SU-8 GM1060 (Gestetec) is spin-coated on the cleaned wafer for 10 s at 500 rpm (ramp 100 rpm/s), 10 s at 1500 rpm (ramp 100 rpm/s), 1 s at 2500 rpm and finally 6 s at 1500 rpm to produce a substrate height of 14 μm. The wafer is baked on a hotplate for 30 min at 130°C, then 25 min at 30°C. The wafer is then exposed on a MA6 mask aligner for 13.2 s (power 360 mJ/cm<sup>2</sup>, intensity 10mW/cm<sup>2</sup>, broad-spectrum lamp, hard contact exposure mode). The exposed wafer is developed manually by bathing in PGMEA twice for 1.5 min, then rinsed in isopropanol and dried with an air gun.

**Device fabrication.** Prior to PDMS casting, both the flow and control layer wafers are subjected to vapor deposition

of trimethylchlorosilane (TMCS, EMD Millipore Corp.) for at least 30 min by placing them within a sealed plastic container with a small dish containing 0.25 ml liquid TMCS. TMCS treatment is repeated for at least 15 min before all subsequent PDMS casting rounds.

The control layer wafer is placed into an aluminum foil-lined glass Petri dish, and 60g of Sylgard elastomer (5:1 mix of elastomer base and curing agent, respectively) is mixed for 1 min at 2000 rpm (400×g) and degassed for 2 min at 2200 rpm (440×g) in a centrifugal mixer. The elastomer mixture is poured on top of the control layer in the Petri dish, and degassed in a vacuum dessicator for 20 min at ambient temperature.

For the flow layer, 21 g of PDMS mixture is prepared at the ratio of 20:1 (base:curing agent), then mixed and degassed in a centrifugal mixer according to the same speeds and times as the control layer. The flow wafer is carefully centered on top of a spin-coater platform using wafer tweezers, and the flow layer PDMS mixture is poured in the center, taking care not to create any bubbles. The mixture is spin-coated onto the wafer with a 15 s ramp and 35 s spin at 2800 rpm.

The degassed PDMS on the control layer wafer is removed from the vacuum chamber. Residual bubbles are removed with a scalpel and any pieces of dust are carefully removed from the control channel grid using the tip of the scalpel blade. Both the control and flow layers are then placed into an oven at 80°C for 28–30 min. After baking, both Petri dishes are removed from the oven and briefly allowed to cool. The control layer is then cut with a scalpel in a rectangle around each pattern replicate, and each rectangle of cured PDMS is carefully peeled away from the silicon wafer. Holes are punched through each of the control line input channels on the patterned side of the PDMS block. The patterned side of the control layer is cleaned twice with Scotch Magic Tape to remove dust and debris then quickly placed on top of the flow layer replicates. A stereomicroscope is used to precisely align the features of the control layer so that they overlap with the chambers visible on the flow layer. Once aligned, the assembled device is bonded at 80°C for 90–180 min. The bonded devices are removed from the oven and briefly allowed to cool. A scalpel is guided around the outer edge of the control layer PDMS block to cut the thin flow layer. Then each individual device is gently peeled from the flow layer wafer, and holes are punched through the patterned side inlets and outlet of the flow layer. Each device is then cleaned with Magic Scotch tape and trimmed to fit within the boundary defined by the glass slide-holding cartridge of the microarray scanner. The assembled device is aligned with a printed microarray on an epoxy-silane glass slide using a stereomicroscope and bonded overnight at 80°C before use.

The flow layer mold is cleaned of residual polymerized PDMS by pouring on another layer of mixed PDMS (this can be leftover control- or flow-layer PDMS mixtures prepared earlier; to stall cross-linking for several hours, store the PDMS mixture at 4°C), and baked at 80°C for at least 1 h. The resulting thicker layer of PDMS can be easily peeled away from the flow layer, resulting in a clean surface to repeat the process. Both the cleaned flow wafer and control wafer are cleaned with high pressure (100 psi) compressed

air gun to dislodge pieces of dust or PDMS before being treated with TMCS.

**Device setup.** Assembled MITOMI chips (Supplementary Figure S4) bonded to microarray printed glass slides were stored at 40°C following an overnight bonding at 80°C until used. To begin an experiment, control line tubing is filled with PBS using a syringe, and pins are placed into the appropriate locations to feed into the control valves of the microfluidic device. The control lines are actuated at low pressure (10 psi) to begin filling the control lines of the microfluidic device. Once all of the control lines are filled, the sandwich valves and button valves are deactivated, and the pressure is increased to 20–22 psi to ensure complete closure of all other valves.

