THE PROTEIN SOCIETY

A *de novo* protein confers copper resistance in *Escherichia coli*

Kenric J. Hoegler[1,2] and Michael H. Hecht[1,2]*

[1]Department of Chemistry, Princeton University, Princeton, New Jersey 08540
[2]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08540

Received 27 October 2015; Accepted 30 December 2015
DOI: 10.1002/pro.2871
Published online 00 Month 2016 proteinscience.org

Abstract: To survive environmental challenges, biological systems rely on proteins that were selected by evolution to function in particular cellular and conditional settings. With the advent of protein design and synthetic biology, it is now possible to construct *novel* proteins that are not biased by eons of selection in natural hosts. The availability of these sequences prompts us to ask whether natural biological organisms can use naïve—non-biological—proteins to enhance fitness in stressful environments. To address this question, we transformed a library of DNA sequences encoding ~1.5 × 10^6 binary patterned *de novo* proteins into *E. coli,* and selected for sequences that enable growth in concentrations of copper that would otherwise be toxic. Several novel sequences were discovered, and one of them, called Construct K (ConK), was studied in detail. Cells expressing ConK accumulate approximately 50% less copper than control cells. The function of ConK does not involve an oxidase, nor does it require two of the best characterized copper efflux systems. However, the ability of ConK to rescue cells from toxic concentrations of copper does require an active proton motive force. Further selections for growth in higher concentrations of copper led to the laboratory evolution of variants of ConK with enhanced levels of activity *in vivo*. These studies demonstrate that novel proteins, unbiased by evolutionary history in the natural world, can enhance the fitness of biological systems.

Synopsis: Living systems evolve to adapt to potentially lethal environmental changes. This normally involves repurposing existing genetic information (i.e. sequences that were selected by billions of years of evolution). Here we show that a completely *de novo* protein, not derived from nature, can enable *E. coli* cells to grow in otherwise toxic concentrations of copper, demonstrating that living systems also have the capacity to incorporate and *protopurpose* entirely novel genetic information.

Keywords: protein design; polar/nonpolar patterning; binary code; four helix bundle; synthetic biology; copper resistance; metal; protein evolution; protopurpose; de novo; molecular evolution

Introduction

A central goal of synthetic biology is to improve upon natural biological systems.[1,2] One particular aim is to enhance the tolerance of organisms to environmental toxins. Toward achieving this objective, one can envision two general approaches: The first takes pre-existing genes, proteins, and/or regulatory elements from natural systems and inserts them into the chosen organism. While these natural sequences may be re-engineered or modified, fundamentally, this approach draws upon molecules "borrowed from nature." A second approach goes beyond sequences provided by nature, and asks whether genes and proteins designed entirely *de novo* can be used to improve upon nature.

The first approach—borrowing from nature—has been used to engineer organisms to tolerate a range of environmental toxins.[3–6] While these efforts have been fruitful, they are intrinsically limited by a

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NSF; Grant number: MCB-1409402.

*Correspondence to: Michael H. Hecht; Department of Chemistry, Princeton University, Princeton, NJ, 08540. E-mail: hecht@princeton.edu

toolbox of genes and proteins that already existed in nature and evolved over billions of years to perform specific functions in particular cellular and environmental niches. To explore beyond the boundaries of this toolbox of natural sequences, one can devise collections of sequences entirely *de novo,* and screen these for desired chemical and/or biological properties.

Previously, we reported the design and construction of large collections of novel proteins.[7–9] Our strategy combines elements of rational design with an ability to construct large libraries. As described previously, this strategy posits that the binary patterning of polar and nonpolar amino acids in the sequence of a protein plays a dominant role in dictating its 3-D structure. For example, a sequence of polar (P) and nonpolar (N) residues with the pattern PNPPNNP incorporates a nonpolar side chain every 3 or 4 residues, matching the α-helical repeat of 3.6 residues/turn. When such sequences are placed in water, the thermodynamic driving force to bury hydrophobic surface area drives the formation of amphiphilic α-helices, which assemble into a tertiary structure that buries nonpolar helical faces against one another. Because the binary code strategy specifies the *type* of amino acid (polar vs. nonpolar), but not the identity of the side chain, it is well suited to combinatorial approaches. Combinatorial mixtures of amino acids can be encoded by degenerate DNA codons, such as NTN (N=A,T,C,G) to encode five nonpolar amino acids (Phe, Leu, Ile, Met, Val), and VAN (V=A,C,G) to encode six polar amino acids (His, Glu, Gln, Asp, Asn, Lys).

We have used the binary code strategy to design several libraries of α-helical and β-sheet proteins, and have demonstrated that proteins from the α-helical libraries fold into well-ordered stable structures.[7–13] Proteins from these libraries bind small molecules, including cofactors, and catalyze reactions.[14–19] Most importantly, several of these *de novo* α-helical proteins function *in vivo* to provide essential activities necessary to sustain the growth of living cells.[20,21]

In this study, we investigate whether proteins designed *de novo* can be used to push the limits of tolerance to an environmental toxin. To address this question, we transformed *E. coli* cells with a library of synthetic genes encoding ~10^6 *de novo* proteins and subjected the transformed cells to selection for survival in toxic concentrations of copper, a well-known bactericide and environmental contaminant.[22,23]

These selections led to the isolation of a novel protein, Construct K (ConK), which allows *E. coli* to grow in otherwise lethal concentrations of copper. Subsequent directed evolution of ConK led to ConK-95b2, an *de novo* sequence that allows *E. coli* to grow in concentrations of copper as high as those tolerated by natural variants of *E. coli* that evolved in agricultural settings in response to bactericidal levels of copper.[24,25]

