OPEN

REP-X: An Evolution-guided Strategy for the Rational Design of Cysteine-less Protein Variants

Kevin Dalton[1,4], Tom Lopez[2], Vijay Pande[1,3] & Judith Frydman[1,2]*

Site-specific labeling of proteins is often a prerequisite for biophysical and biochemical characterization. Chemical modification of a unique cysteine residue is among the most facile methods for site-specific labeling of proteins. However, many proteins have multiple reactive cysteines, which must be mutated to other residues to enable labeling of unique positions. This trial-and-error process often results in cysteine-free proteins with reduced activity or stability. Herein we describe a general methodology to rationally engineer cysteine-less proteins. Briefly, natural variation across orthologues is exploited to identify suitable cysteine replacements compatible with protein activity and stability. As a proof-of-concept, we recount the successful engineering of a cysteine-less mutant of the group II chaperonin from methanogenic archaeon *Methanococcus maripaludis*. A webapp, REP-X (Replacement at Endogenous Positions from eXtant sequences), which enables users to design their own cysteine-less protein variants, will make this rational approach widely available.

Modification of a unique cysteine residue with a maleimide or iodoacetamide conjugated probe is a widely used method of producing site-specifically labeled proteins. These two chemistries have a long history in the literature starting from the first description of the reaction between iodoacetamide and thiols in 1933[1,2]. The value of this reaction to protein biochemistry was reified shortly thereafter in the first published reports of iodoacetamide's reaction with purified, native proteins[3,4]. The promise of maleimide chemistry was confirmed in 1949 when the reaction between substituted maleimides and thiols was shown to proceed at room temperature in water on the minute timescale with a nearly stoichiometric yield[5]. Shortly thereafter, the reaction of N-ethylmaleimide with actin was reported constituting the first published study to demonstrate that substituted maleimides react specifically with purified proteins[6]. In the following decade, creative applications of maleimide chemistry began appearing. The first deployment of bis-maleimides as a crosslinking agent was published in 1956[7]. Four years later, the first chromophore conjugated maleimide was used as an indicator for cysteine residues in proteolysis fragments of serum albumin[8]. Because of their high reaction yields under mild conditions, early biochemists appreciated the potential of substituted iodacetamides and maleimides to label proteins natively.

Today, cysteine directed probes are ubiquitous in protein science. Common substituents include crosslinkers, fluorophores, affinity tags such as biotin or digoxigenin, nitroxide spin labels, and silanes for surface immobilization. Such thiol-reactive probes are used in a range of applications including bulk kinetic measurements, single molecule biophysics, surface plasmon resonance, nuclear magnetic resonance and electron spin resonance spectroscopy[9]. A growing number of thiol-reactive probes are commercially available. As of writing, a search of Thermo Fisher's online catalog alone contains over 90 thiol-reactive labeling kits and at least 13 thiol-reactive crosslinkers. Site-specific labeling has benefited the study of many protein families. However, in order to label thiols, one must first generate single cysteine mutants of a protein of interest. Typically, cysteines are replaced by serines or alanines, in an empirical trial-and-error process that often yields modified cysteine-less mutants that have reduced activity or stability (Fig. 1A). Hence, a rational method to produce active cysteine-less protein variants would be quite desirable to contemporary protein science.

Along these lines, we sought to generate a cysteine-less variant of a well studied group II chaperonin, MmCpn. Chaperonins are ATP-driven components of the cellular protein folding machinery and are essential in all free-living organisms. They comprise two structurally related chaperonin families generally called group I and

[1]Biophysics Program Stanford University, Stanford, California, USA. [2]Department of Biology Stanford University, Stanford, California, USA. [3]Department of Chemistry, Stanford University, Stanford, California, USA. [4]Present address: Department of Molecular and Cellular Biology Harvard University, Cambridge, MA, 02138, USA. *email: jfrydman@stanford.edu

1
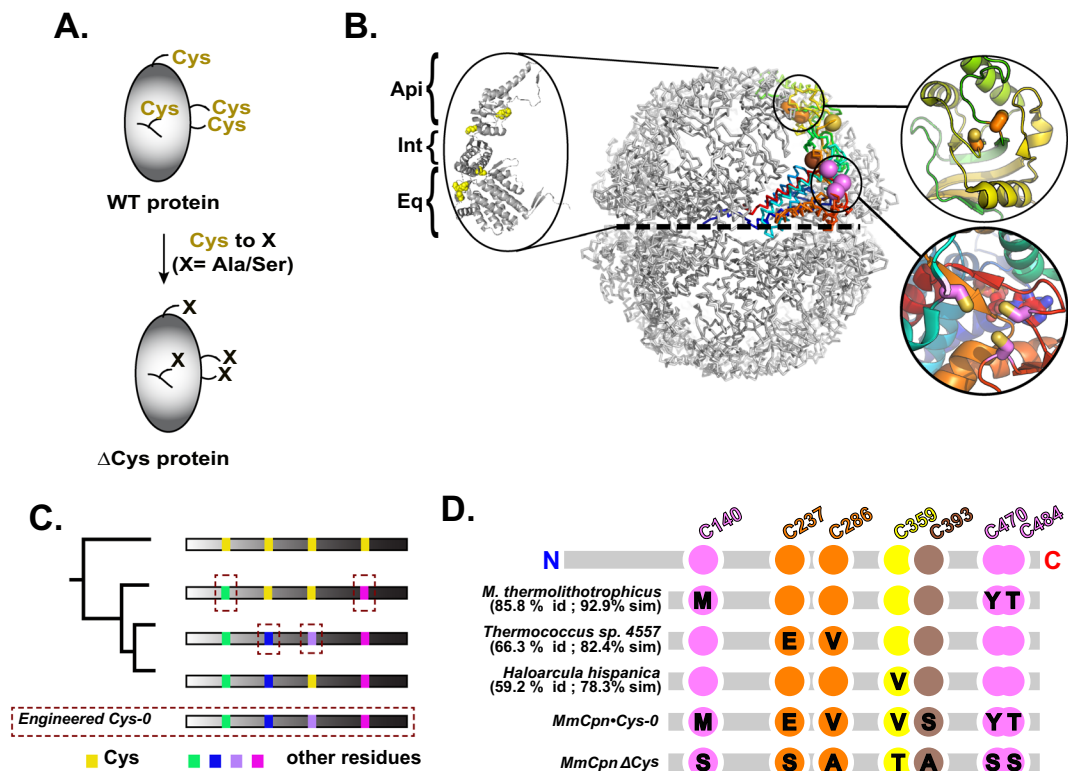
Figure 1. Design of Mm-Cys-0, a cysteine-less chaperonin variant. (**A**) Strategy for the design of Cysteine-less protein variants. (**B**) Molecular architecture of group II chaperonin Mm-Cpn. The crystal structure of the closed state of MmCpn-RLS mutant bound to ADP-AlFx (PDBID: 3RUW[24]). Highlighted are the equator of the complex (dashed line) and the three structural domains. A single subunit is shown in color with the seven cysteine residues rendered as spheres. Inset: cysteine clusters in the apical (top) and equatorial domains (bottom) (**C**) Design strategy of a cysteine-less MmCpn. Positions of seven Cys residues in Mm-Cpn WT are highlighted in yellow. (**D**) Extant sequences yielding the substitutions used to create the Cys-0 variant are shown along with the corresponding percentage identity and similarity between their sequence and MmCpn (see complete alignments in Supplemental Data). The sequence of MmCpn●Cys-0 (this work) is shown as well as the sequence of MmCpnΔCys[16].