**Surface derivatization, protein synthesis, binding assay and device readout.** Biotinylated-BSA (2 mg/ml; Thermo Fisher Scientific) is flowed through the device for 15 min at 3.5 psi. The chip is then washed with 0.01% Tween20 in PBS for 5 min to wash away unbound biotin-BSA. Next, neutravidin (1 mg/ml; Thermo Scientific) is flowed for 15 min followed by 0.01% Tween20 in PBS for 5 min. The button valves are then activated and biotin-BSA is again flowed across the chip for 10 min, blocking all of the neutravidin binding sites except those protected under the area of the button valve. The chip is again washed with 0.01% Tween20 in PBS for 5 min. Then a solution containing 0.5  $\mu$ l biotinylated antibody to GFP (1 mg/ml stock, Abcam ab6658) in 100  $\mu$ l 1% BSA in PBS is flowed across the chip for 5 min, the button valve is deactivated, and the antibody solution is flown for 15 min, allowing the antibody to bind to the available neutravidin under the button valve. Then, the chip is flushed with 0.01% Tween20 in PBS for 5 min, and with PBS for 5 min. The button valve is then activated, and ITT mixture (*in vitro* transcription/translation, NEB PURExpress, 10  $\mu$ l SolutionA, 7.5  $\mu$ l SolutionB, 0.5  $\mu$ l RNase Inhibitor (Roche), 7  $\mu$ l PCR grade water) is flowed for 10 min. The exit valve is activated for 2 min while the ITT is being flowed on-chip to build up pressure. The neck valve is deactivated, and the ITT mixture is allowed to fill the DNA chambers for 1–2 min. Once the DNA chambers are filled, the neck valves are activated, the exit is opened, and fresh ITT is allowed to flow across the chip for 10 min. Then the sandwich valve is activated while flowing ITT mix during the last minute of ITT washing. Once the sandwich valves are partially closed, the button valve is deactivated, the neck valve is deactivated, the exit is closed, and the flow of ITT is stopped. The inlet tree valve controlling entry to the chamber array is closed, and the inlet tree is briefly flushed with 0.01% Tween20 in PBS. Then the entire chip is placed on a flatbed thermal cycler set to 37°C, and incubated for 3–5 h. During this time, the DNA array spots are rehydrated in the ITT mix, transcription and translation occur, synthesized ZF-EGFP fusion protein diffuses and is bound by the anti-GFP moiety located under the button valve, and target DNA diffuses and interacts with the various ZF DNA binding domains. After incubation, the chip is placed into an ArrayWoRx microarray scanner, and an image is taken in three fluorescent channels (A488/GFP, Cy3, Cy5) to determine relative amounts of solution phase target DNA in the