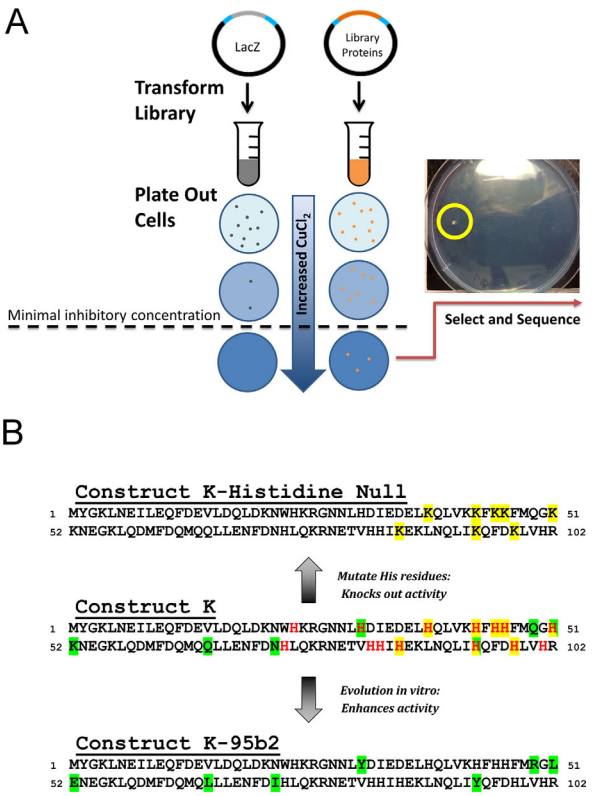

Figure 1. (A) Selection of novel proteins that confer resistance to copper. A library comprising approximately 10^6 unique sequences was transformed into *E. coli* and plated on LB agar spiked with increasing concentrations of $CuCl_2$. Using LacZ as a negative control, we selected colonies that grew above the MIC. The top hit, ConK, (pictured and circled in yellow) served as the subject of this study. (B) Amino acid sequences of the novel protein Construct K (ConK), the inactive mutant Construct K-Histidine Null (ConK-HN), and the evolved variant, Construct K-95b2. ConK confers *E. coli* cells with resistance to copper. ConK-HN is a mutant with eight His → Lys mutations. ConK-95b2 was evolved from ConK, and enables *E. coli* growth in notably high levels of copper. Histidine residues are shown in red font. Residues mutated in ConK-HN are highlighted in yellow, and residues mutated in ConK-95b2 are highlighted in green. Histidines 51 and 94 were modified in both the ConK-HN and ConK95b2.

Results

Selection of novel proteins that confer resistance to copper

Copper is an essential trace element; however, in excess amounts, it is toxic to all organisms. The majority of soluble copper in the environment exists as Cu^{2+}. In *E. coli*, Cu^{2+} likely enters into the periplasmic space via porins[26]. Once in the periplasm, copper fluctuates between the Cu^{2+} and Cu^{1+} states. The Cu^{1+}state can diffuse into the cytoplasm where it can interact with a range of copper-sensitive targets.[27–29]

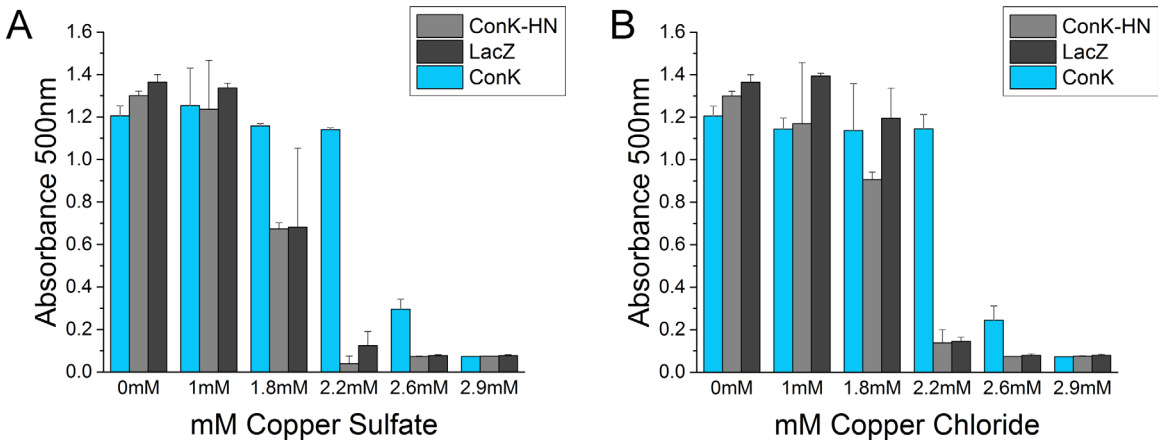

Figure 2. The *de novo* protein ConK confers resistance to toxic levels of copper in liquid culture. *E. coli* strain BW25113 was grown in various concentrations of (A) $CuSO_4$ or (B) $CuCl_2$ for 20 hours at 37°C. The OD_{500} was then measured. Cells expressing ConK grow at concentrations of copper above the MIC for cells expressing the LacZ or ConK-HN controls.

Prior to selecting novel proteins that confer resistance to copper, we determined the minimal inhibitory concentration (MIC) of $CuCl_2$ for *E. coli* strain BW25113. As a control, we used cells expressing β-galactosidase (LacZ) from the same vector as our library of novel proteins. Cells were plated on LB plates containing 100 μ*M* IPTG (to induce expression), and increasing amounts of $CuCl_2$. Plates were incubated for 24 h, and the MIC—defined as the concentration of $CuCl_2$ at which cells fail to grow—was determined to be 4.4 m*M* [Fig. 1(A)].

To isolate novel proteins that confer resistance to copper, we searched through a library of binary patterned sequences. The sequences used for the current study were from a 3rd generation library designed to form 4-helix bundles.[10] The collection is estimated to contain $\sim 1.5 \times 10^6$ different sequences. We transformed this library of *de novo* sequences into *E. coli* strain BW25113, and plated on LB containing 100 μ*M* IPTG and concentrations of $CuCl_2$ above the MIC. To our surprise, hundreds of colonies developed on plates containing concentrations of $CuCl_2$ above the MIC. Plasmids from the 10 colonies that grew at the highest concentration of $CuCl_2$ were isolated, and re-transformed to confirm their ability to rescue cells from copper toxicity. The translated amino acid sequences of all hits are shown in Supporting Information Figure 1(A).

Two of the 10 hits were found twice, thus there are eight unique sequences. The three sequences that rescued at the highest concentration of $CuCl_2$ showed no more sequence identity (~54%) than non-rescuing members of the library. The sequence that rescued at the highest concentration of $CuCl_2$ (5.4 m*M*) was isolated from two independent transformations. This hit was named Construct K (ConK), and its sequence is shown in Figure 1(B). ConK was re-cloned into the expression vector to confirm that its sequence (and not other sequences on the plasmid or in the host strain) is responsible for resistance to $CuCl_2$.

A mutant of ConK fails to confer resistance to copper

Because ConK is expressed in the cytoplasm, we presumed it interacts with copper in the Cu^{1+} state.

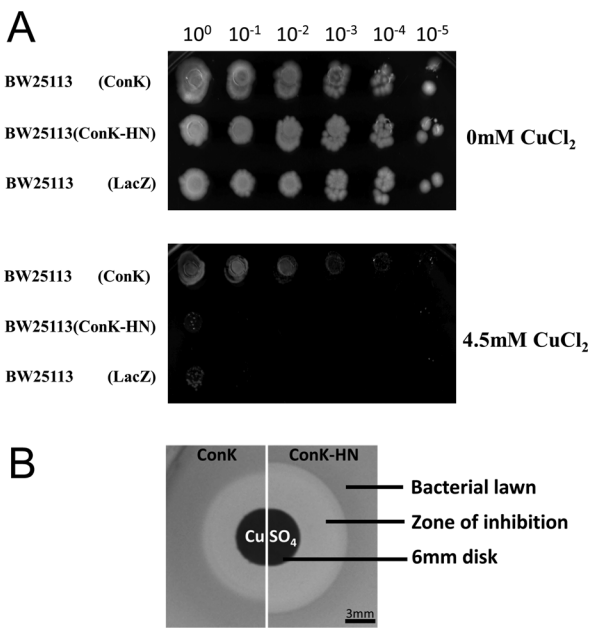

Figure 3. The *de novo* protein ConK confers resistance to toxic levels of copper on solid medium. (A) Tenfold dilutions of E coli BW25113 expressing the indicated proteins were stamped onto LB/Agar with and without CuCl2. Plates also contained 100 μ*M* IPTG and 30 μg/mL chloramphenicol. Cells were allowed to grow for 20 h at 37°C. Cells expressing ConK survived on plates containing 4.5 m*M* $CuCl_2$, while cells expressing the LacZ or ConK-HN controls failed to grow. (B) Comparison of the zone of inhibition for cells expressing either ConK or the ConK–HN control. Cells expressing ConK displayed enhanced tolerance to copper, and grow distinctively closer to the 6 mm disk saturated with $CuSO_4$.