group II chaperonins. Group I chaperonins are present in prokaryotes and the endosymbiotic organelles, chloroplasts and mitochondria[10]. Group I chaperonins are also, albeit comparatively rarely, found in archaeal genomes. The study of the archetypal group I chaperonin, GroEL from Escherichia coli, has long benefitted from the availability of a cysteine-less variant[11] that enabled many mechanistic studies. Even for this successful example, the cysteine-less GroEL variants exhibit impaired activity compared to wild type[11]. Comparable studies have not been possible thus far in the group II chaperonins.

The group II chaperonins are found exclusively in the archaeal and eukaryotic cytosol[12]. Broadly, they are a family of ATP-driven hexadecameric chaperones with two stacked, eight membered rings per complex. These rings are related by a two-fold symmetry axis, which lies along the equator of the complex (Fig. 1B, dashed line). Orthogonal to the two-fold axis is the eight-fold axis, which defines the intra-ring symmetry of the complex. The eukaryotic group II chaperonin, termed TRiC or CCT comprises eight parologous subunits of which four bind ATP. By contrast, the archaeal group IIs are prototypically $\alpha_8\beta_8$ complexes. Other configurations have been reported including group IIs with 9-fold symmetry[13,14]. Here we will discuss the homohexadecameric group II chaperonin from *Methanococcus maripaludis*, MmCpn, which has proven itself to be a very useful model group II chaperonin.

All group II chaperonins share a conserved subunit architecture. The equatorial domain contains the nucleotide binding pocket. The remaining two structural domains of the monomer are the intermediate and apical domains arranged in ascending order from the complex equator (Fig. 1B). The ring interface of MmCpn is formed by the equatorial domain of the chaperonin. The catalytic residue for ATP hydrolysis, D386 in MmCpn, is located in the intermediate domain. The apical domain binds misfolded client proteins. ATP hydrolysis leads to the release of client from the apical domains into the central cavity of the chaperonin. At this point in time, the apical domains close over the client forming an 8-fold symmetric iris and shielding the client from the bulk cytosol such that it my fold in isolation.

MmCpn monomers are 543 residues long and contain seven cysteines distributed throughout the primary sequence in all three structural domains of the enzyme. Cys-140, 470, & 484 are located in the equatorial domain

near the ATP binding pocket. There are two cysteines in the intermediate domain, Cys-359 & 393. Finally, the apical domain contains a pair of cysteines, residues C237 & 286. The three equatorial cysteines are in steric contact with one another (Fig. 1B bottom inset). Additionally the two apical cysteines are also in close proximity (Fig. 1B top inset).

Previous efforts demonstrated that a cysteine-less mutant of MmCpn could be produced from a combination of serine, alanine and threonine point mutations[15,16]. This MmCpn mutant, termed ΔCys (MmCpnΔCys), assembles correctly and can bind unfolded proteins. It can also close its lid upon the addition of ADP-AlFx, which acts as a transition state analog for the ATP hydrolysis reaction. However, MmCpnΔCys is very poorly expressed in *E. coli*, and thus unsuitable for applications requiring milligram quantities of protein. To engineer a stable and highly expressed cysteine-less MmCpn variant for structural and biophysical work, we sought an alternative strategy to identify position-specific cysteine substitutions. We developed a rational approach using natural variation among extant ortholog sequences, which yielded a well expressed chaperonin with wild-type client folding activity. This effort provides proof-of-concept for our strategy to engineer amino acid replacements in proteins.

Results

A rational strategy to design a cysteine-less chaperonin variant.
In order to design a functional, well-expressed cysteine-less chaperonin mutant we sought to integrate available structural information about MmCpn with phylogeny, using extant sequence information for homologous archaeal chaperonins. We reasoned that amino acid substitutions derived from close MmCpn homologs would be well tolerated by this chaperonin. To this end we amassed 2981 group II chaperonin sequences from the NCBI non-redundant protein sequence database using BLAST[17]. Each of these extant chaperonin sequences was aligned to the WT MmCpn sequence using the program Water from the EMBOSS Suite[18]. Water implements the Smith-Waterman sequence alignment algorithm which guarantees optimal pairwise local sequence alignment given an affine gap penalty[19]. To better interpret the information from the alignments, the sequences in the database were then rank ordered by their percentage identity with MmCpn.

