The Degreaser: spot cleaning sequences to maximize protein secretion

Jing Yang Wang

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Neil King, Chair

David Baker

Dan Stetson

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

The Degreaser: spot cleaning sequences to maximize protein secretion

Jing Yang Wang

Chair of the Supervisory Committee:

Neil King

Department of Biochemistry

As computational protein design enables the creation of novel features, forms and functions, such as designed protein nanoparticles, expression and secretion from eukaryotic cells becomes advantageous for the generation of new biologics. However, many designed proteins secrete poorly. Because hydrophobic interfaces are designed into otherwise soluble proteins, many of these proteins gain hydrophobic segments that may be interpreted by cellular membrane insertion machinery to be transmembrane domains. To address this, we develop a computational method based on a transmembrane insertion prediction model and Rosetta: the Degreaser. We use the Degreaser to identify cryptic transmembrane domains and design them away without

compromising the originally-designed protein. Retroactive application of the Degreaser to previously designed nanoparticle components and nanoparticles considerably improves secretion. Modular integration of the Degreaser into design pipelines and incorporation of the Degreaser in large-scale protein design results in proteins that secrete robustly. These secretion-optimized proteins represent the first generation of proteins that are specifically designed for maximal secretion. Future generations of designed proteins can incorporate other secretion optimization features, especially if they can be selected from high-throughput, multiplex screens.

A version of portions of this thesis are currently in preparation as a journal article entitled:

Improving the secretion of designed protein assemblies through negative design of cryptic transmembrane domains

Jing Yang (John) Wang, Alena Khmelinskaia, William Sheffler, Marcos C. Miranda, Aleksandar Antanasijevic, Andrew J. Borst, Susana Vazquez Torres, Chelsea Shu, Yang Hsia, Una Nattermann, Daniel Ellis, Carl Walkey, Maggie Ahlrichs, Sidney Chan, Alex Kang, Hannah Nguyen, Claire Sydeman, Banumathi Sankaran, Mengyu Wu, Asim K. Bera, Lauren Carter, Brooke Fiala, Michael Murphy, David Baker, Andrew B. Ward, and Neil P. King

List of figures

Acknowledgements

I have been extremely fortunate to be able to arrive to where I am today. There are more people than I could possibly acknowledge on this page, but I will attempt to highlight individuals chronologically. First, I'd like to acknowledge Kenneth Higa and Yiling Dai, who helped start my career in research as an undergraduate. I'd like to acknowledge Jim Bowie, Nick Woodall, Tyler Korman, Paul Opgenorth, Meaghan Valliere, Robert Jefferson, and Duyoung Min, who were Bowie lab members that were supportive when I was there, and of my decision to leave UCLA. I'd like to acknowledge Nina Salama, Rich Gardner, and Maia Low, who took a chance on a transfer graduate student and helped provide a home department here at University of Washington. I'd like to thank David Baker and Neil King for taking me under their wings for rotations. I'd also like to specifically thank Neil for the mentorship, time spent in his office hashing out experiments, looking over data, and poring over manuscripts. I'd also like to thank my committee members, Jesse Bloom, Alex Merz, and Dan Stetson for guiding advice.

I'd like to thank the academic community near and far throughout my graduate career. Karla-Luise Herpoldt, Brooke Fiala, Cassie Ogohara, Daniel Ellis, Quinton Dowling, Kate Dapron, Adam Wargacki, and Isaac Sappington provided so much mentorship and were pioneers for the burgeoning King lab. I'd like to thank Thad Huber, a former member, for great time spent discussing basically everything. I'd also like to thank Chloe Adams, Naveen Jasti, Jake Kraft, Audrey Olshefsky, Aaron Sciore, Cara Chao, Jung-Ho Chun, Annie Dosey, Dan Humphrys, Geoff Hutchinson, Mark Langowski, Marti Tooley, Erin Yang, and Dane Zambrano, current members of the King lab who enrich the environment every day. I'd like to specifically thank Vikram Mulligan, Morgan Nance,

Seygey Lyskov, and Rocco Moretti, who I've interacted with (some only virtually) who are literally building Rosetta up constantly. I'd like to thank our collaborators at the VRC, Masaru Kanekiyo, Julia Lederhofer, Rebecca Gillespie, and Seyhan Boyoglu-Barnum who taught a biochemist all about virology and vaccinology. Our critical core members, Lauren Carter, Marcos Miranda, Deleah Pettie, Michael Murphy, Sidney Chan, Maggie Ahlrichs, and others, who make the production of the molecules we work with, which would be so difficult elsewhere, a routine. Special shoutout to Kandise VanWormer, Ratika Krishnamurty, Lance Stewart, and the rest of IPD administration who keep the wheels turning on this rollercoaster.

Alena Khmelinskaia and Chelsea Shu have been essential to this work: Alena for the conceptualization and design of so many protein nanoparticles and characterization work, and Chelsea for a bioinformatic look at the data that we were analyzing. These two amazing women have represented both mentor and mentee in my life, providing experience that I honestly cannot say I would have been able to get from anyone else.

Bryan Andrews, Brooke Angell, Cindy Wei, Roger Tu, Courtnee Clough, Sean Evans, Steve Shen, Henry Chan, Michael Hughes, and Daniel Reveles have been close friends in grad school and close to graduate students that really helped understand and work through the graduate school experience.

Finally, I would like to thank my wife and partner, Taylor Wang, who has made every hour of life outside of work an absolute blast, for sticking it through with me through this crazy process of getting a PhD, for weathering an unprecedented global viral pandemic together, and for always bringing a smile to my days.

Dedication

This work is dedicated to the people throughout my life

who encouraged me to never stop asking questions.

Chapter 1 - Introduction

1.1 The eukaryotic secretory pathway

Nearly 20% of the human proteome consists of secreted proteins [1]. The ability for a cell to transmit cargo into the intracellular environment is critical for communication and underlies multicellular biology [2]. In mammals, antibodies, hormones, cytokines, and growth factors represent just some of the variety of secreted proteins that effect biological functions and are desirable to imitate or target by therapeutics. Membrane-anchored proteins comprise key receptors and structural features that many secreted proteins engage with. In eukaryotes, the secretory pathway, spanning the early endoplasmic reticulum (ER) to the late Golgi, serves as a major thoroughfare for the proper generation, targeting, and maintenance of secretory and membrane proteins [3]. Recombinant secretion of proteins is also mandatory for proteins whose organism of origin are eukaryotic, although some prokaryotic or *in vitro* methods are being engineered to circumvent this [4,5]. Furthermore, the secretory pathway houses specific machinery for post-translational modifications of these proteins, such as glycosylation or furin-mediated proteolytic cleavage. For example, furin cleavage is necessary for the maturation of parathyroid hormone (PTH) in mammals [6,7]. The traversal of a protein through the secretory pathway is critical to understand the pathway itself.

Secreted and membrane proteins typically start with signal peptides, which target them for the secretory pathway. Up to the first 30 residues of the protein constitute the signal peptide, which recruits the signal recognition particle (SRP) and is inserted into the ER

membrane via the Sec translocon [8–10]. The translating polypeptide is then fed through the translocon into the ER, and hydrophobic segments partition into the ER membrane, which after Golgi processing fuses with the plasma membrane, carrying the embedded transmembrane domain [11]. Secreted proteins translocate entirely into the ER, which then mature through secretory vesicles that eventually fuse with the plasma membrane to eject their cargo into the extracellular space [8,12]. This maturation of secreted proteins is a highly regulated process, with strict quality control and maintenance proteins that actively remove any misfolded or otherwise poorly behaving proteins [3,13]. For secreted proteins, the signal peptide is cleaved soon after the rest of the protein begins translocating, while for membrane proteins, the first transmembrane domain serves as the signal peptide, but is not cleaved [14,15]. As more of the protein is translocated and folds, enzymes that recognize glycosylation motifs are recruited to the nascent protein and begin attaching glycans [16]. These glycans help quality control machinery keep track of a protein's folded and trafficking status, and a significant proportion of secreted proteins are glycosylated [17]. Maturing protein is then targeted to trafficking vesicles that finalize the protein's locale, which can be the plasma membrane or even resident within the secretory pathway.

Proteins that trigger a response from quality control machinery, typically when the protein is poorly folded, which can lead to aggregation, or when certain glycosylation cycle checks fail, are removed from the secretory pathway through a process known as endoplasmic reticulum-associated degradation, or ERAD [18,19]. Because the secretory pathway itself does not contain proteasomes, which degrade much a cell's protein,

these misbehaving proteins are retrotranslocated through Sec translocon complexes and out of the secretory pathway [20,21]. Once that protein is removed from the ER, ERAD complexes recruit downstream proteins (such as ubiquitin ligases) that target a polypeptide for proteasomal degradation. Thus, many proteins that are not stable or soluble tend not to accumulate in cells attempting to secrete them; they are likely to be degraded rather than allowed to linger within the cell. This degradation can confound experiments examining protein secretion, as the absence of protein detection can be explained by either active degradation or overall poor expression.

1.2 The Sec translocon and prediction of transmembrane domains

Many studies of the Sec translocon (SecYEG) are conducted in prokaryotic models, but the proteinaceous channel that resides within membranes that allow for polypeptides to partition into those membranes is conserved across all domains of life (Sec61 in eukaryotes) [22]. Itself a complex of transmembrane proteins, it serves as the gatekeeper of either nascent polypeptides or fully folded transmembrane proteins. The translocon also serves as a hub for chaperone engagement, quality control protein recruitment, and post-translational modification machinery [10]. In eukaryotes, the translocon directly engages with ribosomes that are complexed with the SRP to allow the passage of translating polypeptides into the ER lumen. Structural and biochemical studies have elucidated the mechanisms underlying its main function: the translocon must be able to determine the hydrophobicity of the peptides that pass through it [23–25]. For nascent polypeptides, this hydrophobicity sensing is a passive process, wherein the segment’s biophysical properties are sufficient to allow it to exit the translocon core. Molecular dynamics models have been used to very precisely find

explanations to bulk effects that are observed for transmembrane proteins [26,27]. Most analysis of the Sec translocon focuses on transmembrane proteins, and while secreted proteins are not as well-understood, the secretomes of various organisms have begun to be characterized.

To understand in-depth the effect of a protein's sequence on its biochemical and biological behavior requires both an understanding of the biology and some working model for its underlying mechanisms. For protein partitioning by the Sec translocon, biochemical studies of protein hydrophobicity have a history of using chemical intuition and properties to inform such models. Hydrophobicity scales for individual amino acids, for model peptides of certain length, and even of model proteins have been generated to understand the partitioning of polypeptides into membranes [28–31]. These models are typically either bottom-up, where fundamental physical principles are used to design experiments that query specific effects, or top-down, where empirical models that incorporate many potential contributing factors are used to build a more general model. For the prediction of transmembrane protein insertion, many such models have been developed to accurately identify transmembrane domains and engineer transmembrane domains. These models have resulted in the availability of several websites for transmembrane domain prediction and transmembrane structure prediction [32–35]. Because the majority of characterized transmembrane domains are alpha-helical, most models use alpha-helix-based calculations. A 2005 study by Hessa *et al.* established an *in vitro* translation model for measuring transmembrane partitioning in a top-down manner [30]. In 2007, the same group followed up their previous study by describing a

model that was based on their measurements that represented every amino acid identity at every position within a model alpha-helix [36]. Their model is symmetric and takes the form of a Gaussian equation, which is readily encodable for computational methods and rapid to calculate. Other hydrophobicity scales were generated more recently but may not measure position-specific contributions, such as the Moon-Fleming scale [31], or were difficult to describe mathematically and thus encode for computational design purposes [37,38]. However, this does not preclude their future usage for design applications.

The generation of hydrophobicity scales from so many different sources and assays underscores the importance of an empirical model to guide the prediction of transmembrane domains in known sequences. However, also critical is the usage of these models to guide computational protein design, where an arbitrarily large number of sequences can be generated. Therefore, models that can account for as many features within a model transmembrane segment as possible but are still simple enough to describe when encoding in programs are ideal. For this work, the model from Hessa *et al.* was chosen to underpin the computational method generated to design proteins away from transmembrane domains. Computational comparison of the results from multiple different transmembrane prediction models was possible, but outside the scope of the intended applications of this work.

1.3 Computational design of protein-protein interfaces and self-assembling protein nanoparticles

De novo designed proteins are a rapidly growing platform that allows the generation of novel protein sequences that are not explored by the already diverse world of natural proteins [39]. Computational design enables the sampling of sequence space in a way that is not limited by evolutionary constraints [40,41]. The challenge, then, is the downselection from a near-infinite number of sequences to a set of sequences that have some desired biochemical property. For example, the generation of a massive set of miniproteins and subsequent multiplex protease-based stability assay revealed sequence determinants of thermodynamic stability, even when the examined proteins were much smaller than natural proteins [42]. This and other phenotypes can be designed with specific arrangements of the design protocols available within and without Rosetta. Rosetta, at its core, enables Monte Carlo-based sampling of protein space, choosing at each iteration the most stable sequence available to a structural conformation given a guiding scoring function [43]. With a bit of biochemical intuition, such as the designation of core, boundary, and surface regions of a designed protein, Rosetta has been used to design a wide variety of proteins with a diverse range of functions [44–48]. Many recently designed proteins take advantage of hydrophobic packing to enable protein-protein interactions that result in specific binding of a target, and also to generate proteins with new geometries, such as megadalton-scale nanoparticles [49,50].

The design of protein nanoparticles begins with geometric docking of one or two input structures, which are treated as rigid bodies. These rigid bodies are allowed to rotate and translate against each other, but unlike protein-protein docking, the proteins are confined to a desired geometry, such as tetrahedral (T), octahedral (O), or icosahedral (I). This symmetry designation reduces the degrees of freedom available to the inputs, decreasing the amount of sampling necessary to the computational protocol. Originally pioneered by Neil King and Will Sheffler, tcdock was a protocol that resulted in the creation of several two-component protein nanoparticles [51]. Extension of this type of work has resulted in rpxdock, the latest available rigid body docking protocol for nanoparticle design [Sheffler, Yang, Hsia, Dowling *et al.*, unpublished] . After the input proteins are docked into a desired configuration, Rosetta interface design is used to allow the two proteins to interact. These designed interfaces are typically fairly hydrophobic because the layer-based design that generates them allows only nonpolar residues at interfacial cores. Then, when evaluating design metrics such as ddG of interface formation, models with more hydrophobic interactions at those cores tend to score more favorably.

This hydrophobic interface design, then, is directly at odds with the efficient secretion of the proteins that harbor such interfaces. However, because *E. coli* was used for the expression of all of the proteins previously, transmembrane insertion was typically not an issue. Thus, we term segments of high hydrophobicity that exist in otherwise soluble proteins cryptic transmembrane domains. Proteins that were designed with hydrophobic surfaces were likely rejected during initial protein screening, as such proteins are likely

to be insoluble or may not even be expressed. Proteins that included hydrophobic segments but were overall soluble were likely to be well-tolerated in a bacterial expression model, especially when both components of a two-component nanoparticle were expressed simultaneously, because small hydrophobic surfaces could be compensated for by the interaction partner. However, eukaryotic secretory expression of these proteins becomes more difficult, because more stringent quality control mechanisms are in place. Furthermore, hydrophobic segments play a role in transmembrane insertion where they were unlikely to in bacterial culture, because of the signal sequences appended to the expressed proteins that target them to the secretory pathway. Finally, these protein pairs were not expressed simultaneously in eukaryotic culture, where nanoparticle assembly could potentially confound expression results, yet this could be advantageous with respect to rapidly screening away poorly-behaved designs.