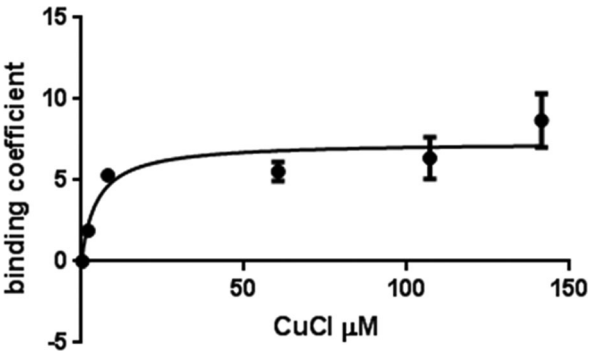

Figure 4. Equilibrium dialysis shows protein ConK binds copper. Because ConK is expressed in the *E. coli* cytoplasm, where copper exists in the +1 state, equilibrium dialysis was performed with Cu^{1+} under anaerobic conditions. Data analysis using GraphPad Prism suggests ConK binds approximately 7 copper ions, with a total apparent binding constant of approximately 4.5 μM. (binding coefficient = [Ligand]bound/[Protein]total)

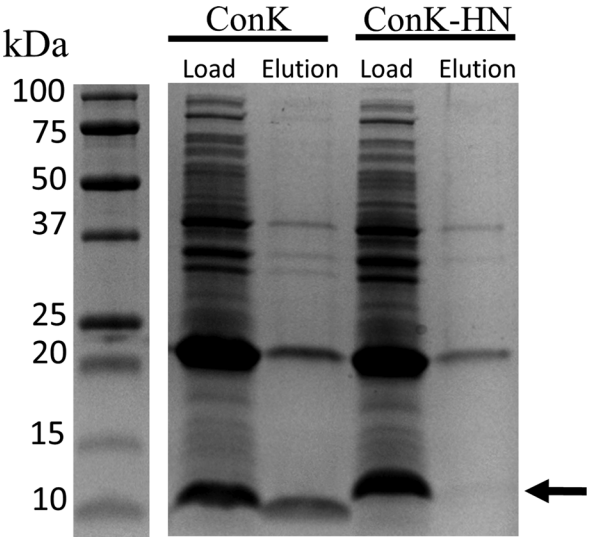

Figure 5. Protein ConK binds immobilized Cu^{2+}. Clarified cell lysates from cells expressing either ConK or ConK-HN were incubated with Cu^{2+} immobilized on iminodiacetic acid sepharose beads. After stringent washes, the amount of protein loaded onto the beads (Load) was compared to proteins retained on the beads (Elution), by electrophoresis and Coomassie blue staining. The synthetic proteins have a mass of approximately12 kD (indicated by arrow). The mutant protein ConK-HN does not bind immobilized Cu^{2+}. It is also important to note the similar expression levels of ConK and ConK-HN (see "Load" lanes). The intense band between 20 and 25 kD is choloramphenicol acetyl transferase.

For this state of copper, the relative affinities of the amino acid side chains have been reported as Cys > His > Met.[30] Since our libraries do not contain cysteine, we surmised that histidine side chains in ConK may be important for function. The sequence of ConK contains 14 histidines [Fig. 1(B)]. Six of these histidines are in constant regions of the binary patterned template and occur in all members of the library, including those that fail to confer resistance to copper. To confirm that (at least some of) the eight histidines in the variable region of ConK are important for the phenotype, we mutated these residues to lysines. We named this 8-fold mutant Construct K-Histidine Null (ConK-HN).

We compared the abilities of ConK and ConK-HN to confer resistance to copper. Assays were performed both in liquid culture and on plates. As shown in Figure 2, ConK enables growth in liquid media containing concentrations of $CuCl_2$ or $CuSO_4$ that prevent growth of the LacZ control. In contrast, ConK-HN fails to provide resistance to copper toxicity; like the LacZ control, cells expressing the mutant only grow at concentrations at or below the MIC.

We also compared the abilities of ConK and ConK-HN to confer resistance to copper toxicity on solid media. We first grew liquid cultures in the absence of added copper, and then stamped serial dilutions of these cultures onto LB plates containing 100 μ*M* IPTG and toxic concentrations of $CuCl_2$. As shown in Figure 3, ConK rescues *E. coli* cells from copper toxicity, while ConK-HN fails to rescue. Similar results were observed when the experiment was repeated in different strains of *E.* coli (Supporting Information Fig. 2). The failure of the ConK-HN mutant to confer resistance, in principle, could derive from reduced expression levels, however, ConK-HN expresses at the same level as the parental ConK sequence (see the "Load" lanes in Fig. 5).

ConK confers resistance to copper, but not to other divalent metals

To determine if ConK confers general resistance to divalent metals, we measured zones of inhibition around toxic concentrations of several metal salts. For these studies, cells were plated in top agar on

Table I. *ConK Confers Resistance to Copper, but not to Other Divalent Metals*

Toxin	ConK (mm)	LacZ (mm)	ConK-HN (mm)
$CuCl_2$	14.2 ± 0.3	16.7 ± 0.3	17.0 ± 0.0
$CuSO_4$	12.7 ± 0.3	16.0 ± 0.0	16.7 ± 0.0
$ZnCl_2$	22.5 ± 0.0	22.0 ± 0.0	22.5 ± 0.0
$CoCl_2$	22.0 ± 0.0	22.0 ± 0.0	22.0 ± 0.0
$NiCl_2$	20.7 ± 0.3	20.7 ± 0.3	20.5 ± 0.7
H_2O_2	26.0 ± 0.0	26.0 ± 0.0	26.0± 0.0
H_2O	0 ± 0.0	0 ± 0.0	0 ± 0.0

Zones of Inhibition (in millimeters) of various toxins. Lawns of *E. coli* were plated in top agar on LB agar and allowed to grow for 16 h in the presence of 6 mm disks saturated with various metal salts ($CuCl_2$, $CuSO_4$, $ZnCl_2$, $CoCl_2$, or $NiCl_2$) or H_2O_2. Zones of inhibition indicate the distance around the disk where cell growth is inhibited. A smaller zone of inhibition indicates cells are more resistant to the toxin. ConK was more effective than the controls (ConK-HN and LacZ) for $CuCl_2$ and $CuSO_4$ (highlighted), but not for any of the other metal salts. ConK also had no effect on the toxicity of hydrogen peroxide. Measurements of the zones diameters were taken and compared [for visual see Fig. 3(B)]. Duplicate experiments were measured and processed with Image J.