The overall strategy of our approach to identify compatible amino acid replacements for each cysteine is outlined in Fig. 1C. We carried out alignments to query the chaperonin sequence database for the most closely related sequence with a substitution at each cysteine position. These analyses informed the choice of residues to substitute to create a cysteine-less functional variant. One challenge in the implementation of this strategy was presented by sets of cysteines in steric contact, which in MmCpn are present in the equatorial and apical domains. Each of these sets consisted of a cluster of several cysteines in steric contact in the MmCpn structure. We were concerned that mutation of one of these residues would require compensatory mutations in the other(s) that maintained the structural integrity of their contact. Accordingly, when the database was queried for the equatorial cysteines (Fig. 1B,D Magenta), we restricted hits to sequences with substitutions at all three positions. Likewise, for the two cysteines in the apical domain (Fig. 1B,D orange), we restricted hits to chaperonin sequences with both residues substituted.

Searching our database for non-cysteine variants of the equatorial cysteine positions C140, 470 & 484 returned a sequence from *Methanococcus thermolithotrophicus* with overall 85.8% identity to MmCpn but with non-cysteine residues at all three cysteine positions (Fig. 1D, Magenta; see alignments in Supplemental Data). Notably, this was the most closely related sequence in the database with non-cysteine variants at any of the three equatorial cysteine residues, suggesting that a mutation at any one of these positions necessitates a compensatory mutation at the other sites. The identities of the corresponding mutations in MmCpn are C140M, C470Y and C484T (Fig. 1D). Similarly, searching for coincident non-cysteine variants of the two apical Mm-Cpn cysteines yielded a sequence containing the substitutions C237E and C286V from *Thermococcus sp. 4557* which is 66.3% identical to MmCpn (Fig. 1D; see alignments in Supplemental Data).

The remaining two cysteines are in the intermediate domain. Cys-359 was found to be mutated to valine in *Haloarcula hispanica* which is 59.2% identical to MmCpn. This was the most closely related organism with any substitution at that location. Finally, Cys-393 exhibited several amino acid substitutions in closely related sequences (Figure S1; see alignments in Supplemental Data). As such, we chose to incorporate serine at this position because it is both present in extant sequences and is isosteric with cysteine. Notably, MmCpnΔCys incorporates an alanine at this position, which is also well represented in the database of homologs. All the above substitutions were introduced into MmCpn to generate a new cysteine-less variant herein termed Cys-0 (MmCpn●Cys-0).

The cysteine-less MmCpn●Cys-0 is well expressed and supports ATP-driven cycling.
While the previously derived MmCpnΔCys could assembled correctly, it did not express well in *Escherichia coli*. This is illustrated by experiments where wild type (WT) and ΔCys variants of MmCpn were expressed from pET21a plasmids in *E. coli* BL-21. Following induction with IPTG, expression was monitored by SDS-PAGE of total lysates (Fig. 2A). WT MmCpn was highly expressed in soluble form and clearly visualized by Sypro Ruby staining, while the previously described MmCpnΔCys mutant protein was barely perceptible in the lysate. Importantly, the same experiment demonstrated that the rationally designed MmCpn●Cys-0 variant was produced at much higher expression levels, comparable to WT (Fig. 2A). In a further test of assembly competence, we examined whether the MmCpn variants remained soluble after treatment of the lysate with 55% ammonium sulfate (AS), as observed for the WT MmCpn (Fig. 2B). Indeed, both WT and Cys-0 variants, but not the ΔCys variant, were very soluble in 55% ammonium sulfate (Fig. 2B). Accordingly MmCpn●Cys-0 could be purified with the MmCpn purification protocol established for WT.