1.4 Recombinant secretion of exogenous proteins

Several designed nanoparticle proteins have been expressed from mammalian cell culture for use in multivalent vaccine candidates [48,52–54]. These, as yet, are purified and then assembled with bacterially-purified assembly partners. Even so, these materials have been shown to be exceptionally valuable as immunogens, and they have potential as delivery vehicles *in vivo*. However, mammalian secretion of protein nanoparticles is still difficult for some designs. Proteins that are not well-behaved *in vitro* tend not to be expressed well *in vivo*. Optimization of these proteins for stability and circulation time has been performed, but for only one protein nanoparticle [55]. This process of multiple rounds of evolution is time intensive for multiple different proteins,

and thus design of new proteins that can have given phenotypes, such as robust secretion, is desirable. When engineering proteins for robust secretion, there are several methods that are available to maximize the purified yield of a given protein. Cell line and strain engineering, plasmid vector construction, and optimization of expression conditions are all ways to extract as much protein as possible for a given construct [56–58]. However, because we would rather design more proteins with these qualities, and because for applications such as genetic delivery of vaccine materials where host engineering is impractical, the tailored design of the protein sequence itself is the chosen approach. This provides several limitations to any methods development to optimize secretion, but, if attained, enables a generalizable strategy that accommodates creation of new and better design methods.

Given the success of the King lab flagship nanoparticle immunogen, DsCav1 / I53-50, which has been shown to elicit potent immune responses relative to unassembled protein immunogens, we chose to proceed with human cell culture [52]. We also chose to redesign existing protein nanoparticles, and devised a new computational protocol, the SecretionOptimizationMover, or Degreaser, to facilitate the secretion of nanoparticle proteins. We hypothesized that canonical Rosetta-based protein design introduces hydrophobic segments into proteins that lead to poor secretion, and that the mechanism underlying such poor secretion is the failure of the protein to translocate efficiently into the ER lumen during expression. This gave us (1) a sequence-level engineering method for protein design, amenable to platforms such as genetic delivery, (2) a general computational method that is compatible with any protein design, not just nanoparticle

interface design, and (3) insights into the determinants of protein secretion and surface display that can serve as bases for future assay development.

Chapter 2: The Degreaser: a computational method for streamlined Rosetta-compatible design of secretion-optimized proteins

2.1 Introduction

Given the burgeoning interest in eukaryotically-expressed proteins, as well as the myriad novel applications available to *de novo* designed proteins, new computational methods are necessary to meet rising demand. The generation of a method that can incorporate sequence-level metric evaluation into Rosetta's existing energy calculations requires several features that did not previously exist within Rosetta architecture. Furthermore, this method should be compatible with other design protocols and methods, as well as robust to changes in Rosetta code over time. Finally, it should also be straightforward and simple enough to use such that even without dedicated developer support, users (that is, protein designers) can achieve the intended outcomes of the method. These are the greater considerations for the Degreaser, though there are many more minor considerations that will be addressed, as well.

In order to introduce sequence-based transmembrane insertion potential prediction into Rosetta, the equations from the 2007 model constructed by Hessa *et al.* are directly coded into the Mover [36]. This is one of the advantages of the Hessa model: the symmetric form of the Gaussian equations used to describe transmembrane insertion potential, as well as their fitted coefficients for each equation and amino acid allows rapid calculation of individual dG_{ins} contribution. Then, with a designation for a sliding window of amino acids (nominally 19), a $dG_{ins,pred}$ vs sequence index plot, like the ones generated by the Hessa webserver, can be created for any input protein structure. The

outputs from the Rosetta-integrated and webserver calculations differed by less than 0.001 kcal/mol, which indicated successful recapitulation of previously-described work. With the capability to rapidly calculate $dG_{ins,pred}$ for an arbitrary segment of amino acids, a simple logical workflow for the Degreaser and its actions on an entire protein structure are built to enable up- or downstream compatibility.

One of the strengths of Rosetta-based protein design is the modularity of its individual components, as well as the relative ease with which they can be brought together. This modularity and compatibility was one of the main characteristics designed into the Degreaser. Movers, as a subset of existing Rosetta protocols, are unique in that they cause some perturbation to an input protein structure, juxtaposed most frequently with Filters, which may use *in situ* perturbations or other changes to the structure to evaluate metrics of an input protein structure, but ultimately effecting no change on that structure [43,59,60]. With this in mind, Movers need to inherit, or at least be capable of inheriting, other user-designated options for what perturbations are and are not allowed. These designations typically come in the form of TaskOperations, which enable users to specify such parameters as which residues are allowed to change, to what residues they can change, and how far away from their initial position they are allowed to move, to name a few. Therefore, the Degreaser needs to be able to handle TaskOperations, and keep those designations without overriding them, but still be able to independently evaluate which changes to make during its own runtime.

To this end, the Degreaser inherits any TaskOperations from the protocol or script using it, but also calculates a second, internal TaskOperation when redesigning residues. Because the perturbation of a residue may lead to small atomic collisions (clashes) or other steric incompatibilities, the Mover needs a way to accommodate any changes that it introduces, in addition to its calculations trying to decide the best residue to place in a particular position. With code initially developed by former Baker lab member Sarel Fleishman and extended by Neil King, the Degreaser incorporates a symmetry-aware DesignAround TaskOperation generator, which allows one to designate shells of a particular distance (e.g. 8 Angstroms), around which particular residues are allowed to move or change. In the case of the Degreaser, the perturbed residue is chosen as the center around which other residues are allowed to move, but are forbidden from changing. This enables the incorporation of amino acid changes that minimize the overall perturbation to the stability of the original protein structure. The initial selection of variant residue and rotamer position of surrounding residues is handled by Rosetta's PackRotamersMover, which uses a Monte Carlo-based sampling for determining the best choice [43].

With Rosetta-based energy calculation and sequence-level transmembrane insertion potential evaluation, a small amount of logic and some decision-making are required for the Degreaser to be run in a "set it and forget it" mode. The key to generate variants that actually significantly increase $dG_{ins,pred}$ is exhaustive sampling of candidate positions residues, the results of which are discussed in section 2.2.1. With this, the other critical decision for generating an output variant is to choose the single variant that

maximally influences the overall protein $dG_{ins,pred}$. In a high-throughput design scenario, inspection of individual candidates becomes intractable, and so the Degreaser, by default, selects the variant that most increases $dG_{ins,pred}$ in the region of lowest initially-identified $dG_{ins,pred}$ (Figure 2.1). Although this selection could potentially include variants that are non-ideal, such as introducing polar residues into hydrophobic cores, it most closely approximates manual selection of Degreaser candidates. Also available is the selection of the variant that least impacts Rosetta score, serving as a proxy for the least perturbed variant that still passes other Degreaser evaluation. The choice of either of these decisions enables fully automated redesign of proteins. Computational outcomes are discussed in this section, while experimentally characterized outcomes are described in Chapter 3.

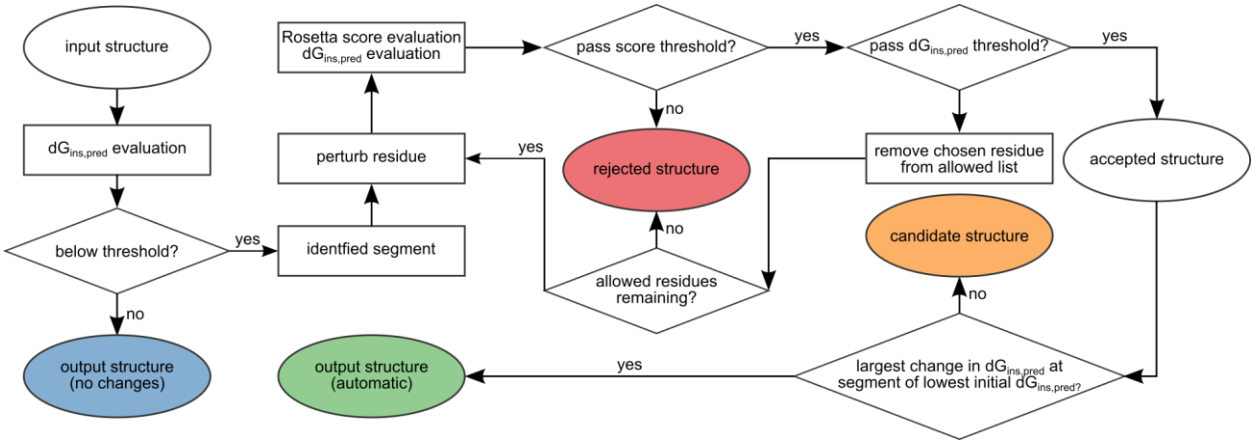


Figure 2.1. Flowchart of Degreaser decision-making. Color-filled ovals represent possible outputs for the Degreaser. Rejected structures (red) are for positions that are considered to be non-mutable, candidate structures (orange) are for positions that are extra variants for evaluation when inputs are limited. Output structures (green, Degreased; blue, unchanged) can be considered "safe" for secretion, but some Degreased structures can still have low $dG_{ins,pred}$ if the input structure has extremely low $dG_{ins,pred}$ that cannot be increased to above a secretion threshold.

2.2 Results

2.2.1 The Degreaser exhaustively samples variants, opposing Rosetta's Monte Carlo approach

Because the sequence space sampled by the Degreaser is not as large as that sampled for *de novo* protein design, we reasoned that, for a given residue within a given identified region of low $dG_{ins,pred}$, every possible variant could theoretically be sampled in a reasonable amount of time. For one position, allowing as many polar residues as possible (DEKRQNSTYH), a maximum of ten cycles would be sufficient, and repacking

of nearby residues (for an 8 Angstrom repack distance, usually about 12 residues [Figure 2.2, left]), is very fast for Rosetta-based design. Thus, for one identified segment of low $dG_{ins,pred}$, only 190 cycles maximum are necessary. This case of 190 cycles without choosing a residue takes about 140 seconds, which is a small amount of computational time relative to the time required to design a protein before that point (Figure 2.2, right). Examination of Degreaser runtime messages and outputs confirms that, at each position, every allowed input residue is evaluated when the $dG_{ins,pred}$ change threshold and score threshold are set arbitrarily high, effectively ensuring that there is no real convergence for the Mover.

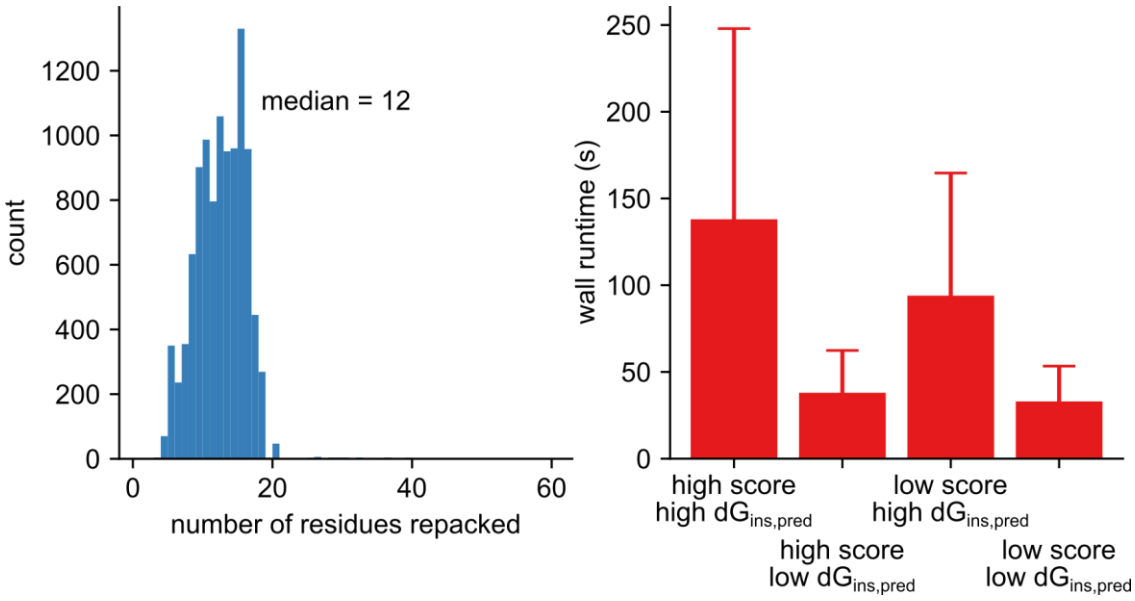


Figure 2.2. Degreaser runtime metrics. (left) Residues repacked per step within representative Degreaser runs. (right) Representative output times against parameter conditions. Seven designed nanoparticle protein models and two natural protein models were used to generate these data. The low score, low $dG_{ins,pred}$ condition most closely represents actual Degreaser application.

This exhaustive (in the case of a high $dG_{ins,pred}$), or at least, near-exhaustive (in more typical use cases) opposes Rosetta design's core principle of sampling sequence space in a Monte Carlo manner. However, because the Degreaser is still sufficiently fast, this sampling is desirable because it increases the chances of identifying variants that significantly increase $dG_{ins,pred}$. Exhaustive sampling is also present in other protocols in Rosetta, such as the HBNet Mover, which exhaustively samples side chain coordinates for the identification of potential hydrogen bonding networks, such as within the core of *de novo* designed helical bundles [44]. This bolsters the concept of combining non-

random sampling to gain some desired phenotype with Rosetta's design methods to stabilize proteins in order to enable application-tailored, model-guided protein design.

2.2.2 Degreaser-identified variants significantly perturb transmembrane insertion potential

With sequence-based dG_{ins} prediction and either redesign or in-line design of proteins of interest, simple re-evaluation of $dG_{ins,pred}$ (as perturbations are made) may not be sufficient to actually influence secretion. For example, a leucine to aspartate perturbation is one of the largest changes in $dG_{ins,pred}$ possible if the residue lies in the middle of a candidate segment, but has only a modest effect near the edges. Furthermore, with sparsely available expression-to-$dG_{ins,pred}$ correlation information, the determination of what constitutes a sufficient perturbation is difficult. Drawing on available secretion data from designed proteins from the King lab, primarily designed protein nanoparticle components, as well as confidently-identified transmembrane and non-transmembrane structures from the PDB, we see a discriminatory threshold of about +2.7 kcal/mol (Figure 2.3, left). Incidentally, this would correspond to a predicted 100 to 1 ratio of translocated to inserted protein according to the Hessa model, but there need not be a stoichiometric relationship between translocation and overall secretion. With this threshold, a 0.27 kcal/mol delta($dG_{ins,pred}$) threshold was chosen for the Degreaser, somewhat arbitrarily. This threshold is intended to prevent incorporation of variants that had little to no effect on $dG_{ins,pred}$--for example, a native valine could be redesigned to threonine, but because threonine is not much more polar than valine, the net effect on $dG_{ins,pred}$ is likely to be very small.