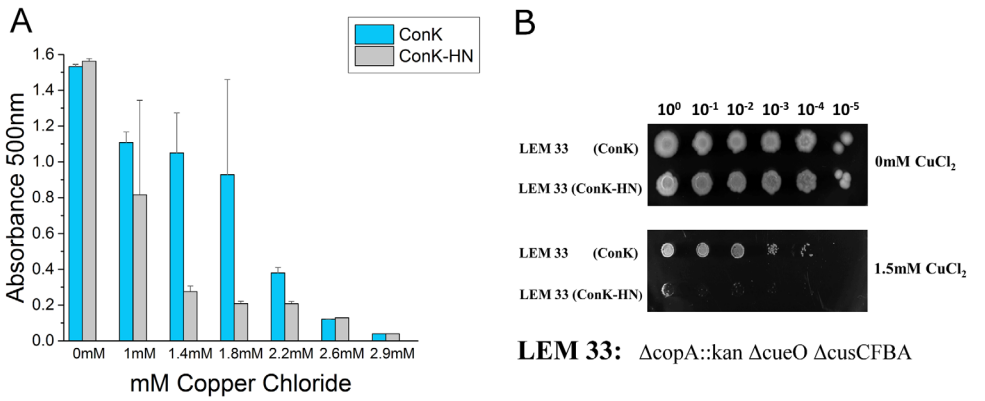

Figure 6. ConK confers resistance to copper in liquid or solid media in cells deleted for three endogenous detoxification systems. (A) The triple knockout strain copA::kan DcueO DcusCFBA (LEM33) expressing either ConK or ConK-HN was grown for 36 h at 37°C in liquid cultures in the presence of increasing concentrations of $CuCl_2$. (B) Tenfold dilutions of *E. coli* Lem33: copA::kan ΔcueO ΔcusCFBA cells expressing the indicated proteins were stamped onto LB/Agar with and without $CuCl_2$. In both liquid and solid media, copA::kan ΔcueO ΔcusCFBA cells expressing ConK grow in the presence of copper, while cells expressing the ConK-HN mutant failed to grow.

LB plates, and a small disk of filter paper saturated with a toxic concentration of a divalent metal salt was placed on the plate. Cells that are far away from the disk grow into a lawn of bacteria. However, closer to the disk, one observes a zone of inhibition due to radial diffusion that produces a gradient of toxin from the disk to the nearby area. The more resistant a cell is to the toxin, the closer the lawn will approach the disk, and the smaller the inhibition zone. An example of such an experiment is shown in Figure 3(B) comparing the resistance to $CuSO_4$ for ConK versus ConK-HN. Zone of inhibition studies were performed using disks saturated with salts of various divalent metals. Measurements of the zone of inhibition show that ConK conferred resistance to $CuCl_2$ and $CuSO_4$, but not to $ZnCl_2$, $CoCl_2$, or $NiCl_2$. These results are summarized in Table I. Growth in liquid cultures confirmed that ConK conferred resistance to $CuCl_2$ and $CuSO_4$, but not to $ZnCl_2$, $CoCl_2$, $NiCl_2$, or $AgNO_3$.

ConK binds copper with moderate affinity

We used equilibrium dialysis to probe the ability of ConK to bind Cu^{1+} *in vitro*. Purified ConK protein in standard buffer was placed in chamber 1 of a dialysis apparatus (Harvard Bioscience Inc.), while increasing amounts of Cu^{1+} were placed in chamber 2 (see Materials and Methods section). After equilibration, the concentration of Cu^{1+} in chamber 2 was quantified using Phen Green™ FL (Thermo Fischer Scientific). Analysis of the data suggests that ConK binds as many as 7 Cu^{1+} ions, with an apparent total dissociation constant of ~4.5 μ*M*. (Fig. 4). These data indicate that ConK has a moderate affinity for copper; however, the apparent stoichiometry suggests that binding may not occur via well-defined pre-organized sites (see discussion).

Equilibrium dialysis was also performed for the ConK-HN mutant, and no binding was observed. Furthermore, ConK readily binds to immobilized metal affinity columns (IMAC) loaded with copper, while ConK-HN fails to bind (Fig. 5). These findings show that the histidine side chains in ConK that are necessary to confer resistance to copper toxicity *in vivo* (Fig. 3) are also important for copper binding *in vitro* (Fig. 5).

E. Coli *cells expressing ConK retain less copper*

Previous studies have shown that the toxicity of copper can be suppressed in *E. coil* or yeast by overexpressing a natural metal binding protein.[31,32] In both cases, copper ions were bound and sequestered inside the cell, presumably preventing them from aberrantly interacting with sensitive targets. Since ConK binds Cu^{1+} *in vitro,* we reasoned that metal sequestration might account for rescue *in vivo*. Therefore, we compared the amount of copper in cells expressing ConK relative to cells expressing the control proteins LacZ or ConK-HN. Unexpectedly, quantification of intracellular copper by ICP-MS (inductively coupled plasma mass spectrometry) revealed there was *less* copper in cells expressing ConK compared to cells expressing the controls ConK-HN or LacZ (Table II).

Thus, cells expressing ConK do *not* accumulate copper ions. Instead, the presence of ConK either reduces the amount of copper that enters cells, or assists in removing copper from cells.

Several copper homeostatic systems are not required for copper resistance by ConK

E. coli has several well-studied copper detoxification systems, including the CueO multicopper oxidase[33], the CusCFBA RND efflux system[34], and the CopA ATPase efflux pump.[35] Strains lacking these system are extremely sensitive to copper.[28]

