Community-Wide Experimental Evaluation of the PROSS Stability-Design Method

Yoav Peleg, Renaud Vincentelli, Brett M Collins, Kai-En Chen, Emma K Livingstone, Saroja Weeratunga, Natalya Leneva, Qian Guo, Kim Remans, Kathryn Perez, et al.

Community-wide experimental evaluation of the PROSS stability-design method

Yoav Peleg[1*], Renaud Vincentelli[2], Brett M. Collins[3], Kai-En Chen[3], Emma K. Livingstone[3], Saroja Weeratunga[3], Natalya Leneva[3], Qian Guo[3], Kim Remans[4], Kathryn Perez[4], Gro E. K. Bjerga[5], Φlvind Larsen [5], Ondřej Vaněk[6], Ondřej, Skořepa[6], Arnaud Poterszman[7], Sophie Jacquemin[7], Svend Kjaer[8], Evangelis Christodoulou[8], Shira Albeck[1], Orly Dym[1], Elena Ainbinder[1], Tamar Unger[1], Anja Schuetz[9], Matthes Susann[9], Bader Michael[9,10], Ario de Marco[11], Paola Storici[12], Marta S. Semrau[12], Peggy Stolt-Bergner[13&], Christian Aigner[13], Sabine Suppmann[14], Adi Goldenzweig[15], Sarel J Fleishman[15*]

1. Department of Life Sciences Core Facilities (LSCF), Weizmann Institute of Science, Rehovot, 7610001, Israel.
2. Unité Mixte de Recherche 7257, CNRS Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille 13009, France.
3. The University of Queensland, Institute for Molecular Bioscience, St. Lucia, Queensland 4072, Australia.
4. European Molecular Biology Laboratory (EMBL), Protein Expression and Purification Core Facility, Meyerhofstrasse 1, 69117 Heidelberg, Germany.
5. NORCE Norwegian Research Centre, Postboks 22 Nygårdstangen, 5038 Bergen, Norway.
6. Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030/8, 12840 Prague, Czech Republic.
7. Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Université de Strasbourg.
8. Structural Biology Science Technology Platform, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK.
9. Max-Delbrück Center for Molecular Medicine (MDC), Robert-Rössle-Straße 10, 13125 Berlin-Buch, Germany.
10. University of Lübeck, Institute for Biology, Ratzeburger Allee 160, 23562 Lübeck, Germany; Charité University Medicine, Charitéplatz 1, 10117 Berlin, Germany; German Center for Cardiovascular Research (DZHK), Partner Site Berlin, Berlin, Germany.
11. Laboratory for Environmental and Life Sciences, University of Nova Gorica, Slovenia.
12. Elettra Sincrotrone Trieste - SS 14 - km 163,5 in Area Science Park, 34149 Basovizza, Trieste, Italy.
13. Vienna Biocenter Core Facilities GmbH, Dr. Bohr-gasse 3, 1030 Vienna, Austria.
14. Max-Planck Institute of Biochemistry, Biochemistry Core Facility, Am Klopferspitz 18, 82152 Martinsried, Germany.
15. Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, 7610001, Israel.

[&]Present address: Medicinal Chemistry, Boehringer Ingelheim RCV GmbH & Co KG, Vienna, Austria.

*** To whom correspondence should be addressed. E-mail:**

yoav.peleg@weizmann.ac.il, sarel.fleishman@weizmann.ac.il

Abstract

Recent years have seen dramatic improvement in protein-design methodology. Nevertheless, most methods demand expert intervention, limiting their widespread adoption. One exception is the PROSS algorithm for improving protein stability and heterologous expression levels that has been applied to a range of challenging enzymes, receptor binding proteins and vaccine immunogens. Here, we benchmark the application of PROSS as a stand-alone tool for protein scientists with no significant experience in modeling. Twelve laboratories from the Protein Production and Purification Partnership in Europe (P4EU) professional network independently challenged the PROSS algorithm with fourteen unrelated targets without support from the PROSS developers. Up to six automatically generated designs were evaluated in each case for expression, stability, and, in some cases, protein function. In eleven cases, the designs exhibited increased heterologous expression levels and/or thermal stability relative to the parental protein. In two prime examples, the human Stem Cell Factor (hSCF) and human Cadherin-Like Domain (CLD12) from the RET receptor, the wild type proteins were not expressible as soluble proteins in *E. coli*, yet the PROSS designs exhibited high expression levels in *E. coli* and HEK293 cells and showed improved thermal stability. This study demonstrates the strengths of community-wide efforts to probe the generality of new methods and recommends areas for future research to advance practically useful algorithms for protein science.

Introduction

The production of recombinant proteins in heterologous systems has become a routine practice [1,2]. A wide range of approaches has been developed to improve heterologous protein expression, including the development of prokaryotic and eukaryotic expression strains, co-expression with chaperones, fusion to solubility tags, and adjustment of the target protein's DNA coding sequence to the host's codon usage [1,3–7]. Despite these important developments, however, many proteins remain difficult to express due to their intrinsically low stability, resulting in insolubility, misfolding, aggregation and degradation. In such cases, the above strategies may result in insufficient protein expression levels or half-life [8].

Computational protein design is an exciting field that holds the promise of revolutionizing protein engineering. Specifically, design methods may provide a universal route to optimize protein expression levels and stability, thus streamlining research and applications. The design of large proteins with complex folds is still a formidable challenge though and typical workflows for designing new or improved variants of natural proteins rely on visual inspection of structures to guide mutational analysis [9–11], as well as on laborious cycles of genetic randomization and selection [12]. Indeed, the long-standing difficulties in optimizing protein stability using purely automated methods has even led to a general pessimism that large proteins are too complex for computational methods to reliably improve [12,13]. These difficulties were usually ascribed to the low accuracy of the energy calculations [13], tradeoffs between stability and activity [14] and potential misfolding due to the designed mutations [15].

To address this shortcoming of protein-design methodology, we recently developed and validated the PROSS design method [16]. PROSS combines phylogenetic analysis with Rosetta atomistic design calculations based on experimentally determined protein structures [16,17] or homology models [18]. PROSS applies a two-step filtering approach: first eliminating mutations that are rare among homologs; and second, by eliminating mutations that are predicted to destabilize the native state according to Rosetta atomistic calculations. This two-step filtering defines a drastically reduced sequence space (compared to the theoretical sequence space of

a large protein) which is moreover expected to comprise few mutations that impair stability or activity owing to natural selection. In the last step, PROSS combinatorially designs several multipoint variants within the filtered sequence space to optimize the native-state energy, each variant encoding up to 10% mutations relative to the parental protein (in some cases, >50 mutations compared to the parent). PROSS has shown some remarkable cases of success in improving heterologous expression levels; for instance, in several large enzymes (300-1,000 amino acids) [16,17,19,20] and malaria and HIV vaccine immunogens [21,22]. PROSS variants exhibited large gains in thermal stability of 15-20℃ and orders of magnitude of improvement in heterologous expression levels while maintaining wild type activity levels. To enable its widespread adoption, PROSS was implemented as an online server (http://pross.weizmann.ac.il/) that has attracted >1,500 academic users so far and led to several publications from other groups that demonstrated substantially stabilized protein variants [19,20,23–25]. Though these results are encouraging, they do not represent a systematic analysis of PROSS's scope, and it is likely that experiments that indicated small improvements or a deterioration in stability following design would typically not be published.

To probe PROSS's generality, we launched a community effort by 12 laboratories (Table 1) that are part of the Protein Production and Purification Partnership in Europe (P4EU) professional network (https://p4eu.org). Each lab selected 1-4 protein target(s) for protein expression for a total of 14 different targets across the benchmark. The PROSS designs were initially tested for soluble expression in *E. coli*, and successful designs were further analyzed for their thermal stability and activity levels. In several cases, the variants were also expressed in eukaryotic hosts, thus providing, to the best of our knowledge, a benchmark of unprecedented breadth and depth for the reliability of a protein-design method. To make the study as representative as possible of a real-world scenario, the PROSS developers (AG and SJF) were only involved in the analysis of the results and not in the choice of targets or in any of the computational design steps, unless technical errors arose.

Materials and Methods

Target proteins

Some of the proteins selected for the benchmark (Table 1) are difficult to express as soluble proteins in an *E. coli* host (HSD17B1, RET(CLD12), Klk6, hSCF, CDK7, UAF-1, Munc18c and Trypsin). The remaining proteins (Vps26A, PTPN3, Munc18a, hLIF, TPH1 and HER2-Nb) can be expressed as soluble proteins in *E. coli* and were selected to assess whether improvements in protein stability and/or in expression levels can be achieved.