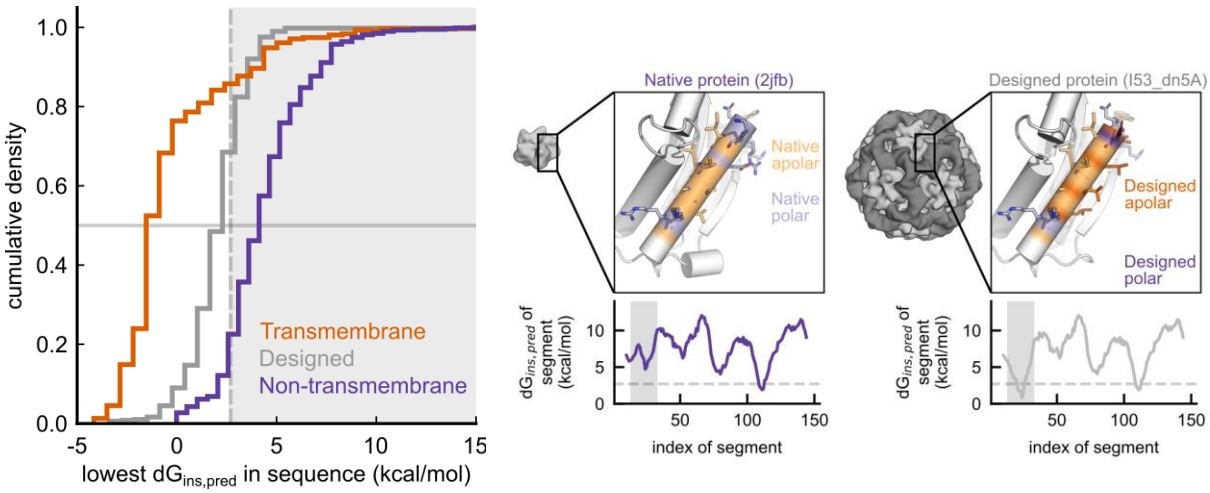


Figure 2.3. dG$_{ins,pred}$ comparison of designed proteins from [50,51] (grey) with transmembrane (orange) and non-transmembrane proteins (purple) in the PDB. (left) The +2.7 kcal/mol threshold (vertical dashed line) discriminates between natural transmembrane and non-transmembrane proteins (85% and 16%, respectively). 61.3% of nanoparticle proteins contained segments below this threshold. A representative example (right) highlights the introduction of a cryptic transmembrane domain into a natural protein after design (grey plot); the designed protein is soluble in bacterial expression. Figure generated with data collected and analyzed by Chelsea Shu.

After running the Degreaser on a small set of previously-designed protein nanoparticle components that were of interest for secretion, we found that many single amino acid changes significantly (>0.5 kcal/mol) increased dG$_{ins,pred}$. This was consistent with our model for otherwise soluble proteins containing hydrophobic segments. Because these proteins are not true transmembrane proteins and therefore do not require membrane spanning segments, there are positions within these segments that are amenable to perturbation and are near the center of each segment. Because a polar residue at the

center of a lipid bilayer is highly unfavorable, a single apolar to polar transition at the center of such a segment can strongly increase $dG_{ins,pred}$. Furthermore, these polar residues are typically well-tolerated in these proteins because the secondary structure elements present within these proteins are not long, hydrophobic helices; rather, they tend to be relatively short helices with residues in turns or loops that were designed to be hydrophobic to assist in protein stabilization or interface formation and are therefore good candidates for mutation to polar residues. Indeed, most of the Degreaser-identified variants contain mutations in relatively solvent-exposed regions of the initial protein or designed interface (Figure 2.4).

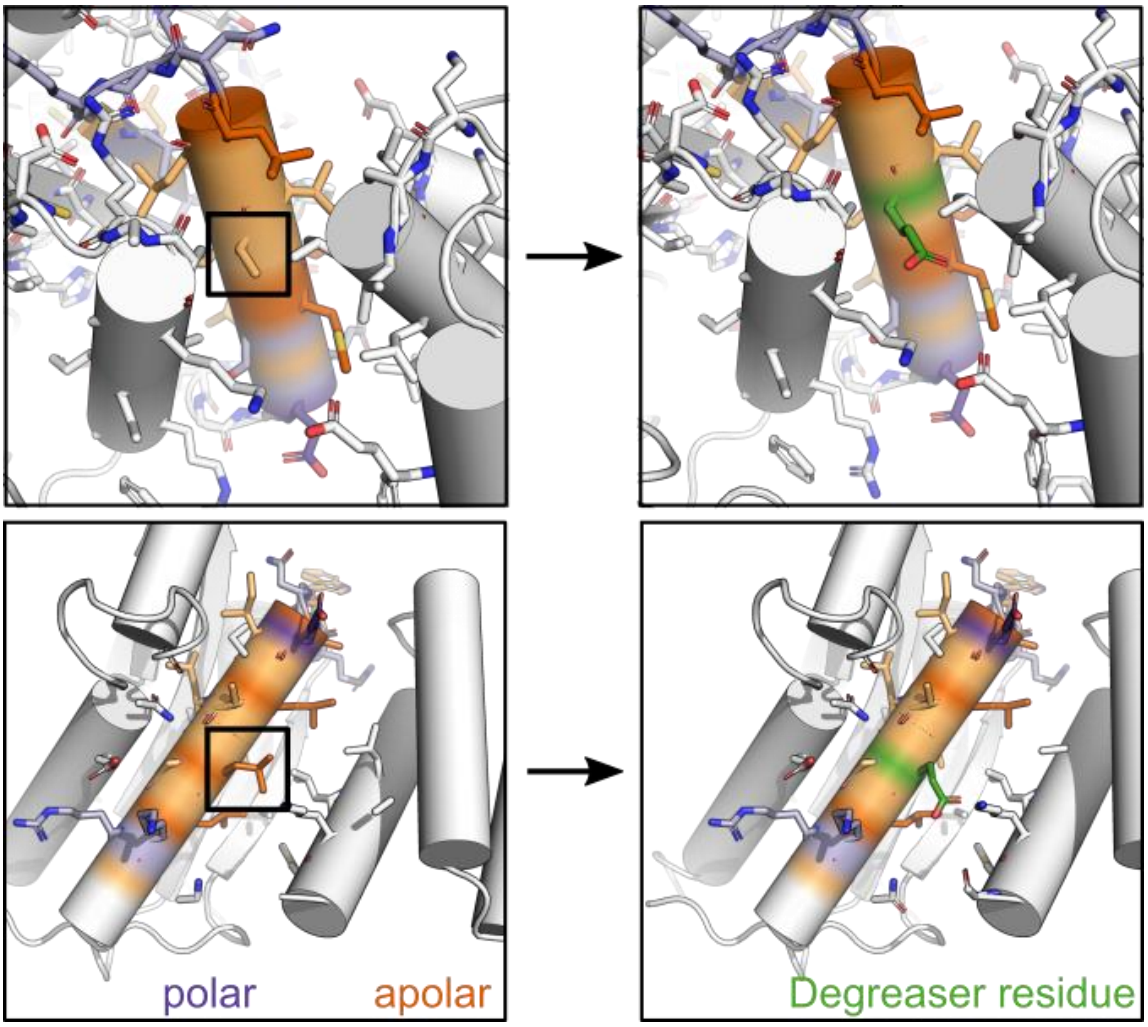


Figure 2.4. Degreaser-identified residues are relatively solvent-exposed, allowing for incorporation of polar residues that significantly increase $dG_{ins,pred}$. Naturally occurring polar (dark purple) and apolar (dark orange) residues create environments where the introduction of designed apolar (light orange) and designed polar (light purple) residues result in a cryptic transmembrane domain. The Degreaser identifies residues within these segments (boxed residues, left), even within the context of their designed interfaces. The mutation of these residues to polar residues (green, right) typically do not completely bury polar or charged side chains within hydrophobic cores.

We also assessed Degreaser redesign for a one-component designed protein nanoparticle, I3-01, which was poorly secreted from mammalian cell culture but robustly expressed in bacterial culture [49]. The designed interface in this nanoparticle introduces a hydrophobic segment toward the N-terminus, which oblates secretion. We deduced that this interface is critical for redesign because the nanoparticle has high sequence identity with its cognate natural scaffold, PDB 1WA3. Several single amino acid variants were identified for I3-01. Two of these appeared well-suited for boosting secretion, H35D and F41Q, which increased $dG_{ins,pred}$ of that segment by 0.71 and 0.86 kcal/mol, respectively. Many variants were also identified in the C-terminal modestly hydrophobic segment, which is also present in the natural protein. This region was also more tolerant of mutation, and so was not regarded as critical with respect to increasing the $dG_{ins,pred}$ of the overall protein.

Heartened by these findings for a small set of designed proteins, we applied the Degreaser to larger sets of designed proteins. Because the repertoire of soluble, secreted designed protein nanoparticles was small, King lab member Alena Khmelinskaia started with the design of one-component protein nanoparticles with scaffolds based on *de novo* designed proteins. Two-component protein nanoparticles were still inconsistent with respect to their response to Degreaser application--some nanoparticle components could be strongly secreted while others expressed poorly independent of the Degreaser--and so were left for future design and characterization attempts. Retrospectively, Degreaser-based design of secreted two-component nanoparticles should be straightforward (relative to bacterially-expressed) because the

design success rates of Degreased and non-Degreased one-component nanoparticles were comparable (further discussed in section 2.2.5). Within this set of Degreased one-component nanoparticles, we found that even in the most conservative of approach allowing only one Degreaser mutation per design, we could shift the distribution of $dG_{ins,pred}$ higher, resulting in an increase in median $dG_{ins,pred}$ of 0.40 kcal/mol (Figure 2.5). Notably, this set of designs were based on scaffolds that already had relatively high $dG_{ins,pred}$, meaning that a large proportion of designed nanoparticles did not need to be perturbed. However, those that were changed by the Degreaser saw an average increase in $dG_{ins,pred}$ of 1.41 kcal/mol. The findings from this design set suggested that more aggressive application of the Degreaser, that is, more mutations allowed per input structure, is not only tolerable but desirable for seeing larger differences in secretion outcomes.

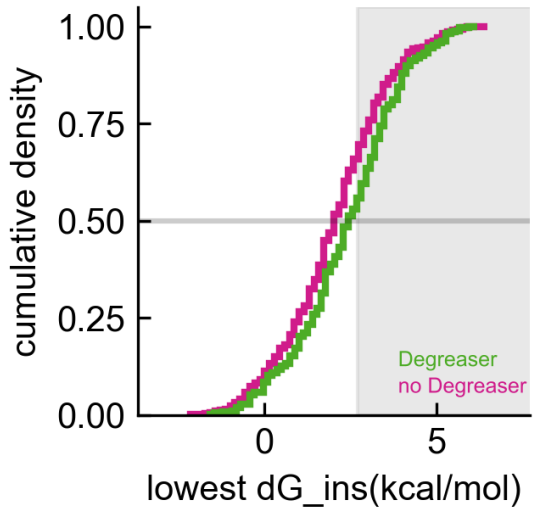


Figure 2.5. Application of the Degreaser (green) to a one-component nanoparticle design set increases median $dG_{ins,pred}$ by 0.40 kcal/mol relative to without the Degreaser (magenta). Although only 44% of designs were mutated by the Degreaser, this was enough to shift the distribution to the right, increasing the proportion of designs above the +2.7 kcal/mol threshold (vertical grey shading).

The Degreaser can also be applied to proteins intended for yeast surface display, which takes advantage of C-terminal fusion of a protein of interest to Aga2p, a protein that must be translocated into the ER lumen so that that protein can be displayed extracellularly [61]. Degreaser application allowing more mutations and larger Rosetta score increases revealed that designed helical peptide-binding proteins, designed by Baker lab member Susana Vazquez Torres, underwent $dG_{ins,pred}$ changes of several kcal/mol (Figure 2.6). These mutations were typically identified outside of the designed binding interface, suggesting that the Degreaser is sensitive enough to changes in score that those positions are not chosen to mutate. This highlights one of the

advantages of design-time calculation and redesign of hydrophobic segments: by modeling these binding proteins in complex with their targets, the total complex score is preserved while hydrophobic segments are designed away. Therefore, variants can be identified that bring candidate designs from unlikely to translocate and secrete to predicted robust secretion, even while preserving its original designed function. In addition, even if a particular candidate design is non-ideal with respect to its design metrics or $dG_{ins,pred}$, because more designs are above a “safe” threshold, the pool from which these proteins are chosen to characterize experimentally is dramatically expanded. This provides two main advantages for protein designers: first, there are more candidates from which to downselect that already have the desired property (high $dG_{ins,pred}$), and second, less computational time is required to generate the candidate pool. For example, with generation of 10,000 proteins and filtering on $dG_{ins,pred}$ without the Degreaser, one might be left with 1,000 proteins from which to choose those passing design metrics, but with the Degreaser, one could generate 5,000 proteins in a similar amount of computational time.

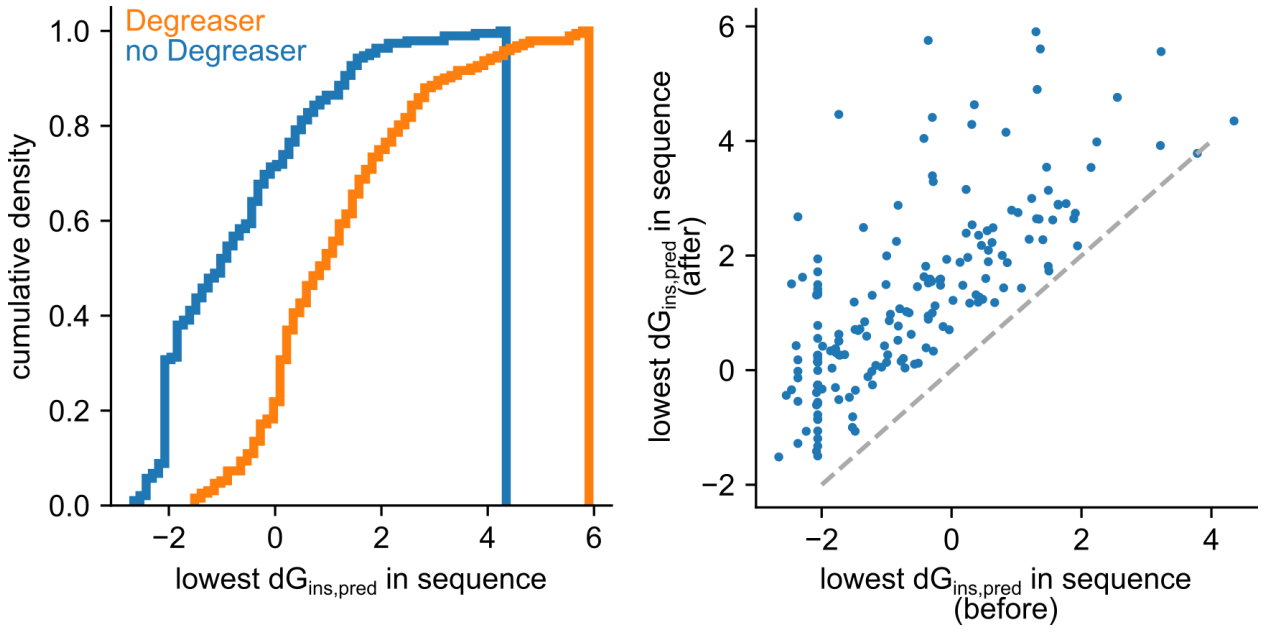


Figure 2.6. Application of the Degreaser to designed helical peptide-binding proteins increases dG$_{ins,pred}$. (left) Distributions of dG$_{ins,pred}$ of a paired set of designed binders (blue) with a substantial increase after Degreaser application (orange). (right) Scatter plot of individual redesigns, with dG$_{ins,pred}$ of a design after Degreaser application on the y-axis, with a y=x line in dashed grey. The median difference is +1.80 kcal/mol, while the median of individual differences is +1.92 kcal/mol.