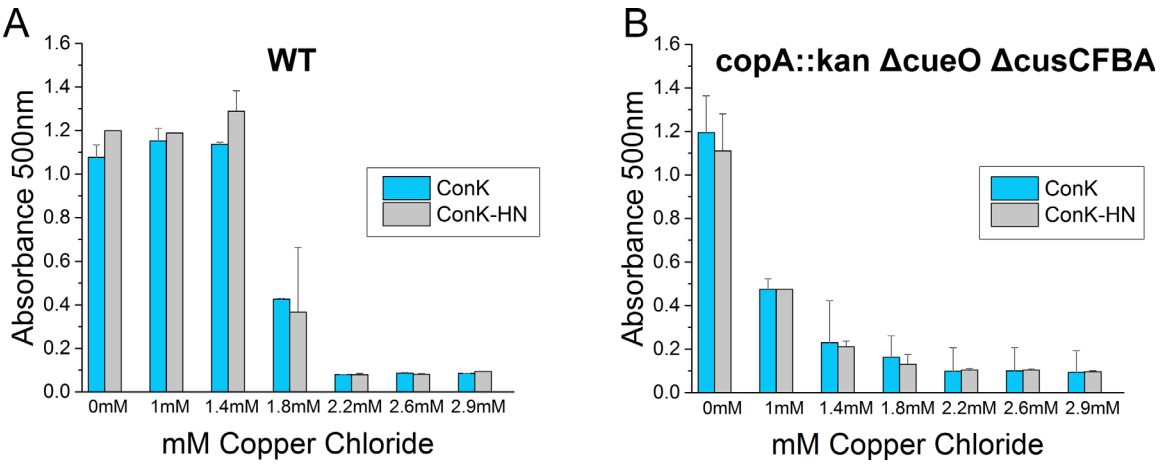

Figure 7. ConK does not function in cells treated with CCCP, a drug that uncouples the proton motive force. Addition of a tolerated amount of carbonyl cyanide m-chlorophenyl hydrazone (5 μ*M* CCCP) to liquid cultures eliminated the ability of ConK to enable growth in toxic concentrations of copper for both (A) WT and (B) copA :: kan ΔcueO ΔcusCFBA cells.

We probed the possible involvement of multicopper oxidase activity using both biochemical and genetic approaches. Several biochemical assays have been developed to measure oxidase activity.[33] Two of these assays use the chromogenic reagents, DHB (dihydroxybenzoic acid) and ABTS (2,2′-azinobis-(3-ethylbenzothiazoline-6-sulphonate)). We performed assays using both reagents and found that cells expressing ConK had no more oxidase activity than controls. This demonstrates that ConK neither functions as a copper oxidase itself, nor does it induce endogenous copper oxidase activity in *E. coli*.

Next we used genetic experiments to test whether any of the three copper detoxification systems listed above are required for the copper resistance phenotype mediated by ConK. For these studies, we obtained the Lem33 strain in which all three systems had been knocked out. This strain—copA::kan DcueO DcusCFBA (LEM33) was transformed with the plasmid expressing ConK and exposed to increasing concentrations of $CuCl_2$. Although this strain is considerably more sensitive to copper than wild-type *E. coli*, ConK still enhanced resistance to copper relative to the ConK-HN control in both liquid and solid media (Fig. 6). ConK also rescues strains lacking each individual mutation (*data not shown*).

Table II. *Cells Expressing ConK Accumulate Less Copper Than Controls*

Strain (plasmid)	ng/mg of dry cells
A	
W3110 F (Conk)	296.2 ± 11.5
W3110 F (LacZ)	646.8 ± 7.5
W3110 F (Conk-HN)	656.2 ± 12.5
B	
copA::kan ΔcueO ΔcusCFBA (Conk)	245.8 ± 0.1
copA::kan ΔcueO ΔcusCFBA (LacZ)	453.8 ± 0.2
copA::kan ΔcueO ΔcusCFBA (Conk-HN)	548.2 ± 11.9

Amount of copper in cells. Cells were grown for 12 h in concentrations of $CuCl_2$ below their MIC. Copper content relative to dry cell mass was measured using ICP-MS (inductively coupled plasma mass spectrometry) in (A) a standard strain of *E. coli*, W3310, grown in 1.4 m*M* $CuCl_2$ and (B) stain copA::kan ΔcueO ΔcusCFBA, which is knocked out for three major copper homeostasis systems, was grown in 0.14 m*M* $CuCl_2$. In both strains, cells expressing ConK accumulated approximately half as much copper as cells expressing the controls, ConK-HN or LacZ. Experiments were performed in duplicate.

CopA::kan ΔcueO ΔcusCFBA cells expressing ConK also accumulate less copper than the same cells expressing the control proteins ConK-HN or LacZ (Table II-B).

These results demonstrate that although these three well-characterized copper homeostatic systems are needed for resistance to copper at concentrations above 4.4 m*M*, they are not required for the ConK rescue phenotype.

ConK requires a functional proton gradient to confer resistance to copper

Our finding that cells expressing ConK retain less copper than controls led us to postulate that ConK either prevents copper from entering cells, or assists in removing copper from cells. Many efflux systems rely on proton gradients, either directly or indirectly (i.e. proton gradients themselves or the ATP produced by coupling to proton gradients). To probe whether a functional proton gradient is required for rescue by ConK, we grew cells in the presence of Carbonyl Cyanide *m*-Chlorophenyl Hydrazone (CCCP), which is known to decouple the proton gradient. As shown in Figure 7, CCCP abrogates the ability of ConK to confer resistance to copper toxicity. Elimination of the ConK rescue phenotype by CCCP was observed in both WT and copA::kan ΔcueO ΔcusCFBA strains [Fig. 7(A,B)]. These results

demonstrate that a functional proton gradient is required for ConK to confer resistance to copper.

An evolved variant of ConK enables growth of E. coli *in 7 m*M *$CuCl_2$*

Directed evolution has been used in a wide range of protein engineering experiments. In virtually all of those previous examples, the target protein was 'borrowed' from nature—i.e. it was based on an amino acid sequence that had already experienced a long evolutionary trajectory in natural biological systems. In the current situation, however, we had the opportunity to start with a sequence that never existed in nature, and to apply selective pressure in the laboratory to direct its evolution toward a desired outcome.

To select a *de novo* protein that confers resistance to higher levels of copper, we subjected ConK to several cycles of random mutagenesis, followed by selection of transformed cells on plates spiked with increasing amounts of $CuCl_2$. As noted above, the MIC for strain BW25113 expressing the LacZ control is 4.4 m*M* $CuCl_2$, while ConK's MIC was 5.4 m*M* $CuCl_2$. Following one round of mutagenesis and selection, we isolated a new sequence, called ConK85a, which confers resistance to 6.0 m*M* $CuCl_2$. Additional rounds of evolution led to ConK95b2, which enables growth at 7.0 m*M* $CuCl_2$ and $CuSO_4$.

The sequence of ConK95b2 is presented in Figure 1, and the sequences of the evolutionary intermediates are shown in Supporting Information Figure 1(B). In the evolved variant, ConK95b2, three histidines mutated to other residues. His31 and His87 were replaced by tyrosines, and His51 was replaced by lysine. This reduction in exposed histidines might improve the specificity of a metal binding site.

Discussion

Evolution requires living systems to repurpose genetic information and adapt it toward new functions.[36–38] Similarly, synthetic biology reengineers genes and proteins toward functions that are innovative and useful.[39] Both natural biology and synthetic biology typically base these revisions on sequences that arose through eons of natural selection. In the current study, however, we asked whether living organisms can also assimilate *novel* genetic information, designed in the laboratory, and purpose it toward new biological functions. To address this question, we investigated whether artificial proteins—neither evolved in nature, nor explicitly designed for any particular function—can enhance the fitness of a living organism. Our findings show that *E. coli* can incorporate ConK into its proteome, and *protopurpose* this novel protein toward a function that enables growth under conditions that would otherwise be lethal.

