PNAS
www.pnas.org

Supporting Information

Primordial emergence of a nucleic acid binding protein via phase separation and statistical ornithine-to-arginine conversion

Liam M. Longo[a,1], Dragana Despotović[a,1], Orit Weil-Ktorza[b,1], Matthew J. Walker[c], Jagoda Jabłońska[a], Yael Fridmann-Sirkis[d], Gabriele Varani[c], Norman Metanis[b,2], Dan S. Tawfik[a,2]

[a]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel;

[b]Institute of Chemistry, Hebrew University of Jerusalem, Jerusalem 9112001, Israel;

[c]Department of Chemistry, University of Washington, Seattle, WA 98195;

[d]Life Sciences Core Facility, Weizmann Institute of Science, Rehovot 7610001, Israel

[1]L.M.L., D.D., and O.W.-K. contributed equally to this work.
[2]To whom correspondence may be addressed. Email: metanis@mail.huji.ac.il or dan.tawfik@weizmann.ac.il

This PDF file includes:

Tables S1 to S5

Figures S1 to S12

Methods

Other supplementary materials for this manuscript include the following:

Dataset 1

Supporting Tables

Type	Experiment(s)	Length	Forward Sequence
dsDNA	SPR, ELISA	101 bp	5'biotin-CCGTCCGTAATCATGGTCATAGCTGTTTCGT TTAAAATGAAGATACGGCGCGATGATACGCGTCGG GTTGTCTCTCTGTTGATACAGAGATACTAGATGTA-3'
dsDNA	SPR, ELISA	29 bp	5'biotin-CCGTCCGTAATCATGGTCATAGCTGTTTC-3'
dsDNA	CD titration	29 bp	5'-CCGTCCGTAATCATGGTCATAGCTGTTTC-3'
dsDNA	NMR titration	12 bp	5'-TAGATCGATCGC-3'
ssDNA	ELISA	29 bases	5'biotin-CCGTCCGTAATCATGGTCATAGCTGTTTC-3'

Table S1: DNA sequences used for the binding assays.

Construct Name	Sequence	ds/ssDNA Binding	Design Notes
HhH-1 Duplicated	M**RSKRTLRSELDDIPGIGPKTAKAL LKHF**GSVE**RSKRTLRSELDDIPGIGP KTAKALLKHF**LEHHHHHH	-/nd	Duplication of the first HhH subdomain of Ancestor-$(HhH)_2$.
HhH-2 Duplicated **Symmetric-$(HhH)_2$**	M**KIKKASVEELTEVPGIGPKLAKKI YEHF**GSVE**KIKKASVEELTEVPGIGP KLAKKIYEHF**LEHHHHHH	++/-	Duplication of the second HhH subdomain of Ancestor-$(HhH)_2$.
Sym100_short	M**TSVEELTEVPGIGPKTAKKILKHF** GSVEKIKK**TSVEELTEVPGIGPKTAK KILKHF**LEHHHHHH	++/-	Most probable amino acid at each position after summing the posterior probabilities of the aligned sites from the ancestral sequence reconstruction. The 4-residue N-terminal poly-basic tail was also truncated.
Ancestor of $(HhH)_2$ Halves	M**NIKKASLEELAKVPGIGPKTAKKI YDYL**GSVE**NIKKASLEELAKVPGIG PKTAKKIYDYL**LEHHHHHH	-/nd	Ancestor of the HhH subdomains derived from an alignment of the HhH subdomain sequences.

Table S2. Symmetrization of Ancestor-$(HhH)_2$. For all constructs summarized in the table, the UvrC linker (underlined) was used to connect two identical HhH subdomains (bold). Note that for chemically synthesized proteins, ASVE was used as the linker sequence instead of GSVE. Binding to DNA was assessed by ELISA (see **Supporting Methods**) and is classified as strong (+++), moderate (++), weak (+), undetectable (-), or not determined (nd).

Construct Name	Soluble Expression Relative to Symmetric-$(HhH)_2$	ds/ssDNA Binding	Design Notes
Symmetric-$(HhH)_2$ G29A	Same	++/nd	Linker mutation to simplify total protein synthesis.
Symmetric-$(HhH)_2$ F28L/F60L	Increased	+/nd	Phe replaced with Leu, which is the second most probable amino acid at position 60 in the inferred ancestor.
Symmetric-$(HhH)_2$ H27A/H59A	Decreased	-/nd	Substitution of His with the simplest residue with high helical propensity.
Symmetric-$(HhH)_2$ H27S/H59S	Decreased	-/nd	Substitution of His with the simplest polar residue.
Symmetric-$(HhH)_2$ H27E/F28L/H59E/F60L	Same	-/nd	Substitution of His with Glu, the second most common residue at position 59 in the consensus sequence.
Symmetric-$(HhH)_2$ H27K/H59K	Same	++/-	Substitution of His with another basic amino acid.
Symmetric-$(HhH)_2$ H27R/H59R	Same	++/nd	Substitution of His with another basic amino acid.
Symmetric-$(HhH)_2$ K1R/K3R/K4R/K33R/K35R/K36R	Decreased	++/nd	Six of the 12 Lys residues converted to Arg.
Symmetric-$(HhH)_2$ K1R/K3R/K4R/K22R/K23R/K33R/K35R/K36R/K54R/K55R	Decreased	+++/nd	Ten of the 12 Lys residues converted to Arg.
Symmetric-$(HhH)_2$ K1R/K3R/K4R/K19R/K22R/K23R/K33R/K35R/K36R/K51R/K54R/K55R [Symmetric-$(HhH)_2$-Arg]	Decreased	+++/-	Complete exchange of Lys to Arg.
Symmetric-$(HhH)_2$-Arg V13I/Y25L/H27R/F28L/G29A/V45I/Y57L/H59R/F60L [**Primordial-$(HhH)_2$-Arg Version 1**]	Decreased	+/-	Primordial variant with alternative core packing (V13I and V45I). Combines the above simplification mutations, as well as exchange of Tyr to Leu, which is the most probable amino acid at position 25 in the inferred ancestor.
Primordial-$(HhH)_2$-Arg Version 1 I13V/I45V	Decreased	+/-	Reversion of V13I and V45I mutations
Symmetric-$(HhH)_2$-Arg Version 1 I13V/V31I/I45V [**Primordial-$(HhH)_2$-Arg**]	Decreased	+/-	Primordial-$(HhH)_2$-Arg Version 1 plus linker optimization, to give the final primordial design (sequence listed in **Fig. 1**).

Table S3: Alphabet simplification of Symmetric-$(HhH)_2$. Starting with Symmetric-$(HhH)_2$ (**Table S2**), the non-abiotic amino acids (Phe, His, Tyr, and Lys) were systematically replaced with abiotic amino acids. Mutations were made simultaneously at both symmetry-related sites. Numbering is based on the synthetic constructs (*i.e.*, without methionine as the first residue). Binding to DNA was assessed by ELISA (see **Supporting Methods**) and is classified as strong (+++), moderate (++), weak (+), undetectable (-), or not determined (nd).

Construct Name	Soluble Expression Relative to Parent	dsDNA Binding	+*Design Notes
Symmetric-$(HhH)_2$ I16K/I48K	Increased	-	Aliphatic to charged mutation to deform the binding loop.
Symmetric-$(HhH)_2$ G17E/G49E	Increased	-	Mutation of conserved Gly to a larger acidic amino acid to block the binding surface.
Symmetric-$(HhH)_2$ G17Q/G49Q	Increased	-	Mutation of conserved Gly to a larger polar amino acid to block the binding surface.
Primordial-$(HhH)_2$-Arg I16K/I48K	No expression	nd	Aliphatic to charged mutation to deform the binding loop.
Primordial-$(HhH)_2$-Arg G17E/G49E	No expression	nd	Mutation of conserved Gly to a larger charged amino acid to block the binding surface.
Primordial-$(HhH)_2$-Arg G17Q/G49Q	No expression	nd	Mutation of conserved Gly to a larger polar amino acid to block the binding surface.

Table S4: HhH motif inactivation mutants. To confirm that dsDNA binding is mediated by the PGIGP binding loop characteristic of the HhH motif, several loop mutations were tested. Numbering is based on the synthetic constructs (*i.e.*, without methionine as the first residue). Binding to dsDNA was assessed by ELISA (see **Supporting Methods**) and is classified as strong (+++), moderate (++), weak (+), undetectable (-), or not determined (nd). The I-to-K mutations were introduced to exclude the possibility of binding by nonspecific electrostatic interactions. Mutations in the Primordial-$(HhH)_2$-Arg background resulting in complete loss of expression.

Construct Name	Sequence	Design Notes
Precursor-Arg	RIRRASVEELTEVPGIGPRLARRILERLA	All basic residues are Arg (29 residues total length)
Precursor-Orn	OIOOASVEELTEVPGIGPOLAOOILEOLA	All basic residues are Orn (29 residues total length)
Precursor-Arg, Scrambled 1	GEPAIRSEPIRLRLEGVRALRTEVGLRIRA	All positions of Precursor-Arg scrambled (30 residues).
Precursor-Arg, Scrambled 2	GRVRRSEIPLGIELEAGTLRLARRPEIRVA	All positions except Arg in Precursor-Arg scrambled (30 residues).

Table S5: Synthetic polypeptides. For the scrambled variants, an N-terminal Gly was added to promote labeling with a fluorescent dye. The amino acid ornithine is denoted by the single-letter code O. All peptides except Precursor-Orn were custom synthesized by Synpeptide Co. Ltd (see **Supporting Methods**).