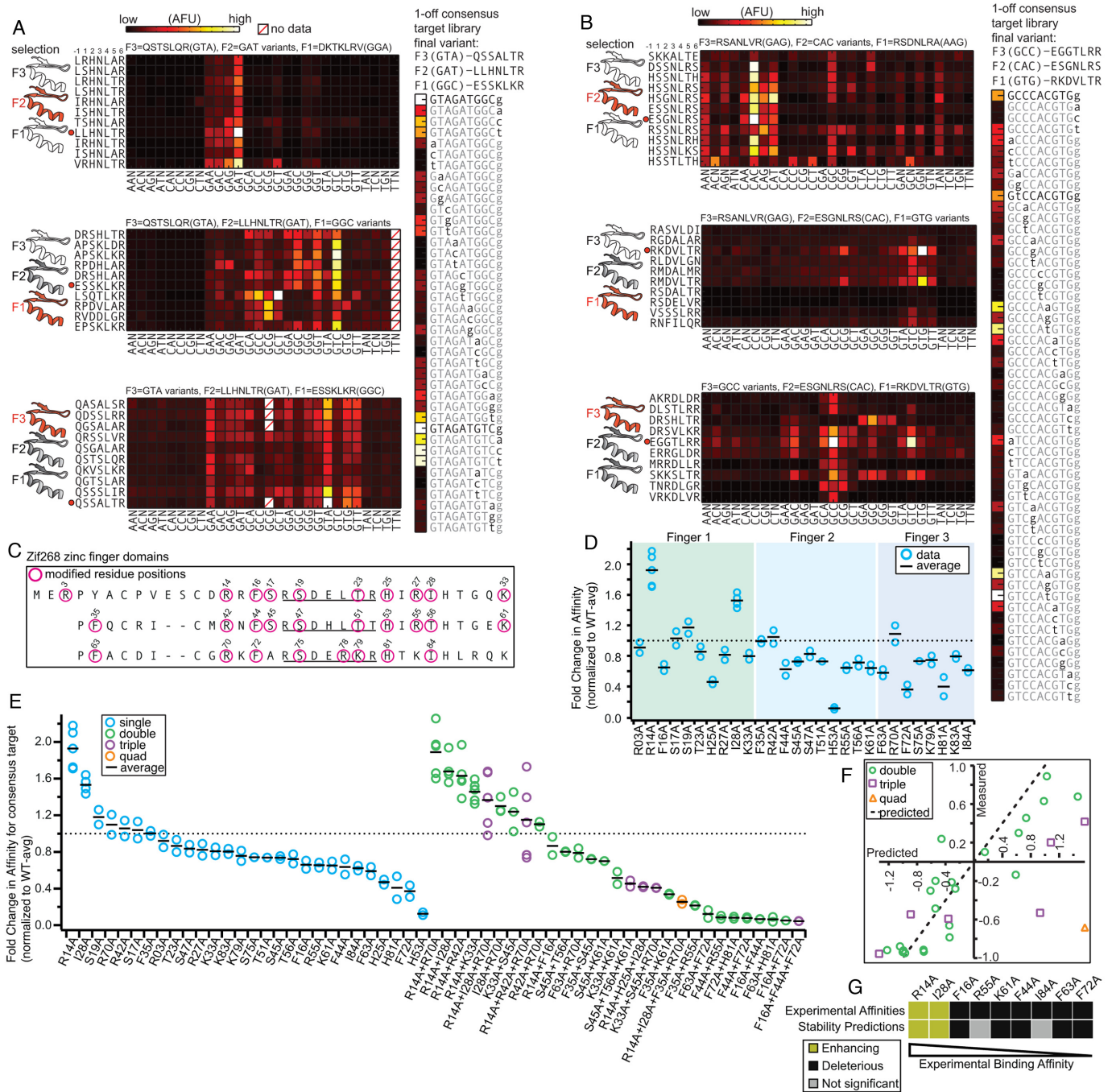
MITOMI chamber. The button valves are then activated, the sandwich valves are deactivated, and the neck valve is activated. The flow space is washed with 0.01% Tween20 in PBS for 5 min to remove unbound target DNA and protein, then the chip is scanned again in the three fluorescent channels, giving the total protein signal (EGFP/A488 signal) and the relative amount of target bound (Cy5 signal). Each ZF-EGFP fusion template was tagged with Cy3 during the final PCR step, and though signal from this channel was captured in each scan, it was not factored into downstream binding-specificity analyses, primarily because little Cy3 signal was detected as being bound by the ZF proteins and normalization for protein amount was performed with the EGFP fluorescence.

**Image analysis and affinity value calculations.** Images acquired from the experiment were processed using a 1024 unit detection array in GenePix v6.0. Raw tif files from the ArrayWoRx scanner were loaded into the GenePix software, and using the grid detection tool, an array of 1024 circular areas was snapped onto the EGFP spots detected in the A488 scan after washing. Small adjustments to the grid were performed by hand for poorly detected locations. Mean and median fluorescence measurements were taken in each fluorescent channel before and after washing. In addition, local background measurements were taken by dragging the detection array off the button valve locations into the space just outside the reaction chamber. For each scan, each circular area in the array was background corrected using its own local background measurement. Data points were filtered to ensure that EGFP levels were at least 500 AFU or higher and Cy5 target levels at least 1000 AFU or higher. These filtered, and background-corrected data points were used to calculate relative affinity values:

$$\text{relative affinity (reported in AFU)} = \frac{\text{Cy5}_{\text{bound}}(\text{after wash})}{\text{Cy5}_{\text{total}}(\text{before wash})} \cdot \frac{1}{\text{A488}(\text{EGFP after wash})}$$

These relative affinity values were used to compare ZF affinity to various targets across different experiments. In general, at least 2 data points were averaged together to arrive at a single affinity value (as in the heatmaps generated in Figures 1F and 2A, B).