Table 1. Participating labs and their PROSS stability design targets

Partner	Protein		Source organism	PDB entry[a]
	Full Name	Abbreviation		
Institute for Molecular Bioscience, The University of Queensland (Brisbane)	Vacuolar protein sorting 26A	Vps26A	zebrafish	6MD5
	Protein tyrosine phosphatase N3	PTPN3	human	4QUM
	Munc18c / Munc18-3	Munc18c	mouse	3PUK
	Munc18a/ Munc18-1	Munc18a	rat	3C98
Vienna BioCenter Core Facilities (Vienna)	human Leukemia Inhibitory Factor	hLIF	human	1PVH_B, 2Q7N_B
Department of Biochemistry, Faculty of Science, Charles University (Prague)	17-β-hydroxysteroid dehydrogenase 1	HSD17B1	human	1FDV
Structural Biology Science Technology Platform, Francis	Cadherin-Like Domain (1-2) of RET	RET(CLD12)	human	2X2U

Crick Institute (London)				
Protein Expression and Purification Core Facility, EMBL (Heidelberg)	**Kallikrein 6**	**Klk6**	**human**	**1LO6**
Structural Proteomics Unit, Weizmann Institute (Rehovot)	**human Stem Cell Factor**	**hSCF**	**human**	**1EXZ**
Institute of Genetic, Molecular and cellular biology, (Strasbourg)	**Cyclin-dependent kinase 7**	**CDK7**	**human**	**1UA2**
Protein Facility, Elettra Synchrotron (Trieste)	**USP1-associated factor 1**	**UAF1**	**human**	**5L8E**
Protein Production and Characterization Platform, Max Delbrück Center (Berlin)	**Tryptophan hydroxylase 1**	**TPH1**	**human**	**1MLW**
NORCE Norwegian Research Centre (Bergen)	**Trypsin**	**Trypsin**	***Streptomyces griseus***	**1OS8**
Laboratory for environmental and life sciences, University of Nova Gorica (Nova Gorica)	**Single-domain camelid antibody (VHH) targeting HER2**	**HER2-Nanobody (HER2-Nb)**	**A synthetic library of humanized VHHs**	**A model generated by SWISS-MODEL**

[a] The PDB entry for each target was used for PROSS stability calculations, except for HER2-Nb for which a model was generated.

PROSS stability design

Experimental or model structures of the protein targets were submitted to the PROSS web server (http://pross.weizmann.ac.il/) using the default settings. In all cases, positions proximal to ligands or active sites were specified by the users to be excluded from the design process to retain the protein's molecular activity (Supplement). Each group selected 1-6 PROSS designs for experimental analysis. All the experimentally tested protein sequences (the parent and the PROSS designs) are listed in the Supplement.

Construct design

For the purposes of standardization, for each target expressed in *E. coli*, the genes encoding the wild type and the PROSS designs were codon optimized for *E. coli* expression by each group. DNA synthesis of most genes was performed by TWIST Bioscience (unless otherwise specified in the Supplement). Each group used its own preferred cloning approach and expression vectors. In order to have a uniform assay for high-throughput screening, all constructs, apart from the Trypsin constructs using the arabinose system, were based on a T7 promoter-operator system for transcriptional regulation. Additionally, in order to streamline and facilitate the purification process, all targets were fused to a hexahistidine tag (at the N- or C-terminus), apart from the Trypsin and hSCF constructs, which harbor a deca- and tetradeca-histidine tag at the N-terminus, respectively. Several targets contain, in addition to the histidine tag, a solubility-enhancing tag such as Sumo, GST, MBP and Trx (details are provided in the Supplement, as well as the target-specific cloning strategy).

High-throughput expression screen

Parallel high-throughput expression screens were performed for most of the PROSS designs and wild type proteins submitted by the individual groups (total of 37 constructs) following a standardized protocol [26] at AFMB, CNRS, Marseille to evaluate whether the PROSS designs would exhibit improved soluble expression compared to the wild type under different culturing conditions. Each plasmid (except for the Trypsin expression plasmids, see description below) was transformed into three *E. coli* strains: T7 Express (NEB), supplemented with the pLysSRARE plasmid extracted from Rosetta(DE3)pLysS (Novagen); Rosetta(DE3)pLysS (Novagen), and BL21(DE3)pLysS (Novagen) and grown in three different media. The protein soluble

yields of all the constructs were determined after purification from the nine culture conditions generated per construct. Trypsin expression was performed in *E. coli* strain MC1061 [27]. Cells were grown for 2 hours at 37ºC in 2 ml TB and LB medium in 24-well plates, induced with 0.5 mM IPTG (and, in the case of Trypsin, with 0.1% L-arabinose) then the sample was cooled to 17°C and cultures were grown overnight (approximately 24h of cultures in total). Cells were harvested and frozen in the lysis buffer. Expression was also assessed (except for Trypsin) in NZYTech auto-induction media (NZYTech, Lisbon, Portugal) in the same conditions (2 hours at 37°C followed by overnight culture at 17°C, harvest and freezing in lysis buffer). After thawing the pellets, cells were processed and analyzed as previously described [26]. The level of soluble protein in the elution fractions of the Nickel 96 purifications were quantified using a labchip GXII system (Perkin Elmer).

Protein expression, purification and analysis

Each group selected its preferred expression strain(s) and cultivation conditions. Protein purification and subsequent analysis (stability and activity) were performed as described for each target in the Supplement.

Results

Target proteins selected for the benchmark

The 14 targets in this benchmark are mostly of eukaryotic origin, with the exception of a microbial Trypsin (Table 1). This reflects the general observation that eukaryotic proteins are more challenging for heterologous production and therefore of greater interest for stability design. Targets were selected due to known difficulties in expression in *E. coli* hosts, as well as to enhance the expression and stability of expressible proteins (details in Supplement). The targets include a number of different enzymes, such as HSD17B1, which is responsible for the conversion of estrone to estradiol and is thus implicated in estrogen-dependent cancers [28]; PTPN3/PTPH1, which specifically dephosphorylates the epidermal growth factor substrate 15 [29] and mitogen-activated protein kinase 12 [30]; TPH1 which catalyzes the initial step in the biosynthesis of the neurotransmitter serotonin [31]; bacterial Trypsin which is part of a commercial mixture of proteases (Pronase®) used for protein hydrolysis in research applications, and human Klk6, a serine protease which

may play a role in Alzheimer's disease [32]. We also selected two cytokines: hLIF which maintains cells in an undifferentiated state and hSCF which plays a role in the formation of blood and germ cells. hSCF is manufactured by Amgen (under the trade name STEMGEN®) for use in combination with NEUPOGEN® for certain cancer patients undergoing transplantation [33,34]. Proteins involved in the formation of large protein complexes were also selected, such as UAF1, which is part of the deubiquitinating enzyme heterodimeric complex USP1-UAF1 [35]; The Cyclin Dependent Kinase 7 (CDK7), a subunit of the transcription/DNA repair factor TFIIH [36], is both an effector kinase which phosphorylates both RNA Polymerase II and other transcription factors and a Cdk Activating Kinase (CAK) for essential CDKs [37]; VPS26A, which is a subunit of the core 'Retromer' complex that mediates endosomal protein sorting and trafficking [38]; and Munc18a and Munc18c which regulate SNARE-complexes that mediate vesicle fusion [39]. We also targeted the CLD12 fragment of the RET receptor tyrosine kinase, which comprises two extracellular cadherin-like domains. Interestingly, the folding of the extracellular domain of RET is sensitive to mutations in Hirschsprung's disease [40,41], which causes protein misfolding and retention of the receptor in the endoplasmic reticulum. Finally, a nanobody targeting HER2 (HER2-Nb) [42] was submitted to PROSS. Since there was no available structure of the HER2-Nb, the sequence was first submitted to SWISS-model to generate a model structure [43]. In a first submission (1st), the multiple sequence alignment was generated by PROSS (as recommended for most targets). However, PROSS mutated several positions known to distinguish camelid from conventional antibodies to the amino acid identities that characterize conventional antibodies due to the very high prevalence of conventional antibodies in public sequence databases. To avoid the bias toward conventional antibodies, HER2-Nb was submitted to PROSS again (2nd) using a custom-made alignment containing only camelid antibody sequences.