Supporting Figures

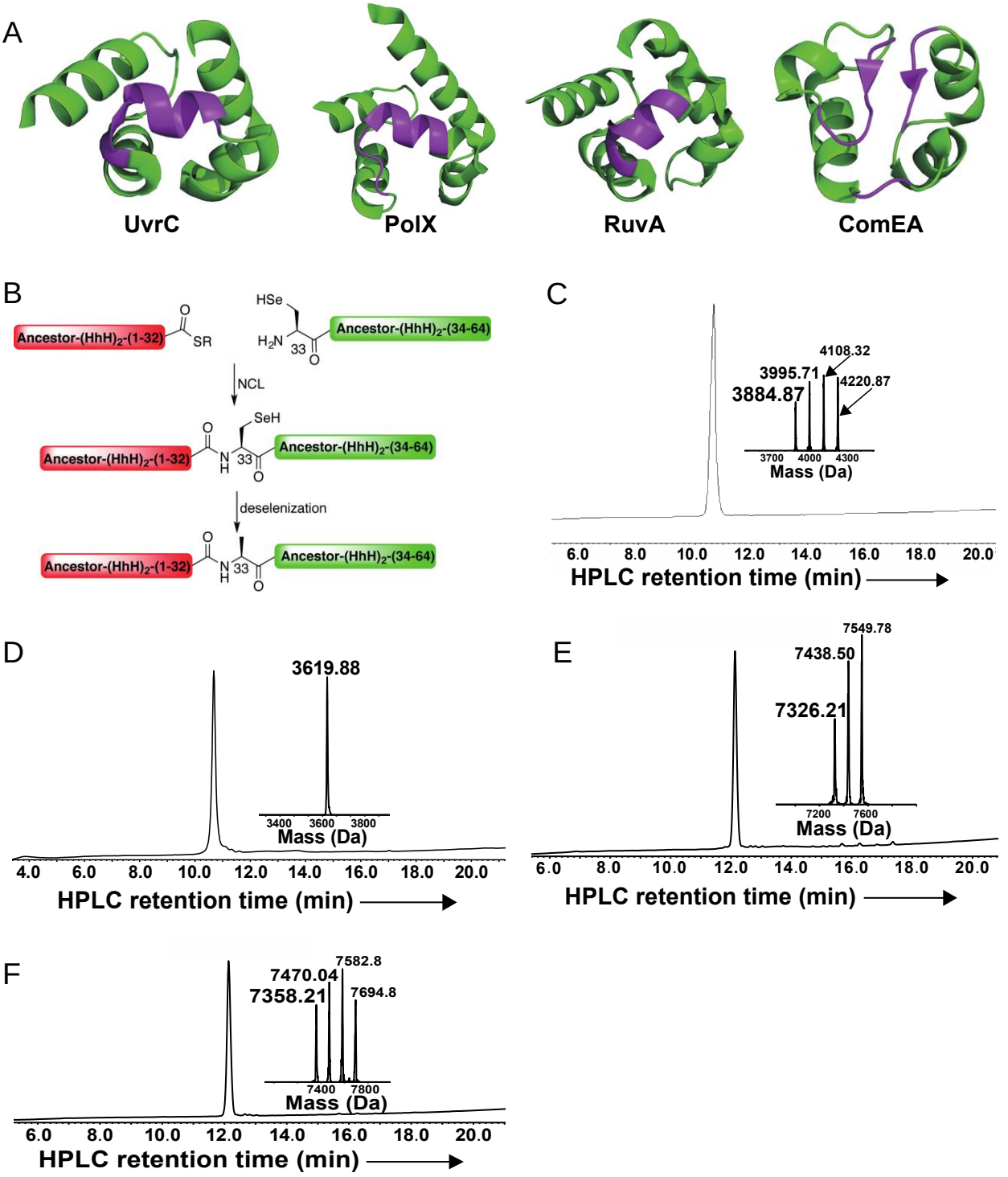

Fig. S1. Reconstruction and synthesis of Ancestor-(HhH)$_2$. A. The different linker regions of the contemporary (HhH)$_2$ families. Three out of these four families, UvrC, PolX, and RuvA, use a short α-helix to link the two HhH subdomains and cap the hydrophobic core (PDB accession codes 2nrt, 3b0x, and 2ztc, respectively). ComEA (PDB accession code 2duy), however, caps the

hydrophobic core with a short intervening loop that connects the HhH subdomains and N- and C-terminal extensions that adopt a short β-sheet-like conformation. The linker for Ancestor-$(HhH)_2$ was chosen to be UvrC-like because of its dominance among the contemporary families and its simplicity. Additionally, using the UvrC linker increased the internal sequence similarity of the inferred ancestor, thereby promoting further simplification (see main text for more details). **B**. Chemical protein synthesis scheme for Ancestor-$(HhH)_2$. Because the protein was too long for a single solid phase synthesis reaction, two half-peptides were synthesized and then joined using a native chemical ligation (NCL) and a deselenization approach. The N-terminal half-peptide bears the C-terminal thioester surrogate *N*-acyl urea, Nbz (1) moiety (NHalf-COSR) and the C-terminal peptide bears an N-terminal selenocysteine (Sec, U) residue (Sec-CHalf). After peptide ligation was complete, the selenocysteine residue was deselenized to yield alanine. **C**. HPLC chromatogram and mass spectrum demonstrating successful synthesis of Ancestor-$(HhH)_2$-NHalf-COSR (M_{calc} = 3885.56 Da, M_{obs} = 3884.87±0.74 Da, $[M+TFA]_{calc.}$ = 3997.56, $[M+TFA]_{obs}$ = 3995.71±0.64 Da, and other TFA salts). **D**. HPLC chromatogram and mass spectrum demonstrating successful synthesis of fragment Ancestor-$(HhH)_2$-Sec-CHalf (M_{calc} = 3619.13 Da, M_{obs} = 3619.88±0.99 Da). **E**. HPLC chromatogram and mass spectrum demonstrating successful NCL of the half peptides to yield Ancestor-$(HhH)_2$-(A33U), in which residue 33 is selenocysteine (M_{calc} = 7327.52 Da, M_{obs} = 7326.21±1.54 Da, $[M+TFA]_{calc.}$ = 7439.53 Da, $[M+TFA]_{obs}$ = 7438.50±1.88 Da). **F**. HPLC chromatogram and mass spectrum demonstrating successful deselenization of Ancestor-$(HhH)_2$-(A33U) to yield Ancestor-$(HhH)_2$ ($[M+TFA]_{calc}$ = 7360.58 Da, $[M+TFA]_{obs}$ = 7358.21±1.60 Da, $[M+2TFA]_{calc}$ = 7472.58 Da, $[M+2TFA]_{obs}$ = 7470.04±2.03 Da, and other TFA salts). Masses were determined with ion trap MS, and the deviations from the theoretical masses are within the error of the measurement.

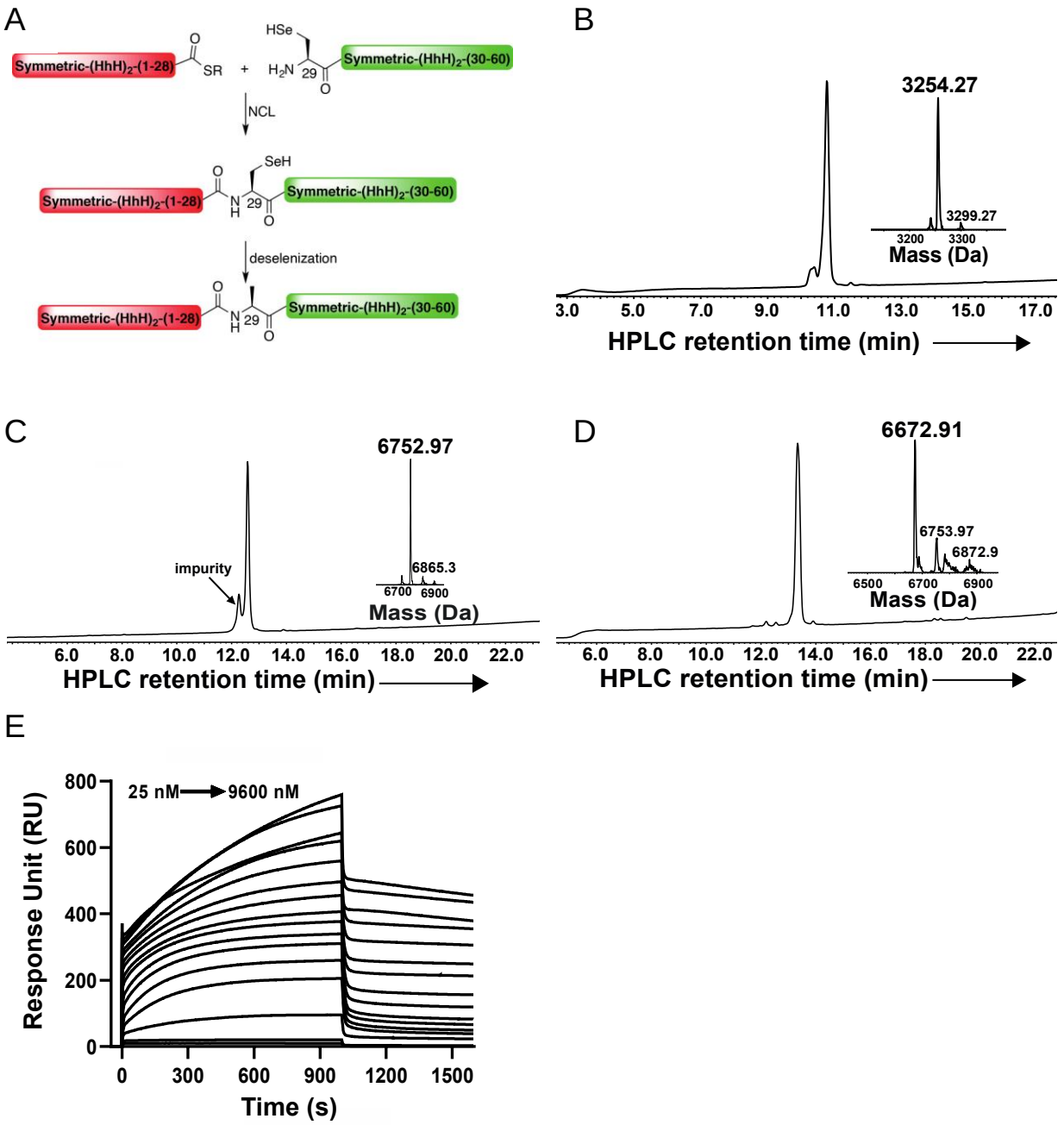

Fig. S2. Synthesis and characterization of Symmetric-(HhH)$_2$. A. Chemical protein synthesis scheme for Symmetric-(HhH)$_2$. Because the protein was too long for a single solid phase synthesis reaction, two half-peptides were synthesized and then joined using an NCL and deselenization approach. The N-terminal half-peptide bears the C-terminal thioester surrogate *N*-acyl urea, Nbz moiety (NHalf-COSR), and the C-terminal peptide bears an N-terminal selenocysteine (Sec, U) residue (Sec-CHalf). After peptide ligation, the selenocysteine residue is deselenized to yield alanine. **B**. HPLC chromatogram and mass spectrum demonstrating successful synthesis of Symmetric-(HhH)$_2$-NHalf-COSR (M_{calc} = 3255.92 Da, M_{obs} = 3254.27±0.46 Da). **C**. Note that Symmetric-(HhH)$_2$-Sec-CHalf is equivalent to Ancestor-(HhH)$_2$-Sec-CHalf (**Fig. S1D**). HPLC

chromatogram and mass spectrum demonstrating successful NCL of the half peptides to yield Symmetric-$(HhH)_2$-(A29U) (M_{calc} = 6754.88 Da, M_{obs} = 6752.97±0.78 Da). **D**. HPLC chromatogram and mass spectrum demonstrating successful deselenization of Symmetric-$(HhH)_2$-(A29U) to yield Symmetric-$(HhH)_2$ (M_{calc} = 6675.92 Da, M_{obs} = 6672.91±0.60 Da). Masses were determined with ion trap MS, and the deviations from the theoretical masses are within the error of the measurement. **E**. Binding of synthesized Symmetric-$(HhH)_2$ to 101 bp dsDNA as measured by SPR (25ºC, 20 μL/min flow rate). As with Ancestor-$(HhH)_2$, the kinetics are biphasic, and 1000 s contact time was insufficient to reach steady state for some concentrations.