**Protein modeling methods.** The Rosetta software suite (35) was used to model the Zif268 (PDBid-1A1L (36)) alanine mutants in the presence of the DNA target consensus sequence. First, we computed differences in Rosetta energy units regarding the protein-DNA interaction (37) caused by the mutations ( $\Delta\Delta G_{\text{bind}} = \text{Rosetta Binding Energy}_{\text{mut}} - \text{Rosetta Binding Energy}_{\text{WT}}$ ), the Rosetta Binding Energy for intermolecular interactions is computed by the difference of between the energy of the complex and the sum of the energies of the two molecules apart ( $\text{Rosetta Binding Energy} = \text{Rosetta Energy}_{\text{complex}} - (\text{Rosetta Energy}_{\text{partner1}} + \text{Rosetta Energy}_{\text{partner2}})$ ). To perform such calculations we used the Rosettascripts application (38) with which mutants were generated and side chains in contact with DNA base pairs were subjected to energy minimization by sampling small deviations of the dihedral angles of the rotamers present in the crystal structure. The stability calculations for



**Figure 2.** Engineering ZF TF specificity and affinity. (A, B) To the left of each heatmap is a cartoon of the 3 ZF domains in the TF, where ‘hollow’ ZF domains indicate the location of unaltered RHs, red ZF domains indicate the location of the RH variants being tested, and gray ZF domains indicate locations where RH variants have already been screened. RH amino acid variants are listed and a red dot indicates the RH that was chosen and used in the subsequent selection round. The vertically-oriented heatmap gives the results of the 1-off DNA consensus target library with the final ZF TF design. Values represent averages of 2–10 independent measurements. (A) RH variant characterization across three rounds of screening to engineer a ZF TF that binds the sequence 5'-GTA GAT GGC. (B) RH variant characterization across three rounds of screening to engineer a ZF TF that binds the sequence 5'-GCC CAC GTG. (C) Amino acid alignment of the ZF domains in Zif268 where magenta circles indicate residues that were changed to alanine. (D) Results of single residue alanine scan at the residue positions listed and the resulting fold-change in binding affinity relative to WT Zif268 towards the WT consensus target. (E) Rank ordered fold-change in binding affinity relative to WT Zif268 towards the WT consensus target for single and multiple mutant variants. (F) Comparison of measured versus predicted fold-change in binding affinity to determine if a simple additive model can explain the change in affinity caused by multiple mutations. Each data point (circle) in panels D–F represents a dilution series consisting of 7 measurements while the black bars show the mean value of 2–5 independent experiments. (G) Comparison between experimentally measured Zif268-DNA affinities and *in silico* predicted protein stabilities. The subset of mutants with the most pronounced binding changes (excluding histidine to alanine mutants) are shown. Rosetta energies were only considered significant if the absolute changes were above 0.75.



Zif268 were also performed using Rosettascripts the differences in stability were computed through the difference in Rosetta energy between the mutants and the wild-type proteins ( $\Delta\Delta G_{\text{stab}} = \text{Rosetta Energy}_{\text{mut}} - \text{Rosetta Energy}_{\text{WT}}$ ). These simulations were performed in the presence of the zinc ions and all side chains were allowed to sample different rotamers and where energy minimized by sampling small deviations in the dihedral angles. For each mutant, 50 structural models were generated and the average of Rosetta energies is reported. Detailed command lines are presented in Supplementary Figure S5.

## RESULTS

Gene synthesis by APE (Figure 1B) is a restriction enzyme- and ligase-free method that operates without the need for transformation, sequencing, and PCR/oligomer purification steps. By coupling APE with cell-free protein expression and MITOMI we reduce the time between protein design and characterization to 3–4 days per iteration cycle (Figure 1C). For ZF TF engineering, this pipeline consists of four stages: (i) solid-phase gene assembly from oligonucleotides, (ii) fusion and overhang extension PCR to generate expression ready linear templates, (iii) microfluidic device programming and (iv) cell-free protein expression, purification and DNA-binding characterization (Figure 1C).