The 14 proteins tested in the benchmark vary in size, with lengths ranging from 122 (HER2-Nb) to 594 (Munc18a) amino acids. The proteins belong to diverse folds as indicated by the CATH database (https://www.cathdb.info) with six proteins classified as mainly β (Fig. 1, first row and HER2-Nb and RET(CLD12)), two as mainly α (hLIF and hSCF) and the remaining as either α+β or α/β. Two targets are dimeric (hSCF

and HSD17B1). The diversity in size, fold and oligomeric state of the target set provide an unprecedented variety for benchmarking a protein-design method.

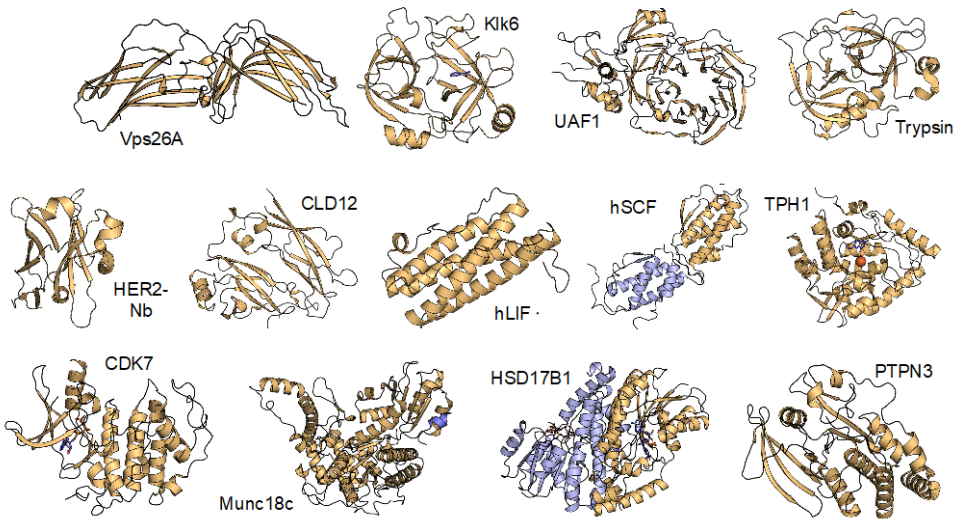

Figure 1. Structural classification of the proteins selected for the PROSS-benchmark. The 14 proteins tested in the benchmark are structurally classified using the CATH database (https://www.cathdb.info). According to CATH, 6 of the 14 proteins are classified as mainly β (all the proteins in the first row, HER2-Nb and RET(CLD12) in the second row), two are mainly α (hLIF and hSCF) and the other six are classified as either α+β or α/β. Two targets are dimeric (hSCF and HSD17B1, one of the monomers is shown in light blue). Munc18c and Munc18a belong to the same fold family and hence only Munc18c is shown.

Orders of magnitude increase in soluble expression in HTP screening

To obtain a uniform and quantitative analysis of protein expression across most of the tested designs, we performed a high-throughput protein solubility screen in *E. coli* under standardized experimental conditions. Expression was tested in three *E. coli* strains and three types of medium (Materials and Methods). Out of 14 targets, 10 were analyzed by HTP screening for soluble expression in *E. coli* (Table 2). We observed soluble expression for four of these ten (hLIF, HSD17B1, hSCF and TPH1), enabling quantification of expression levels for the wild type protein and/or at least one of the PROSS variants (Table 2 and Fig. 2). For TPH1, the most significant expression level was detected only in T7 Express (NEB) cells using TB medium, likely due to the requirement of sufficient iron supplementation. D1 variant exhibited comparable expression to the wild type, whereas two other designs exhibited slightly

reduced expression (Table 2, Fig. 2 and Supplement). By contrast, for hLIF, hSCF and HSD17B1, a marked increase in the solubility of at least one PROSS design compared to the wild type protein was observed (Table 2, Fig. 2 and Supplement). For hLIF and hSCF up to 78- and 112-fold increases in soluble expression levels were observed, respectively (corresponding to high yields of approximately 50 mg/L for both, Fig. 2).

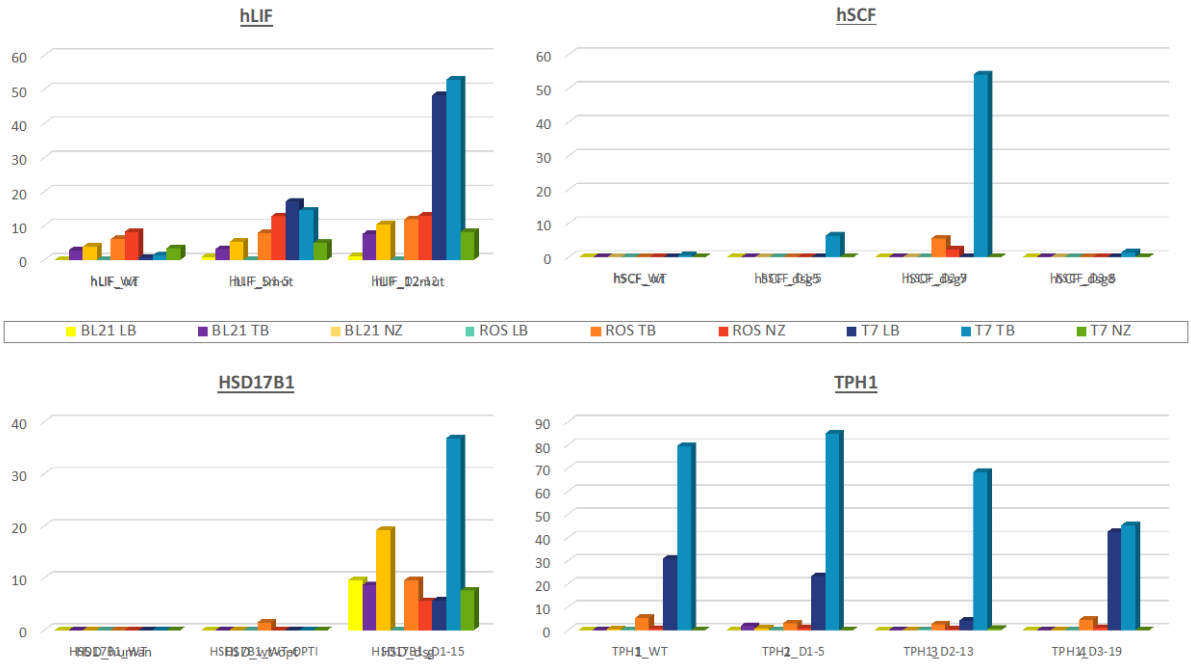

Figure 2. HTP expression of soluble targets in *E. coli*. Soluble expression levels of hLIF, HSD17B1, hSCF and TPH1, are presented in mg of proteins per liter of culture (mg/L). The different colors represent the expression conditions (combination of *E. coli* strain and medium): Yellow- BL21(DE3) pLysS in LB medium; purple- BL21(DE3) pLysS in TB medium; light orange- BL21(DE3) pLysS in NZ medium; light green- Rosetta(DE3) pLysS in LB medium; orange- Rosetta(DE3) pLysS in TB medium; red- Rosetta(DE3) pLysS in NZ medium; dark blue- T7 express in LB medium; light blue- T7 express in TB medium; green- T7 express in NZ medium.

Remarkably, for HSD17B1, no soluble expression was observed for both versions of the wild type protein (both codon-optimized and wild-type DNA sequences were tested in this case), while a high level of 36.8 mg/L soluble protein was detected for the single PROSS variant tested (Table 2, Fig. 2 and Supplement). The other targets tested in HTP screening (Table 2; CDK7, RET(CLD12), UAF1, Trypsin, Klk6 and HER2-Nb-1st) showed no soluble *E. coli* expression for the wild type protein and the

PROSS designs under all tested conditions (Table 2). We noted that the HTP solubility screen gave the best results for the *E. coli* strain T7 Express grown in TB or LB medium (Fig. 2), while under all other conditions expression was considerably lower and often below detection levels (Fig. 2). Thus, HTP screening showed three targets with orders of magnitude improvement in *E. coli* soluble expression, another exhibited no change, and six exhibited no expression in either the wild type or the PROSS designs.