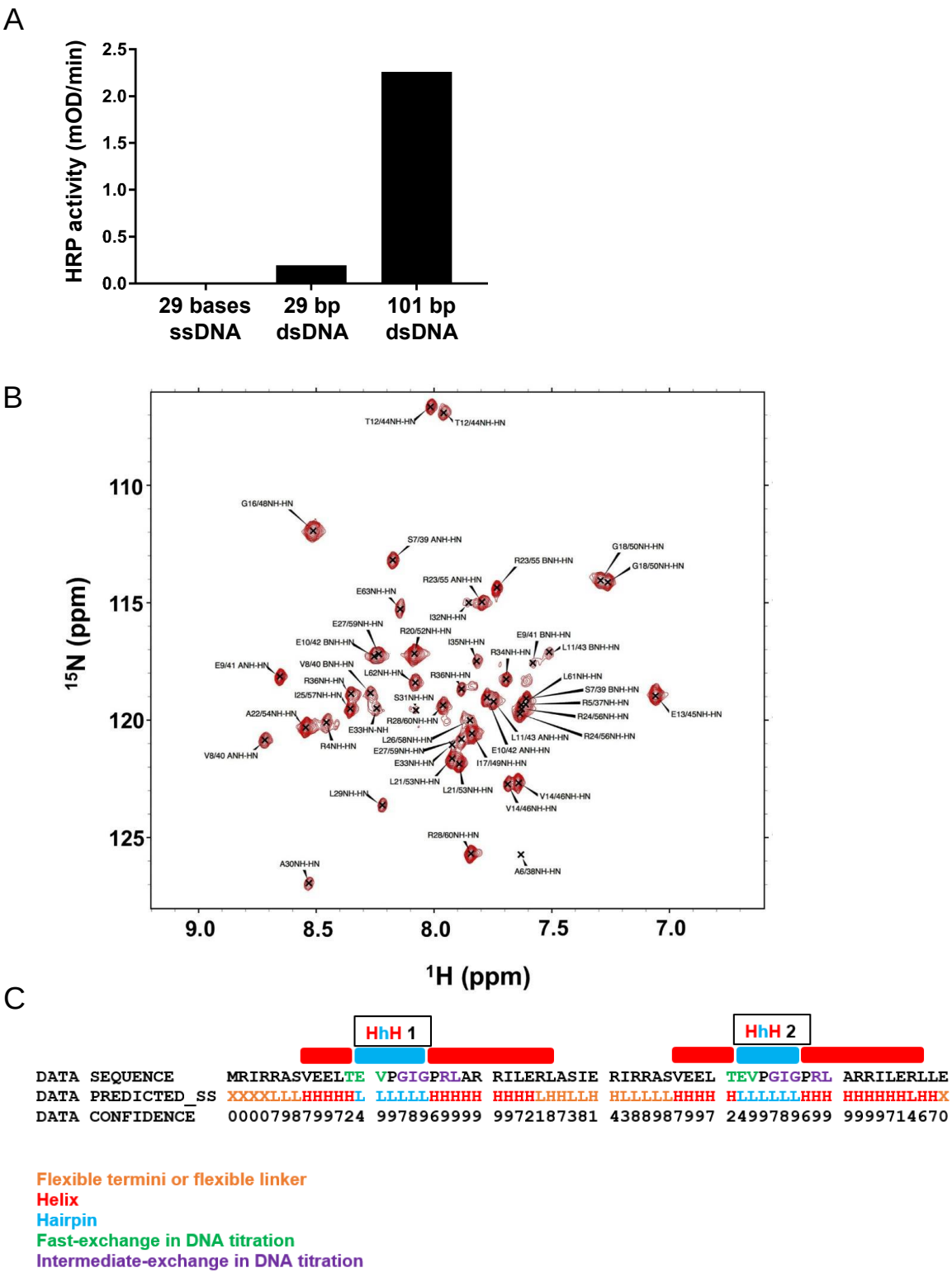

Fig. S3. Characterization of Primordial-(HhH)$_2$-Arg. **A**. Binding of *E. coli* expressed, 6xHis-tagged Primordial-(HhH)$_2$-Arg to various dsDNA fragments as assayed by ELISA. **B**. NMR-HSQC spectrum reporting peak assignments of *E. coli* expressed, tag-free Primordial-(HhH)$_2$-Arg. **C**. TALOS+ neural network secondary structure prediction based on backbone chemical shift

assignments. Secondary structure prediction is missing or low confidence for the linker and the first four N-terminal residues due to broadened spectra from flexible conformations.

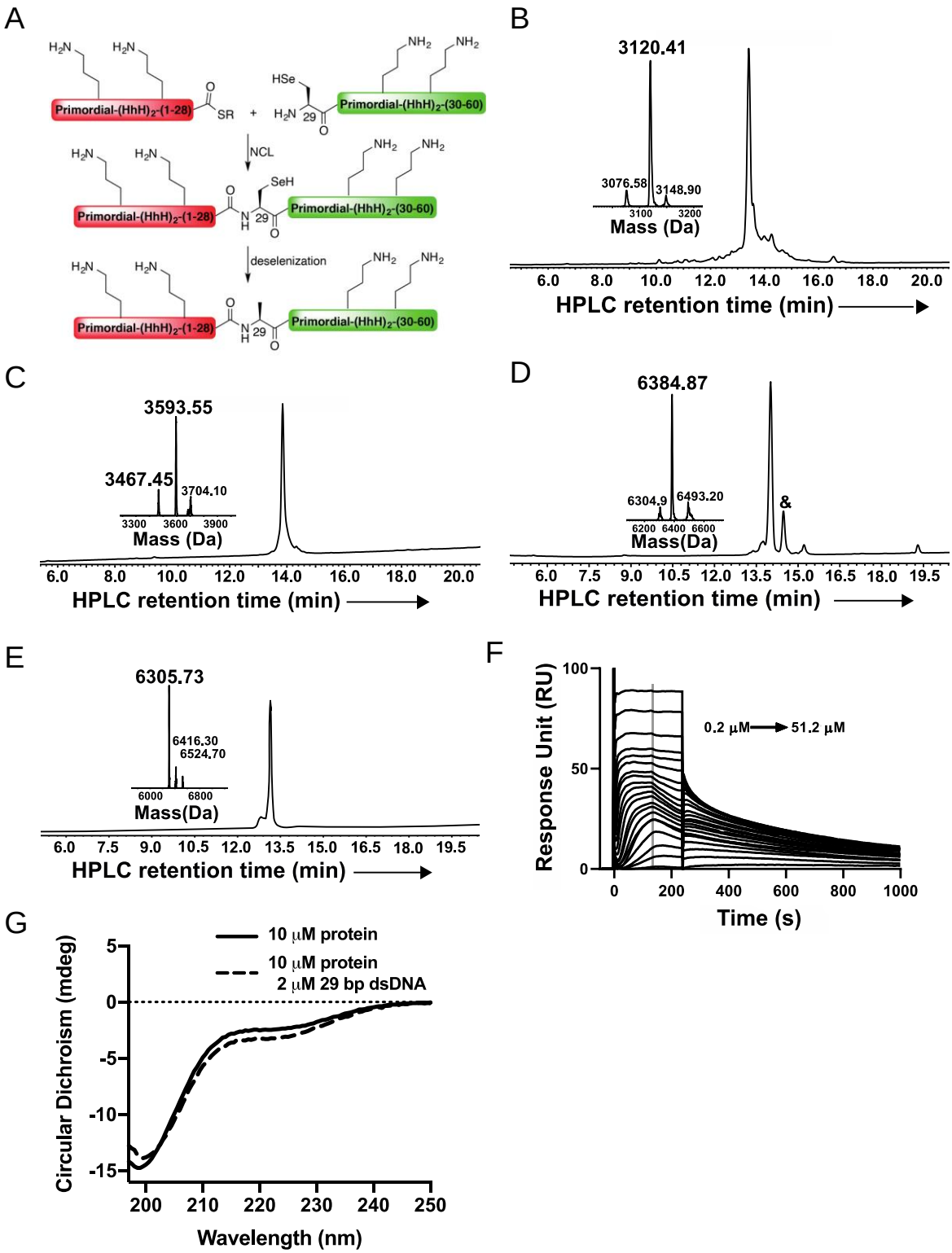

Fig. S4. Synthesis and characterization of Primordial-(HhH)$_2$-Orn. **A**. Chemical protein synthesis scheme for Primordial-(HhH)$_2$-Orn. Because the protein was too long for a single solid phase synthesis reaction, two half-peptides were synthesized and then joined using an NCL and deselenization approach. In this case, the N-terminal half-peptide bears the C-terminal thioester

surrogate Nbz moiety (NHalf-COSR) and the C-terminal peptide bears an N-terminal selenocysteine (Sec, U) residue (Sec-CHalf). After peptide ligation, the selenocysteine residue is deselenized to yield alanine. **B**. HPLC chromatogram and mass spectrum demonstrating successful synthesis of Primordial-$(HhH)_2$-Orn-NHalf-COSR (M_{calc} = 3121.72 Da, M_{obs} = 3120.41±1.91 Da). **C**. HPLC chromatogram and mass spectrum demonstrating successful synthesis of fragment Primordial-$(HhH)_2$-Orn-Sec-CHalf ($[M+Na]_{calc}$ = 3464.96 Da, $[M+Na]_{obs}$ = 3467.45±2.07 Da; $[M+TNP]_{calc}$ = 3596.11 Da, $[M+TNP]_{obs}$ = 3593.55±2.24 Da). **D**. HPLC chromatogram and mass spectrum demonstrating successful NCL of the half peptides to yield Primordial-$(HhH)_2$-Orn-(A29U) (M_{calc} = 6386.53 Da, M_{obs} = 6384.87±0.58 Da). Column impurities are denoted with an ampersand (&). **E**. HPLC chromatogram and mass spectrum demonstrating successful deselenization of Primordial-$(HhH)_2$-Orn-(A29U) to yield Primordial-$(HhH)_2$-Orn (M_{calc} = 6307.57 Da, M_{obs} = 6305.73±0.56 Da). Masses were determined with ion trap MS, and the deviations from the theoretical masses are within the error of the measurement. **F**. Binding of synthesized Primordial-$(HhH)_2$-Orn to 101 bp dsDNA as measured by SPR (25 ºC, 20 μL/min flow rate, 250 s contact time). The grey line denotes the time points taken to generate the binding curves in **Fig. 3C** in the main text. **G**. A small yet consistent increase in the CD signal that reports α-helicity is observed upon mixing Primordial-$(HhH)_2$-Orn with 29 base pair dsDNA, suggesting folding upon binding. This interpretation is bolstered by the more notable increase in α-helicity observed for the statistically guanidinated variants, which bind dsDNA with greater avidity than Primordial-$(HhH)_2$-Orn (**Fig. 3E**). Plotted spectra are background-subtracted to remove the signal associated with dsDNA.