We next examined whether MmCpn●Cys-0 has ATPase activity and can undergo the ATP-driven conformational change typical of Group II chaperonins. ATP hydrolysis promotes a conformational change from an open, client binding conformation to a closed configuration, characterized by formation of a built-in lid over the
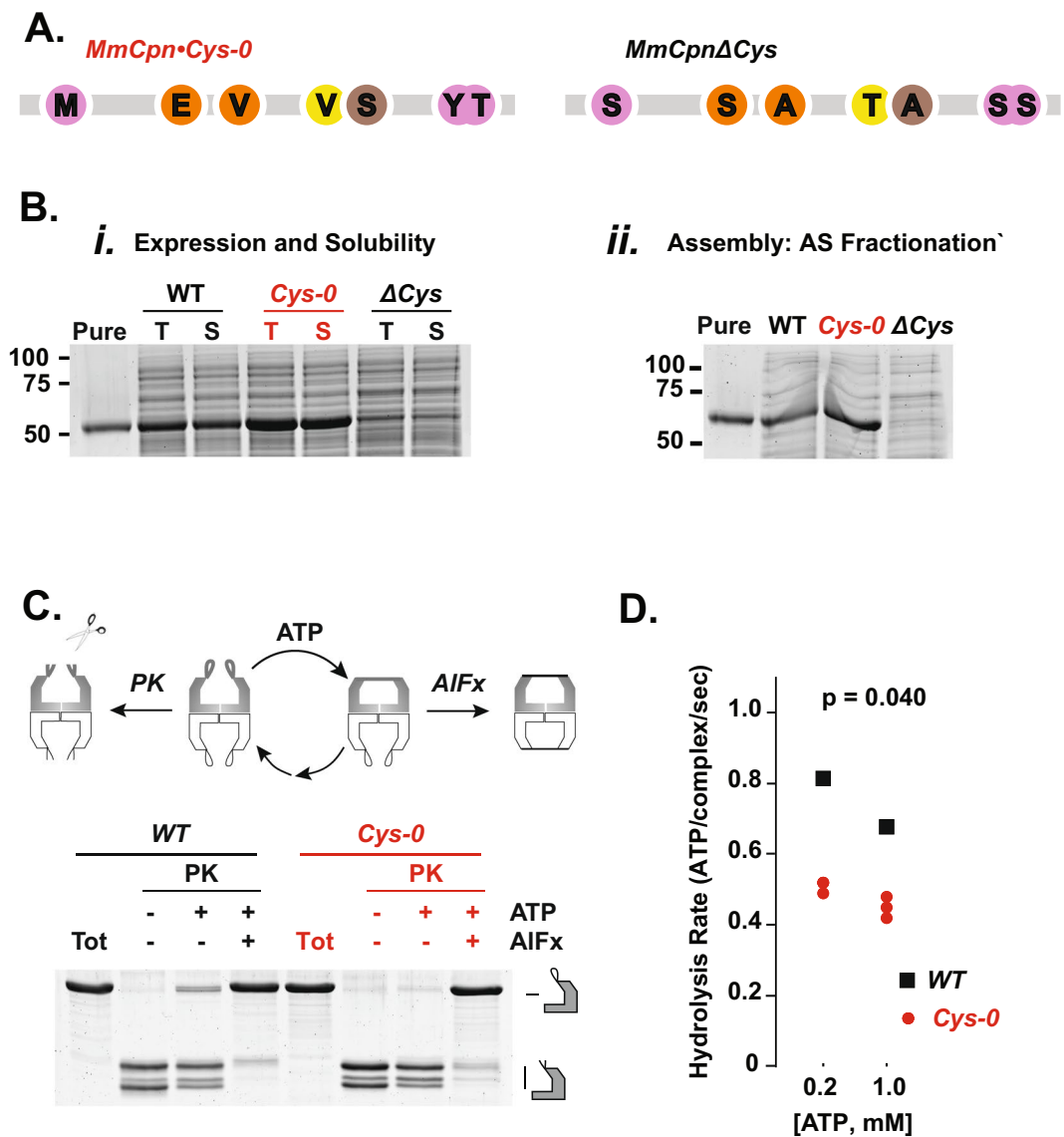
Figure 2. MmCpn•Cys-0 is well expressed and has ATPase activity. (**A**) Cys replacements in MmCpn•Cys-0 and MmCpnΔCys. (**B**) MmCpn variant expression and Assembly. *i*. Test expression of chaperonin variants. Lane 1, 1 µg of purified WT MmCpn. Other lanes contain 2 µl of each sample. T indicates total lysate while S indicates soluble proteins following centrifugation at 22,000 × g for 30 minutes. *ii*. Ammonium sulfate (AS) fractionation as a test of complex assembly. SDS-PAGE of 2 µl of test expressions following a 55% ammonium sulfate (AS) fractionation. (**C**) Proteinase K (PK) protection assay for purified chaperonin mutants. All gels stained with Sypro Ruby (**D**) ATPase values for WT and MmCpn•Cys-0.

central chamber of the complex (Fig. 2C). To this end, we exploited the observation that, in the open state, the lid segments are sensitive to non-specific proteolytic digestion by proteinase K, while in the closed state these lid segments are strongly protease-protected (Fig. 2C). Inclusion of AlFx, a transition state analogue, in the ATP hydrolysis reaction stabilizes the closed conformation, rendering the chaperonin resistant to Proteinase K. This assay demonstrated that incubation with ATP and AlFx yielded similar levels of protection to both WT and Cys-0 chaperonins, indicating the MmCpn•Cys-0 mutant can fully attain the closed conformation in response to ATP (Fig. 2C). As described below (Fig. 3A), native agarose gel electrophoresis confirmed this conclusion and provided additional support for the ATP-induced formation of the closed MmCpn conformation (Fig. 3A bottom panel).