Several observations allow us to speculate on the function of ConK. First, ConK binds copper (Fig. 4). While this binding occurs with moderate affinity (~4.5 μ*M*), the apparent stoichiometry suggests that binding does not occur via well-defined pre-organized sites. This is reminiscent of several natural proteins involved in metal homeostasis, which have abundant histidine residues, and bind metal ions with moderate affinity and variable stoichiometry.[40,41]

Mutation of several of the histidine residues in ConK destroys copper binding. The loss of binding *in vitro* simultaneously eliminates the rescue phenotype *in vivo* (Fig. 2). Thus, copper binding appears to play a role in the rescue mechanism. This finding led us to consider whether sequestration of copper by ConK might be responsible for its ability to enable growth in otherwise toxic levels of copper. This hypothesis was bolstered by studies showing that overexpression of natural metal binding proteins can alleviate copper toxicity in *E. coli*, presumably by sequestering copper ions in the cell. Curiously, however, cells expressing ConK do not sequester copper, but in fact, accumulate *less* copper (Table II). Thus, although copper binding may be necessary for rescue by ConK, sequestration is *not* the mechanism of rescue.

Next we considered the possibility that ConK might alleviate copper toxicity by oxidizing Cu^{1+} to the less toxic Cu^{2+} state. However, several experiments showed the *de novo* protein is not a copper oxidase, nor does it induce endogenous copper oxidase activity.

We also tested whether any of the three best-characterized copper detoxification systems in *E. coli* were involved in ConK function. We found that ConK still alleviates toxicity in a strain deleted for CueO, which encodes the multicopper oxidase. Likewise, ConK rescue was still observed in strains deleted for CusCFBA and CopA, which encode efflux systems. Moreover, cells expressing ConK, but lacking the natural copper efflux proteins copA and CusCFBA also contain less copper (Table II). Thus, neither of these efflux systems is required for ConK's ability to reduce cellular accumulation of copper.

To probe if other efflux systems may be involved, cells were treated with CCCP, a drug that decouples the proton motive force and would be expected to interfere with most efflux systems. Treatment with CCCP eliminates rescue by ConK, suggesting that rescue involves an efflux system driven by the proton motive force.

Based on these findings, ConK seems to mediate rescue by a novel mechanism, perhaps by forming transient complexes with Cu^{1+} ions, which are

ultimately passed along to non-specific efflux systems that have been implicated in copper resistance.[42–44] This would be the first example of a cytoplasmically expressed soluble protein directly assisting in the removal of copper from an *E.coli*. Further studies will be required to support or disprove this hypothesis; however, irrespective of the exact mechanism, it is clear that the *de novo* protein, ConK, can be assimilated into the *E. coli* proteome and used to enable life under conditions that are otherwise lethal.

Finally, we asked if directed evolution could be used to enhance this *protopurposed* protein's function. ConK95b2, an evolved variant of ConK, confers resistance to 7.0 m*M* $CuCl_2$. This is similar to the level of resistance conferred by the Pco system, which was isolated from strains of *E. coli* living in the guts of pigs raised on feed containing high levels of copper.[24,25] It is interesting to compare the evolutionary history of the Pco system to that of ConK95b2. The Pco system arose in a natural organism, *E. coli*, which had evolved over many millions of years, and was then subjected to selection for toxin resistance for an additional few years. In contrast, ConK95b is a *de novo* sequence, which was evolved in the laboratory over the course of several weeks from a starting sequence, ConK, that never existed in nature, and was isolated for the first time last year.

While this is not the first example of an organism engineered to tolerate an environmental toxin, our approach is substantially different from other efforts in synthetic biology.[3–6] Traditional work in synthetic biology, although fruitful, is intrinsically limited by its reliance on natural genes and proteins, which are biased by billions of years of selection to perform a specific function in a defined cellular setting. In contrast, the current study shows that the fitness of a living organism can be improved using unevolved/unbiased sequences designed *de novo*. Furthermore, once these completely *de novo* proteins are incorporated into living cells, they can be evolved to enhance the activity of a sequence that did not originate in a living system, but which can nonetheless support life.

Materials and Methods

Chemicals, strains, and growth conditions

Carbonyl Cyanide *m*-Chlorophenyl Hydrazone was obtained from Sigma-Aldrich. Proteins were expressed from a modified pCA24N vector containing a chloramphenicol (Cam) resistance gene, as described previously.[10] Electrocompetent and chemically competent cells were made according to standard procedures. Luria broth was supplemented with 30 μg/mL Cam and/or 30 μg/mL kanamycin. The working concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG) was 100 μ*M*. Bacterial strains are listed in Supporting Information Figure 3.

Protein expression and purification

Cells transformed with plasmids encoding the protein of interest were grown overnight at 37°C on LB plates containing 30 μg/mL Cam. Single colonies were picked and grown in liquid cultures (LB with 30 μg/mL Cam) for 12 h at 37°C. Expression was induced at $OD_{600} = \sim 0.5$ by addition of IPTG, and cells were further cultured for 16 h at 37°C. Cell lysates were isolated using an EmulsiFlex French press (Avestin). Protein was isolated in two steps. First, lysate was run over a His Trap HP nickel column (GE Healthcare) washed with wash buffer (200 m*M* Na_2HPO_4, 500 m*M* NaCl, pH7.4) and eluted with elution buffer (200 m*M* Na_2HPO_4, 500 m*M* NaCl, 250 m*M* imidazole, pH7.4). Even without a His-tag, ConK binds to the nickel column, presumably because of the high percentage of histidine residues (14%). Second, fractions containing ConK were run over a HiLoad 26/60 superdex 75 size exclusion column (GE Healthcare) in buffer containing (40 m*M* HEPES, 300 m*M* NaCl).

Selection of proteins that confer resistance to copper

LB plates containing antibiotics and IPTG were made using the working concentrations specified above. Prior to pouring the plates, liquid LB agar was spiked with various amounts of $CuCl_2$, ranging from 0.4 to 6.8 m*M* $CuCl_2$. Cells were transformed via electroporation: 50 ng of the 3G library10 was added to 50 μL of competent cells and electroporated. Cells were brought up in Super Optimal broth with Catabolite repression (SOC) medium and shaken for 1 h at 37°C. Cells expressing 3G Library or LacZ control were spread on plates and grown overnight at 37°C. Plasmids from the 10 colonies that grew at the highest concentrations of $CuCl_2$ were isolated, retransformed into fresh competent cells, and streaked out at toxic concentrations to confirm rescue.