Low-throughput screening

Table 2 summarizes the results of low-throughput testing of the 14 benchmark targets. Notably, low-throughput expression compares favorably to the HTP expression above, as 12 of the 14 target proteins could be solubly expressed (wild type and/or PROSS designs; Table 2), compared to only four in HTP (Fig. 2 and Table 2). Still, the patterns observed in HTP screening for the four targets which were soluble in HTP screening (TPH1, hLIF, HSD17B1 and hSCF) were similar in low-throughput experiments (Table 2). For four targets for which no soluble expression was detected in the HTP screening (KlK6, UAF1, Trypsin and HER2-Nb), soluble expression was observed in low-throughput experiments through the use of optimized experimental conditions (alternative *E. coli* strains or the use of solubilization tags) that were not tested in the HTP screening (Table 2). For instance, Klk6 expression in the *E. coli* SHuffle T7 Express strain yielded high soluble expression of two PROSS variants relative to no detectable expression of wild-type KIk6. Similarly, the T7 promoter-driven expression of Trypsin as a genetic fusion to disulfide-bond isomerase C (DsbC) in the SHuffle T7 Express strain led to soluble expression, and UAF1 expression in BL21-Codon-Plus(DE3)-RP cells with His-TRX and His-GST fusions yielded detectable expression levels of both the wild type and PROSS designs relative to no detectable expression for any construct in the HTP screening. The HER2-Nb, which showed no soluble expression in the HTP screen, showed soluble expression in BL21(DE3) co-transformed with sulfhydryl oxidase (Table 2 and Supplement). It should be noted that the HER2-Nb designs from the first submission showed less soluble expression than the wild type nanobody (Table 2, HER2-Nb, 1st), while the designs from the second submission exhibited similar yields to the wild type protein (Table 2, HER2-Nb, 2nd).

Four targets that were not tested in the HTP screen (PTPN3, Munc18a, Munc18c and Vps26A) all showed significantly higher expression levels (of up to 10-fold) for several PROSS variants relative to wild type expression levels (Table 2). We note that in the HTP screen, the expression of hSCF fused to His-Sumo was monitored whereas an additional step of cleaving the fusion target was performed during the low-throughput expression study (Table 1 and Supplement), resulting in no soluble expression of the wild type protein and high solubility for design D2 (Figs. 2). CDK7 can be expressed at mg scale in baculovirus-infected insect cells [44,45] but none of the tested PROSS variants (D1, D2 and D3) yielded soluble protein in *E. coli* (Supplement, CDK7, Fig. 1, lanes 3 to 8). In conclusion, only two targets, RET(CLD12) and CDK7, showed no soluble expression in *E. coli* for both wild type and PROSS designs under all tested conditions by the individual groups (Table 2).

Taken together, the high-throughput expression screening and the efforts by the individual groups provide a compelling demonstration that both approaches are complementary. PROSS improved expression levels in the vast majority of the cases, although PROSS does not necessarily enable a “one-size-fits-all” experimental protocol. Rather, the use of specialized expression vectors, strains, and fusion constructs is, in most cases, still required to achieve high expression levels. It is also worth noting that despite the promising initial results of the above-mentioned targets achieved by the individual groups, for four targets, UAF1, HER2-Nb, CDK7 and Trypsin, low expression levels and/or inconsistency in the expression results, prevented further work on the designs.

The initial scope of the benchmarking was to explore how PROSS design affects protein expression in *E. coli*. However, several participants further probed the expression of PROSS designs in eukaryotic systems. For HSD17B1, the single PROSS variant could be expressed solubly in *E. coli,* yielding 19 mg/L relative to no detectable wild type expression (Table 2 and Supplement). In an attempt to obtain even higher expression levels, the wild type HSD17B1 and the single PROSS design were expressed in HEK293 cells. Intracellular expression of the wild type protein in transiently and stably transfected HEK293 cells yielded about 2 and 7 mg/L protein, respectively. Expression of the HSD17B1 PROSS design in transiently transfected HEK293 cells resulted in 10-fold yield improvement reaching 27 mg/L. For

RET(CLD12), secreted expression was demonstrated in transiently transfected HEK293 cells (Table 2 and Supplement), with an approximately 10-fold increase in the secreted expression levels for PROSS design D1 compared to the RET(CLD12) wild type construct (Table 2 and Supplement). Increased expression levels were also observed for the two additional PROSS designs D2 and D3. Thus, for two target proteins for which PROSS designs failed to improve expression in bacterial cells, it nevertheless resulted in substantial increases in eukaryotic expression yields.

Table 2. PROSS designs expression levels in *E. coli* HTP vs. low-throughput screening

Target name	Number of variants and mutations[a]	Protein length[b]	Expression levels in HTP screening[c]	Expression levels in low-throughput screening[d]
Vps26A	**5 variants** **D1 -9** **D2- 12** **D3- 14** **D4- 18** **D5- 21**	**319**	**ND[e]**	**Improved soluble expression using BL21(DE3) for all variants, primarily for D1**
PTPN3	**5 variants** **D1- 5** **D2- 8** **D3- 11** **D4- 17** **D5- 24**	**282**	**ND[e]**	**Expression of D1 variant similar to wild type, D2, D3, D4, D5 up to 10-fold increase**
Munc18c	**6 variants** **D1- 11** **D2- 14** **D3- 20** **D4- 28** **D5- 39** **D6- 47**	**592**	**ND[e]**	**~10 fold increase for variant D3**
Munc18a	**2 variants** **D1- 19** **D2- 35**	**594**	**ND[e]**	**Similar levels of expression for the PROSS variants as for wild type protein**
hLIF	**2 variants** **D1- 5**	**180**	**Improved expression for PROSS variants.**	**Improved soluble expression for both PROSS variants**

	D2- 12		**D12 variant expressed >80 fold better than wild type**	
HSD17B1	**1 variant D1- 15**	**328**	**No expression for wild type constructs. High solubility for the PROSS variant**	***E. coli*: No expression for wild type. High expression levels for the single PROSS variant (19 mg/L)**
				HEK293: Moderate expression for wild type in transient and stable transfection. High expression for PROSS variant in transient transfection
RET(CLD12)	**3 variants[f] D1- 5 D2- 8 D3- 18**	**240**	**No soluble expression**	***E. coli*: No soluble expression**
				HEK293: Up to 10-fold soluble expression for variant D1
Klk6	**3 variants[g] D1- 6 D2- 12 D3- 16**	**223**	**No soluble expression**	**Low solubility in BL21(DE3) pLysS. High soluble expression in SHuffle T7 Express for D1 and D3 variants**
hSCF	**3 variants[h] D1- 5 D2- 9 D3- 5**	**141**	**His-Sumo-hSCF was monitored. Mutants expressed better than wild type. D2 variant expressed >112 fold better than the wild type**	**High solubility for D2 variant following removal of the His-Sumo tag**
CDK7	**3 variants D1- 3 D2- 9 D3- 12**	**345**	**No soluble expression**	**No soluble expression**
UAF1	**2 variants D1- 7 D2- 14**	**572**	**No soluble expression**	**Low expression levels as fusion proteins**
TPH1	**3 variants D1- 5 D2- 13 D3- 19**	**443**	**High soluble expression for wild type and PROSS variants. D1 expression**	**Results are in agreement with the HTP screen. No improved expression for variants**

			is similar to wild type and expression of D2 and D3 lower than wild type	
Trypsin	**4 variants D1- 2 D2- 3 D3- 8 D4- 12**	**223**	**No soluble expression**	**Solubility observed for several PROSS variants only when fused to DsbC, but the results were inconsistent**
HER2-Nb[i]	**1st attempt: 4 variants D1.1- 3 D2.1- 4 D3.1- 5 D4.1- 9**	**122**	**No soluble expression**	**Expression in BL21(DE3) co-expressed with sulfhydryl oxidase. Lower expression levels for PROSS variants than wild type**
	2nd attempt: 1 variant D1.2- 5		**ND[e]**	**Similar expression levels to wild type**

[a]The number of PROSS variants that were experimentally tested and the number of mutations incorporated per variant (D1, D2, etc.). Protein sequence and mutation sites are provided in the Supplement. A few protein variants for which expression levels were low in the initial screening test (Table 2) have not been used for further purification and characterization.

[b]The number of amino acids for each protein is indicated. In the case of hSCF, since the wild type failed to express as a soluble protein we used a commercially available protein with a length of 165 amino acids. For further details refer to footnotes in Tables 2 and 3 and Supplement.

[c]High-throughput screening in three *E. coli* strains in three media.