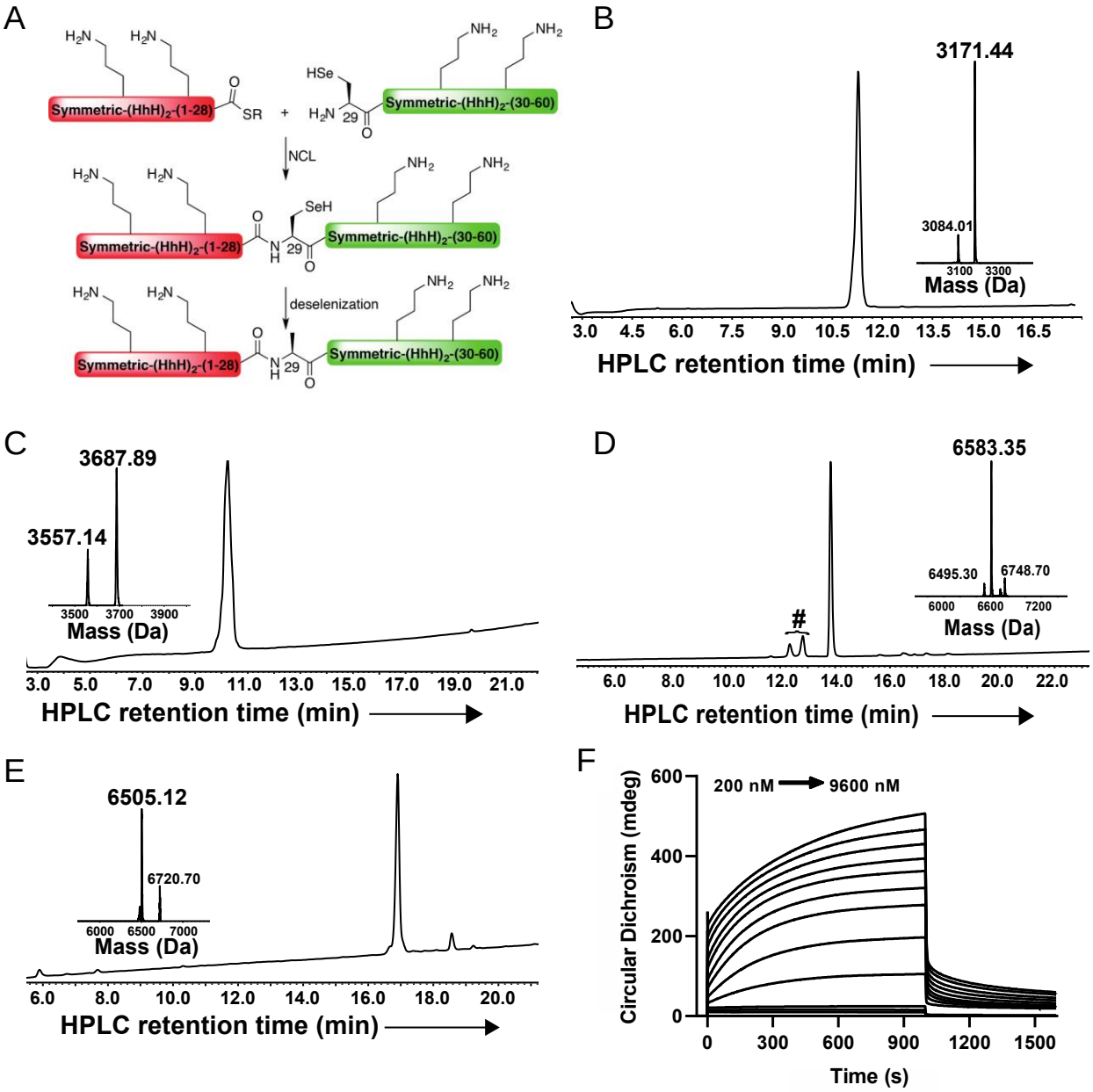

Fig. S5. Synthesis and characterization of Symmetric-(HhH)$_2$-Orn. **A**. Chemical protein synthesis scheme for Symmetric-(HhH)$_2$-Orn. Because the protein was too long for a single solid phase synthesis reaction, two half-peptides were synthesized and then joined using an NCL and deselenization approach. The N-terminal half-peptide bears the C-terminal thioester surrogate Nbz moiety (NHalf-COSR), and the C-terminal peptide bears an N-terminal selenocysteine (Sec, U) residue (Sec-CHalf). After peptide ligation, the selenocysteine residue is deselenized to yield alanine. **B**. HPLC chromatogram and mass spectrum demonstrating successful synthesis of Symmetric-(HhH)$_2$-Orn-NHalf-COSR (M_{calc} = 3171.75 Da, M_{obs} = 3171.44±1.25 Da). **C**. HPLC chromatogram and mass spectrum demonstrating successful synthesis of fragment Symmetric-(HhH)$_2$-Orn-Sec-CHalf ($[M+Na]_{calc}$ = 3557.96 Da, $[M+Na]_{obs}$ = 3557.14±0.56 Da; $[M+TNP]_{calc}$ = 3689.11 Da, $[M+TNP]_{obs}$ = 3687.89±0.84 Da). **D**. HPLC chromatogram and mass spectrum

demonstrating successful NCL of the half peptides to yield Symmetric-$(HhH)_2$-Orn-(A29U), in which position 29 is selenocysteine (M_{calc} = 6586.56 Da, M_{obs} = 6583.35±0.39 Da). Column impurities are denoted with a hash (#). **E**. HPLC chromatogram and mass spectrum demonstrating successful deselenization of Symmetric-$(HhH)_2$-Orn-(A29U) to yield Symmetric-$(HhH)_2$-Orn (M_{calc} = 6507.60 Da, M_{obs} = 6505.12±0.81 Da). Masses were determined with ion trap MS, and the deviations from the theoretical masses are within the error of the measurement. **F**. Binding of synthesized Symmetric-$(HhH)_2$-Orn to 101 bp dsDNA as measured by SPR (25 ºC, 20 µL/min flow rate, 1000 s contact time).

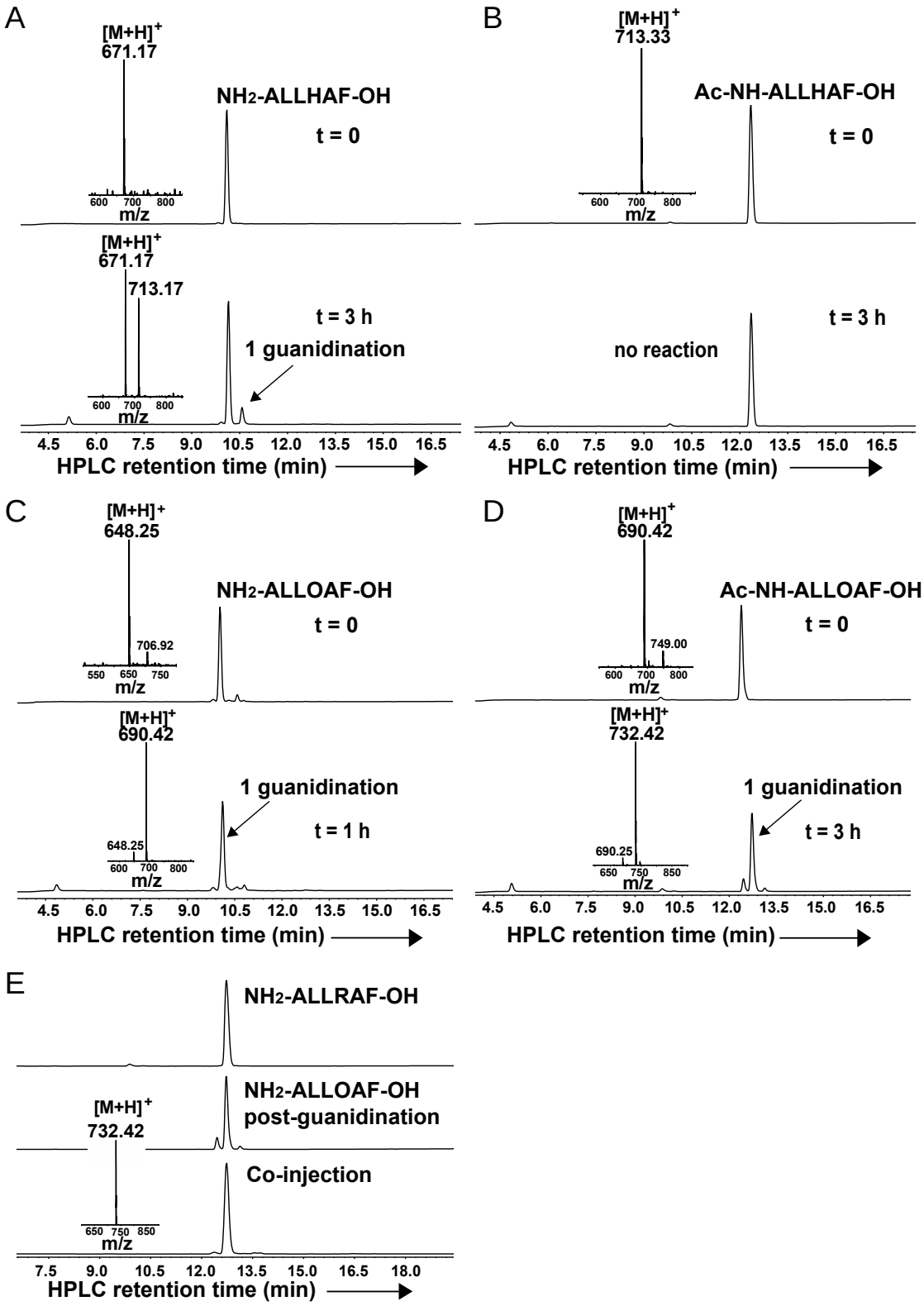

Fig. S6. Assessing the kinetics, efficiency, and specificity of the guanidination reaction. **A**. To determine the reactivity of the N-terminus, and confirm that His residues are not modified by the guanidination reaction, NH_2-ALL**H**AF-OH was synthesized (upper HPLC chromatogram;