Without the addition of AlFx, the chaperonin will freely hydrolyze ATP and cycle between the open and closed conformation in the process. The amount of proteinase K protection observed in this state thereby reports on the fraction of its nucleotide cycle the chaperonin spends in the open, proteinase sensitive state. Under cycling conditions, the Cys-0 variant was less proteinase-protected than WT, suggesting that MmCpn•Cys-0 spends more time in the open state during its nucleotide cycle (Fig. 2C). This observation is consistent with a decreased rate of chamber closure or an increase in the rate of chamber re-opening. To evaluate the likelihood of each explanation,
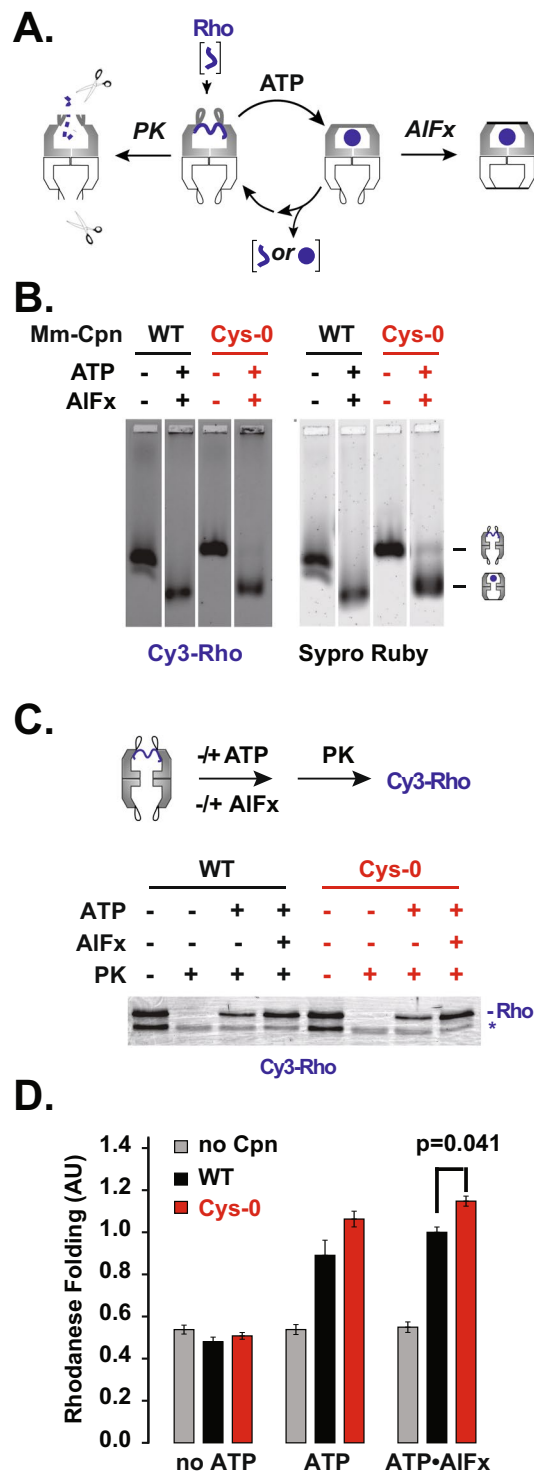
Figure 3. MmCpn●Cys-0 can effectively mediate substrate folding. (**A**) Scheme of chaperonin folding cycle (**B**) Non-denaturing agarose electrophoresis to monitor binding MmCpn and the substrate protein rhodanese. Cy3-maleimide labeled rhodanese co-migration with MmCpn in different nucleotide states (top). Total protein stained with Sypro Ruby. (**C**) Protease protection of Cy3-rhodanese by MmCpn in different nucleotide states. Rho: Rhodanese; *: background band. (**D**) Rhodanese folding of WT and Cys-0 MmCpn. The y-axis represents the amount of cyanide converted to thiocyanate by rhodanese in 10 minutes relative to the mean amount converted by the wild type sample with ATP-AlFx. Cys-0 and WT do not significantly differ under actively cycling conditions (p-value by two sided t-test for ATP: $p = 0.134$) but Cys-0 folds rhodanese with slightly higher yield than wild type chaperonin under ATP-AlFx conditions (p-value for ATP*AlFx: $p = 0.0409$). This is based on $n = 3$ measurements for both conditions.

we measured the ATP hydrolysis rate of MmCpn●Cys-0 compared to WT. For all ATP concentrations tested, the Cys-0 variant hydrolyzed ATP at comparable, albeit slightly lower rates that WT MmCpn (Fig. 2D). This result strongly suggests that the decreased level of proteinase protection seen in cycling Cys-0 is due to a decreased rate of closure and not faster re-opening which would manifest and increased ATPase rate.

Of note, both WT and Cys-0 chaperonins exhibit the previously described inhibition of ATP hydrolysis rates at high ATP concentrations (compare 200 and 1000 micromolar ATP in Fig. 2D, Supplemental Table 1). This indicated that MmCpn●Cys-0 preserves the allosteric regulation of ATP hydrolysis characteristic of group II chaperonins.

MmCpn●Cys-0 is fully functional in ATP-dependent substrate folding.

Next, we examined the ability of MmCpn●Cys-0 to mediate protein folding. Bovine rhodanese, responsible for detoxifying cyanide ions *in vivo*, is a well characterized model substrate for MmCpn. Once denatured, rhodanese cannot fold without chaperone assistance, forming insoluble aggregates instead. However, denatured Rhodanese diluted into buffer containing WT MmCpn binds to the chaperonin, forming a characteristic high molecular weight complex (Fig. 3A). Upon addition of ATP or ATP-AlFx to induce lid closure, MmCpn encapsulates rhodanese within its central chamber. Rhodanese folds to the native active state within the central chaperonin chamber, and is released upon reopening of the chamber following ADP and phosphate release (Fig. 3A)[20]. With ATP-AlFx lid reopening is blocked, and the encapsulated substrate remains locked within the closed chaperonin chamber.

To compare the abilities of WT and Cys-0 Mm-Cpn variants to fold Rhodane in an ATP-dependent manner, we dissected substrate engagement at different steps of the ATP-driven MmCpn folding cycle (Fig. 3B–D). We first examined substrate binding using denatured cy3-labeled rhodanese. MmCpn binding was monitored by comigration with Rhodanese on non-denaturing gel electrophoresis. Protein staining visualized the slowly migrating MmCpn (Sypro Ruby, Figs 3B and S2) while fluorescence imaging visualized Cy3-rhodanese (Fig. 3B). Both variants formed the characteristic Rhodanese-Cpn complex in the absence of ATP, with similar efficiency. Upon incubation with ATP-AlFx, both WT and Cys-0 similarly changed their mobility to the faster migrating lid-closed state. Since WT and Cys-0 comparably retained the bound substrate in this analysis, we concluded MmCpn●Cys-0 effectively encapsulated substrate upon closure.

The ability of Cys-0 to encapsulate the bound Rhodanese was tested directly using a protease sensitivity assay. A rhodanese-Cpn complex was treated with Proteinase K following incubation in the absence or presence of ATP or ATP-AlFx (Fig. 3C). In the open state, the Cpn-bound non-native protein is unstructured and extremely protease sensitive. However, upon ATP-induced lid closure, the encapsulation of the substrate within the chamber results in full protection from proteolysis. As expected, rhodanese bound to either WT or Cys-0 MmCpn was protease sensitive. Incubation of either WT or Cys-0 with ATP-AlFx or with ATP led to full protection of the bound substrate from proteinase K digestion. This indicates that ATP leads to efficient encapsulation of the bound substrate within the closed central chamber both for WT MmCpn and for the Cys-0 variant (Fig. 3C).