2.2.3 Degreaser-identified variants do not perturb canonical design metrics

While the Degreaser effectively raises the dG$_{ins,pred}$ of proteins with strongly hydrophobic segments, simply changing arbitrary nonpolar residues to aspartate could potentially achieve a similar effect. However, “aspartate scanning” would likely introduce charges into hydrophobic cores, which are differentially tolerated by proteins. Some proteins may be able to accommodate the burial of a charged residue, but most computationally-designed protein-protein interfaces, be it for binding of a specific target or for protein

oligomerization, have snugly fitted hydrophobic cores that typically do not tolerate the burial of charged residues. Hydrophobic to charged (aspartate, glutamate, arginine, or lysine) amino acid changes are commonly used to break cognate protein-protein interactions. Thus, the Degreaser must be able to make such changes but not arbitrarily. Using Rosetta's Packer, even a naive scorefunction accurately identifies boundary regions or other regions that are solvent exposed for incorporation of polar residues. Also, many marginally boundary regions acquire uncharged polar residues during Degreaser design, even when other residues are allowed, suggesting that the final residues identified best preserve the initial structure's stability.

As-written, the Degreaser does not take into account other commonly-used design metrics, because these are typically evaluated by dedicated Filters. This was to keep the Degreaser as modular as possible, such that other metrics could be evaluated before or after Degreaser application. Still, designed proteins with and without Degreaser application can be compared because there is otherwise minimal difference between the design sets. These differences were first evaluated computationally for a select set of designed protein nanoparticle components, then for designed nanoparticles, and finally for designed protein binders. Variants that passed computational screening were then characterized experimentally, which is further discussed in Chapter 3.

Two designed protein nanoparticles, I53_dn5 and O43-38, designed by Baker lab members George Ueda and Una Nattermann, respectively, were chosen for Degreaser

redesign [48]. The King lab flagship I53-50 was also chosen for Degreaser redesign; however, the trimeric component (I53-50A) was shown to secrete robustly and was thus unlikely to need redesign, and the pentameric component (I53-50B) was known to be poorly behaved with respect to expression and purification, even in bacterial systems. This pentamer was ultimately recalcitrant to redesign, but is not precluded from secretion optimization, because its cognate natural scaffold, PDB 2OBX, can be solubly secreted. For I53_dn5, the pentameric component (I53_dn5A) was well-expressed in bacterial systems but very poorly expressed in mammalian cell culture. Interestingly, the trimeric component (I53_dn5B), expressed and secreted robustly when fused to influenza HA, but expressed and secreted poorly without that fusion. For O43-38, neither component could be secreted from mammalian cell culture.

Degreaser redesign of I53_dn5A quickly revealed that an N-terminal portion of the protein, where the nanoparticle hydrophobic interface was designed, could accommodate polar residues. The $dG_{ins,pred}$ of that region was lowered to +0.779 kcal/mol from +4.71 kcal/mol in its natural scaffold, indicating that design resulting in a highly hydrophobic segment need not originate from a marginally hydrophobic segment (Figure 2.3, right). Rather, the inclusion of a small number of key nonpolar residues, especially on the surface of the protein, is sufficient to create stretches that can abolish protein secretion. This could be a feature of such scaffolds chosen for protein nanoparticle design, because if the initial scaffold has some hydrophobic residue “seeds,” the final designed nanoparticle may have better design metrics than other comparable candidates. Nonetheless, the Degreaser identified several candidate

positions in I53_dn5A for redesign, and PDB position 16, a tryptophan in the natural scaffold and designed component, could be changed to a glutamate with little to no penalty in Degreaser runtime Rosetta score, as well as re-evaluated Rosetta ddG of interface formation (Figure 2.7). This tryptophan is critical to the natural protein's function as a lumazine synthase but not necessary for nanoparticle formation. It is also a solvent-exposed residue, and was an ideal candidate for characterization. In parallel, King lab member Dan Ellis worked to improve the expression and purification stability of I53_dn5A, identifying several regions of the protein that could be redesigned by Rosetta or consensus design. The Degreaser-identified substitutions were also compatible (but not identified by) this parallel design, providing support for the modularity and orthogonality of the Degreaser.

O43-38 proved to be a more complex case study, wherein both components of the nanoparticle contained hydrophobic segments to be redesigned. The tetrameric component, O43-38B, yielded several promising variants, but none could be secreted. Diving more deeply into the expression of its cognate natural scaffold and Degreaser variants, we found that the natural scaffold was well-expressed in mammalian culture but that the protein was aberrantly glycosylated, resulting in poor secretion. Indeed, there was a N-glycosylation motif (NxS/T) at position 29, and when N29 was mutated serine, the natural protein could be secreted. This substitution did not significantly affect the design metrics of the nanoparticle, and was used for all subsequent variants. As with I53_dn5A, redesign to improve protein stability and solubility was carried out in parallel to Degreaser redesign. In total, a set of ten O43-38B variants were assessed for

protein secretion, but not all were re-evaluated for nanoparticle interface metrics (Figure 2.7). Several O43-38A variants were also identified by the Degreaser, with no additional redesign. These variants preserved the initially-designed nanoparticle interface, and had comparable ddG of interface formation to that of the original design, suggesting that these variants would retain their assembly competency.

For I3-01, the considerations with respect to its designed interface were more cautious relative to other designed proteins. The contact angle of the components within the nanoparticle is somewhat oblique, leading to a quite small designed interface. We reasoned that this interface may be particularly sensitive to mutation. However, calculated protein interface metrics for I3-01 suggested that none of the variants would perturb interface formation (Figure 2.7). This also provided evidence that the Degreaser's simple usage of Rosetta score for overall protein stability is valid even after more complex design metrics are calculated.

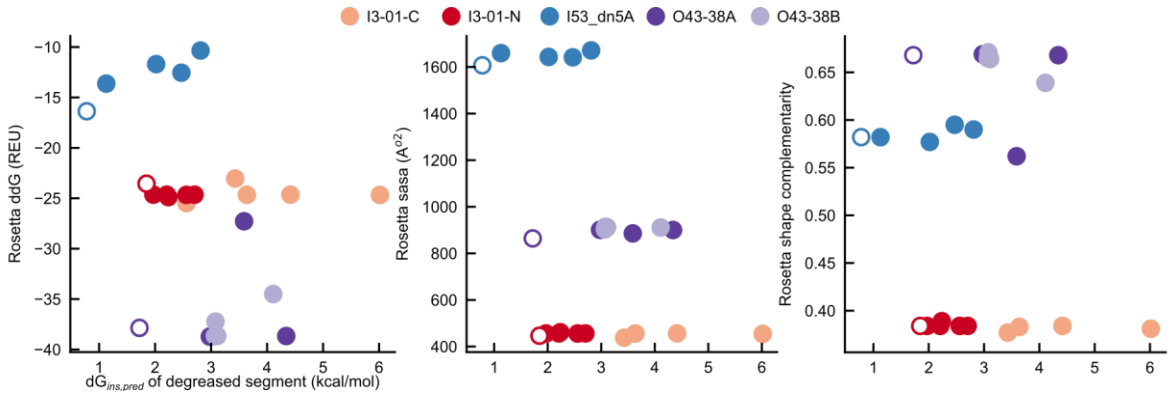


Figure 2.7. Canonical design metrics of I53_dn5A, O43-38, and I3-01 variants. The initial design is the empty marker of each series. Most Degreaser variants do not significantly perturb these metrics, with an exception for O43-38A, in which a buried valine is changed to an aspartate residue. However, this was still within Degreaser score tolerance, and so was chosen for evaluation.

Although redesign of small sets of proteins provided a proof-of-concept for the Degreaser, assessment of more proteins to bolster the cause-and-effect relationship between $dG_{ins,pred}$ and protein secretion required larger sets of designed proteins. The aforementioned one-component protein nanoparticles for mammalian expression and designed helical peptide-binding proteins for yeast display were ideal candidates for Degreaser benchmarking. These design sets not only provided valuable metric data for how the Degreaser influences the design process, but also insights about how in-line application of the Degreaser could be set up.

2.2.4 Degreaser setup and application

This section is intended to serve as a reference for how the Degreaser can be added to existing design protocols or to redesign already-characterized proteins, as well as a guide to interpret the results when considering ordering designs.

For already-designed proteins, or for cases where an input number of proteins is limiting (e.g. constrained to a certain number of natural proteins), running a RosettaScripts protocol with only the Degreaser is sufficient to generate several candidates. With respect to parameters, turning on “dump_singles” and “dump_multis” is recommended because every possible output is to be examined. The other parameters that are important to consider and that are useful to deviate from default are “score_tolerance” and “max_mutations.” By default, “score_tolerance” is set to 15 REU, which was a rule-of-thumb value chosen for nanoparticle interface design. However, for other applications, such as oligomers, monomeric protein binders, or natural proteins, larger values of REU changes is likely well-tolerated. To be rigorous, “score_tolerance” can be incrementally scanned from, for example, +15 REU to +150 REU, but +45 REU has been a reasonable value that still yields well-tolerated outputs. max_mutations can be scanned from 1 to 5 to obtain multi-mutation variants that maximally increase $dG_{ins,pred}$, and is worth considering depending on how many changes within a segment are tolerable for the intended application. Thus, a one-line Mover instantiation for a RosettaScripts script for the Degreaser would look like:

```
<SecretionOptimizationMover name="degrease_protein"
score_tolerance="45" max_mutations="3" dump_singles="true"
dump_multis="true" aas_allowed="DEKRQNSTY" />
```

And within the protocol:

```
<add_mover name="degrease_protein" />
```

The resulting output pdb that accompanies the scorefile will be the "chosen" structure; that is, the one that maximally increases $dG_{ins,pred}$ at the region of lowest initial $dG_{ins,pred}$. This could be a multiple-substitution variant, or a single-substitution variant. However, the dumped single variants and multi variants is still a rich source of information and potential candidates. The Degreaser outputs are best interpreted in descending order of their effects on $dG_{ins,pred}$, upon the region of lowest $dG_{ins,pred}$, in order to see the candidates that still do perturb $dG_{ins,pred}$ but not as much as the Mover-chosen variant. The single variants are identified independently of one another, and so they are not always directly combinable. For example, a variant at position 18 that increases $dG_{ins,pred}$ and a variant at position 19, when combined, may break the secondary structure of the protein or cause another deviation in metric that makes that particular combination undesirable. Because the pdbs of all of the possible variants are output, manual inspection of each variant is possible, primarily to determine the location of the substitution and whether or not it is compatible with biochemical properties of the protein. For example, the partial burial of an asparagine residue near a designed interface may be tolerable, but the total burial of a glutamate within the core of a protein may be too destabilizing for that protein. Even so, some variants within a segment are additively combinable, because if they are not within 8 Angstroms of each other, they are effectively independent. However, for only three possible single variants, there are

six combinations possible, and the number of combinations increases dramatically with the number of variants to consider. Therefore, it is recommended to experimentally characterize as many single variants as possible before combining them. As yet, there is not a definitive threshold of $dG_{ins,pred}$ or residue identity that corresponds to maximal secretion, though negatively charged residues (aspartate, glutamate) appear frequently among well-secreted variants.

For large-scale design, the Degreaser can be included before or after any design step. Typically, interface design protocols or neural network-based protocols make hydrophobic cores a bit too hydrophobic, and so it is recommended to add the Degreaser as a polishing step if the protein is intended for eukaryotic surface display or secretion. For other applications, the original scaffold protein may contain hydrophobic segments (e.g. from natural proteins), and so Degreaser application before any design can broaden the secretion-amenable outputs. In these cases, dumping of the variants during Degreaser runtime is prohibitive due to the sheer number of outputs generated, and so the Mover instantiation would look like:

```
<SecretionOptimizationMover name="degrease_protein"
score_tolerance="45" max_mutations="3" dump_singles="false"
dump_multis="false" aas_allowed="DEKRQNSTY"
task_operations="some_task_operation,some_other_operation" />
```

Where the Degreaser can inherit design TaskOperations such that residues that are prohibited from changing can still be prohibited from changing. If, for some reason, the stability of the initial protein is critical to the user's application, the Degreaser can also be run in a mode where it accepts the change with the smallest change in overall

Rosetta score, while still effecting a $dG_{ins,pred}$ change above the designated threshold. So far, this has not been a necessary use case because many of the initially-designed proteins are highly stable. This approach could be more useful, for example, for redesign of otherwise metastable proteins:

```
     <SecretionOptimizationMover name="degrease_protein"
lowest_dscore="true" largest_ddg="false" score_tolerance="45"
max_mutations="3" dump_singles="false" dump_multis="false"
aas_allowed="DEKRQNSTY" />
```

In essence, the Degreaser can be run without specification of any options by the user and should still produce reasonable results for large-scale design sets to boost secretion or display success rate.

index	sequence	dG_ins	score	ddG_ins	dscore	mutants
51	TTESLLNAIENLRNAIDLL	7.740	-414.01	5.226	37.958	L64N, L61N, L67D
51	TTESLLNAIELLRNAILLL	4.480	-430.84	1.967	21.126	L64N
51	TTESLLNAIENLRLAILLL	4.458	-430.52	1.945	21.444	L61N
51	TTESLLNAIELLRLAIDLL	3.907	-431.12	1.394	20.843	L67D
51	TTESLLNAIELQRLAILLL	3.752	-432.20	1.239	19.762	L62Q
51	TTESLLNAIELLRLAILQL	3.703	-428.64	1.190	23.321	L68Q
51	TTESLLNASELLRLAILLL	3.537	-431.79	1.023	20.171	I59S
51	TTESLTNAIELLRLAILLL	3.351	-439.17	0.837	12.794	L56T
51	TTESLLNAIELLRLAILLN	3.258	-441.75	0.745	10.216	L69N
51	ETESLLNAIELLRLAILLL	3.002	-449.64	0.488	2.327	T51E
51	TTESTLNAIELLRLAILLL	2.994	-441.47	0.480	10.492	L55T
51	TTESLLNAIDLLRLAILLL	2.976	-435.41	0.463	16.551	E60D
51	TTESLLNSIELLRLAILLL	2.972	-442.12	0.458	9.843	A58S
51	TTESLLNAIELLRLSILLL	2.963	-418.44	0.450	33.527	A65S
51	TTEDLLNAIELLRLAILLL	2.935	-445.25	0.422	6.716	S54D
51	TTESLLDAIELLRLAILLL	2.889	-443.73	0.375	8.233	N57D
51	TTESLLNAIELLRLATLLL	2.839	-430.20	0.325	21.769	I66T
51	TTESLLNAIELLELAILLL	2.795	-434.91	0.281	17.056	R63E
51	TTESLLNAIELLRLAILLL	2.513	-451.96	0.000	0.000	

Figure 2.8. Sample Degreaser output, sorted by delta($dG_{ins,pred}$). The top five in this list can be chosen for further screening in a small-scale case, or the top one can be chosen in a large-scale case.