$[M+H]^+_{calc}$ = 671.39 Da, $[M+H]^+_{obs}$ = 671.17 Da) and subjected to the guanidination reaction (lower HPLC chromatogram). Slow, single guanidination of the peptide was observed after three hours ($[M+H]^+_{calc}$ = 713.40 Da, $[M+H]^+_{obs}$ = 713.17 Da). **B**. No guanidination of Ac-NH-ALL**H**AF-OH was observed after 3 hours, as indicated by analytical HPLC and ESI-MS analysis ($[M+H]^+_{calc}$ = 713.40 Da, $[M+H]^+_{obs}$ = 713.33 Da), indicating that the slow, single guanidination of the peptide described in Panel A occurs at the amino-terminus. **C**. Synthesis of NH_2-ALL**O**AF-OH (upper HPLC chromatogram; $[M+H]^+_{calc}$ = 648.41 Da, $[M+H]^+_{obs}$ = 648.25 Da). Near complete guanidination was achieved within one hour, resulting in a single major peak ($[M+H]^+_{calc}$ = 690.43 Da, $[M+H]^+_{obs}$ = 690.42 Da). Preferential guanidination of ornithine relative to the N-terminus is likely due to steric effects (2–4). **D**. Synthesis of Ac-NH-ALL**O**AF-OH (upper HPLC chromatogram; $[M+H]^+_{calc}$ = 690.43 Da, $[M+H]^+_{obs}$ = 690.42 Da). A single major peak corresponding to the conversion of ornithine to arginine was observed (lower HPLC chromatogram; $[M+H]^+_{calc}$ = 732.44 Da, $[M+H]^+_{obs}$ = 732.42 Da). **E**. The peptide NH_2-ALL**R**AF-OH (*i.e.*, the expected guanidination product of NH_2-ALL**O**AF-OH) was synthesized (upper HPLC chromatogram), and showed virtually identical retention time as the guanidination product of NH_2-ALL**O**AF-OH (middle HPLC chromatogram). When the two samples were mixed and co-injected, the resulting HPLC chromatogram contained a single major peak, the mass of which matched that of NH_2-ALL**R**AF-OH ($[M+H]^+_{calc}$ = 732.44 Da, $[M+H]^+_{obs}$ = 732.42 Da), indicative of efficient and selective conversion to arginine. Analytical HPLC was performed on an XSelect C18 column. For short peptides only the Z = 1 was observed, no error for calculations.

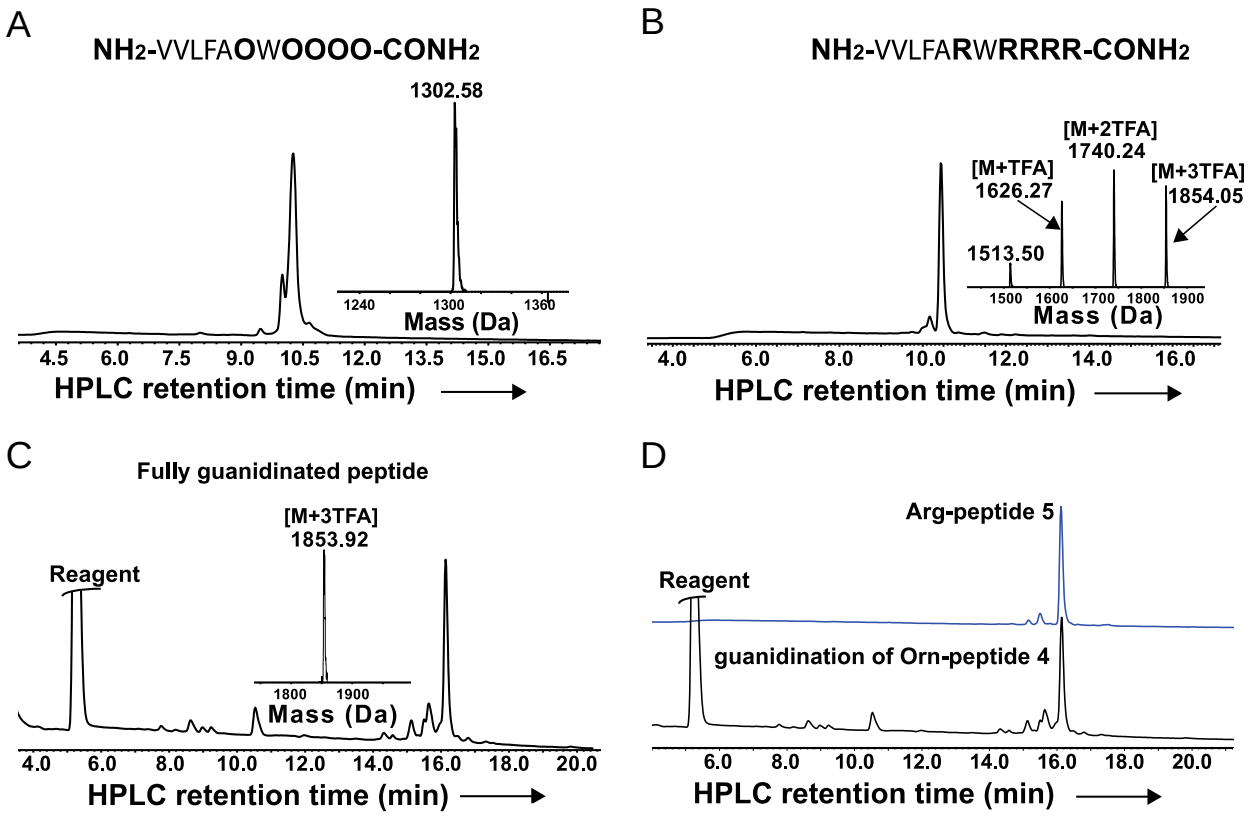

Fig. S7. Optimizing the guanidination of polypeptides with multiple ornithine residues. **A**. To confirm that guanidination of adjacent ornithine residues can be achieved, NH_2-VVLFA**O**W**OOOO**-OH was synthesized (M_{calc} = 1302.83 Da, M_{obs} = 1302.58±0.22 Da). The small shoulder peak is the result of incomplete coupling between ornithine and tryptophan. **B**. The corresponding arginine containing peptide, NH_2-VVLFA**R**W**RRRR**-OH, was synthesized for comparison (M_{calc} = 1513.87 Da, M_{obs} = 1513.50±0.01 Da, as well as TFA salts). **C**. After 2 hours of reaction, the ornithine peptide (NH_2-VVLFA**O**W**OOOO**-OH) showed complete guanidination, with all of its 5 ornithine residues converted to arginine ($[M+3TFA]_{calc}$ = 1852.92 Da, $[M+3TFA]_{obs}$ = 1853.92±0.11). **D**. The post-guanidination ornithine peptide and the synthetic arginine peptide have the same retention times on an HPLC XSelect C18 column. Masses were determined with ion trap MS, and the deviations from the theoretical masses are within the error of the measurement.

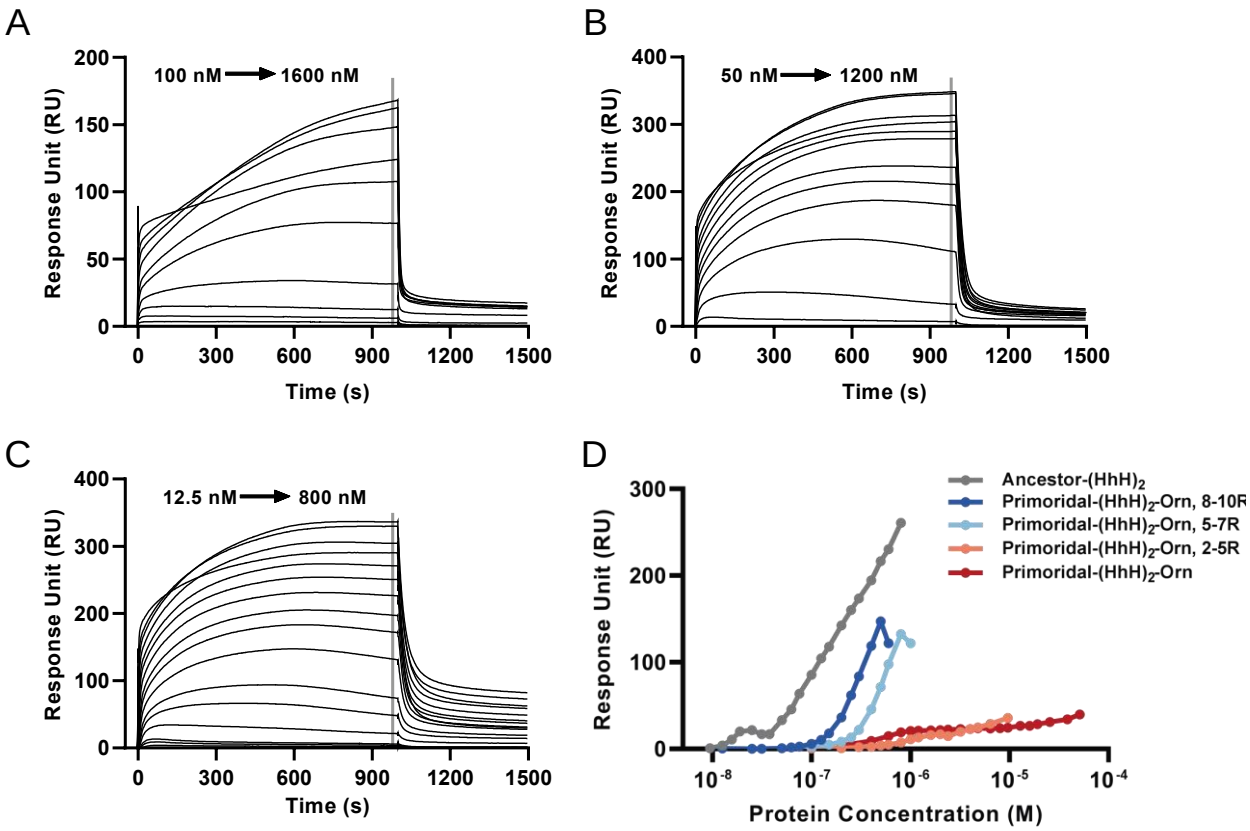

Fig. S8. Binding of guanidinated variants of Primordial-(HhH)$_2$ to dsDNA as measured by SPR. **A**. Sensograms of binding of Primordial-(HhH)$_2$, 2-5R to 101 base pair dsDNA. **B**. Binding of Primordial-(HhH)$_2$, 5-7R to 101 base pair dsDNA. **C**. Binding of Primordial-(HhH)$_2$, 8-10R to 101 base pair dsDNA. As with other constructs, the association kinetics are biphasic, and in some runs (**Panels A-C**) steady state was not reached within 1000 s (see legend of **Fig. 1**). The grey lines denote the time points taken to generate the steady-state binding curve in **Fig. 3C**). **D**. Steady-state binding of guanidinated variants of Primordial-(HhH)$_2$ to 29 base pair dsDNA as measured by SPR (raw sensograms not shown). SPR experiments were performed at 25 ºC with a 20 µL/min flow rate and 1000 s contact time, except for Primordial-(HhH)$_2$-Orn, which required a contact time of 250 s.