Liquid cultures

Overnight cultures were started from single colonies and grown for 12 h at 37°C in LB containing antibiotics. To determine comparative minimal inhibitory concentrations (MIC), these cultures were diluted 1:1000 in LB containing antibiotics and IPTG, and combined 1:1 with DI water spiked with increasing concentrations of the metal salt. Cultures were grown overnight at 37°C. OD_{500} was taken to minimize copper ion absorbance.

Zone of inhibition

Overnight cultures were diluted to an OD600 of 0.5 and plated in top agar onto LB plates containing Cam and IPTG. 6 mm disks, were saturated with metal salts dissolved in H2O: 0.5*M* $CuCl_2$, 0.5M $CuSO_4$, 0.5*M* $ZnCl_2$, 0.183*M* $CoCl_2$, 0.5*M* $NiCl_2$ or 0.3% H_2O_2 and then placed on the plates. Lawns were allowed to grow for 16 h. Measurements were taken with Image J. Averages and standard deviations were derived from duplicate experiments.

Immobilized metal affinity chromatography

Overnight cultures were grown in LB containing Cam and IPTG. 1.5 ml of overnight culture was spun-down and re-suspended in 250 μL of Bug-Buster Protein Extract Reagent (Novagen). After incubation for 20 min, samples were spun for 30 min at 13,000 rpm. Supernatants were incubated with Cu^{2+} immobilized on Iminodiacetic Acid Sepharose Beads (Sigma-Aldrich) for 1 h at 4°C. Beads were washed 4 x for 4 min with stringent wash buffer (50 m*M* Tris HCl, 500 m*M* NaCl, 50 m*M* Imidazole, pH 7.6.) Beads were then suspended in 2 x Laemmli sample buffer (Bio-Rad) containing 5% β-mercapto ethanol, heated to 95°C for 5 min, and spun for 2 min at 13,000rpm. Supernatants were normalized and compared to initial lysates by SDS-PAGE gel (Bio-Rad).

Plate stamping assays

Overnight cultures were diluted to an OD_{600} of 0.5 and serially diluted in 10-fold steps. Using a sterile 48 pin stamp, diluted samples were stamped onto LB plates containing antibiotics, IPTG, and increasing concentrations of $CuCl_2$. Plates were allowed to grow for 24 h at 37°C and imaged in a gel box.

ICP-ms

Overnight cultures were diluted 1:1000 in 50 ml of LB containing antibiotics, IPTG and $CuCl_2$. To enable cell growth, $CuCl_2$ concentrations were kept below the experimentally determined MICs: 1.4 m*M* for W3110 F and 0.14 m*M* for the copA::kan ΔcueO ΔcusCFBA strain. Cells were grown for 12 h at 37°C, spun down and washed twice with 50 m*M* sodium phosphate buffer containing 5 m*M* EDTA. Next they were washed twice with 50 m*M* sodium phosphate to remove EDTA. Cells were dried overnight at 70°C. The mass of dried pellets was determined. Cells were then re-suspended in 3.5% HNO_3 and heated to 95°C for 30 min. Insoluble material was removed by spinning at 13,000 rpm for 1 h, and supernatant was diluted 1:10 with HNO_3. Copper content was determined using an ELEMENT 2^{TM} ICP-MS (Thermo Scientific). Experiments were performed in duplicate.

Equilibrium dialysis

DispoEquilibrium DIALYZER TM MWCO 5000 Da (Harvard Apparatus) was used to analyze copper binding. All buffers and reagents were bubbled with nitrogen overnight to remove oxygen, and aerobic indicators (Oxoid—Thermo Scientific) were used to verify. 350μM ascorbic acid was used to reduce $Cu^{2+}$to Cu^{1+}, which was verified by the disappearance of a peak at 630 nm. Ten micromolar pure ConK in standard buffer (40 m*M* HEPES, 300 m*M* NaCl, pH 7.6) was placed in chamber 1 of the dialysis apparatus, and increasing amount of Cu^{1+} diluted in standard buffer were placed in chamber 2. The dialysis apparatus was rocked for 36 h at room temperature. Concentration of copper in chamber 2 was measured using Phen Green TM FL (Thermo Fischer Scientific). This was done by first devising a standard with known copper concentrations. The linear relationship between Cu^{1+} binding and fluorescent quenching enabled us to determine the concentration of Cu^{1+} in chamber 2. Calculations were performed using Harvard Apparatus standard protocols, and the binding curve was fit using GraphPad Prism software. To confirm that protein remained soluble when bound to Cu^{1+} the contents of chamber 1 were spun at 13,000 rpm for 1 min then monitored by electrophoresis.

Directed evolution

Directed evolution experiments were done using error prone PCR.[45] Forward and reverse primers, CGTCTTCACCTCGAGAAATC and GACCTCAGAACTCCATCTGG, were used to amplify the 306 bp sequence encoding a rescue protein. 30 cycles of amplification were carried out in the presence of manganese chloride, introducing an average of 4–5 nucleotide substitutions per sequence. After amplification, fragments were digested and re-cloned into fresh vectors. These mutant libraries were transformed into cells and plated on concentrations of copper sulfate above ConK's MIC in an identical manner to our original library selection [Fig. 1(A)]. To confirm rescue, sequences that enabled growth were retransformed into *E. coli* and streaked out on plates containing concentrations of copper above ConK's MIC.

Acknowledgments

We thank John A. Higgins and Elizabeth Lundstrom of the Princeton Geosciences Department for assistance with ICP-MS, and James A. Imlay of the University of Illinois for his generous gift of *E. coli* strains knocked out for copper homeostasis mechanisms. *Conflict of Interest*: The authors declare that there are no conflicts of interest.