2.2.5 Large-scale Degreaser applications

Redesign of individual components is a limiting case with respect to Degreaser application. For those proteins, because they have already been screened for a

phenotype (assembly competency), there is great pressure to find optimal versions of those proteins for secretion. With large-scale design, we are not as limited in number of available proteins, and can thus use less conservative parameters in the Degreaser. We can also see the bulk effect of the Degreaser on protein design. The two sets of proteins examined in depth are one-component protein nanoparticles (with Alena Khmelinskaia) and designed helical peptide-binding proteins (with Susana Torres).

We sought to expand the repertoire of one-component nanoparticles, but chose to focus our efforts on modularly-generated *de novo* designed proteins. The chosen scaffolds are a class of proteins known as HFuses, designed by Baker lab member Yang Hsia, which consist of two main portions [62]. The oligomeric core of each protein is a *de novo* designed helical bundle, which are designed as soluble proteins. These sometimes contain hydrophobic segments if the core of the protein does not contain hydrogen bonding networks (HBNets, designed by Scott Boyken), or if the protein is very long, which sometimes leads to repeated segments of hydrophobics due to the secondary structure of the helical bundle. The other portion of these proteins consists of designed helical repeat proteins (DHRs), which are also designed as soluble proteins. DHRs are typically quite polar because their hydrophobic cores are relatively small, especially when their constituent repeats are short. Both helical bundles and DHRs have been previously characterized to be soluble and highly stable, and were attractive use as modular building blocks [40,62]. HFuses take advantage of a similar dock-and-design strategy as is used to generate protein nanoparticles, with the design of a hydrophobic interface between the helical bundle unit and the DHR unit in order to

create a protein that is shaped like a propeller. This hydrophobic interface was generally the target of Degreaser redesign. Because there are a large number of available helical bundles and DHRs, the combination of any pair allows the generation of a huge number of potential scaffolds. 1,094 of these design models were used as the starting point for one-component nanoparticle design into three target architectures: icosahedral (I3), octahedral (O3), and tetrahedral (T3).

Alena Khemlinskaia spearheaded the design protocols and downselection of canonically-designed particles, and the addition of the Degreaser was done in parallel, with identical design protocols. These particles were named Khmelinskaia-Wang one-component assemblies, or KWOCAs. This generated two main sets of designed proteins: those that included the Degreaser, and those that did not. Comparison of design metrics between these two sets of proteins revealed that there was little to no perturbation of most conventional design metrics, but ddG of interface formation was slightly increased. This was to be expected because the introduction of polar residues near hydrophobic interfaces likely decreases the strength of that interface. Nonetheless, there were still ample design models for further filtering and manual inspection before experimental characterization. To make the fairest comparisons possible while maximizing the number of secreted proteins from these sets, a subset of the non-Degreaser designed set was filtered to have $dG_{ins,pred}$ above 2.7 kcal/mol, termed the ND set. The Degreaser designed set was filtered as well, termed the DG set. Within the DG design set, 420 of the 1,048 designs (40%) were actually mutated by the Degreaser, while mutations meeting the Degreaser criteria were not identified for 18

designs and these were rejected. Notably, there was an increase in $dG_{ins,pred}$ of the Degreaser-guided design models before any sequences were filtered on $dG_{ins,pred}$ (47% > +2.7 kcal/mol compared to 36% for the OG design models). After filtering, DG designs that were not mutated by the Degreaser had an average $dG_{ins,pred}$ of +3.97 kcal/mol, while those bearing mutations had an average $dG_{ins,pred}$ of +3.38 kcal/mol. Because sequences with originally high $dG_{ins,pred}$ are not mutated, the lower average $dG_{ins,pred}$ of sequences with mutations is due to the low original $dG_{ins,pred}$ of those sequences. In total, 99 designed proteins were selected for characterization among three sets, 57 OG (not Degreaser-designed or filtered), 19 ND, and 23 DG (Figure 2.9).

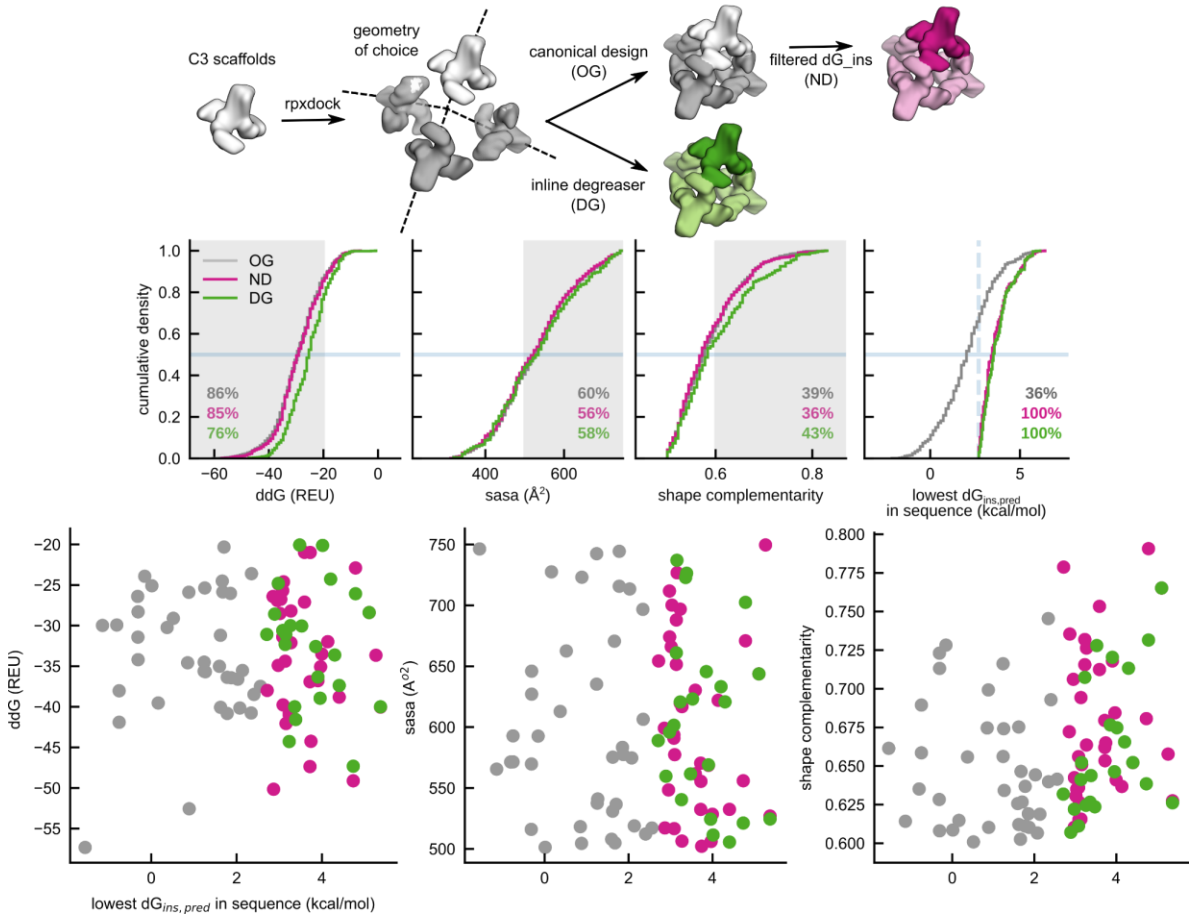


Figure 2.9. Design process and metrics of OG, ND, and DG KWOCAs. (top) Trimeric (C3) scaffolds are docked into desired geometries, then interface design protocols are used without (grey) and with (green) the Degreaser. ND (magenta) designs are a subset of conventional designs filtered on $dG_{ins,pred}$ > 2.7 kcal/mol, and the subset of DG designs chosen are also filtered on $dG_{ins,pred}$. (middle) Design metrics used to choose proteins for experimental characterization do not differ significantly for the different design sets and subsets (threshold for passing a metric in grey). The distribution of $dG_{ins,pred}$ is a subset of designs shown in Figure 2.5. (bottom) Of the designs chosen for experimental characterization, $dG_{ins,pred}$ is not correlated with those design metrics.

Susana Vazquez Torres, while designing helical peptide-binding proteins, noted that many such proteins displayed poorly on the surface of yeast, but that those that could be displayed had high affinities for their targets. Because proteins displayed on the surface of yeast are expressed as C-terminal fusions to Aga2p, a secreted protein that becomes outer membrane anchored by disfulfide bonds to Aga1p, they require efficient translocation, or the entire construct may be misfolded and degraded. Furthermore, because these proteins were generated via *de novo* computational protein design, $dG_{ins,pred}$ was not an explicit consideration. Degreaser application allowing up to three mutations and with a score tolerance of 45 REU was used to try to increase the $dG_{ins,pred}$ of these proteins. However, binding of these proteins to their targets needed to be preserved, which may not be guaranteed by the score evaluation of the Degreaser. Therefore, the Degreased designs were also evaluated for their predicted binding affinities alongside the original designed proteins. Strikingly, even though the $dG_{ins,pred}$ distribution median could be shifted by +1.80 kcal/mol, other design metrics, such as ddG of interface formation, were shifted but still viable with respect to metric thresholds (Figure 2.10). Interestingly, sets of designs targeting different peptide binders had different $dG_{ins,pred}$ distributions. This suggests that the origin of generation of *de novo* proteins can determine those proteins' hydrophobicities. However, this large-scale design set provides further support that the Degreaser's usage of Rosetta score as a proxy for overall protein stability (in this case, protein-ligand complex stability) is sufficient relative to other more complex design metrics.

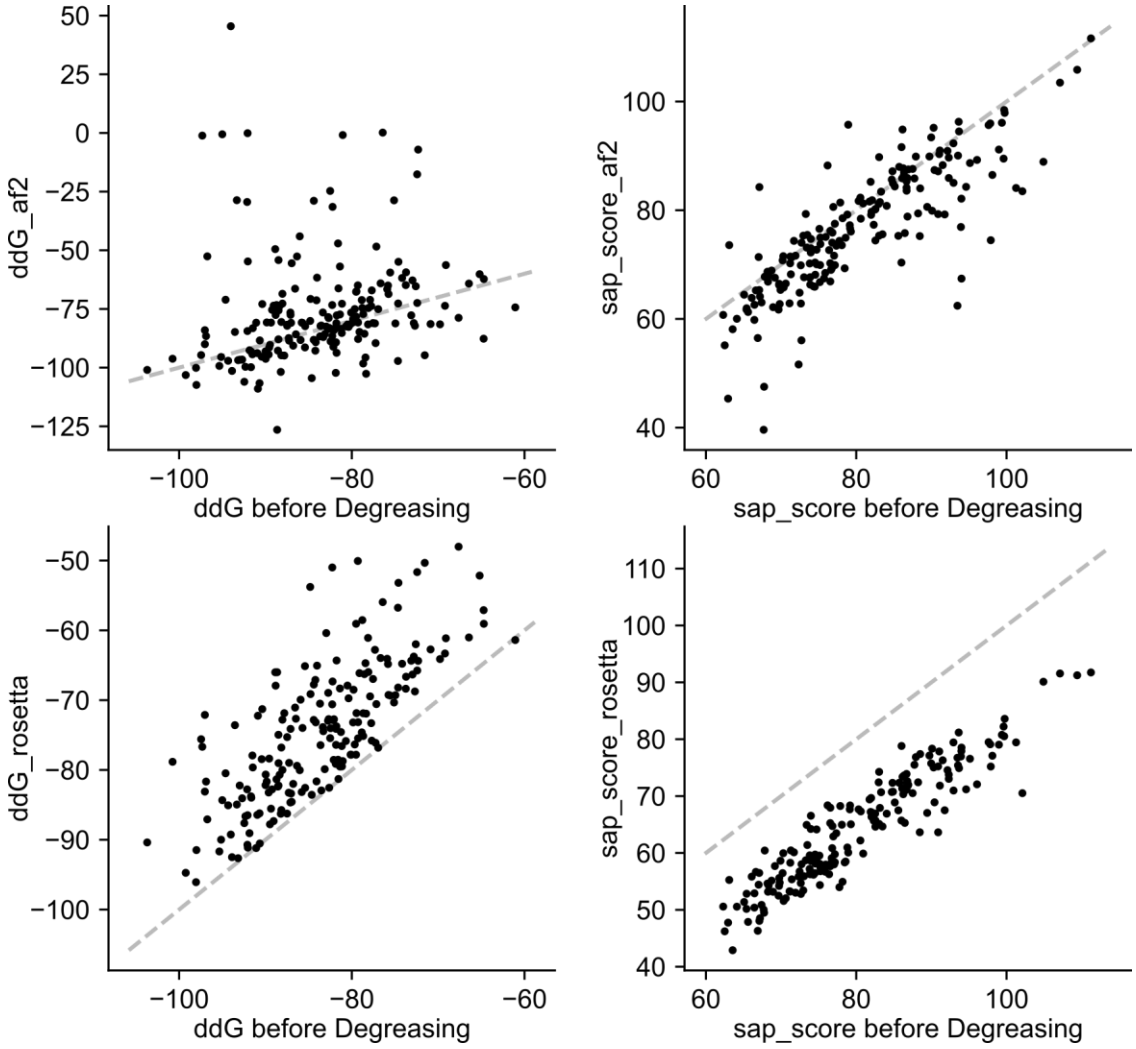


Figure 2.10. Design metrics of Degreased and non-Degreased helical peptide-binding proteins. Two examined metrics are Rosetta ddg with (ddG_af2) and without (ddG_rosetta) AlphaFold2 structure prediction as well as Rosetta-calculated sap_score [63–66]. Figures generated with designs and metrics generated by Susana Vazquez Torres.

2.3 Discussion

Computational evidence for the viability of Degreaser outputs, variants of otherwise well-expressed, stable, soluble proteins that have higher $dG_{ins,pred}$, is sufficient to justify its usage for any application where proteins are to be secreted or displayed on outer membranes of eukaryotic cells. Even though the insights used for establishing the computational protocol are derived from a small number of proteins, the data bolster the hypothesis that poor secretion is linked to $dG_{ins,pred}$. The Degreaser, when applied to large sets of proteins, still reflect the same trend that some computationally designed proteins introduce hydrophobic segments that can prevent secretion.

Some features of these design sets that are worth noting suggest that there are other factors that can also influence protein secretion. First, the distributions of $dG_{ins,pred}$ for many of these proteins are still quite positive. Given that within the Hessa model, a $dG_{ins,pred}$ of zero would be a 50% partitioning into membranes, it is striking that many proteins that are not well-secreted have such large positive $dG_{ins,pred}$. Other models that also predict transmembrane partitioning could increase the accuracy of such predictions, and reflect a different threshold for secreted versus transmembrane proteins. Further studies from the same group showed a correlation of the *in vitro* scale with an *in vivo* one, but the values do not align perfectly [67]. A simple possibility is that the entire data set used to generate this model is shifted toward positive values, which could arise if, for example, for some reason the two glycosylation sites used to calculate partitioning are not identically glycosylated. Nonetheless, large, negative values within this model do correspond with poorly secreting proteins, while large positive values

correspond with strongly secreting proteins. Therefore, large, positive changes to proteins that have low $dG_{ins,pred}$ are still valuable. This does not preclude other calculations for transmembrane propensity that could further inform protocols such as the Degreaser. Second, even though the Degreaser does not incorporate the measurement of other design metrics such as ddG of interface formation, its score calculation is sufficient to preserve the overall stability of most proteins and complexes. This may also mean that more “aggressive” approaches to increasing $dG_{ins,pred}$ are tolerable. Parameters such as score_threshold, dG_ins_threshold, or even the $dG_{ins,pred}$ on which to filter candidate designs for secretion could theoretically be much higher. Thus, for downstream users, sweeping these parameters until there are designs that are intolerable could be an approach to maximize secretion success rate. Finally, these results are after only one round of Degreaser application, where some cases of proteins may not be totally and efficiently redesigned. For example, if a protein has multiple segments of low $dG_{ins,pred}$, the Degreaser will only redesign the segment of lowest initial $dG_{ins,pred}$. Workarounds and future applications that address all of these potential pitfalls are addressed in Chapter 4.