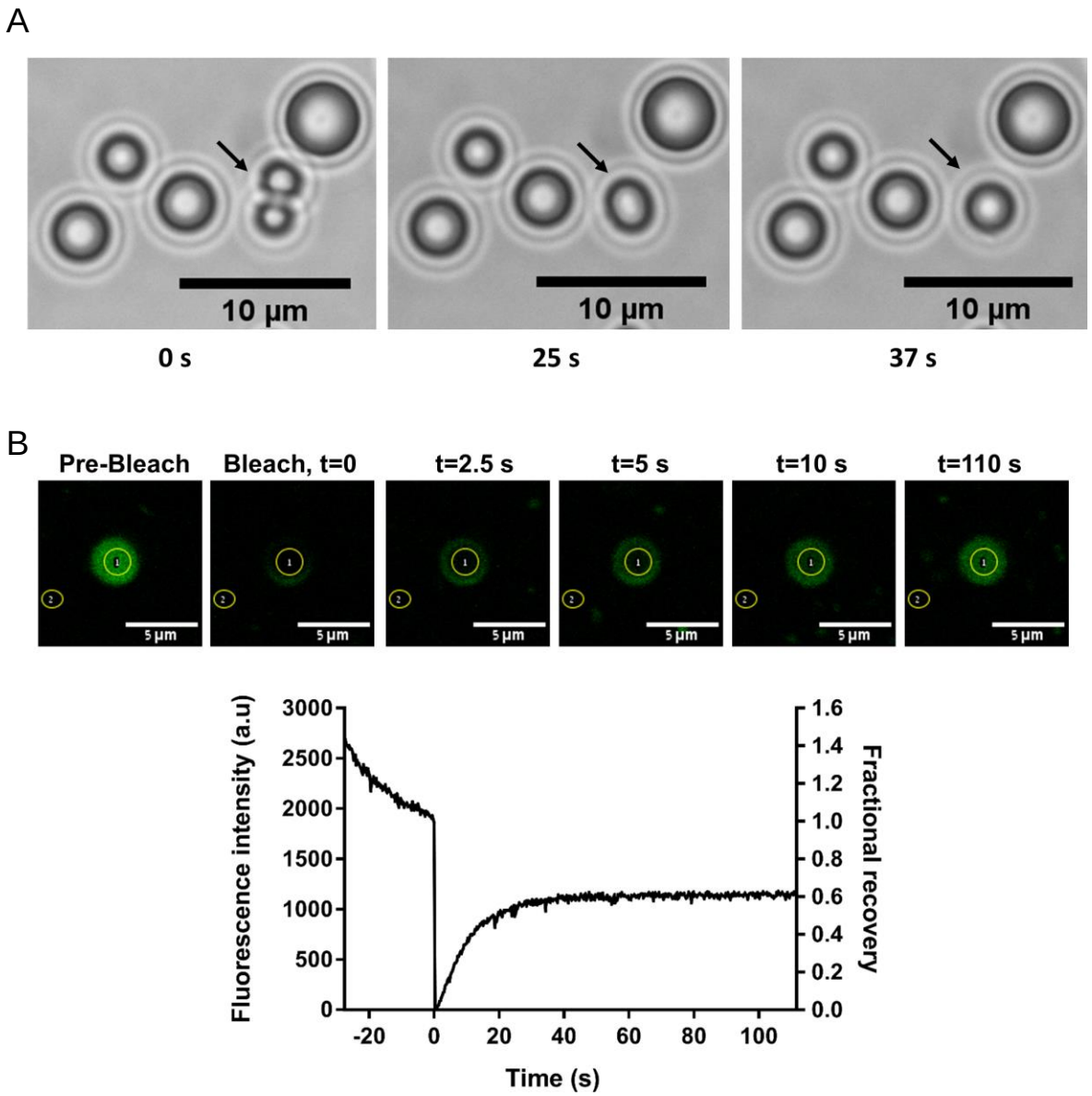

Fig. S9. Phase separation of Precursor-Arg with polyU. **A**. Coalescence of two droplets (indicated by arrow) occurred over short time scales, confirming their fluid nature, as observed by transmission microscopy using a 100x oil immersion objective. Reaction conditions are the same as in **Fig. 4** in the main text. **B**. FRAP analysis of a fluorescent droplet using an inverted confocal microscope (Olympus IX81) with 60x oil immersion objective. Bleaching was performed inside the droplet, and recovery was followed over time. Fluorescence intensity was background normalized by subtracting the signal outside the droplet (area 2) from the signal inside the droplet (area 1). Fractional recovery was normalized by the signal immediately prior to bleaching. The reaction mixture contained 1.0 mg/mL polyU and 208 µM Precursor-Arg (200 µM unlabeled protein plus 8 µM fluorescein-labeled protein).

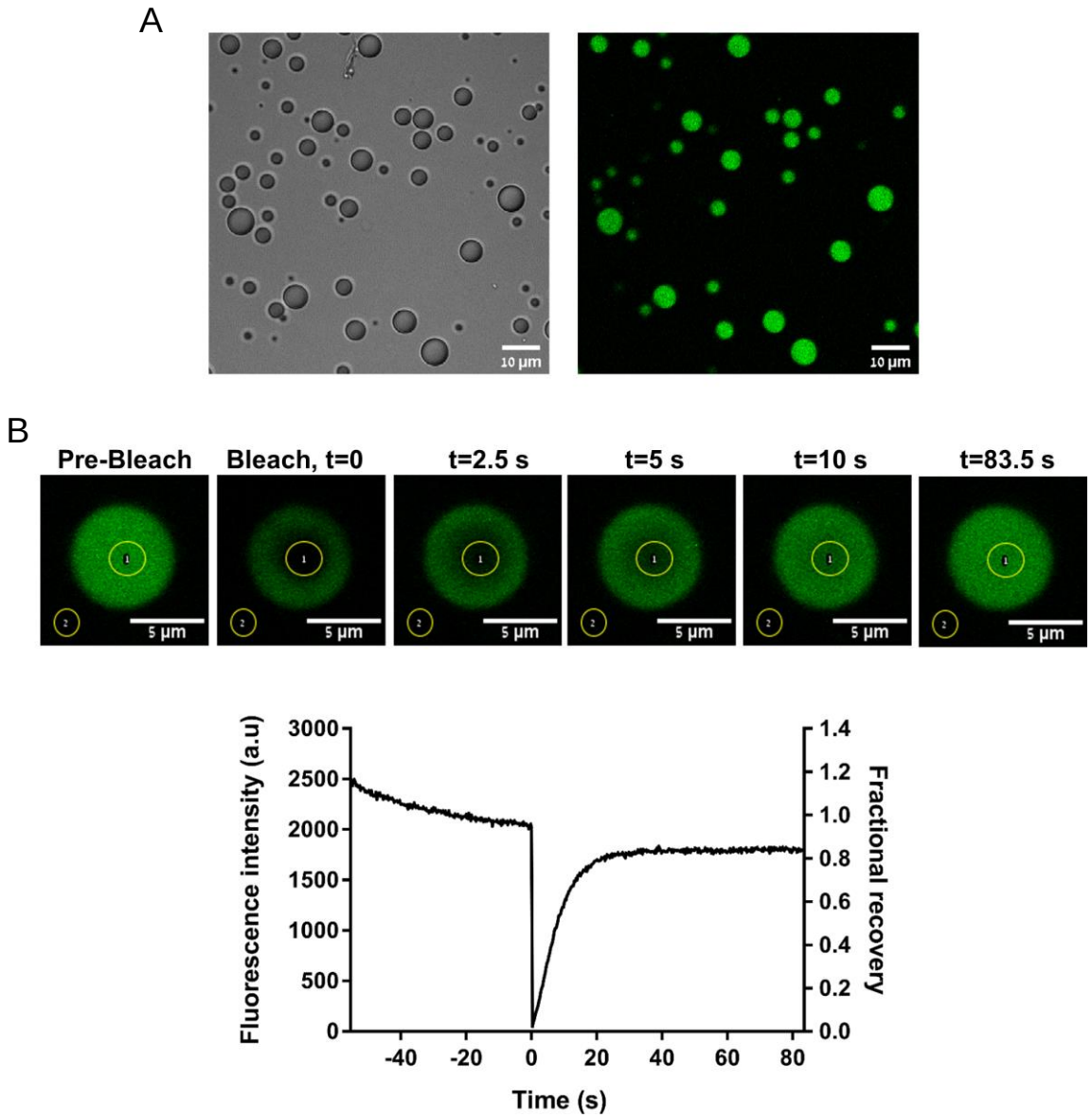

Fig. S10. Phase separation of Ancestor-(HhH)$_2$. **A**. Ancestor-(HhH)$_2$ (200 µM unlabeled protein plus 4 µM fluorescein labeled protein) and polyU (1 mg/mL) form droplets upon mixing, as observed by transmission and fluorescence microscopy. **B**. FRAP analysis of a fluorescent droplet on an inverted confocal microscope with a 60x oil immersion objective. Bleaching was performed inside the droplet and recovery was followed over time. Fluorescence intensity was background normalized by subtracting the signal outside the droplet (area 2) from the signal inside the droplet (area 1). Fractional recovery was normalized by the signal immediately prior to bleaching. Images were taken using inverted confocal microscope (Olympus IX81) with a 60x oil immersion objective.