References

1. Cameron ED, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology. Nat Rev Microbiol 12:381–390.
2. Steensels J, Snoek T, Meersman E, Nicolino MP, Voordeckers K, Verstrepen KJ (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. FEMS Microbiol Rev 38:947–995.
3. Chong H, Yeow J, Wang I, Song H, Jiang R (2013) Improving acetate tolerance in Escherichia coli by rewiring its global regulator cAMP receptor protein (CRP). PLoS One 8:e77422
4. Jin H, Chen L, Wang J, Zhang W (2014) Engineering biofuel tolerance in non-native producing microorganisms. Biotechnol Adv 32:541–548.
5. Nies DS (1999) Microbial heavy-metal resistance. Appl Microbiol Biotechnol 51:730–750.
6. Freeman JL, Persans MW, Nieman K, Salt DE (2005) Nickel and cobalt resistance engineered in Escherichia coli by overexpression of serine acetyltransferase from the nickel hyperaccumulator plant *Thlaspi goesingense*. Appl Environ Microbiol 71:8627–8633.
7. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262:1680–1685.
8. Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) De novo proteins from designed combinatorial libraries. Protein Sci 13:1711–1723.
9. Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, Hecht MH (2003) Stably folded de novo proteins from a designed combinatorial library. Protein Science 12:92–102.
10. Bradley LH, Kleiner RE, Wang AF, Hecht MH, Wood DW (2005) An intein-based genetic selection enables construction of a high-quality library of binary patterned de novo sequences. Protein Eng Des Sel 18:201–207.
11. Wei Y, Kim S, Fela D, Baum J, Hecht MH (2003) Solution structure of a de novo protein from a designed combinatorial library. Proc Natl Acad Sci USA 100: 13270–13273.
12. Go A, Kim S, Baum J, Hecht MH (2008) Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. Protein Sci 17:821–832.
13. Arai R, Kobayashi N, Kimura A, Sato T, Matsuo K, Wang AF, Platt JM, Bradley LH, Hecht MH (2012) Domain-swapped dimeric structure of a stable and functional de novo four-helix bundle protein,WA20. J Phys Chem B 116:6789–6797.
14. Moffet DA, Certain LK, Smith AJ, Kessel AJ, Beckwith KA, Hecht MH (2000) Peroxidase activity in heme proteins derived from a designed combinatorial library. J Am Chem Soc 122:7612–7613.
15. Moffet DA, Case MA, House JC, Vogel K, Williams R, Spiro TG, McLendon GL, Hecht MH (2001) Carbon monoxide binding by de novo heme proteins from a designed combinatorial library. J Am Chem Soc 123: 2109–2115.
16. Wei Y, Hecht MH (2004) Enzyme-like proteins from an unselected library of designed amino acid sequences. Protein Eng Des Sel 17:67–75.
17. Patel SC, Bradley LH, Jinadasa SP, Hecht MH (2009) Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins. Protein Sci 18:1388–1400.
18. Das A, Wei Y, Pelczer I, Hecht MH (2011) Binding of small molecules to cavity forming mutants of a de novo designed protein. Protein Sci 20:702–711.
19. Cherny I, Korolev M, Koehler AN, Hecht MH (2012) Proteins from an unevolved library of de novo designed sequences bind a range of small molecules. ACS Synth Biol 1:130–138.
20. Fisher MA, McKinley KL, Bradley LH, Viola SR, Hecht MH (2011) De novo designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth. PLoS One 6:e15364
21. Smith BA, Mularz AE, Hecht MH (2015) Divergent evolution of a bifunctional de novo protein. Protein Sci 24:246–252.
22. Elguindi J, Hao X, Lin Y, Alwathnani HA Wei G, Rensing (2011) Advantages and challenges of increased antimicrobial copper use and copper mining. Appl Microbiol Biotechnol 91:237–249.
23. Rensing C, Grass G (2003) Escherichia coli mechanisms of copper homeostasis in a changing environment. FEMS Microbiol Rev 27:197–213.
24. Tetaz TJ, Luke RK (1983) Plasmid-controlled resistance to copper in *Escherichia coli*. J Bacteriol 154: 1263–1268.
25. Lee SM, Grass G, Rensing C, Barrett SR, Yates CJ, Stoyanov JV, Brown NL (2002) The Pco proteins are involved in periplasmic copper handling in *Escherichia coli*. Biochem Biophys Res Commun 295:616–620.
26. Rademacher C, Masepohl B (2013) Copper-responsive gene regulation in bacteria. Microbiology 158:2451–2464.
27. Macomber L, Rensing C, Imlay JA (2007) Intracellular copper does not catalyze the formation of oxidative DNA damage in *Escherichia coli*. J Bacteriol 189:1616–1626.
28. Macromber L, Imlay JA (2009) The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. Proc Natl Acad Sci USA 106:8344–8349.
29. Tan G, Cheng Z, Pang Y, Landry AP, Li J, Lu J, Ding H (2014) Copper binding in IscA inhibits iron-sulphur cluster assembly in *Escherichia coli*. Mol Microbiol 93: 629–644.
30. Rubino JT, Chenkin MP, Keller M, Riggs-Gelasco P, Franz KJ (2011) A comparison of methionine, histidine and cysteine in copper(I)-binding peptides reveals differences relevant to copper uptake by organisms in diverse environments. Metallomics 3:61–73.
31. Culotta VC, Joh HD, Lin SJ, Slekar KH, Strain JA (1995) Physiological role for *Saccharomyces cerevisiae* copper/zinc superoxide dismutase in copper buffering. J Biol Chem 270:29991–29997.
32. Shiraishi N, Nishikimi M (1998) Suppression of copper-induced cellular damage by copper sequestration with S100b protein. Arch Biochem Biophys 357:225–230.
33. Grass G, Rensing C (2001) CueO is a multi-copper oxidase that confers copper tolerance in *Escherichia coli*. Biochem Biophys Res Commun 286:902–908.
34. Mealman TD, Blackburn NJ, McEvoy MM (2012) Metal export by CusCFBA, the periplasmic Cu(I)/Ag(I) transport system of *Escherichia coli*. Curr Top Membr 69:163–196.
35. Fan B, Rosen BP (2002) Biochemical characterization of CopA, the *Escherichia coli* Cu(I)-translocating P-type ATPase. J Biol Chem 277:46987–46992.
36. Jacob F (1977) Evolution and tinkering. Science 196: 1161–1166.
37. Kunin V, Ouzounis CA (2003) The balance of driving forces during genome evolution in prokaryotes. Genome Res 13:1589–1594.
38. Soucy S, Huang J, Gogarten JP (2015) Horizontal gene transfer: building the web of life. Nat Rev Gen 16:472–482.
39. Blomberg R, Kries H, Pinkas DM, Mittl PR, Grütter MG, Privett HK, Mayo SL, Hilvert D (2013) Precision

is essential for efficient catalysis in an evolved Kemp eliminase. Nature 503:418–412.
40. Ge R, Watt RM, Sun X, Tanner JA, He QY, Huang JD, Sun H (2006) Expression and characterization of a histidine-rich protein, Hpn: potential for Ni^{2+} storage in *Helicobacter pylori*. Biochem J 393:285–293.
41. Watt RK, Ludden PW (1998) The identification, purification, and characterization of CooJ. A nickel-binding protein that is co-regulated with the Ni-containing CO dehydrogenase from *Rhodospirillum rubrum*. J Biol Chem 273:10019–10025.
42. Nishino K, Yamasaki S, Hayashi-Nishino M, Yamaguchi A (2010) Effect of NlpE overproduction on multidrug resistance in *Escherichia coli*. Antimicrob Agent Chemother 54:2239–2243.
43. Nikaido H (1996) Multidrug efflux pumps of gram-negative bacteria. J Bacteriol 178:5853–5859.
44. Koowatananukul C, Chansong N, Chuanchuen R (2010) The contribution of active efflux in reduced susceptibilities to copper sulfate and zinc chloride in *Escherichia coli* isolates from swine. Thai J Vet Med 40: 317–321.
45. Wilson D, Keefe AD (2001) Random Mutagenesis by PCR, Chapter 8, Unit 8.3. In: Current Protocols in Molecular Biology. John Wiley & Sons, Inc. http://dx.doi.org/10.1002/0471142727.mb0803s51
46. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008