2.4 Materials and methods

Writing and testing of C++, bash, and Rosetta scripts were performed with Vim. Python 2.7 and 3.7 were used for other calculations and plotting of various $dG_{ins,pred}$ data. Chelsea Shu used the Uniprot database, the PDB, and generated data using Python 3.7. Visualization and generation of protein structure models were performed with

PyMOL. Python packages used include NumPy, matplotlib, and FlowCal. Proteins were designed with RosettaScripts, rpxdock, MPNN, and evaluated with AlphaFold.

Chapter 3: Recombinant secretion of Degreaser-identified variants

3.1 Introduction

With the computational results obtained for the $dG_{ins,pred}$ values of several proteins of interest, we sought to both boost the secretion of these proteins as well as reinforce our understanding of how translocation and overall secretion are related. The selection of eukaryotic model was driven by availability and downstream intended application; the suspension culture mammalian expression system HEK 293F is a standard for recombinant secreted proteins [68]. Careful construct design considerations were also made to ensure that the comparisons of secretion from protein to protein that we observed were fair and unbiased, because without such uniformity, the measurement of bulk secretion could be influenced by a host of underlying factors. And, as with the computational portion, the results from working with a small set of proteins were used to guide experiments and construct design for larger sets of proteins. However, true multiplex, high-throughput efforts were not available to examine library-scale sets of designed proteins.

Among the previously-discussed systems available for characterization of eukaryotic protein secretion, we chose the HEK 293F platform for its robustness with respect to overall yields of other recombinant proteins. Kanekiyo *et al.* used this system to secrete influenza haemagglutinin (HA)-bearing protein nanoparticles, which consisted of a genetic fusion of the viral protein to a naturally-occurring nanoparticle, *H. pylori* ferritin [69,70]. We reasoned that designed protein nanoparticles are to be compared to these

naturally-occurring nanoparticles, and given the relatively high yields of protein obtained from this expression system, it was ideal for characterization of other recombinantly secreted proteins. Notably, the yields obtained for nanoparticle proteins were typically orders of magnitude lower than those obtained for non-assembling proteins. Although CHO expression systems were also attractive for their robust yields, they were more typically used for monoclonal antibody expression and were thus not chosen [71]. Another advantage of the HEK platform was the ease of transfection and culture, as well as amenability to genetic manipulation--several methods for transient expression, stable cell line generation, and transcriptional-level tools are readily available. Furthermore, because we are screening a large number of constructs, transient transfection was the most expeditious route to characterization of many constructs in a short amount of time. Each construct was genetically fused to a murine immunoglobulin kappa (IgK) signal peptide sequence, which was previously described to be robust for recombinant secretion. Other signal sequences are possible, but for determining the effects of the Degreaser on secretion, we choose one to use across all constructs. We chose to measure protein content in cell culture supernatant from conditioned medium after transient transfections, and rather than allowing cells to die to obtain maximal protein yields, we chose to harvest the transfections while cells have high viability, though yields are overall lower. Because some constructs expressed very poorly, which would necessitate impractical culture volumes to obtain precise measurements using purification and standard quantification methods (e.g. UV-Vis spectroscopy), we opted for indirect measurement of protein using immunoassays. ELISAs were first attempted as a way to rapidly screen a large number of constructs, but the cell culture conditions

interfered with sensitive detection of our proteins, and so instead Western blots were chosen for quantification. Quantitative Western blots have the advantage of isolation of a protein of interest by electrophoretic separation before quantification and very sensitive limits of detection, but have the drawback of small dynamic range [72,73]. We proceeded with a myc peptide tag (EQKLISEEDL) on each of our protein constructs and quantified protein in bulk cell populations and clarified supernatants, providing a rough fractionation of each culture (Figure 3.1). Although this is a measurement of bulk secretion and not exactly translocation efficiency, which would be what $dG_{ins,pred}$ really predicts, it is sufficient and practical for our purposes. These quantitations are shown in subsequent figures to represent the yields of characterized proteins.

Proteins that were well-secreted were then transfected at larger scales for purification and biochemical characterization. Because most of the proteins characterized were designed to form nanoparticles, the assembly competency of these proteins served as a readout for the preservation of the original protein structure during redesign. This garnered two main advantages: first, the proteins could be easily separated from other material after purification by size-exclusion chromatography, because the assembled nanoparticles are much larger than most soluble proteins and complexes. Secondly, rather than analyze protein stability by means of unfolding or aggregation experiments, a binary “assembly-competent” or “assembly incompetent” could be assigned to each nanoparticle component (or nanoparticle protein in the case of one-component nanoparticles). This characterization had the added benefit of deepening our understanding of protein interface design metrics on nanoparticle formation, though that

is not the focus of this particular body of work. Finally, secreted protein nanoparticle components were practically useful for the multivalent display of other eukaryotically-expressed proteins, particularly those used as immunogens for novel vaccine candidates.

The scales at which the constructs in this work are characterized is still quite small relative to what is possible with multiplex, high-throughput assays. In order to fully gain a systematic understanding of $dG_{ins,pred}$ or other design metrics and their effects on protein secretion, we need ways to manipulate and engineer cells and constructs to report on secretion with high fidelity. Two case studies will be presented in this section: one regarding the redesign of a model protein to incorporate segments of varying $dG_{ins,pred}$, and one using the designed peptide-binding proteins from Susana Vazquez Torres' work. Taken together, they represent a small portion of what could be extended to studies that allow for massively parallel characterization of protein translocation and secretion, but not all of the requisite tools are available. Work toward such tools will be discussed in Chapter 4.

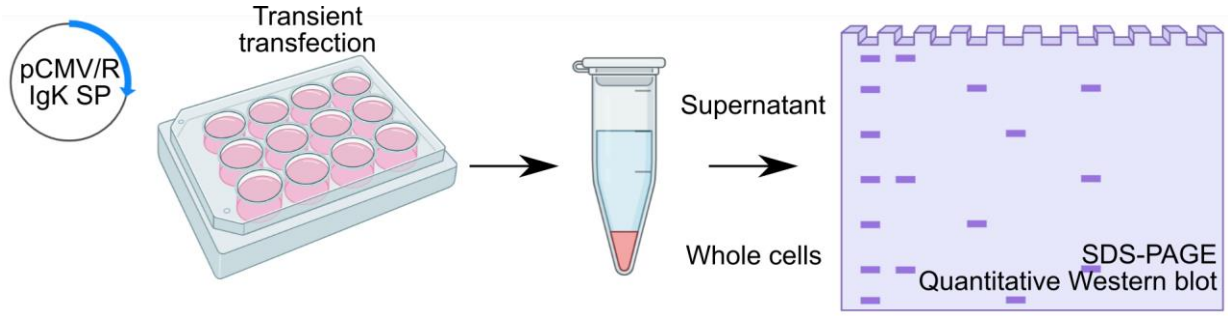


Figure 3.1. Graphical representation of secretion characterization workflow. Transient transfection enables capture of protein content in cell supernatant and in whole cells while viability is high so that degradation or other noise does not skew measured protein levels. Separation of supernatant from whole cells gives secreted yields, but does not capture *in vivo* trafficking effects within the cell. Quantitative Western blot enables specific detection of tagged proteins.

3.2 Results

3.2.1 Characterization of Degreased, previously designed proteins

Three main nanoparticles and their component proteins were analyzed in-depth to understand what features of the Degreaser would be most desirable for large-scale applications. Also examined were a viral fusion glycoprotein and fusions of viral glycoproteins to Degreased nanoparticle components. These represent the first systematic view of the secretion of several designed protein nanoparticles and their glycoprotein fusions, as well as the first characterizations of these secreted nanoparticle components. Other nanoparticle components were also examined but were not further analyzed, but are presented here as well.

I53-50 is one of the most-used designed protein nanoparticles from the King lab. I53-50A, the trimeric component, is designed from PDB 1WA3, a thermophile-derived enzyme. This component has been robustly expressed, soluble, and maintains its assembly competency over long periods of time, and has been evolved to have increased circulation time *in vivo* [55]. I53-50B, the pentameric component, is designed from PDB 2OBX. This component has proven difficult to work with, with respect to soluble expression, *in vitro* stability, and retention of assembly competency over time. Although 2OBX can be secreted from mammalian cell culture, I53-50B and its Degreaser variants cannot even be expressed. Therefore, unfortunately, any nanoparticles intended for use as immunogen materials must use bacterially-purified I53-50B, and only fusion to I53-50A is amenable, though this has proven robust with respect to vaccine candidate outcomes [52–54].

I53_dn5 is another icosahedral nanoparticle, but was specifically designed with the scaffolding of immunogen proteins in mind. Its trimeric component, I53_dn5B, consists of a *de novo* designed cyclic oligomer that has an N-terminus that faces directly radially outward from the cage center so that it can best scaffold trimeric viral proteins. In this case, I53_dn5B fused to influenza HA is very highly expressed, much more so than the nanoparticle component alone. Although it is difficult to draw conclusions from one case, it is tempting to speculate that the HA is responsible for the strong expression of that protein, especially when HA is so highly expressed elsewhere. I53_dn5A, the pentameric component, is a relatively stable and well-expressed protein in bacterial

culture, but is minimally secreted. However, this protein does have long-term stability issues. The Degreaser identified variant W16E dramatically increases the secretion of this protein, with no significant effect on the overall expression, which could be interpreted as no significant effect on stability. Dan Ellis' redesigns of I53_dn5A dramatically increase expression as well as stability, but only the variants including mutations at position 16 away from the original W are secreted. The final variant with all mutations merged, I53_dn5A.2, has mutations at both positions 16 and 23, which are part of the initially-identified hydrophobic segment. Indeed, this variant secretes robustly, but importantly, when these positions are reverted to their original identities, W and A, respectively, $dG_{ins,pred}$ and secretion are dramatically reduced. Thus, for this protein, we see a strong correspondence between secretion level and $dG_{ins,pred}$ (Figure 3.2). This variant retains robust assembly competency, as evidenced by nanoparticle formation when it is mixed with bacterially-purified I53_dn5B. With the advent of secreted I53_dn5A.2, a fully-secreted two-component nanoparticle is envisionable, and is discussed in Chapter 4.

O43-38 is a two-component octahedral nanoparticle designed by Una Nattermann of the Baker lab. Its trimeric component, O43-38A, is also designed based on 1WA3, which is known to secrete robustly. Three Degreaser variants were characterized for this protein, and only one, L31N, greatly increased $dG_{ins,pred}$ and secretion levels. However, this mutation abolished the assembly competency of this protein. The other two variants, V19D and L29E, did not significantly increase secretion but preserved assembly competency (Figure 3.2). This component and its variants provide an

interesting case study, wherein some variants expected to break cage formation did not, but the less aggressive mutation did. It is possible that the cage interface is slightly off-target from the design model, resulting in effects that are unpredictable when the protein is changed. This protein is worth revisiting, considering that the scaffold is well-behaved, many proteins are designed based upon it, and other nanoparticle components based on it can be secreted. The tetrameric component, O43-38B, is less well-behaved in bacterial expression systems, especially after purification. However, some variants were identified that boost secretion. As with I53_dn5A, parallel redesign identified variants that would increase protein solubility, stability, and expression. Of these variants, the best-secreting and best-behaved one, called DEO5, contained Degreaser variants A141E and N146D. Also like with I53_dn5A.2, reversion of these positions to their original identities abolished secretion. These two cases provided strong evidence that increasing $dG_{ins,pred}$ is directly responsible for increasing protein secretion.

I3-01, a one-component icosahedral nanoparticle, could be boosted in its secretion by one mutation in its N-terminal region, H35D. Interestingly, no amount of C-terminal mutation increased overall I3-01 secretion in the absence of any N-terminal mutations, but a combination of four mutations, H35D and three in the C-terminal region, L171Q-S177E-V180N, boosted secretion even further than that of H35D alone. This provided evidence that the approach of increasing $dG_{ins,pred}$ at the region of lowest initial $dG_{ins,pred}$ is critical so that variants that could increase secretion are not missed due to the presence of other hydrophobic segments within a protein. It also highlighted one of the

weaknesses of the Degreaser, in that one application was not sufficient to maximize secretion yield, though the mutation could be combined with no apparent detrimental effects. All of the characterized I3-01 variants showed preservation of assembly, as shown by nsEM and DLS measurements. This represented the first secreted designed protein nanoparticle. Carl Walkey, a member of the Baker and King labs, independently redesigned wild-type I3-01 to bear negative charged residues on the interior of the nanoparticle. This negative-interior I3-01 could be expressed much more strongly than its wild-type counterpart, leading to an incidental increase in secretion, but the negative-interior variant with the four secretion mutations secreted exceptionally well, rivaling that of the natural nanoparticle ferritin. These negatively-charged variants also retained their assembly competency, reflected as monodisperse particles by DLS and nsEM. I3-01 NS (negative / secretion) is the flagship secreted particle for the King group.