Finally, we compared the ATP-dependent production of folded Rhodanase when bound by MmCpn●Cys-0 versus MmCpn WT (Fig. 3D). Rhodanese folding results in active enzyme that can be measured using a colorimetric assay. The rhodanese-Cpn complex was incubated in the absence or presence of ATP or ATP-AlFx, and the extent of rhodanese folding was measured using an endpoint assay. Importantly, the cysteine-less MmCpn●Cys-0 promoted rhodanese folding as effectively as WT MmCpn both under cycling conditions, i.e. in the presence of ATP, or when the complex is locked in the closed state with ATP-AlFx. Together, these experiments indicate that our approach enabled the design of a well folded, assembled and active cysteine-less chaperonin complex.

Discussion

We here describe a rational bioinformatic approach to engineer active cysteine-less variants of proteins using an analysis of extant sequences. Our premise is to exploit natural variation to identify amino acid substitutions compatible with a protein's structure and activity. Unless cysteines are essential for protein catalysis or to generate a specific structural or functional motif, such as a Zn-finger, it is likely that examination of orthologues of a protein of interest will identify suitable cysteine replacements compatible with its structure and activity. We documented the success of our strategy by engineering a cysteine-less variant of the group II chaperonin MmCpn. To our great surprise, this technique allowed us to define seven simultaneous point mutants that together could replace the cysteines in the protein while retaining a functional, well expressed complex. The unintuitive nature of the point mutations selected by this approach is worth emphasizing.

The common practice to eliminate cysteines by mutation to alanine or the isosteric amino acid serine often fails to yield active or stable protein variants. This is illustrated by analysis of the previously obtained cysteine-less MmCpn mutant ΔCys, which is less stable and less well expressed, indicating this approach is not the best choice for the chaperonin system. Furthermore, to enumerate and experimentally explore all possible, 20^7 (1.28 trillion), amino acid combinations is too onerous to be feasible. A remarkable feature of our approach is its ability to identify spatially or structurally linked replacements. For instance, this allowed us to replace cysteine clusters that are in steric contact in the folded protein. By identifying sets of non-cysteine residues that maintained the interaction in orthologous proteins, our approach permitted the replacement of the entire cysteine cluster without loss of protein integrity or activity. To make this engineering approach available to other investigators we developed a software tool which we named Replacement at Endogenous Positions from eXtant sequences (REP-X). The current (alpha) version of the software is available at (http://github.com/kmdalton/rep-x). The basic design of the application is summarized in Fig. 4. We predict this software will enable other protein scientists to rapidly and rationally design cysteine-less protein variants, opening the way to introducing cysteines at specific positions for site specific labeling.

While Rep-X should be broadly applicable for proteins which the cysteines are not essential for functions, the approach relies on there being some amount of diversity at a given position. Thus, Rep-X cannot provide
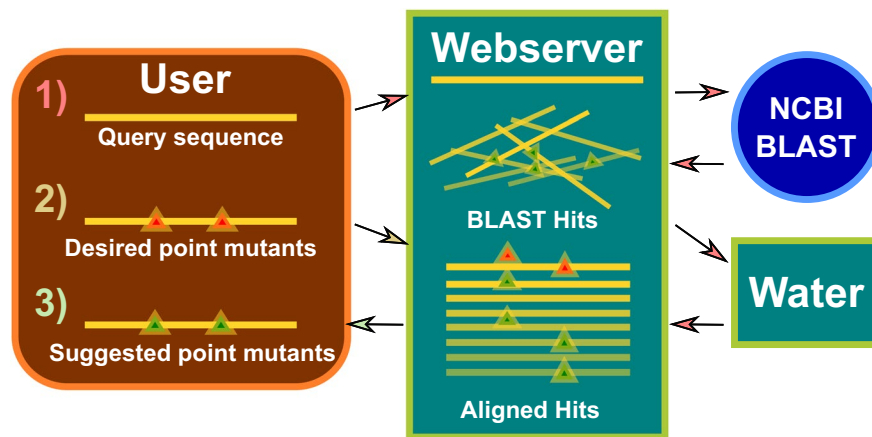
Figure 4. Design of the Rep-X Webserver. (**1**) Query sequences are submitted to the webserver by users. BLAST[17] hits are retrieved by the NCBI server and aligned to the query sequence using Water from the EMBOSS Suite[18]. (**2**) Point mutants are requested by the user, and (**3**) point mutants are suggested by the webserver.

alternatives for residues that are 100% conserved across all available sequences, as may be the case of cysteines participating in essential structural motifs, such as Zinc-fingers, or involved in active site catalysis.

An important consideration for our approach, is that there is no intrinsic limitation on the amino acid type being replaced. REP-X can equally aid in the design of single tryptophan mutants for intrinsic fluorescence studies, be used to generate single lysine variants for site specific labeling by N-hydroxysuccinimide conjugates or to generate specific single amino acid variants for position specific labelling with stable isotopes for NMR experiments. Our results suggest that this technique can be particularly powerful in systems for which structural data are available in order to identify cysteines which are in close proximity. Future iterations of REP-X may include novel features, for instance allowing to predict sites for cysteine incorporation in a protein.

Materials and Methods

MmCpn expression constructs. MmCpn WT and MmCpnΔCys plasmids derive from previous work. Briefly, the sequence of the chaperonin was cloned into the Nde1 and BamHI of pET21a+ (EMD Chemicals) yielding pET21MmCpnWT[21]. MmCpnΔCys was obtained from D. Hoersch and T. Kortemme[16]. The C-terminal polyhistidine tag of MmCpnΔCys was removed by the introduction of a stop codon using the QuickChange II site-directed mutagenesis kit (Agilent Technologies). Codon optimized MmCpn•Cys-0 in the pJ414 vector was obtained from DNA2.0. All relevant coding sequences are included in the Supplemental Fasta file.