APE is a solid phase synthesis approach based on asymmetric primer extension (Figure 1B). We succeeded in performing up to nine consecutive rounds of extension (Supplementary Figure S6), whereas previous methods achieved up to four consecutive extension steps (39). Gene assembly on magnetic beads allows pooling of identical reactions before fractionation to incorporate unique oligomers in subsequent extension cycles, and sequential addition of each oligomer reduces the risk of forming chimeric products that result from one-pot assembly reactions. In this study we implemented APE assembly using manual bench-top methods but the method can be automated with standard high-throughput liquid-handling robotic platforms. Our APE gene synthesis exhibited a low error rate of 0.78 errors/kb, which is as good or better than other current gene synthesis approaches based on ligation or PCR assembly (Figure 1D; Supplementary Tables S3 and S6). We have generated synthetic products with APE ranging over a length of 480–610 bp with 6–9 rounds of extension, respectively. We fuse our APE synthesized gene products with GFP to form a 1029 bp long final gene product, which is then rendered expression ready using a standard two-step extension PCR reaction. These lengths are sufficient to cover a number of proteins and protein domains that are of interest to protein engineering and synthetic biology such as: transcription factor binding domains (Zn-finger, bHLH, bZip, homeodomains, etc.), scFVs (700–800 bp), single-chain antibodies, single-domain antibody/nanobodies (300–400 bp) and protein–peptide binding domains such as SH3 (150–200 bp), PDZ (200–300 bp), bZIP (180–240 bp). For designed ZF TFs, five DNA oligonucleotides are required to assemble the ZF ‘core’ (239 bp), with lengths between 60 and 77nt, and overlaps of 25–28 bp (Supplementary Figure S2, Table S1) consisting of three unique oligomers coding for the three recognition helices, and two universal ‘link-

ing’ oligomers. Finally, APE permits the synthesis of genes with considerable internal sequence redundancies such as Zn-fingers, and allows the synthesis of a large number of gene variants which have high-sequence homology, which is generally problematic when using pooling based synthesis approaches.

We first synthesized 64 different ZF TFs and quantified binding of each against a library of the corresponding 64 predicted consensus DNA targets (Figure 1E, F; Supplementary Tables S7 and S8). APE-MITOMI successfully expressed all 64 ZFs and 90.6% (58/64) of these bound DNA (Figure 1E). Of those ZFs that bound DNA, 89.7% (52/58) bound the expected consensus target within the top 4 highest-affinity targets and 58.6% (34/58) bound the expected target with highest affinity. To verify that failure to bind and that sequence specificity was unaffected by our approach we cloned and sequence verified the CCC and DDD ZF TF variants generating the plasmid based versions CCCp and DDDp. Both plasmid-based ZF TFs gave identical results when compared to APE generated variants. The heat map in Figure 1F shows the quantitative binding specificity of each of the 64+2 ZF variants. This approach can be used to rapidly identify orthogonal ZF TFs for use in the design and implementation of synthetic genetic networks (14) and to eliminate ZFs with sub-optimal binding characteristics such as low affinity or extensive non-specific binding. For example, we observed that ZFs containing a ‘C’ finger in position F2 exhibited high levels of non-specific binding, especially when combined with finger variants A and B in position F3.

We next applied APE-MITOMI to engineering ZF specificity. We decided to engineer a ZF with a consensus sequence of ‘GTA GAT GGC’, taking advantage of the relatively well-populated selection of GNN-binding RHs in the Zinc Finger Consortium database (Figure 2A). We first modified F2 of Zif268 by replacing it with 16 RHs (Figure 2A and Supplementary Figure S9) listed as ‘GAT’ binders. Characterizing these variants showed that there was considerable variability in the specificity and binding affinity of the 16 RHs tested. One RH, LLHNLTR, exhibited the highest affinity and specificity for ‘GAT’, and we thus selected this helix for the next step, which involved substituting RH variants into the F1 position to target ‘GGC’. Selecting a helix with good specificity from these RHs proved challenging since nearly all RHs preferred ‘GTC’ and exhibited considerable non-specific binding (Supplementary Figure S10). A similar specificity profile was observed when we tested the highest affinity F1 RHs in a different framework, which reduced the non-specific background (Supplementary Figure S11). We selected ESSKLR as the best ‘GGC’ binder in position F1 considering its level of specificity for ‘GGC’ relative to ‘GTC’. We then substituted F3 with RHs reported to bind ‘GTA’. The RH that displayed the highest ‘GTA’ binding affinity and specificity, QSSALTR, was generated by selecting the highest frequency residue in each position derived from all ‘GTA’ RH variants listed in the database (Supplementary Figure S12). In order to characterize the actual specificity of our synthetic ZF TF (F1 ESSKLR, F2 LLHNLTR, F3 QSSALTR), a 1-off target library was prepared for the consensus sequences: ‘GTA GAT GGC’ and ‘GTA GAT GTC’. Given the relatively

poor performance observed during the F1 selection step, the engineered ZF TF exhibited a surprisingly high affinity and specificity for the intended target consensus sequence ('GTAGATGGC'). In an attempt to improve the specificity for the intended target, we performed an additional screen for designed RH variants which were anticipated to bind 'GGC' using online DNA binding site predictors (20–23) (Supplementary Figure S13), but nearly all these variants displayed a higher affinity for GTC and thus failed to further improve the specificity of the engineered ZF TF (Supplementary Figure S14).