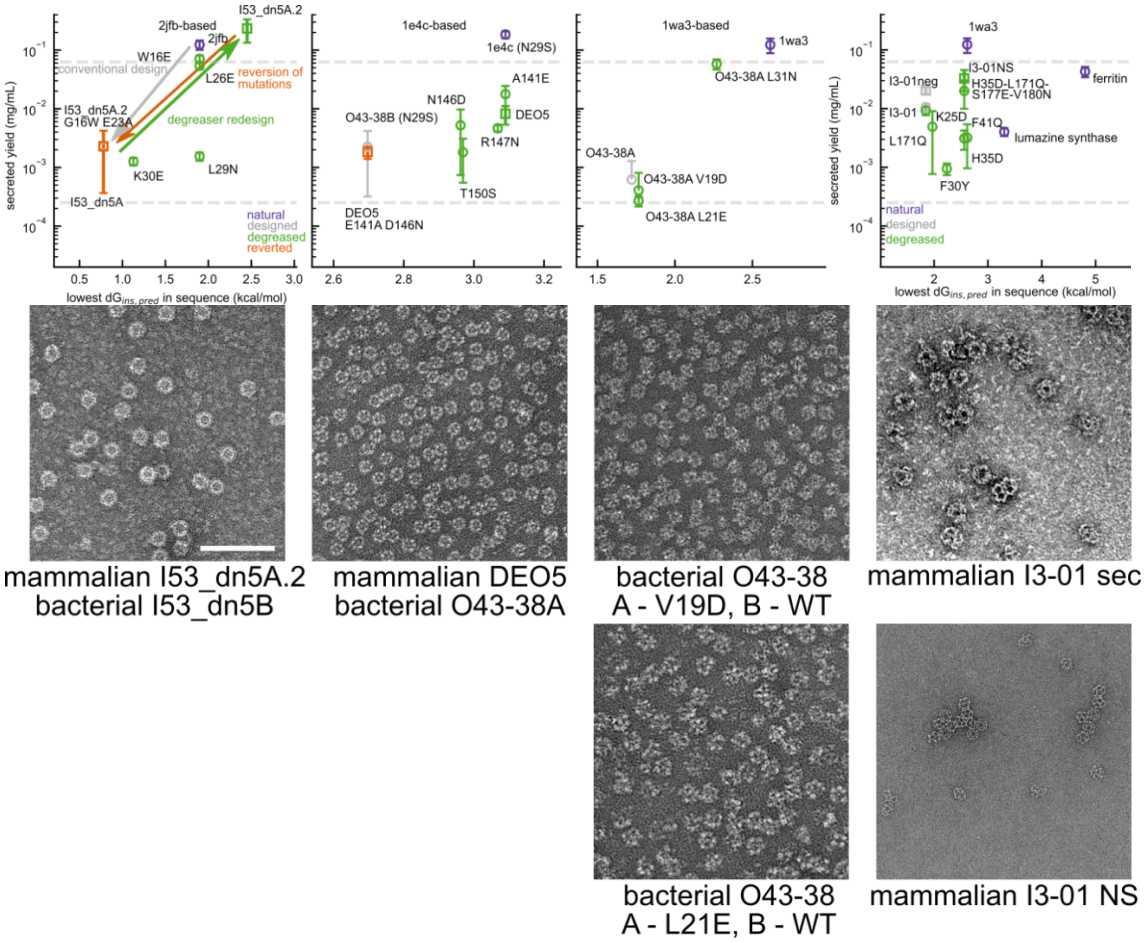


Figure 3.2. Secreted yields of I53_dn5A, O43-38, and I3-01 series (wild type and variants). (left) The progression of design from natural protein (purple points) to original nanoparticle component (grey points and arrow) significantly reduces both $dG_{ins,pred}$ and secreted yield. Redesign (including Degreaser redesign, green) boosts both $dG_{ins,pred}$ and secreted yield. Reversion of only Degreaser mutations (orange) oblates secretion, supporting the idea that it is the $dG_{ins,pred}$ that is critical for high secretion. (right) The best Degreased version of I3-01, I3-01 NS, secretes with yields on par to that of ferritin, which does not contain any cryptic transmembrane domains. (bottom) Representative nsEM micrographs of assembled particles with Degreaser variants.

With secreted protein nanoparticles and components, generation of nanoparticles displaying viral glycoproteins, which must be expressed in eukaryotic culture, was broadened. Before Degreaser redesign, I53-50A was used to display trimeric glycoproteins, such as RSV-F. The generation of I53_dn5 enabled the scaffolding of trimeric HA on I53_dn5B. As only trimeric nanoparticle components were available, assessed glycoproteins were trimeric as well, with some monomers examined. Now, one-component nanoparticles can display twenty copies of a trimeric protein, or sixty copies of a monomeric protein. I3-01 NS has been used to display both Epstein-Barr virus (EBV) gH/gL and SARS-CoV2 receptor binding domain (RBD) [74,75]. O43-38B, in unpublished work, has been used to display the tetrameric influenza neuraminidase. In collaboration with King lab member Aaron Sciore and Vaccine Research Center (VRC) collaborator Julia Lederhofer of the Kanekiyo lab, we generated the first neuraminidase-displaying designed protein nanoparticles. For this case, the robust secretion of O43-38B was critical to express a workable amount of neuraminidase fusion protein. I53_dn5A.2 can be used as fusions to monomeric or potentially to pentameric glycoproteins, and I53_dn5 could be a nanoparticle that displays two different glycoproteins, because both of its constituent components can be secreted. Still, more secreted protein nanoparticles can be designed for specific applications, much as I53_dn5 was designed to display HA, but now these protein components can be more assuredly secreted.

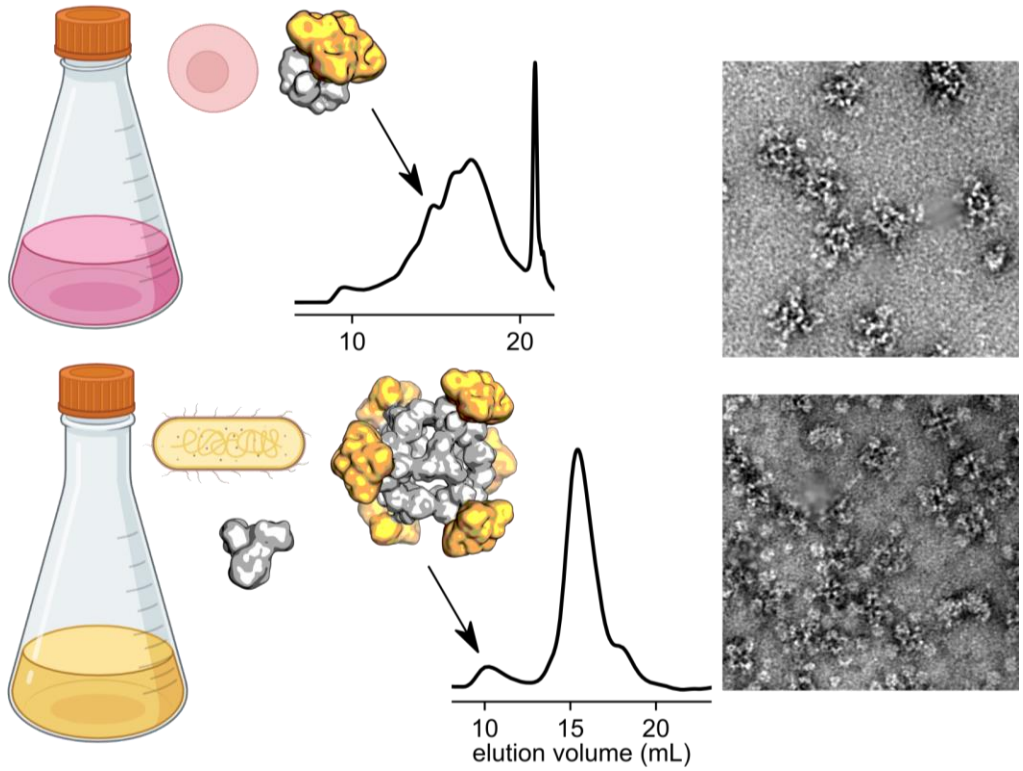


Figure 3.3. Secretion and characterization of a neuraminidase-bearing nanoparticle. (left) Neuraminidase fusions to O43-38B(DEO5) (neuraminidase sequences designed by Julia Lederhofer and Daniel Ellis) are expressed as secreted proteins from HEK 293F, while O43-38A is purified from bacterial culture. (center) O43-38B fusion protein is purified with size exclusion chromatography from cell culture supernatant IMAC eluates, then assembled *in vitro* with O43-38A. Assemblies are also purified with size exclusion chromatography. Although assemblies are not stoichiometrically efficient, negative stain TEM (right) shows on-target assemblies: nanoparticles bearing neuraminidase. Aggregates and degradation products appear to be present in the EM sample, but other characterization shows that the particles are monodisperse after purification (data not shown).

Worth mentioning is also the case of Degreaser redesign of naturally-occurring proteins. Because immunogens for vaccine candidates tend to be viral fusion glycoproteins, and because these fusion proteins may take advantage of a hydrophobic fusion peptide for membrane engagement in its natural function, these sequences tend to contain segments of very low $dG_{ins,pred}$ [76,77]. For vaccine immunogen applications or for expression of these proteins for structural studies, the membrane engagement property of these peptides is usually not necessary, and so are prime targets for the Degreaser. One such example is human metapneumovirus fusion protein (hMPV-F), which has a fusion peptide buried in its pre-fusion conformation [78]. From its structure, we can see that this fusion peptide is actually relatively solvent-exposed even though it is protected by other domains of the protein, and so amenable to mutation by the Degreaser. However, because we do not want to perturb the structure of the pre-fusion protein, we set a relatively low score tolerance as a proxy of stability. This Degreased hMPV-F, which contains four mutations across two identified segments of low $dG_{ins,pred}$ (although, this could be considered one very long segment), which dramatically boosts the secretion of that protein when it is fused to I53-50A. This suggests that further Degreasing of other viral glycoproteins is possible to maximize expression of those proteins, especially when membrane fusion is not a necessary feature.

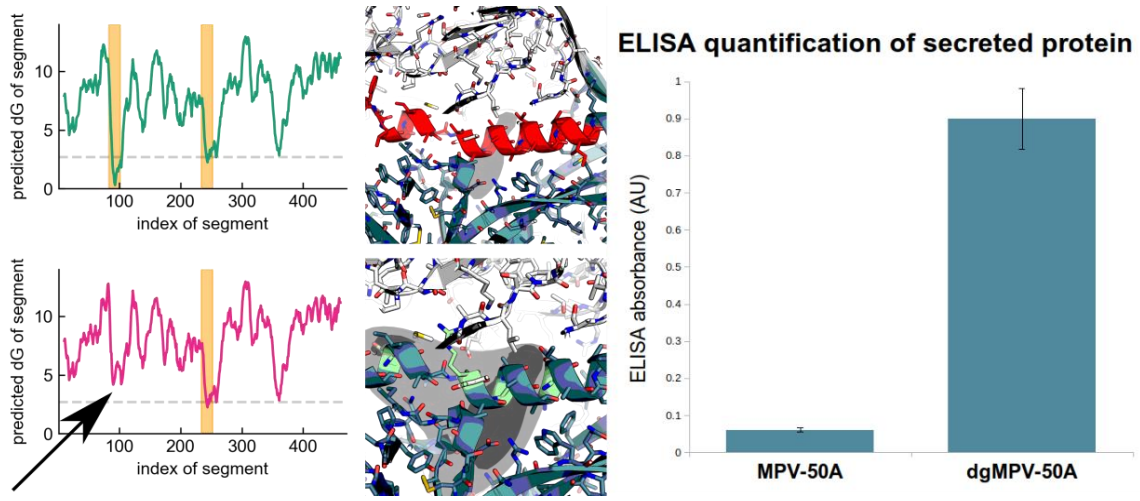


Figure 3.4. Degreasing of hMPV-F. (left top) the original sequence of pre-fusion stabilized hMPV-F contains a very hydrophobic fusion peptide, shown structurally (center top, red). Degreasing of hMPV-F (left bottom) can be performed because there is solvent-exposed space in that core for polar residues (center bottom, green). (right) the Degreased sequence gives much higher ELISA signal than wild-type when cell supernatant from HEK 293F transient transfection of these constructs is performed. Figures generated from data collected by Brooke Fiala.

3.2.2 Characterization of novel secreted protein nanoparticles

Although I3-01 NS is a robust one-component nanoparticle, we sought to expand the features available to secreted nanoparticle proteins. There are not many naturally-occurring secreted protein nanoparticles, and the engineered ferritin and lumazine synthase nanoparticles have some drawbacks, including some shared with I3-01. For example, the termini of these proteins tend to face inward or at an oblique angle relative to the tangent of the radius of each cage, making fusions of viral glycoproteins to these termini difficult [69,79]. Noteworthy for both ferritin and lumazine synthase, however, is

a lack of significantly hydrophobic segments, which is fortunate for protein secretion. Taking lessons learned from the design of I53_dn5, the HFuses used to design KWOCAs are more modular, with the design of nanoparticles with termini at desired locations possible, though not specified in this case. With the 104 characterized KWOCAs, we gained valuable insights about nanoparticle design and secretion.

Because the screening of protein phenotype, in this case, nanoparticle assembly, is much faster and amenable to multiplex methods with bacterial expression and purification than with mammalian transfection and purification, we evaluated each protein in both expression systems. These constructs were codon-optimized for their respective hosts and expressed routinely. Notably, all constructs but one expressed in *E. coli*, reinforcing the stereotype of *de novo* designed proteins' robustness. However, not all of these designs formed their intended architectures: most of the expressed KWOCAs did not form into nanoparticle assemblies. These non-assembling designs are interesting in their own right: many retain their soluble, oligomeric states, express highly, and are quite stable, enabling crystallographic studies that reveal more information about the scaffold protein set. Solved structures of some of these proteins, such as that of KWOCA 39, reflect how even with a small deviation from the design model (RMSD 2.0 Angstroms for one subunit), this imprecision is sufficient to disrupt the next higher layer of design. Thus, even with design 'failure' outcomes, we can build hypotheses about the scaffolds. Here, we posit that the flexibility of the scaffold proteins, both of the helical bundle core and of the DHR 'arms' relative to that core, probably prevents on-target, monodisperse assembly. This notion was reinforced by several of the KWOCAs

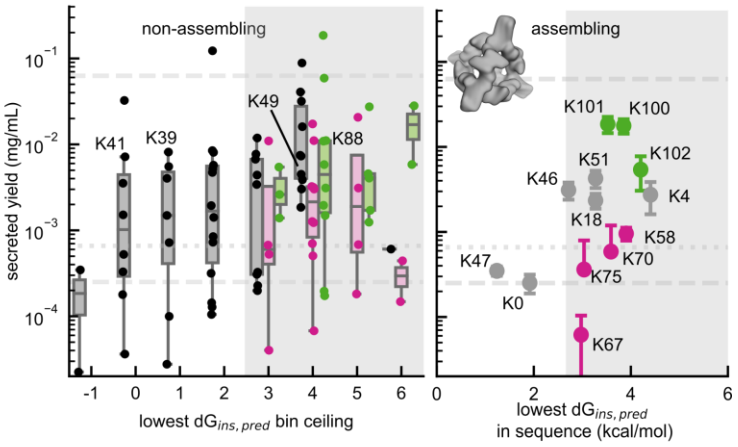
that were shown to assemble by low-resolution characterization (size-exclusion chromatography) but into polydisperse structures. The cases of KWOCA 18 and KWOCA 70, where these nanoparticles form a range of structures, but not including the designed structure, demonstrate ‘partial design failure,’ where the intended designed interface can most likely partially form, but that other factors of the scaffolds prevent it from being exclusive. Then, there are assemblies that form stable, monodisperse structures, but are not the designed one. The case of KWOCA 4, which forms an octahedral nanoparticle but is designed as an icosahedron, is notable because the designed interface is nearly on-target, but that a subtle shifting of the designed interface causes the subunits to contact each other at an angle that favors octahedra. These we considered as ‘design success,’ because a stable nanoparticle was formed, but pedantically could be considered design failures. Finally, we characterized several KWOCAs that form the intended designed structure. Interestingly, most of these are octahedral or tetrahedral assemblies, with only one designed icosahedron forming the intended architecture. This could be due to the scaffold properties or the nanoparticle design protocol, and is worth further exploration.