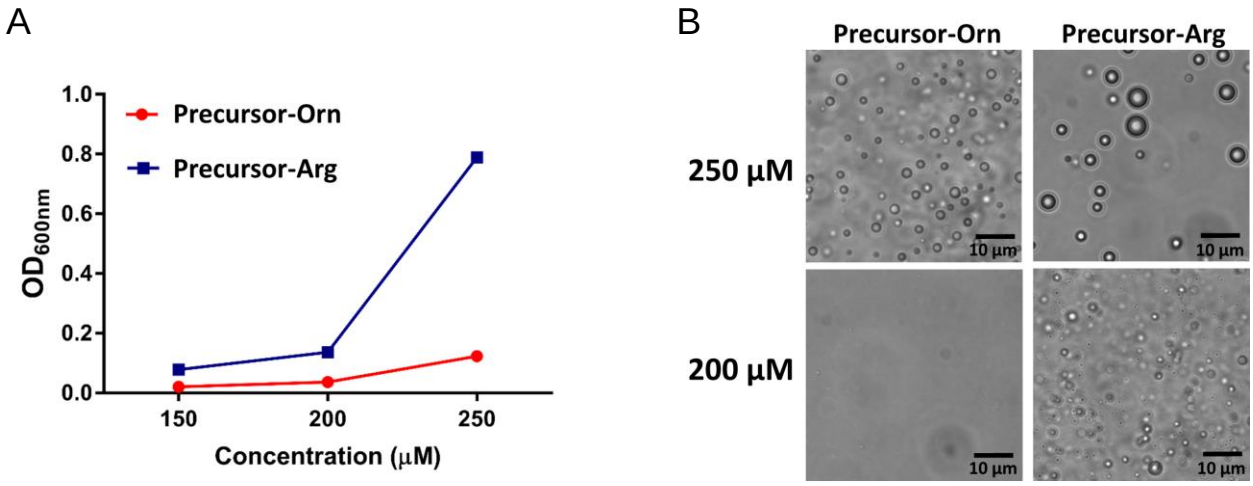

Fig. S11. **Precursor-Arg forms larger droplets at lower concentrations than Precursor-Orn**. **A**. The dependence of phase separation on polypeptide concentration in the presence of 1 mg/mL polyU. The extent of phase separation was assessed by turbidity measurements at 600 nm after 5 minutes of incubation at room temperature. **B**. Transmission micrographs of coacervates formed by 200 μM or 250 μM polypeptide and 1 mg/ml polyU.

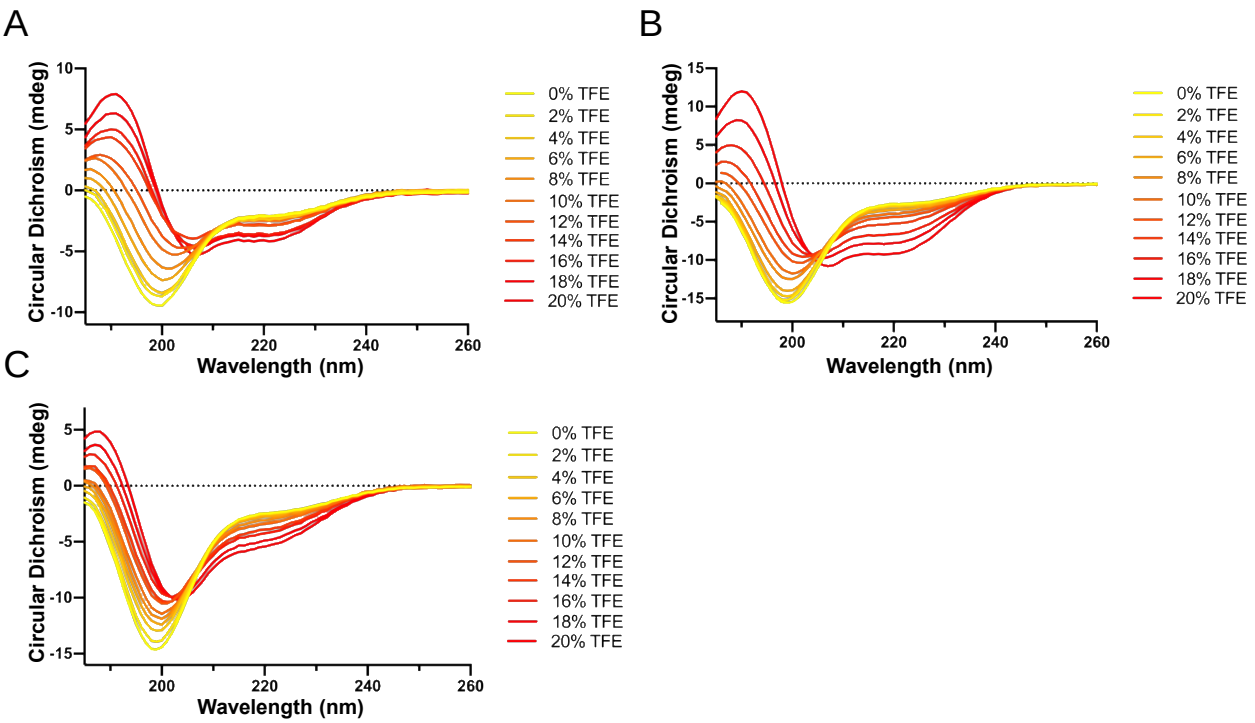

Fig. S12. Circular dichroism spectra from trifluoroethanol (TFE) titrations. **A**. Precursor-Arg spectra. **B**. Scrambled 1 spectra. **C**. Scrambled 2 spectra. Each curve represents the average of two spectra after correction for dilution due to titration with TFE and subtraction of the buffer signal.

Supporting Methods

Protein expression and purification. Synthetic genes were ordered from Twist Bioscience (www.twistbioscience.com) and cloned into a pET21 vector, yielding variants with a C-terminal 6xHis tag. Mutant constructs were generated following Wang and Malcolm (5). All constructs were verified by Sanger sequencing. Transformed BL21(DE3) cells were induced at OD600 = ~0.6 with 1 mM Isopropyl β-D-1-thiogalactopyranoside. Induced cells were shake-incubated at 20 °C overnight. Cell pellets were collected by centrifugation and frozen at -20 °C for storage. The purification protocol was developed throughout this study as more variants were expressed and tested; the optimized protocol is as follows: Cell pellets from 250 mL cultures were resuspended in 30 mL of 100 mM NaCl, 50 mM Tris/HCl pH 7.5, 2.5 U/ml Benzonase (Merck Millipore), 0.2 U/ml rDNase1 (Thermo Fisher Scientific), 0.25x Protease Inhibitor Cocktail for Histidine-tagged proteins (Sigma-Aldrich), and 0.3 mg/mL lysozyme and shake-incubated at 37 °C for 90 minutes. The lysates were then cooled on ice for 15 minutes and sonicated at 30 % power for 5 minutes, with a 30 s cool-down period after every 30 s of sonication. After cell lysis, the samples were spiked with 1 M NaCl and 5 mM imidazole and gently rocked for 15 min at room temperature, as high concentrations of NaCl promoted the solubility of some of the $(HhH)_2$ proteins. Pellets were clarified by centrifugation at 10,000 RCF for 1 hour and passed through a 0.45 µm sterile filter. Clarified lysate was applied to 3 mL packed NiNTA resin pre-equilibrated in 5 mM imidazole, 1 M NaCl, 50 mM Tris/HCl, pH 7.5 ("purification buffer"). After sample application, the resin was washed with 25 mL of purification buffer. Some $(HhH)_2$ constructs were copurified with dsDNA; to remove it, the resin was washed first with 25 mL of purification buffer containing 4M GuHCl, then with 50 mL of purification buffer containing 3 M NaCl. Nonspecifically bound proteins were removed by washing with purification buffer spiked with 25 mM imidazole. The bound $(HhH)_2$

protein was eluted by 5 mL of 850 mM imidazole, 50 mM Tris/HCl, 150 mM NaCl, pH 7.5. Samples were then concentrated to ~0.5 mL and analyzed by SDS-PAGE gel to determine their purity and estimate the protein concentration. The absence of residual dsDNA was confirmed by a Qubit Fluorometer calibrated for dsDNA (ThermoFisher Scientific).

Enzyme-linked immunosorbent assay (ELISA). Binding to DNA was measured by an ELISA assay. DNA sequences used for the binding assay can be found in **Table S1**. The arbitrarily chosen promoter region of the *nrdH* gene (Gene ID: 947161) was amplified from the *E. coli* genome, then subjected to a second round of PCR using a single 5'-biotinylated primer. The 101 base pair PCR product was purified with a QIAquick PCR purification Kit (Qiagen), eluted in water, and the DNA concentration was determined by measuring A_{260nm} (NanoDrop 2000, ThermoScientific). Other dsDNA constructs were ordered as oligonucleotides (Integrated DNA Technologies) and annealed with their complementary oligonucleotide by heating to 95 °C, followed by slow cooling down to room temperature at a rate of 2 °C per min. Streptavidin-coated plates (StreptaWell, Roche) were incubated for 30 min with 100 µl of either 1 µM biotin, 1.5 ng/µl 29 bases of biotinylated ssDNA, 1.5 ng/µl 29 base pair biotinylated dsDNA, or 1.5 ng/µl 101 base pair biotinylated dsDNA (**Table S1**) in binding buffer (50 mM Tris/HCl pH 8.0, 150 mM NaCl). The wells were washed extensively with binding buffer, then blocked with binding buffer spiked with 1 mg/mL BSA. Coated wells were incubated for 45 min with proteins in binding buffer plus 1 mg/mL BSA. The unbound proteins were washed, and 100 µl of 1 µg/ml HRP-labeled mouse anti-His antibody (200 µg/ml, Santa Cruz Biotechnology) was added. Following 45 min incubation, the wells were washed, the substrate 3,3',5,5'-tetramethybenzidine (TMB; ES001, Millipore) was added, and OD_{650nm} was monitored. ELISA binding data are presented as the rate of increase in OD_{650nm} with the background (wells coated with biotin) subtracted (null in nearly all cases).

Materials for total chemical protein synthesis. Buffers were prepared using MilliQ water (Millipore, Merck). Ultrapure guanidinium chloride (Gn·HCl, MP Biomedicals, LLC, France) was used in all ligation reactions. $Na_2HPO_4 \cdot 12H_2O$, tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl), ethanedithiol (EDT), triisopropylsilane (TIPS), *D,L*-dithiothreitol (DTT), 2,2′-Dithiobis(5-nitropyridine) (DTNP) (6), sodium ascorbate and 4-mercaptophenylacetic acid (MPAA) and methyl 3-mercaptopropionate (MMP) were purchased from Sigma-Aldrich (Rehovot, Israel). 1-guanyl-3,5-dimethyl pyrazole nitrate (the guanidination reagent) was purchased from Chem-Impex. All Fmoc-amino acids were obtained from CS Bio Co. (Menlo Park, CA) or Matrix Innovation (Quebec City, Canada), with the following side chain protecting groups: Arg(Pbf), Glu(OtBu), Gly(OtBu), Ser(tBu), Thr(tBu), Tyr(tBu), Lys(Boc), His(Trt), Trp(Boc). (Pbf = 2,2,4,6,7- pentamethyl-2,3-dihydrobenzofuran-5-sulfonyl). Nα-Fmoc-Nδ-Boc-L-Ornithine (shortly Fmoc-Orn(Boc)-OH, three letter code Orn, and one letter code O) was purchased from Chem-Impex. TentaGel® R RAM resin (loading 0.19 mmol/g), Fmoc-Phe-TentaGel® R PHB resin (loading 0.18 mmol/g) and Fmoc-Leu-TentaGel® were purchased from Rapp Polymer GmbH (Germany). 1-[Bis(dimethylamino)methylen]-5- chlorobenzotriazolium 3-oxide hexafluorophosphate, *N*,*N*,*N*′,*N*′-Tetramethyl-O-(6- chloro-1H-benzotriazol-1-yl)uronium hexafluorophosphate (HCTU) and Ethyl cyano(hydroxyimino)acetate (OxymaPure) were purchased from Luxembourg Biotechnologies Ltd. (Rehovot, Israel). All solvents: *N*,*N*-dimethylformamide (DMF), dichloromethane (DCM), acetonitrile (ACN), *N*,*N*-diisopropylethyl amine (DIEA), Trifluoroacetic acid (TFA) and piperidine (Pip) were purchased from Bio-Lab (Jerusalem, Israel) and were peptide synthesis, HPLC, or ULC-grade. Synthesis of Fmoc-Sec(Mob)-OH (selenocysteine three letter code is Sec, and one letter codes is U) was reported previously (7).