We repeated the same sequential selection process to engineer a ZF TF recognizing the sequence 'GCC CAC GTG', which is a more challenging target sequence given the 'CAC' in the F2 position (Figure 2B and Supplementary Figures S15–S17). A number of RHs in positions F2 and F1 were found that bound 'CAC' and 'GTG', respectively. Finding a high affinity, yet specific RH in position F3 for GCC was more challenging and we had to settle for a RH that bound both 'GCC' and 'GTC' (Supplementary Figure S18). Although each individual RH appeared to bind the intended target sequence fairly specifically, when testing the assembled ZF TF using a one-off library the consensus target was 'GTC CAT GTG', but 'GCC CAC GTG' could also be bound albeit with a lower affinity. Other RHs from the initial 'GCC' screen, which displayed lower affinity but higher specificity for 'GCC' were also tested against a one-off library (Supplementary Figure S19), but all of these exhibited higher affinity for other targets than 'GCC CAC GTG'. These results indicate that engineering ZF TFs to bind a particular sequence is relatively easy to achieve using a step-wise selection process, but engineering the precise specificity landscape of a ZF TF is more challenging. The individual specificity profiles of RHs can vary considerably, and in some instances do not reflect their annotated sequence specificity given in the ZF Consortium Database or online binding site predictors. One significant advantage of APE-MITOMI based assembly and characterization is the fact that the method returns information on the precise specificity landscape and affinity of the synthesized transcription factors during all stages of assembly. APE-MITOMI allows the generation of TFs with similar consensus sequences but different specificity and/or affinity profiles and enables the examination of these characteristics for the function of native and synthetic transcriptional regulatory networks.

Focus in ZF engineering has been on rationally designing ZF specificity, but it is unclear to what extent affinity can be tuned and whether affinity can be tuned independently of sequence specificity. The ability to tune ZF affinity is of interest in creating synthetic transcriptional regulatory networks (14) and it would drastically simplify the task of engineering ZF TFs if affinity could be tuned independently of specificity. Based on the Zif268 crystal structure we selected 28 residues that could be involved in determining the protein's affinity to DNA, and changed these residues to alanine (Figure 2C). As expected, most modifications resulted in modest decreases in affinity as compared to wildtype (WT) Zif268 (up to 2-fold). To our surprise, a few modifications led to increased affinity (R14A, I28A, R70A) (Figure 2D). Given that the single substitutions resulted in modest affinity decreases, we tested 22 double sub-

stitutions, 6 triple substitutions and 1 quadruple substitution (Figure 2E). These novel mutants allowed us to extend the dynamic range to 1/5 of WT and allowed us to smoothly tune affinity over the entire accessible range between 2x increased and 5x reduced affinity. We could show that most of the affinity mutants retain their specificity landscape, indicating that affinity can be tuned independently of specificity (Supplementary Figure S20). We also tested whether a simple additive model could predict double, triple and quadruple substitutions (Figure 2F). Overall, the simple additive model captures the trend for double substitutions but performs poorly for triple and quadruple substitutions.

Next, we sought to understand whether the observed differences in affinity could be predicted by energy calculations performed *in silico* based on the available structural information for Zif268. To do so, we modeled single alanine mutants with the Rosetta protein modeling software (35). First, we analyzed the effect of the alanine mutants on the Zif268-DNA binding energy and perhaps not surprisingly, given the low number of atomic interactions established across the interface for the large majority of these residues, we did not find any large differences in the binding energies of the point mutants in comparison with the WT Zif268. Bolstered by the differences in affinity observed in mutants of zinc-coordinating residues (H25A, H53A and H81A), which recognizably play a critical role for structural integrity of the zinc finger fold (40), we asked if other mutations could also cause considerable losses in protein stability and consequent losses in DNA binding affinity. To address this question we performed *in silico*  $\Delta\Delta G$  calculations using Rosetta. Being aware of the limitations of current modeling techniques in accurately describing subtle structural and functional effects, we focused on mutants that exhibited pronounced differences in binding (reductions < 0.7 and improvements > 1.3–9 in total). In this subset of mutants (seven deleterious and two enhancing; Figure 2G) we were able to correctly predict the variations in binding affinity with our stability predictions for seven of the nine mutant proteins, strongly suggesting that the observed binding affinity differences are linked to protein stability.