[d]The individual lab assays were performed in *E. coli* unless stated otherwise.

[e]ND- Not determined.

[f]Each of the RET(CLD12) PROSS-variants tested harbors some additional mutations, C87R, C216S, N98Q and N199Q. These changes include 2 cysteine mutations and 2 mutations of N-linked glycosylation sites. The RET(CLD12) variant harboring these 4 changes has been expressed in CHO cells and used for structure determination (PDB 2X2U) (or details see RET(CLD12) in Supplement).

[g]For each of the Klk6 PROSS-variants tested, an additional variant harboring 2 extra mutations (R78G and R80Q) was tested as well. These mutations have been incorporated to prevent auto-proteolysis (for details see Klk6 in Supplement).

[h]For hSCF D3 variant several PROSS-suggested changes have been removed (see hSCF in Supplement).

[i]A HER2-Nb model generated by SWISS-model was submitted to PROSS twice, in each case using a different multiple sequence alignment as input. In the first submission (1st) the alignment was generated by PROSS with default parameters as recommended for most

targets. In the second run (2nd) a custom-made alignment containing only antibody sequences from the *Camelidae* family (i.e., VHHs) was generated and submitted as input.

Increased thermal stability in 9 out of 10 targets.

For the nine targets in which PROSS designs exhibited increased or similar expression levels compared to the wild type protein, we also tested the relative thermal stability of the newly designed proteins. Remarkably, for nine of these proteins, an increase in the thermal stability was observed for at least one PROSS design compared to the wild type protein (Table 3). The increases in thermal stability compared to the wild type protein varied from 1^{o}C (for Munc18c_D1) to as much as 27^{o}C (for PTPN3_D4). When several PROSS designs of a single target were tested, a correlation between the number of mutations and the stability was found in six cases (Vps26A, Munc18c, Munc18a, hLIF, RET(CLD12) and TPH1) (Table 3). This result suggests an additive and positive relationship between the number of designed mutations and stability gain that has not been demonstrated previously for PROSS designs. However, there are exceptions to this correlation, and in PTPN3, the design with the most mutations (D5 with 24 mutations) resulted in a slight decrease in the thermal stability (of 2^{o}C) compared to the PROSS design D4 with 17 mutations (Table 3). For hSCF and HSD17B1 for which a single PROSS design was tested, an increase in thermal stability of 19.5^{o}C and 18^{o}C compared to the wild type protein was observed, respectively (Table 3). Only in a single case (Klk6), a decrease in the thermal stability was observed for the two PROSS variants tested compared to the wild type protein (Table 3). It is notable that Klk6 is a very challenging protein for design since it comprises six disulfide bonds.

Table 3. Comparison of protein stability

Protein	Variants	Protein stability (Tm, °C)*& Tm	# mutations	Range of change in Tm#
Vps26A	WT	64	0	8-19°C ↑
	D1	72	9	
	D2	75	12	
	D3	77	14	
	D4	79	18	
	D5	83	21	
PTPN3	WT	33	0	14-27°C ↑
	D1	47	5	
	D2	51	8	
	D3	53	11	
	D4	60	17	
	D5	58	24	
Munc18c	WT	47	0	1-10°C ↑
	D1	48	11	
	D2	49	14	
	D3	51	20	
	D4	53	28	
	D5	56	39	
	D6	57	47	
Munc18a	WT	46	0	20-22°C ↑
	D1	66	19	
	D2	68	35	
hLIF	WT	61.5	0	7-9°C ↑
	D1	68.5	5	
	D2	70.5	12	
HSD17B1	WT	51.8	0	18°C ↑
	D1	69.8	15	
RET(CLD12)	WT	48	0	16.1-20.8°C ↑
	D1	64.1	5	
	D2	66.3	8	
	D3	68.8	18	
Klk6	WT	75.1	0	3.7-13.7°C ↓
	D1	61.4	6	
	D3	71.4	16	

hSCF	**WT**	**61.0**	**0**	**19.5°C ↑**
	D2	**80.5**	**9**	
TPH1	**WT**	**43.5**	**0**	**6.3-7°C ↑**
	D1	**49.8**	**5**	
	D2	**50.3**	**13**	
	D3	**50.5**	**19**	

*For each variant (marked as WT, D1, D2., etc.) the Tm (in °C degrees) and the number of mutations incorporated in each PROSS variant are indicated. Protein variants that showed low expression levels in the initial screening test (CDK7, UAF1, Trypsin and HER2-Nb, Table 2) have not been further purified and characterized.

[&]It should be noted that comparison of stability was not always performed against wild type protein purified from *E. coli* because some of the wild type proteins failed to express in *E. coli* (wild type HSD17B1 and Klk6 were expressed in HEK293 and insect cells, respectively). In addition, wild type hLIF exhibited only low levels of expression in *E. coli*. Therefore, for comparison of stability we used a GST-hLIF protein. Another example is the hSCF for which no soluble expression in *E. coli* was detected for the native wild type protein which represents the 141 residues receptor binding core and for which the protein structure was determined (Table 2 and Supplement). Therefore, stability was performed in comparison to a commercial wild type protein, consisting of 165 residues, which is a soluble form of hSCF obtained following refolding of inclusion bodies. Klk6 PROSS-variants tested for stability harbour two extra mutations (R78G and R80Q), which have been incorporated to prevent auto-proteolysis.

[#]Increase in Tm of PROSS variants compared to the wild type protein is indicated by an upward arrow and decrease in Tm by a downward arrow.

Conservation of biological activity

Eight out of the 14 proteins tested in the benchmark were further analyzed for their functional activity. For 6 out of the 8 proteins tested, similar or improved activity was observed for at least one of the PROSS variants compared to the wild type activity (Table 4). Improved enzymatic activity was detected for all variants of PTPN3 and a single variant of TPH1 (Table 4). Activity levels were also maintained for the hSCF, HSD17B1, Munc18a and Munc18c proteins (Table 4). Qualitative activity data for hLIF on inhibition of embryonic stem cells differentiation indicated similar activity for the wild type and the two designs (for details see hLIF in Supplement). By contrast, in line with the results reported above that Klk6 was recalcitrant to design, reduced activity was observed for both PROSS variants of this protein (Table 4).

Table 4. Protein activity

Protein	Activity assay	Activity[#,&]	Remarks
PTPN3	**Phosphatase assay**	**WT- 0.02 µmol/min/µM D1- 0.076 µmol/min/µM D2- 0.087 µmol/min/µM D3- 0088 µmol/min/µM D4- 0.066 µmol/min/µM D5- 0.082 µmol/min/µM**	**Higher specific activity for all PROSS variants**
Munc18c	**Binding assay**	**WT- 1.5 µM D3- 2.5 µM**	**Similar binding as for wild type**
Munc18a	**Binding assay**	**WT- 5.8 nM D1- 10 nM D2- 7.4 nM**	**Activity of PROSS variants is maintained**
HSD17B1	**Oxidation of estradiol**	**WT- 1.52 mmol/min/mg D1- 1.28 mmol/min/mg**	**Wild type produced in HEK293 and PROSS variant (15 mut) in *E. coli***
Klk6	**Hydrolysis assay for serine protease**	**WT- 15.9 u/sec/mM D1- 7 u/sec/mM D3- 2.7 u/sec/mM**	**Reduced activity for both PROSS variants. Wild type produced in insect cells, PROSS variants in *E. coli***
hSCF	**Cell proliferation**	**WT- EC50- 0.27nM D2- EC50- 0.37nM**	**Similar activity for the wild type and D2 PROSS variant**
TPH1	**Formation of 5-Hydroxy-tryptophan**	**WT- 8.88 RFU/s D1- 9.98 RFU/s D2- 0.86 RFU/s D3- 2.93 RFU/s**	**A single variant D1 (5 mut) has higher activity and two additional variants (D2 and D3) have lower activity than wild type**

[#]Details on the activity assays are described in the Supplementary section. Few protein variants for which expression levels were low (CDK7, UAF1, Trypsin and HER2-Nb) in the initial screening test (Table 2) have not been used for further purification and characterization. In addition, for two proteins, RET(CLD12) and Vps26A, for which purification and stability measurements were conducted (Table 3) no activity assays were performed. TPH1 activity in RFU/s, relative fluorescence units per second. hLIF activity is not presented in Table 4 since the data are qualitative (for details see hLIF in Supplement).