Buffer composition. Three standard chaperonin buffers were used in this work. MQA: 20 mM HEPES-KOH (pH 7.4), 50 mM NaCl, 5 mM MgCl$_2$, 0.1 mM EDTA, 10% (v/v) glycerol. MQB: 20 mM HEPES-KOH (pH 7.4), 1 M NaCl, 5 mM MgCl$_2$, 0.1 mM EDTA, 10% (v/v) glycerol. ATPase buffer: 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM MgCl$_2$, 10% (v/v) glycerol). Buffers were supplemented with 1 mM DTT and 0.5 mM PMSF during lysis.

Expression tests of chaperonin mutants. *E. coli* BL-21 pRosettas containing the three expression vectors were cultured in terrific broth with 100 ug/ml ampicillin and 30 ug/ml chloramphenicol overnight then diluted 1:1000 into 50 ml of fresh media. These cultures were grown at 37 C with 250 rpm shaking until reaching an OD600 between 0.9 and 1.0. The cultures were induced with 1 mM IPTG and transferred to a 16 C incubator overnight. The following morning the cells were pelleted by centrifugation at 5,000 × g for 10 minutes and flash frozen in liquid nitrogen.

Frozen cell pellets were resuspended in 10 ml of MQA buffer and lysed in an Emulsiflex-B15 (Avestin). Lysates were cleared by centrifugation at 22,000 × g for 30 minutes. Ammonium sulfate was added to the cleared lysate to a final concentration of 55% saturation. After a 15 minute incubation at 4 C, the insoluble fraction was removed by centrifugation at 22,000 × g for 30 minutes. 2 μl of the lysed, cleared, and ammonium sulfate soluble fractions were loaded onto 15% Tris-glycine SDS Polyacrylamide gels and electrophoresed at 25 mA for 1 hour. Gels were stained with Sypro Ruby (ThermoFisher) and imaged on Typhoon scanner (GE).

Chaperonin purification. Expression, lysis, and ammonium sulfate cuts were performed as described in the previous section excepting that the culture volumes were increased to 1 L. Anion exchange and heparin affinity chromatography were performed as described previously[21]. After heparin chromatography, MmCpn solutions were concentrated in 15 ml centrifugal concentrators with a 100 kDa nominal molecular weight cutoff, buffer exchanged at least 3 times into MQA, concentrated to between 80 and 150 mg/ml, flash frozen in liquid nitrogen and stored at −80 C.

ATPase assay.   A previously described enzyme coupled assay[22] was used to measure the ATPase activity of MmCpn●Cys-0. 180 μl of ATPase buffer containing 10 mM phosphoenolpyruvate, 1.5 mM NADH, 5 units of pyruvate kinase (rabbit muscle type VII buffered aqueous glycerol solution, Sigma-Aldrich), 4.6 units of l-lactic dehydrogenase (bovine heart type XVII buffered aqueous glycerol solution, Sigma-Aldrich), and 0.25 μM MmCpn was warmed to 37 C on a heat block. This was mixed rapidly with the indicated concentration of ATP and pipetted into a quartz cuvette which was pre-equilibrated at 37 C in a Hewlett Packard 8453 spectrophotometer. Conversion of NADH to NAD+ was monitored by a decrease in absorbance at 340 nm for 120 seconds. The linear region of the timecourse was fit and the slope used to calculate the rate of ATP hydrolysis by MmCpn.

Proteinase K assays.   Assays were performed as described previously[21] except that the chaperonin nucleotide reaction incubation time was extended to 60 minutes at 37 C. Briefly, ATPase buffer containing 0.25 μM MmCpn was pipetted onto the appropriate ATP conditions (no ATP, 1 mM ATP, 1 mM ATP + 1 mM AlNO3 + 6 mM NaF), and incubated at 37 C for 1 hour. Subsequently, 400 ng of proteinase K was added and the chaperonin was digested at room temperature for 10 minutes. The reaction was quenched by the addition of 0.05 M phenylmethanesulfonyl fluoride. The samples were electrophoresed on a 15% tris-glycine SDS-PAGE gel and stained with Sypro Ruby (Thermo Fisher).

Cy-3 rhodanese protection was performed just as the standard proteinase K protection assay except that 0.25 μM chaperonin was first added to 1 μM final concentrated rhodanese dissolved in 6 mM guanidium buffered with 100 mM HEPES pH 7.4 for 10 minutes on ice and subsequently centrifuged for 10 minutes at top speed in a microcentrifuge. This sample was applied to the nucleotide conditions, proteinase, and quenched with PMSF as described above.

Rhodanese folding assay.   Rhodanese folding was performed as described previously[20]. Rhodanese was diluted into chaperonin containing buffer with a final molar excess of 4:1 rhodanese to Cpn. Most rhodanese aggregates during this process. While the average number of rhodanese monomers per chaperonin is still poorly characterized, the final complex is approx. 1:1 complex. Briefly, 100 μM denatured rhodanese in 100 mM HEPES pH 7.4 with 6 M guanidinium chloride and 5 mM DTT was diluted 1:100 into ATPase buffer containing 0.25 μM MmCpn. Samples were incubated on ice for 10 minutes followed by a 10 minute spin at top speed in a 4 C refrigerated microcentrifuge to pellet aggregated rhodanese. The samples were pipetted onto nucleotide conditions containing no ATP, 5 mM ATP, or 5 mM ATP+ 1 mM AlNO3+ 6 mM NaF and incubated for 60 minutes at 37 C. Rhodanese activity was then assayed colorimetrically as described previously[23]. A single 10 minute time point was used. 0.5 M EDTA was substituted for 0.4 M CDTA.