## DISCUSSION

We coupled a gene synthesis strategy directly with high-throughput on-chip protein expression, purification and quantitative characterization, showing that it is possible to completely circumvent molecular cloning and cell-based processing steps generally needed in protein biochemical analysis. Integration of these methods reduced the time required between protein design and protein characterization to 3–4 days. In this study we synthesized over 400 ZF TF variants, and characterized the binding specificity of each for a total of over 98 000 measurements. By comparison, the human genome encodes  $\approx 675$  C<sub>2</sub>H<sub>2</sub> ZF TFs (41). We have previously shown that over 400 full-length drosophila transcription factors could be expressed on a single MITOMI device (32) indicating that there is no major bottleneck to on-chip protein expression. Although we synthesized all 400+ genes in this study by hand, the APE gene synthesis method can be readily automated on standard liquid handling robots eliminating this last bottleneck in the process.

It should thus be possible to characterize thousands of synthetic protein variants per week, which in turn will enable exploration of the protein sequence-function relationship in unprecedented detail, aid in the development of accurate computational predictions of protein function, and allow us to rapidly engineer novel proteins.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank members of the Maerkl Lab for helpful discussions, the Deplancke Lab for use of their ArrayWoRx scanner, and EPFL for financial support.

*Author contributions:* M.C.B. and S.J.M. conceived of the synthesis technique, designed the project, and analyzed the data. M.C.B. and E.P. tested early protein variants, optimized the gene synthesis method and determined the error rate. M.C.B. prepared all of the ZF TF templates and conducted all of the MITOMI experiments. B.E.C. performed the *in silico* analysis. M.C.B., E.P., B.E.C. and S.J.M. wrote the paper.

## FUNDING

Funding for open access charge: École Polytechnique Fédérale de Lausanne (EPFL).

*Conflict of interest statement.* None declared.