Klk6 PROSS-variants tested for activity harbor two extra mutations (R78G and R80Q), which have been incorporated to prevent auto-proteolysis.
[&]It should be noted that comparison of activity was not always performed against wild type protein purified from *E. coli* because some of the wild type proteins failed to express in *E. coli* Therefore, wild type HSD17B1 and Klk6 were expressed in HEK293 and insect cells, respectively. In addition, wild type hLIF exhibited only low levels of expression in *E. coli*. Therefore, for comparison of activity we used a GST-hLIF protein. Another example is the hSCF for which no soluble expression in *E. coli* was detected for the native wild type protein which represents the 141 residues receptor binding core and for which the protein structure was determined (Table 2 and Supplement). Therefore, activity was performed in comparison to a commercial wild type protein, consisting of 165 residues, which is a soluble form of hSCF obtained following refolding of inclusion bodies.

Discussion

The marginal stability of many proteins is perhaps the most general problem that obstructs the application of proteins in basic and applied research [46]. This problem has attracted intensive research over several decades resulting in a wide range of computational and experimental methods to address it [8]. PROSS is unique among those in providing a one-step, non-iterative design solution that can be applied to any protein for which a known structure [46] or high-quality model [18] is available. PROSS designs were shown to improve thermal stability as well as functional expression yields in many cases without disrupting molecular activity [16,17,19,21–23,47].

The current study provides the broadest and most in-depth experimental benchmark of a stability-design method. It provides several important bottom lines: (1) PROSS can be successfully implemented by protein scientists without background in protein design, and in general, improvement is observed from screening up to six designs; (2) Achieving high levels of functional expression usually requires tailoring the expression protocol for the target protein, as only 30% (3 out of 10) of the tested targets exhibited improved expression relative to wild type in our standardized high-throughput screen (Table 2 and Fig. 1) compared to 64% (9 out of 14) for low-throughput screening (Table 2); (3) in 90% (9 out of 10) of the cases tested, thermal stability increased by up to 27℃ relative to the parental protein (Table 3); (4) Our findings suggest an additive and positive relationship between the number of designed mutations and the stability gain (Table 3). (5) In 57% of the cases (8 out of 14 cases) in which no, or some, bacterial soluble expression was observed for the wild type protein, the PROSS designs improved the amount of protein produced; (6)

In two cases in which no improvement in bacterial expression was noted in *E. coli*, a substantial improvement was observed in eukaryotic expression of the designs (up to 10-fold, compared to the parent protein); and (7) For 7 out 8 targets for which activity was monitored at least one of the tested designs was unperturbed-demonstrating that tradeoffs between stability and activity can often be averted (Table 4).

Our benchmarking exercise demonstrates that marginal protein stability can be addressed, in the majority of cases, through a one-step computational method. Historically, it was argued that the molecular complexity of proteins, the tradeoffs between stability and activity [14,48,49] and the low accuracy of protein-design calculations [13] would preclude a general method that can be universally applied to stabilize proteins [12]. The automated application of PROSS to very different proteins suggests that on the contrary, none of the arguments above is an insurmountable challenge to design methodology [50]. An important methodological improvement in PROSS relative to previous design methods lies in its combination of atomistic design calculations and evolutionary-conservation analysis; the former optimize the energy of the native state, while the latter mitigate the risk of introducing mutations that may lead to a decrease in activity or foldability [46,50,51] as those are almost certain to have been purged by natural selection. Furthermore, we recently demonstrated that this combination of evolutionary-conservation analysis and atomistic design calculations can automatically address significant challenges in another persistent problem in protein science and engineering, namely in the design of enzymes [52,53] and binders [54,55] that exhibit orders of magnitude improvement in their activity.

Previous community-wide efforts have highlighted important areas for improvement in protein-design methodology [56,57]. The current benchmark demonstrates that stability design has matured to the point where a large fraction of challenging proteins can be optimized automatically. Nevertheless, our benchmark highlights areas that may require new methods, such as maintaining the molecular activity of proteins that harbor multiple disulfide bonds and methods to improve the automated selection of the most relevant homologous sequences in very large protein families. Finally, we conclude that to bring forth the full benefits of PROSS design, one must

consider the expression host, genetic construct, fusion partners and cell culture conditions.

Acknowledgments

We thank T. Dornblut and J. Tischer (MDC Berlin, Germany) for excellent technical assistance. The TPH1 study was funded by the VIP program of the German Federal Ministry for Education and Research (03V0276). We thank the University of Queensland Remote Operation Crystallization and X-ray (UQ ROCX) facility and the staff for their support with crystallization experiments; the staff of the Australian Synchrotron for assistance with X-ray diffraction. B.M.C. was supported by an NHMRC Senior Research Fellowship (APP1136021); E.K.L was funded by a UQ Research Scholarship, an IMB Research Advancement Award, and an AINSE Postgraduate Research Award; S. J. was supported Instruct-ULTRA (Grant 731005); This work was supported by funds from the National Health and Medical Research Council (NHMRC) (APP1099114), the ARRS project N4-0046, and the Australian Research Council (ARC) (DP160101743). Research in the Fleishman lab was supported by the European Research Council (815379), the Israel Science Foundation (1844) and charitable donations from the Milner Foundation and Sam Switzer and family. AG and SJF are named inventors on patents relating to the PROSS method and various designs. This work was supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01 for the AFMB. Experiments in IGBMC benefited from resources of the French Infrastructure for Integrated Structural Biology FRISBI (ANR-10-INBS-05) and of Instruct-ERIC. The VBCF acknowledges public funding from the Austrian Federal Ministry of Education, Science and Research, and the City of Vienna via the Vienna Business Agency; SK and EC were supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001999), the UK Medical Research Council (FC001999), and the Wellcome Trust (FC001999).

Literature

[1] Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, S. Gräslund, P. Nordlund, J. Weigelt, B.M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H.-W. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.-H. Kim, Z. Rao, Y. Shi, T.C. Terwilliger, C.-Y. Kim, L.-W. Hung, G.S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J.L. Sussman, R.C. Stevens, S.A. Lesley, I.A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M.I. Donnelly, W.H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S.K. Burley, J.S. Emtage, J.M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S.C. Almo, J.B. Bonanno, A. Fiser, S. Swaminathan, F.W. Studier, M.R. Chance, A. Sali, T.B. Acton, R. Xiao, L. Zhao, L.C. Ma, J.F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C.K. Ho, D. Wang, H. Wang, M. Jiang, G.T. Montelione, D.I. Stuart, R.J. Owens, S. Daenke, A. Schütz, U. Heinemann, S. Yokoyama, K. Büssow, K.C. Gunsalus, Protein production and purification, Nat. Methods. 5 (2008) 135–146.

[2] R. Assenberg, P.T. Wan, S. Geisse, L.M. Mayr, Advances in recombinant protein expression for use in pharmaceutical research, Curr. Opin. Struct. Biol. 23 (2013) 393–402.

[3] A. de Marco, E. Deuerling, A. Mogk, T. Tomoyasu, B. Bukau, Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli, BMC Biotechnol. 7 (2007) 32.

[4] N.S. Berrow, K. Büssow, B. Coutard, J. Diprose, M. Ekberg, G.E. Folkers, N. Lévy, V. Lieu, R.J. Owens, Y. Peleg, Others, Recombinant protein expression and solubility screening in Escherichia coli: a comparative study, Acta Crystallogr. D Biol. Crystallogr. 62 (2006) 1218–1226.

[5] Y. Peleg, T. Unger, Application of high-throughput methodologies to the expression of recombinant proteins in E. coli, Methods Mol. Biol. 426 (2008) 197–208.

[6] Y. Peleg, T. Unger, Resolving bottlenecks for recombinant protein expression in E. coli, Methods Mol. Biol. 800 (2012) 173–186.

[7] D. Busso, Y. Peleg, T. Heidebrecht, C. Romier, Y. Jacobovitch, A. Dantes, L. Salim, E. Troesch, A. Schuetz, U. Heinemann, G.E. Folkers, A. Geerlof, M. Wilmanns, A. Polewacz, C. Quedenau, K. Büssow, R. Adamson, E. Blagova, J. Walton, J.L. Cartwright, L.E. Bird, R.J. Owens, N.S. Berrow, K.S. Wilson, J.L. Sussman, A. Perrakis, P.H.N. Celie, Expression of protein complexes using multiple Escherichia coli protein co-expression systems: a benchmarking study, J. Struct. Biol. 175 (2011) 159–170.