Non-denaturing agarose gel electrophoresis.   Samples were prepared as in the proteinase protection assays, but rather than being digested, they were loaded onto a 1% agarose gel with 1 mM $MgCl_2$ and 80 mM MOPS pH 7.4, and electrophoresed for 1 hour at 100 volts and optionally stained with Sypro Ruby (Thermo Fisher) before imaging staining on a Typhoon Scanner (GE).

Rhodanese purification.   10 mg of bovine rhodanese (Sigma) was resuspended in 4 ml of buffer A (50 mM acetate pH 5.0 with 20 mM sodium thiosulfate). Undissolved material was removed by centrifugation at top speed in a microcentrifuge. The supernatant was applied to a Mono-S chromatography column and eluted over a linear gradient into buffer B (50 mM acetate, pH 5.0, 20 mM sodium thiosulfate, and 500 mM NaCl). The fraction containing the most rhodanese was further purified by size exclusion chromatography on an Superdex-75 column pre-equilibrated with buffer A. Resulting fractions were dehydrated in a speedvac and resuspended to a final concentration of 100 μM in 6 M guanidinium with 100 mM HEPES at pH 7.4.

Cy3-rhodanese labeling.   After Mono-S chromatography, 400 μl of rhodanese solution was reserved from the purification. This solution was added onto a solution of Cy3-Maleimide dissolved in 20 μl of DMSO and incubated at room temperature for 2.5 hours. The labeling reaction was quenched with 10 μl of 1 M DTT. This solution was then purified by size exclusion and denatured at described above.

Data availability
REP-X available at http://github.com/kmdalton/rep-x

References
1. Dickens, F. Interaction of halogenacetates and SH compounds: The reaction of halogenacetic acids with glutathione and cysteine. The mechanism of iodoacetate poisoning of glyoxalase. *Biochem. J.* **27**, 1141–1151 (1933).
2. Rapkine, L. Sur l'action de l'acide monoiodacétique sur le glutathion et la cystéine. *Compt. Rend. Soc. Biol.* **112**, 790–792 (1933).
3. Anson, M. L. The reactions of iodine and iodoacetamide with native egg albumin. *J. Gen. Physiol.* **23**, 321–331 (1940).
4. Smythe, C. V. The reaction of iodoacetate and of iodoaceta- mide with various sulfhydryl groups, with urease, and with yeast preparations. *The Journal of Biological Chemistry* **114**, 612–601 (1936).
5. Friedmann, E. Spectrophotometric investigation of the interaction of glutathione with maleimide and n-ethylmaleimide. *Biochim. Biophys. Acta* **9**, 65–75 (1952).
6. Tsao, T. C. & Bailey, K. The extraction, purification and some chemical properties of actin. *Biochim. Biophys. Acta* **11**, 102–113 (1953).
7. Moore, J. E. & Ward, W. H. Cross-linking of Bovine Plasma Albumin and Wool Keratin. *J. Am. Chem. Soc.* **78**, 2414–2418 (1956).
8. Witter, A. & Tuppy, H. N-(4-Dimethylamino)-3,5-Dinitrophenyl)maleimide: A coloured sulfhydryl reagent. *Biochim. Biophys. Acta* **45**, 429–442 (1960).

9. Hubbell, J. H. *et al.* Atomic form factors, incoherent scattering functions, and photon scattering cross sections. *J. Phys. Chem. Ref. Data* **4**, 471–538 (1975).
10. Hayer-Hartl, M., Bracher, A. & Hartl, F. U. The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends Biochem. Sci.* **41**, 62–76 (2016).
11. Rye, H. S. *et al.* GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings. *Cell* **97**, 325–338 (1999).
12. Lopez, T., Dalton, K. & Frydman, J. The Mechanism and Function of Group II Chaperonins. *J. Mol. Biol.* **427**, 2919–2930 (2015).
13. Marco, S. *et al.* The molecular chaperone TF55: Assesment of symmetry. *FEBS Lett.* **341**, 152–155 (1994).
14. Chaston, J. J. *et al.* Structural and Functional Insights into the Evolution and Stress Adaptation of Type II Chaperonins. *Structure* **24**, 364–374 (2016).
15. Hoersch, D. & Kortemme, T. A Model for the Molecular Mechanism of an Engineered Light-Driven Protein Machine. *Structure* **24**, 576–584 (2016).
16. Hoersch, D., Roh, S.-H., Chiu, W. & Kortemme, T. Reprogramming an ATP-driven protein machine into a light-gated nanocage. *Nat. Nanotechnol.* **8**, 928–932 (2013).
17. Madden, T. The BLAST Sequence Analysis Tool. (National Center for Biotechnology Information (US), 2003).
18. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
19. Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
20. Douglas, N. R. *et al.* Dual Action of ATP Hydrolysis Couples Lid Closure to Substrate Release into the Group II Chaperonin Chamber. *Cell* **144**, 240–252 (2011).
21. Reissmann, S., Parnot, C., Booth, C. R., Chiu, W. & Frydman, J. Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. *Nat. Struct. Mol. Biol.* **14**, 432–440 (2007).
22. Lopez, T., Dalton, K., Tomlinson, A., Pande, V. & Frydman, J. An information theoretic framework reveals a tunable allosteric network in group II chaperonins. *Nat. Struct. Mol. Biol.* **24**, 726–733 (2017).
23. Weber, F. & Hayer-Hartl, M. Refolding of Bovine Mitochondrial Rhodanese by Chaperonins GroEL and GroES. In Chaperonin Protocols (ed. Schneider, C.) 117–126 (Springer New York).
24. Pereira, J. H. *et al.* Mechanism of nucleotide sensing in group II chaperonins. *EMBO J.* **31**, 731–740 (2012).

Acknowledgements

Author contributions

K.D. wrote the program and carried out most experiments in collaboration with T.L. V.P. and J.F. provided guidance, J.F. directed the project. K.D. and J.F. wrote and all authors edited the M.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-58794-x.

Correspondence and requests for materials should be addressed to J.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.