## REFERENCES

- Packer, M.S. and Liu, D.R. (2015) Methods for the directed evolution of proteins. *Nat. Rev. Genet.*, **16**, 379–394.
- Tinberg, C.E., Khare, S.D., Dou, J., Doyle, L., Nelson, J.W., Schena, A., Jankowski, W., Kalodimos, C.G., Johnsson, K., Stoddard, B.L. *et al.* (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*, **501**, 212–216.
- Correia, B.E., Bates, J.T., Loomis, R.J., Baneyx, G., Carrico, C., Jardine, J.G., Rupert, P., Correnti, C., Kalyuzhnyi, O., Vittal, V. *et al.* (2014) Proof of principle for epitope-focused vaccine design. *Nature*, **507**, 201–206.
- Kosuri, S. and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods*, **11**, 499–507.
- Quan, J., Saaem, I., Tang, N., Ma, S., Negre, N., Gong, H., White, K.P. and Tian, J. (2011) Parallel on-chip gene synthesis and application to optimization of protein expression. *Nat. Biotechnol.*, **29**, 449–452.
- García-Cordero, J.L. and Maerkl, S.J. (2015) Mechanically Induced Trapping of Molecular Interactions and Its Applications. *J Lab Autom.* doi:10.1177/2211068215578586.
- Maerkl, S.J. and Quake, S.R. (2007) A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science*, **315**, 233–237.
- Wolfe, S.A., Nekludova, L. and Pabo, C.O. (2000) DNA Recognition by Cys2His2 Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 183–212.
- Beerli, R.R. and Barbas, C.F. (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.*, **20**, 135–141.
- Tebas, P., Stein, D., Tang, W.W., Frank, I., Wang, S.Q., Lee, G., Spratt, S.K., Surosky, R.T., Giedlin, M.A., Nichol, G. *et al.* (2014) Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.*, **370**, 901–910.
- Gupta, A., Christensen, R.G., Rayla, A.L., Lakshmanan, A., Stormo, G.D. and Wolfe, S.A. (2012) An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods*, **9**, 588–590.
- Gersbach, C.A., Gaj, T. and Barbas, C.F. (2014) Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. *Acc. Chem. Res.*, **47**, 2309–2318.
- Gonzalez, B., Schwimmer, L.J., Fuller, R.P., Ye, Y., Asawapornmongkol, L. and Barbas, C.F. (2010) Modular system for the construction of zinc-finger libraries and proteins. *Nat. Protoc.*, **5**, 791–810.
- Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K. and Collins, J.J. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, **150**, 647–658.
- Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J. and Khalil, A.S. (2014) Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell*, **158**, 110–120.
- Dreier, B., Fuller, R.P., Segal, D.J., Lund, C.V., Blancafort, P., Huber, A., Kokscha, B. and Barbas, C.F. (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **280**, 35588–35597.
- Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protoc.*, **1**, 1329–1341.
- Maeder, M.L., Thibodeau-Beganny, S., Osiaik, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. *et al.* (2008) Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
- Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. *et al.* (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods*, **8**, 67–69.
- Persikov, A.V., Osada, R. and Singh, M. (2009) Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics*, **25**, 22–29.
- Persikov, A.V. and Singh, M. (2013) De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.*, **42**, 97–108.
- Persikov, A.V., Rowland, E.F., Oakes, B.L., Singh, M. and Noyes, M.B. (2014) Deep sequencing of large library selections allows computational discovery of diverse sets of zinc fingers that bind common targets. *Nucleic Acids Res.*, **42**, 1497–1508.
- Persikov, A.V., Wetzel, J.L., Rowland, E.F., Oakes, B.L., Xu, D.J., Singh, M. and Noyes, M.B. (2015) A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res.*, **43**, 1965–1984.
- Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M. *et al.* (2015) C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.*, **33**, 555–562.
- Fu, F. and Voytas, D.F. (2013) Zinc Finger Database (ZiFDB) v2.0: a comprehensive database of C2H2 zinc fingers and engineered zinc finger arrays. *Nucleic Acids Res.*, **41**, D452–D455.
- Ramirez, C.L., Foley, J.E., Wright, D.A., Müller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
- Lam, K.N., van Bakel, H., Cote, A.G., van der Ven, A. and Hughes, T.R. (2011) Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res.*, **39**, 4680–4690.
- Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C. and Jahn, D. (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.*, **33**, W526–W531.
- Wilson, R. (2011) Preparation of single-stranded DNA from PCR products with streptavidin magnetic beads. *Nucleic Acid Ther.*, **21**, 437–440.
- Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeberg, J. and Uhlén, M. (2005) The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis*, **26**, 501–510.
- Bjournson, A.J. and Cooper, J.E. (1992) Band-stab PCR: a simple technique for the purification of individual PCR products. *Nucleic Acids Res.*, **20**, 4675.

32. Rockel,S., Geertz,M., Hens,K., Deplancke,B. and Maerkl,S.J. (2013) iSLIM: a comprehensive approach to mapping and characterizing gene regulatory networks. *Nucleic Acids Res.*, **41**, e52.
33. Nam, Y., Branch,D.W. and Wheeler,B.C. (2006) Epoxy-silane linking of biomolecules is simple and effective for patterning neuronal cultures. *Biosens. Bioelectron.*, **22**, 589–597.
34. Thorsen,T., Maerkl,S.J. and Quake,S.R. (2002) Microfluidic large-scale integration. *Science*, **298**, 580–584.
35. Leaver-Fay,A., Tyka,M., Lewis,S.M., Lange,O.F., Thompson,J., Jacak,R., Kaufman,K., Renfrew,P.D., Smith,C.A., Sheffler,W. *et al.* (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.*, **487**, 545–574.
36. Elrod-Erickson,M., Benson,T.E. and Pabo,C.O. (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure*, **6**, 451–464.
37. Ashworth,J., Taylor,G.K., Havranek,J.J., Quadri,S.A., Stoddard,B.L. and Baker,D. (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.*, **38**, 5601–5608.
38. Fleishman,S.J., Leaver-Fay,A., Corn,J.E., Strauch,E.-M., Khare,S.D., Koga,N., Ashworth,J., Murphy,P., Richter,F., Lemmon,G. *et al.* (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One*, **6**, e20161.
39. Lundqvist,M., Edfors,F., Sivertsson,Å., Hallström,B.M., Hudson,E.P., Tegel,H., Holmberg,A., Uhlén,M. and Rockberg,J. (2015) Solid-phase cloning for high-throughput assembly of single and multiple DNA parts. *Nucleic Acids Res.*, **43**, e49.
40. Struthers,M.D., Cheng,R.P. and Imperiali,B. (1996) Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science*, **271**, 342–345.
41. Vaquerizas,J.M., Kummerfeld,S.K., Teichmann,S.A. and Luscombe,N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.