[8] T.J. Magliery, Protein stability: computation, sequence statistics, and new experimental methods, Curr. Opin. Struct. Biol. 33 (2015) 161–168.

[9] M.C. Crank, T.J. Ruckwardt, M. Chen, K.M. Morabito, E. Phung, P.J. Costner, L.A. Holman, S.P. Hickman, N.M. Berkowitz, I.J. Gordon, G.V. Yamshchikov, M.R. Gaudinski, A. Kumar, L.A. Chang, S.M. Moin, J.P. Hill, A.T. DiPiazza, R.M. Schwartz, L. Kueltzo, J.W. Cooper, P. Chen, J.A. Stein, K. Carlton, J.G. Gall, M.C. Nason, P.D. Kwong, G.L. Chen, J.R. Mascola, J.S. McLellan, J.E. Ledgerwood, B.S. Graham, VRC 317 Study Team, A proof of concept for structure-based vaccine design targeting RSV in humans, Science. 365 (2019) 505–509.

[10] V. Tournier, C.M. Topham, A. Gilles, B. David, C. Folgoas, E. Moya-Leclair, E. Kamionka, M.-L. Desrousseaux, H. Texier, S. Gavalda, M. Cot, E. Guémard, M. Dalibey, J. Nomme, G. Cioci, S. Barbe, M. Chateau, I. André, S. Duquesne, A. Marty, An engineered PET depolymerase to break down and recycle plastic bottles, Nature. 580 (2020) 216–219.

[11] H.J. Wijma, R.J. Floor, P.A. Jekel, D. Baker, S.J. Marrink, D.B. Janssen, Computationally designed libraries for rapid enzyme stabilization, Protein Eng. Des. Sel. 27 (2014) 49–58.

[12] H. Zhao, F.H. Arnold, Directed evolution converts subtilisin E into a functional equivalent

of thermitase, Protein Eng. 12 (1999) 47–53.
[13] D. Baker, What has de novo protein design taught us about protein folding and biophysics?, Protein Sci. 28 (2019) 678–683.
[14] B.K. Shoichet, W. a. Baase, R. Kuroki, B.W. Matthews, A relationship between protein stability and protein function, Proceedings of the National Academy of Sciences. 92 (1995) 452–456.
[15] S.J. Fleishman, D. Baker, Role of the biomolecular energy gap in protein design, structure, and evolution, Cell. 149 (2012) 262–273.
[16] A. Goldenzweig, M. Goldsmith, S.E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R.L. Lieberman, A. Aharoni, I. Silman, J.L. Sussman, D.S. Tawfik, S.J. Fleishman, Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability, Mol. Cell. 63 (2016) 337–346.
[17] M. Goldsmith, N. Aggarwal, Y. Ashani, H. Jubran, P.J. Greisen, S. Ovchinnikov, H. Leader, D. Baker, J.L. Sussman, A. Goldenzweig, S.J. Fleishman, D.S. Tawfik, Overcoming an optimization plateau in the directed evolution of highly efficient nerve agent bioscavengers, Protein Eng. Des. Sel. 30 (2017) 333–345.
[18] J. Zahradník, L. Kolářová, Y. Peleg, P. Kolenko, S. Svidenská, T. Charnavets, T. Unger, J.L. Sussman, B. Schneider, Flexible regions govern promiscuous binding of IL-24 to receptors IL-20R1 and IL-22R1, FEBS J. (2019). https://doi.org/10.1111/febs.14945.
[19] X. Brazzolotto, A. Igert, V. Guillon, G. Santoni, F. Nachon, Bacterial Expression of Human Butyrylcholinesterase as a Tool for Nerve Agent Bioscavengers Development, Molecules. 22 (2017). https://doi.org/10.3390/molecules22111828.
[20] M.H. Hettiaratchi, M.J. O'Meara, T.R. O'Meara, A.J. Pickering, of chondroitinase ABC improves efficacy and stability, (n.d.). https://advances.sciencemag.org/content/advances/6/34/eabc6378.full.pdf.
[21] I. Campeotto, A. Goldenzweig, J. Davey, L. Barfod, J.M. Marshall, S.E. Silk, K.E. Wright, S.J. Draper, M.K. Higgins, S.J. Fleishman, One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 998–1002.
[22] S.K. Malladi, D. Schreiber, I. Pramanick, M.A. Sridevi, A. Goldenzweig, S. Dutta, S.J. Fleishman, R. Varadarajan, One-step sequence and structure-guided optimization of HIV-1 envelope gp140, Current Research in Structural Biology. (2020). https://doi.org/10.1016/j.crstbi.2020.04.001.
[23] A.R. Lambert, J.P. Hallinan, R. Werther, D. Głów, B.L. Stoddard, Optimization of Protein Thermostability and Exploitation of Recognition Behavior to Engineer Altered Protein-DNA Recognition, Structure. (2020). https://doi.org/10.1016/j.str.2020.04.009.
[24] P.S. Georgoulia, S. Bjelic, R. Friedman, Deciphering the molecular mechanism of FLT3 resistance mutations, FEBS J. (2020). https://doi.org/10.1111/febs.15209.
[25] C.M. Buldun, J.X. Jean, M.R. Bedford, M. Howarth, SnoopLigase Catalyzes Peptide–Peptide Locking and Enables Solid-Phase Conjugate Isolation, J. Am. Chem. Soc. 140 (2018) 3008–3018.
[26] N.J. Saez, H. Nozach, M. Blemont, R. Vincentelli, High throughput quantitative expression screening and purification applied to recombinant disulfide-rich venom proteins produced in E. coli, J. Vis. Exp. (2014) e51464.
[27] G.E.K. Bjerga, H. Arsın, Ø. Larsen, P. Puntervoll, H.T. Kleivdal, A rapid solubility-optimized screening procedure for recombinant subtilisins in E. coli, J. Biotechnol. 222 (2016) 38–46.
[28] W. He, M. Gauri, T. Li, R. Wang, S.-X. Lin, Current knowledge of the multifunctional 17β-hydroxysteroid dehydrogenase type 1 (HSD17B1), Gene. 588 (2016) 54–61.
[29] M.-Y. Li, P.-L. Lai, Y.-T. Chou, A.-P. Chi, Y.-Z. Mi, K.-H. Khoo, G.-D. Chang, C.-W. Wu, T.-C. Meng, G.-C. Chen, Protein tyrosine phosphatase PTPN3 inhibits lung cancer cell proliferation and migration by promoting EGFR endocytic degradation, Oncogene. 34 (2015) 3791–3803.
[30] K.-E. Chen, S.-Y. Lin, M.-J. Wu, M.-R. Ho, A. Santhanam, C.-C. Chou, T.-C. Meng, A.H.J. Wang, Reciprocal allosteric regulation of p38γ and PTPN3 involves a PDZ