High performance liquid chromatography (HPLC). Analytical reversed-phase (RP) HPLC analyses were performed on a Waters Alliance HPLC with UV detection (220 nm and 280 nm) using a XSelect C18 column (130 Å, 3.5 μm, 4.6 × 150 mm) or XBridge C4 column (300 Å, 3.5 μm, 4.6 × 150 mm). Preparative and semi-preparative RP-HPLC were performed on a Waters 150Q LC system using a XSelect C18 column (130 Å, 5 μm, 30 × 250 mm) or XBridge BEH300 C4 column (300 Å, 5 μm, 19 × 150 mm). Linear gradients of ACN (with 0.1 % TFA, buffer B) in water (with 0.1 % TFA, Buffer A) were used for all systems to elute bound peptides. The flow rates were 1 mL/min (analytical column heated at 30 °C), 10 mL/min (semi-preparative), and 20 mL/min (preparative).

Electrospray ionization mass spectrometry (ESI-MS). ESI-MS was performed on LCQ Fleet Ion Trap mass spectrometer (Thermo Scientific). Peptide masses were calculated from the experimental mass to charge (m/z) ratios from the observed multiply charged species of each peptide. Deconvolution of the experimental MS data was performed with the help of MagTran v1.03 software.

General procedure for Fmoc-solid-phase peptide synthesis (Fmoc-SPPS). Peptides were prepared with an automatic peptide synthesizer (CS136XT, CS Bio Inc. CA), except otherwise noted) typically on 0.25 mmol scales. Fmoc-amino acids (2 mmol in 5 mL DMF) were activated with HCTU (2 mmol in 5 mL DMF) and DIEA (4 mmol in 5 mL DMF) for 5 min, then allowed to couple for 25 min with constant shaking. Fmoc deprotection was carried out with 20 % piperidine in DMF (2 × 5 min). Selenocysteine was coupled using DIC/OxymaPure activation method (7). After synthesis, the peptide-resins were washed with DMF, DCM and dried under vacuum. The dried peptide-resins were deprotected and cleaved using a TFA/water/thioanisole/

triisopropylsilane/ethanedithiol (92.5:1.5:1.5:1.5:1.5) cocktail for 4 h. The cleavage mixtures were filtered and TFA was evaporated with N_2-bubbling to a minimum volume, to which an eightfold volume of cold ether was added dropwise. The precipitated crude peptide was centrifuged (5000 rpm, 10 min), ether was removed, and the crude peptide was dissolved in ACN/water (1:1) containing 0.1 % TFA, further diluted to ca. 25 % ACN with water, and lyophilized.

Preparation of C-terminal thioester peptides. The C-terminal thioester peptides were synthesized first as Fmoc-Dbz-resin (0.25 mmol scale) on automated peptide synthesizer. Mono-Fmoc-3,4-diaminobenzoic acid (Fmoc-Dbz-OH, 3 equiv) (1) activated with HCTU (3 equiv)/DIEA (6 equiv) in DMF was doubly coupled to the free amine of TentaGel® R RAM resin (0.19 mmol/g, 0.25 mmol scale) for 1 h, followed by double coupling of the first amino acid (2 x 1 h), while the N-terminal peptide was protected with Boc-protecting group prior to the Dbz to Nbz conversion (1). After synthesis completion, the resin was washed with DCM and a solution of *p*-nitrophenyl chloroformate (5 equivalents) in DCM was added, shaken for 1 h at 25 °C and washed with DCM (3 × 5 mL) and DMF (3 × 5 mL) (repeated twice). Following this, the resin was washed with DCM and DMF and a solution of 0.5 M DIEA in DMF was added and shaken for additional 30 min to complete the cyclization and Nbz formation (repeated twice), and washed with DMF (3 × 5 mL). Peptide was deprotected and cleaved as described previously to give the crude peptide acylurea derivative, peptide-Nbz. The crude peptide-Nbz (100 mg, ~3 mM) was dissolved in phosphate buffer (100 mM, 6 M Gn·HCl, pH ~7) and treated with methyl 3-mercaptopropionate (MMP) (5% v/v) for 5-7 hours at room temp. The reaction was monitored by analytical HPLC (XSelect C18 column, 130 Å, 3.5 μm, 4.6 × 150 mm), and ESI-MS, and the peptide was purified by preparative HPLC.

Native chemical ligation (NCL). The $(HhH)_2$ constructs were too long for a single SPPS reaction; thus, two half-peptides were synthesized then joined using native chemical ligation (NCL) and deselenization. In this case, the N-terminal half-peptide bears a C-terminal thioester (or thioester surrogate, such as *N*-acyl urea, Nbz (1)) moiety (NHalf-COSR) and the C-terminal peptide bears an N-terminal selenocysteine (Sec, U) residue (Sec-CHalf). Briefly, the NHalf-COSR peptide was dissolved in argon degassed PB buffer (100 mM NaH_2PO_4, 6 M Gn·HCl, 0.2 M MPAA, 0.05 M TCEP and 0.3 M sodium ascorbate, pH 7.3) and added to the Sec-CHalf peptide, yielding a final peptide concentration of about 1 mM. The progress of the reaction was followed by analytical HPLC (XBridge C4 column, 300 Å, 3.5 μm, 4.6 × 150 mm) with a gradient of 5-70 % buffer B over 20 min. The ligation was quenched after 4 h. The ligation product was purified by semi-preparative HPLC (XBridge BEH300 C4 column, 300 Å, 5 μm, 19 × 150 mm, method 25-50 % B over 45 min). The resulting peptide still contained a selenocysteine residue at the site of ligation.

Deselenization reaction. Sec-containing peptides were dissolved in 1 mL of argon degassed PB buffer (100 mM, 6 M Gn·HCl, 100 equivalents DTT, pH 7.3), and left for 30 min, upon which 100 equivalents of TCEP in 200 μL of the same argon degassed buffer were added (8, 9). The progress of the reaction was followed by analytical HPLC (XBridge C4 column, 300 Å, 3.5 μm, 4.6 × 150 mm) with gradient 5-70 % B over 20 min, and TCEP was added as needed to generate the desired deselenized product.

Guanidination of ornithine-containing polypeptides. Peptides (0.238 μmol) were dissolved in TDW to a concentration of about 2 mg/mL. Stock solution containing 50 mg of 1-guanyl-3,5 dimethyl pyrazole nitrate dissolved in 300 μL TDW was prepared (0.83 M). For guanidination reaction, ~6 equivalents per each ornithine residue present in the peptide were added to the 1 mL

peptide solution. The pH was adjusted to 9.3 using 1 M NaOH, and the reaction was incubated at 37 °C (10). The progress of the reaction was followed by HPLC and ESI-MS, and 10 µL of reaction aliquots quenched with 20 µL of 0.1 % TFA in TDW were injected (XBridge C4 column, 300 Å, 3.5 µm, 4.6 × 150 mm) and eluted using a gradient of 5 % B in eluent A for 2 min then 5-50 % B over 20 min.

Fluorescence recovery after photobleaching (FRAP). Fluorescence recovery after photobleaching was measured for Ancestor-$(HhH)_2$ and Precursor-Arg coacervates, prepared as described in the main text, and loaded into capillary slides. Coacervates were imaged using an Olympus IX81 inverted confocal microscope equipped with a 60x oil immersion objective (UPlanSApo 60x/1.35 Oil, Olympus). Photobleaching was achieved by the Tornado method, in which samples were excited with a 488 nm laser diode at 100 % power for 200 ms. The imaging time was 278.1 ms/frame. The fluorescence intensity (excitation wavelength 488 nm, emission wavelength 520-550 nm) was monitored as a function of time for both the bleached area and the background, using the Fiji platform, and the recovery of the bleached region was normalized against the background. Data were analyzed using Prism (GraphPad Software Inc.).

Dataset 1 (separate file): Sequence alignment of $(HhH)_2$ protein families

SI References

1. J. B. Blanco-Canosa, P. E. Dawson, An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew. Chemie - Int. Ed.* (2008) https:/doi.org/10.1002/anie.200705471.
2. L. C. Allen, T. Viswanatha, Reaction of amino acids with guanidinating agents. *Can. J. Biochem.* (1970) https:/doi.org/10.1139/o70-185.
3. W. L. Cosand, R. B. Merrifield, Concept of internal structural controls for evaluation of inactive synthetic peptide analogs: synthesis of [Orn13,14]apamin and its guanidination to an apamin derivative with full neurotoxic activity. *Proc. Natl. Acad. Sci. U. S. A.* (1977) https:/doi.org/10.1073/pnas.74.7.2771.
4. B. Filippi, *et al.*, Synthetic enkephalins addicting properties and conformational studies in solution. *Int. J. Pept. Protein Res.* (1979) https:/doi.org/10.1111/j.1399-3011.1979.tb01918.x.
5. W. Wang, B. A. Malcolm, Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange(TM) Site-Directed Mutagenesis. *Biotechniques* (1999).
6. A. L. Schroll, R. J. Hondal, S. Flemer, 2,2'-Dithiobis(5-nitropyridine) (DTNP) as an effective and gentle deprotectant for common cysteine protecting groups. *J. Pept. Sci.* (2012) https:/doi.org/10.1002/psc.1403.
7. P. S. Reddy, S. Dery, N. Metanis, Chemical synthesis of proteins with non-strategically placed cysteines using selenazolidine and selective deselenization. *Angew. Chemie - Int. Ed.* (2016) https:/doi.org/10.1002/anie.201509378.
8. L. Dery, *et al.*, Accessing human selenoproteins through chemical protein synthesis. *Chem. Sci.* (2017) https:/doi.org/10.1039/c6sc04123j.
9. N. Metanis, E. Keinan, P. E. Dawson, Traceless ligation of cysteine peptides using selective deselenization. *Angew. Chemie - Int. Ed.* (2010) https:/doi.org/10.1002/anie.201001900.
10. S. Ariely, M. Wilchek, A. Patchornik, Synthesis of poly-L-arginine. *Biopolymers* (1966) https:/doi.org/10.1002/bip.1966.360040110.