domain-modulated complex formation, Sci. Signal. 7 (2014) ra98.
[31] T. Swami, H.C. Weber, Updates on the biology of serotonin and tryptophan hydroxylase, Curr. Opin. Endocrinol. Diabetes Obes. 25 (2018) 12–21.
[32] I. Prassas, A. Eissa, G. Poda, E.P. Diamandis, Unleashing the therapeutic potential of human kallikrein-related serine proteases, Nat. Rev. Drug Discov. 14 (2015) 183–202.
[33] I.H. Chin-Yee, M. Keeney, A.K. Stewart, A. Belch, I. Bence-Buckler, S. Couban, K. Howson-Jan, M. Rubinger, D. Stewart, R. Sutherland, V. Paragamian, M. Bhatia, R. Foley, Optimising parameters for peripheral blood leukapheresis after r-metHuG-CSF (filgrastim) and r-metHuSCF (ancestim) in patients with multiple myeloma: a temporal analysis of CD34(+) absolute counts and subsets, Bone Marrow Transplant. 30 (2002) 851–860.
[34] H.E. Johnsen, C. Geisler, E. Juvonen, K. Remes, G. Juliusson, P. Hörnsten, S. Kvaloy, G. Kvalheim, G.W. Jürgensen, L.M. Pedersen, O.J. Bergmann, A. Schmitz, M. Boegsted, Priming with r-metHuSCF and filgrastim or chemotherapy and filgrastim in patients with malignant lymphomas: a randomized phase II pilot study of mobilization and engraftment, Bone Marrow Transplant. 46 (2011) 44–51.
[35] Z. Yu, H. Song, M. Jia, J. Zhang, W. Wang, Q. Li, L. Zhang, W. Zhao, USP1-UAF1 deubiquitinase complex stabilizes TBK1 and enhances antiviral responses, J. Exp. Med. 214 (2017) 3553–3563.
[36] O. Kolesnikova, L. Radu, A. Poterszman, TFIIH: A multi-subunit complex at the cross-roads of transcription and DNA repair, Adv. Protein Chem. Struct. Biol. 115 (2019) 21–67.
[37] R.P. Fisher, Cdk7: a kinase at the core of transcription and in the crosshairs of cancer drug discovery, Transcription. 10 (2019) 47–56.
[38] S.J. Norwood, D.J. Shaw, N.P. Cowieson, D.J. Owen, R.D. Teasdale, B.M. Collins, Assembly and solution structure of the core retromer protein complex, Traffic. 12 (2011) 56–71.
[39] J.K. Archbold, A.E. Whitten, S.-H. Hu, B.M. Collins, J.L. Martin, SNARE-ing the structures of Sec1/Munc18 proteins, Curr. Opin. Struct. Biol. 29 (2014) 44–51.
[40] Q. Jiang, F. Liu, C. Miao, Q. Li, Z. Zhang, P. Xiao, L. Su, K. Yu, X. Chen, F. Zhang, A. Chakravarti, L. Li, RET somatic mutations are underrecognized in Hirschsprung disease, Genet. Med. 20 (2018) 770–777.
[41] C. Tomuschat, P. Puri, RET gene is a major risk factor for Hirschsprung's disease: a meta-analysis, Pediatr. Surg. Int. 31 (2015) 701–710.
[42] S. Djender, A. Schneider, A. Beugnet, R. Crepin, K.E. Desrumeaux, C. Romani, S. Moutel, F. Perez, A. de Marco, Bacterial cytoplasm as an effective cell compartment for producing functional VHH-based affinity reagents and Camelidae IgG-like recombinant antibodies, Microb. Cell Fact. 13 (2014) 140.
[43] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling, Bioinformatics. 22 (2006) 195–201.
[44] G. Lolli, E.D. Lowe, N.R. Brown, L.N. Johnson, The crystal structure of human CDK7 and its protein recognition properties, Structure. 12 (2004) 2067–2079.
[45] W. Abdulrahman, I. Iltis, L. Radu, C. Braun, A. Maglott-Roth, C. Giraudon, J.-M. Egly, A. Poterszman, ARCH domain of XPD, an anchoring platform for CAK that conditions TFIIH DNA repair and transcription activities, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E633-42.
[46] A. Goldenzweig, S.J. Fleishman, Principles of Protein Stability and Their Application in Computational Design, Annu. Rev. Biochem. 87 (2018) 105–129.
[47] B. Bandyopadhyay, A. Goldenzweig, T. Unger, O. Adato, S.J. Fleishman, R. Unger, A. Horovitz, Local energetic frustration affects the dependence of green fluorescent protein folding on the chaperonin GroEL, J. Biol. Chem. 292 (2017) 20583–20591.
[48] S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D.S. Tawfik, Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein, Nature. 444 (2006) 929–932.

[49] E.M. Meiering, L. Serrano, A.R. Fersht, Effect of active site residues in barnase on activity and stability, J. Mol. Biol. 225 (1992) 585–589.
[50] J. Weinstein, O. Khersonsky, S.J. Fleishman, Practically useful protein-design methods combining phylogenetic and atomistic calculations, Curr. Opin. Struct. Biol. 63 (2020) 58–64.
[51] O. Khersonsky, S.J. Fleishman, Why reinvent the wheel? Building new proteins based on ready-made parts, Protein Sci. 25 (2016) 1179–1187.
[52] G. Lapidoth, O. Khersonsky, R. Lipsh, O. Dym, S. Albeck, S. Rogotner, S.J. Fleishman, Highly active enzymes by automated combinatorial backbone assembly and sequence design, Nat. Commun. 9 (2018) 2780.
[53] O. Khersonsky, R. Lipsh, Z. Avizemer, Y. Ashani, M. Goldsmith, H. Leader, O. Dym, S. Rogotner, D.L. Trudeau, J. Prilusky, P. Amengual-Rigo, V. Guallar, D.S. Tawfik, S.J. Fleishman, Automated Design of Efficient and Functionally Diverse Enzyme Repertoires, Mol. Cell. 72 (2018) 178-186.e5.
[54] R. Netzer, D. Listov, R. Lipsh, O. Dym, S. Albeck, O. Knop, C. Kleanthous, S.J. Fleishman, Ultrahigh specificity in a network of computationally designed protein-interaction pairs, Nat. Commun. 9 (2018) 5286.
[55] S. Warszawski, A. Borenstein Katz, R. Lipsh, L. Khmelnitsky, G. Ben Nissan, G. Javitt, O. Dym, T. Unger, O. Knop, S. Albeck, R. Diskin, D. Fass, M. Sharon, S.J. Fleishman, Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces, PLoS Comput. Biol. 15 (2019) e1007207.
[56] S.J. Fleishman, T.A. Whitehead, E.M. Strauch, J.E. Corn, S. Qin, H.X. Zhou, J.C. Mitchell, O.N. Demerdash, M. Takeda-Shitaka, G. Terashi, I.H. Moal, X. Li, P.A. Bates, M. Zacharias, H. Park, J.S. Ko, H. Lee, C. Seok, T. Bourquard, J. Bernauer, A. Poupon, J. Aze, S. Soner, S.K. Ovali, P. Ozbek, N.B. Tal, T. Haliloglu, H. Hwang, T. Vreven, B.G. Pierce, Z. Weng, L. Perez-Cano, C. Pons, J. Fernandez-Recio, F. Jiang, F. Yang, X. Gong, L. Cao, X. Xu, B. Liu, P. Wang, C. Li, C. Wang, C.H. Robert, M. Guharoy, S. Liu, Y. Huang, L. Li, D. Guo, Y. Chen, Y. Xiao, N. London, Z. Itzhaki, O. Schueler-Furman, Y. Inbar, V. Potapov, M. Cohen, G. Schreiber, Y. Tsuchiya, E. Kanamori, D.M. Standley, H. Nakamura, K. Kinoshita, C.M. Driggers, R.G. Hall, J.L. Morgan, V.L. Hsu, J. Zhan, Y. Yang, Y. Zhou, P.L. Kastritis, A.M. Bonvin, W. Zhang, C.J. Camacho, K.P. Kilambi, A. Sircar, J.J. Gray, M. Ohue, N. Uchikoga, Y. Matsuzaki, T. Ishida, Y. Akiyama, R. Khashan, S. Bush, D. Fouches, A. Tropsha, J. Esquivel-Rodriguez, D. Kihara, P.B. Stranges, R. Jacak, B. Kuhlman, S.Y. Huang, X. Zou, S.J. Wodak, J. Janin, D. Baker, Community-wide assessment of protein-interface modeling suggests improvements to design methodology, J. Mol. Biol. 414 (2011) 289–302.
[57] R. Moretti, S.J. Fleishman, R. Agius, M. Torchala, P.A. Bates, P.L. Kastritis, J.P.G.L.M. Rodrigues, M. Trellet, A.M.J.J. Bonvin, M. Cui, M. Rooman, D. Gillis, Y. Dehouck, I. Moal, M. Romero-Durana, L. Perez-Cano, C. Pallara, B. Jimenez, J. Fernandez-Recio, S. Flores, M. Pacella, K. Praneeth Kilambi, J.J. Gray, P. Popov, S. Grudinin, J. Esquivel-Rodríguez, D. Kihara, N. Zhao, D. Korkin, X. Zhu, O.N.A. Demerdash, J.C. Mitchell, E. Kanamori, Y. Tsuchiya, H. Nakamura, H. Lee, H. Park, C. Seok, J. Sarmiento, S. Liang, S. Teraguchi, D.M. Standley, H. Shimoyama, G. Terashi, M. Takeda-Shitaka, M. Iwadate, H. Umeyama, D. Beglov, D.R. Hall, D. Kozakov, S. Vajda, B.G. Pierce, H. Hwang, T. Vreven, Z. Weng, Y. Huang, H. Li, X. Yang, X. Ji, S. Liu, Y. Xiao, M. Zacharias, S. Qin, H.-X. Zhou, S.-Y. Huang, X. Zou, S. Velankar, J. Janin, S.J. Wodak, D. Baker, Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions, Proteins. 81 (2013) 1980–1987.