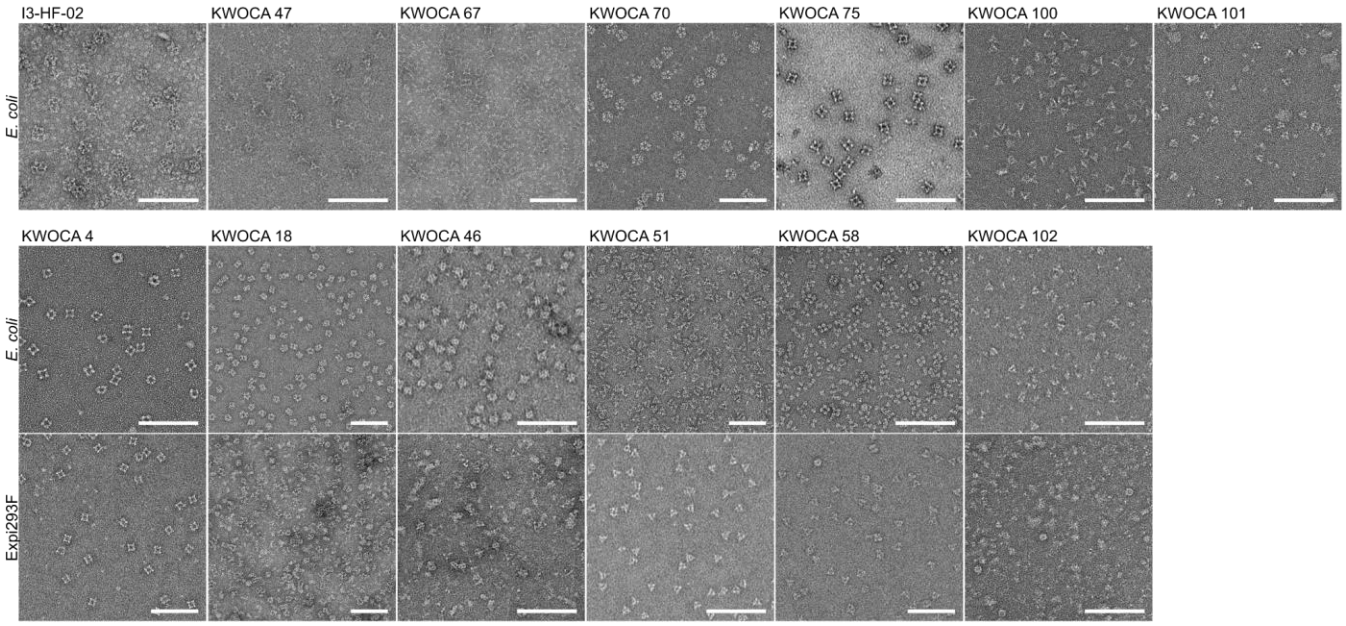
With twenty-two candidates for assemblies, we also examined the secretion properties of the entire KWOCA set. The overall secretion success rate, which we threshold using the secretion yield of wild-type I3-01, a poorly-secreting protein, was relatively high (68%) but much lower than the expression success rate in *E. coli*. The DG KWOCA set’s higher secretion rate (87% compared to 67% and 48% in the OG and ND sets, respectively) confirm the utility of the Degreaser in predicting and improving protein

secretion. This was in stark contrast to those in the ND set, where those proteins were filtered on $dG_{ins,pred}$ as well but not allowed to mutate by the Degreaser. Even more interesting were the KWOCAs that did not express at all in mammalian culture, which suggests that there are other active degradation processes that preclude cellular accumulation of the protein. Nonetheless, several assembling and non-assembling KWOCAs could be secreted, some at quite high levels. However, the highest secreting assemblies had yields lower than those of the highest secreting non-assemblies. Whether this is a feature of assembling proteins or is incidental to these particular designs remains to be determined. Because the assembly success rate of the ND and DG sets were not significantly different to that of the OG set, the filtering on $dG_{ins,pred}$ or application of the Degreaser can be considered safe for design applications.

Comparisons of mammalian-expressed, secreted KWOCAs to bacterially-expressed KWOCAs show that there is no difference between these sets of proteins. Electron micrographs, DLS, and HPLC analysis of each pair of assembling KWOCAs revealed that the purified proteins were identical regardless of expression system. Perhaps the biggest difference is the yield obtained from the two, where several dozens of milligrams of purified protein can be obtained from one liter of bacterial culture, one liter of mammalian culture would yield, for most of these constructs, between one and ten milligrams. However, for cases where eukaryotic expression is mandatory, this yield is much higher than what could be obtained before, which was less than one milligram per liter of mammalian cell culture. These differences in secretion yield were, for the most part, due to initially-high $dG_{ins,pred}$ of proteins within the ND and DG sets, but among the

DG members, many contained Degreaser mutations. Thus, it's possible that with the conservative approach of allowing only one Degreaser mutation, some proteins could not be rescued with respect to $dG_{ins,pred}$, and that allowing more mutations would have made a larger impact. Looking at the distribution of $dG_{ins,pred}$ *post hoc*, we see that even though the Degreaser can effect large changes in $dG_{ins,pred}$ per protein with only one mutation, many have $dG_{ins,pred}$ much lower than 2.7 kcal/mol, even after this increase.





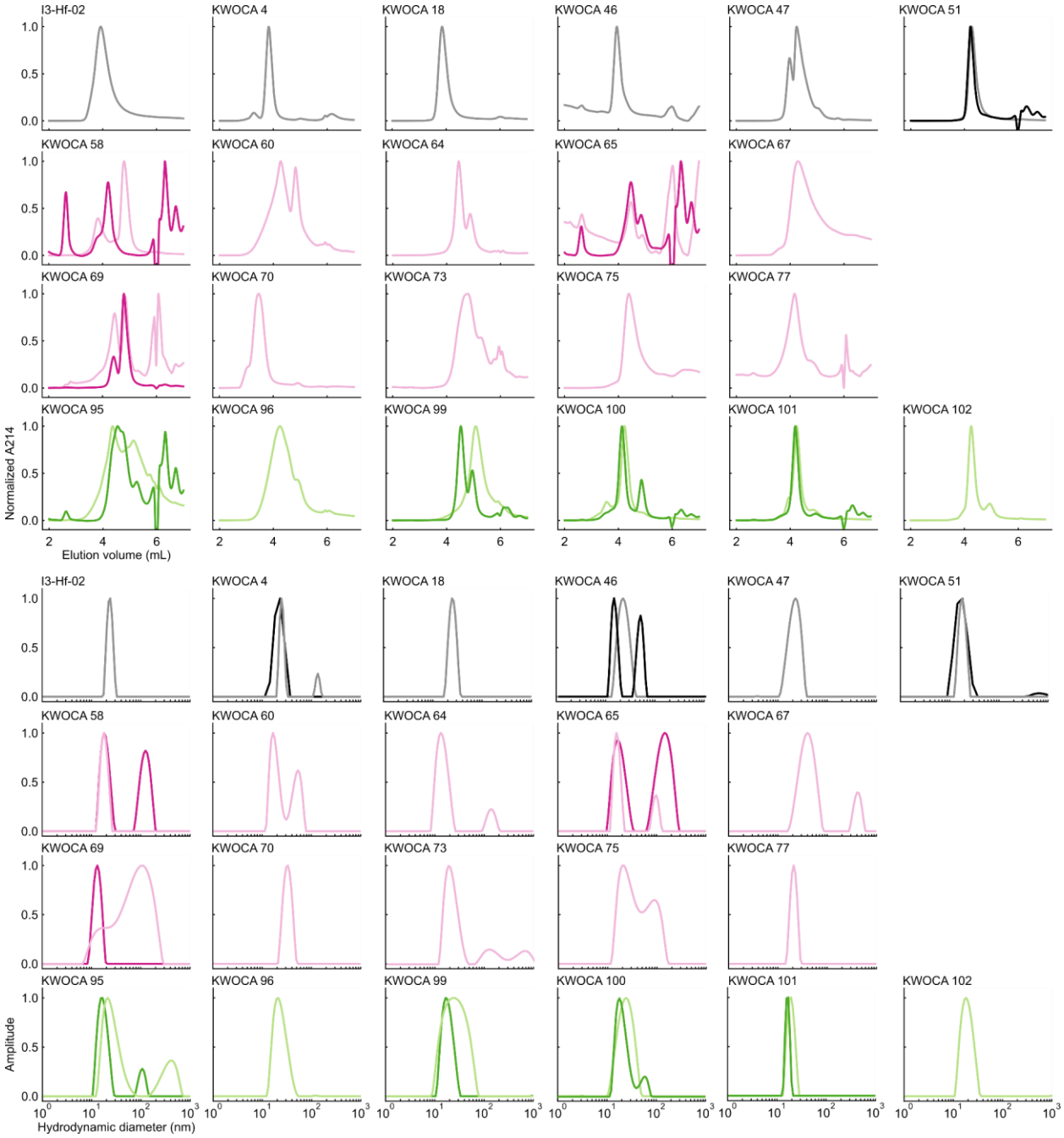

I3-Hf-02
KWOCA 4
KWOCA 18
KWOCA 46
KWOCA 47
KWOCA 51
KWOCA 58
KWOCA 60
KWOCA 64
KWOCA 65
KWOCA 67
KWOCA 69
KWOCA 70
KWOCA 73
KWOCA 75
KWOCA 77
KWOCA 95
KWOCA 96
KWOCA 99
KWOCA 100
KWOCA 101
KWOCA 102
Normalized A214
Elution volume (mL)
Amplitude
Hydrodynamic diameter (nm)

Figure 3.5. Secreted yield and low-resolution characterization of bacterial and mammalian sourced KWOCAs. (top left) Non-assembling KWOCAs are sorted into bins of $dG_{ins,pred}$ windows of 1 kcal/mol each. With these bins, we see that a higher proportion of designs with higher $dG_{ins,pred}$ secrete strongly, with the DG designs (green) secreting the highest among designs. (top right) Assembling KWOCAs even further highlight the $dG_{ins,pred}$ and secretion relationship, with no KWOCAs with $dG_{ins,pred}$ below +2.7 kcal/mol secreting, while those with the highest $dG_{ins,pred}$ secreted the most strongly. (center, bottom) nsEM, HPLC, and DLS confirm KWOCA assembly for both bacterially-sourced (lighter lines) and mammalian-sourced (darker lines) KWOCAs. Figure panels generated by Alena Khmelinskaia.

High-resolution structural characterization by Andrew Borst and Aleksandar Antanasijevic of KWOCAs 4 and 51 by cryoEM revealed potential accuracies and inaccuracies of our design protocol. KWOCA 51, which formed the intended architecture, was well-resolved at 5.1 Angstroms and had a 1.3 Angstrom C-alpha RMSD, demonstrating the formation of on-target nanoparticles. KWOCA 4, on the other hand, was also well-resolved at 6.6 Angstroms, but showed significant deviations from the intended architecture. Interestingly, the C-alpha RMSD of the helical bundle cores of the scaffold proteins were very close to the design model (1.3 Angstrom C-alpha RMSD), but the subunits are rotated with respect to one another, resulting in this off-target assembly. This small physical difference resulting in a large consequence of off-target architecture informs future design, suggesting that even small weaknesses in interaction of the designed interface can lead to totally different results. Even so, these

two nanoparticles are of utility for both understanding of secreted nanoparticle design, as well as any future applications for such proteins, which is discussed in Chapter 4.

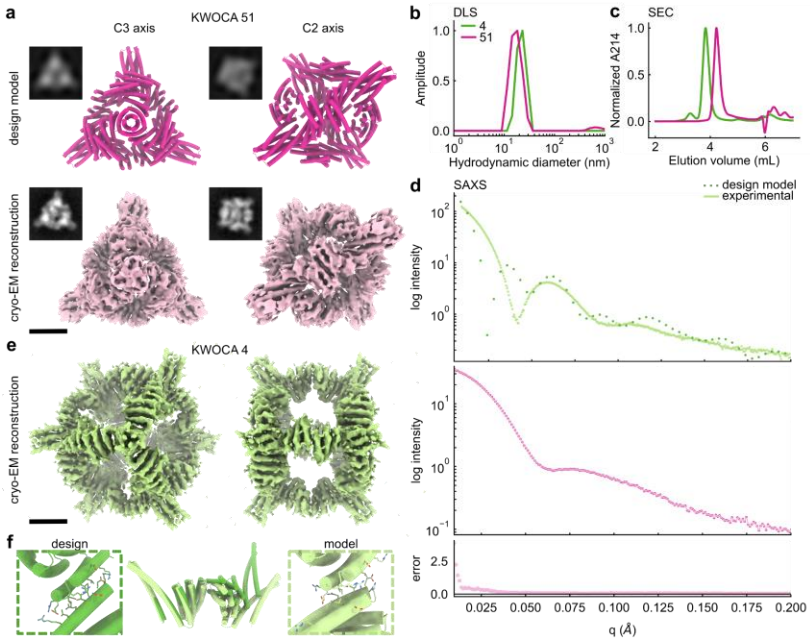


Figure 3.6. High-resolution structural characterization of KWOCAs 4 and 51. Figures generated by Alena Khmelinskaia with data collected by Andrew Borst and Aleksandar Antanasijevic. Reproduced from manuscript in preparation. (a) Design model and cryo-EM density map of KWOCA 51. DLS (b), SEC traces (c), SAXS profiles (d) of KWOCA 51 (pink) and KWOCA 4 (green). (e) Cryo-EM density map of KWOCA 4. (f) Overlay of KWOCA 4 scaffold monomers across the nanoparticle interface highlighting the interface contact angle difference between the design model and the best fitting cryo-EM model, obtained by fitting and relaxing eight copies of the trimeric scaffold into the density map (design model and full best fitting and relaxed model can be found in Supplementary Figure S11b). Theoretical SAXS profiles predicted based on the design models (dotted darker lines) are overlaid in (d) with SAXS profiles predicted from the cryo-EM models. Comparisons to the experimentally obtained SAXS profiles can be found in Supplementary Figure S11c. Scale bar, 5 nm (a,e).

3.2.3 Toward systematic characterization of secretion defects due to transmembrane segments

In order to refine our model of how $dG_{ins,pred}$ impacts overall protein secretion, we'd like to able to characterize systematic sets of proteins that have otherwise similar qualities, but that only differ by $dG_{ins,pred}$. Though the KWOCA set was informative, with only 22 potentially-assembling proteins, and 77 non-assembling ones, we get only a glimpse into how $dG_{ins,pred}$ influences secretion. Furthermore, the overall secretion of a protein is intrinsically tied to the overall expression of that protein, which could be influenced by a large number of factors. For example, unstable proteins may not be expressed well, even if they would otherwise be predicted to secrete efficiently. These would appear as proteins that do not secrete, though they are not informative with respect to $dG_{ins,pred}$. Also, with KWOCAs, we examined both assembling and non-assembling proteins simultaneously, which is a significant phenotype with respect to cellular behavior--large protein complexes could form aggregates or partial assemblies on their way to forming their final architecture, which could trigger unfolded protein responses within the cell's secretory pathway. Thus, we need to be able to examine (1) otherwise uniform sets of proteins, such as designs based on one initial protein, and (2) larger sets of proteins, in the thousands or tens of thousands, so that other elements influencing secretion can be discarded as noise when present.

To this end, we devised a strategy to systematically introduce segments of varying $dG_{ins,pred}$ to natural protein, PDB 1WA3. Because this protein secretes robustly as a wild-type protein, and it has been designed to some nanoparticle components that

secrete and others that don't, we posited that this protein exists at a sensitive threshold for $dG_{ins,pred}$ with respect to overall secretion. Using leucine as a model hydrophobic residue, and because it has the largest negative contribution to $dG_{ins,pred}$ within sequence segments, we generated 42 variants of 1WA3 that had $dG_{ins,pred}$ ranging from +3.37 kcal/mol to +1.09 kcal/mol. This provided us with seven $dG_{ins,pred}$ values at six different positions within the protein (Figure 3.7). A lower threshold of between +1.47 and +1.85 kcal/mol was identified to be critical for robust secretion, with tested sequences below +1.47 kcal/mol secreting with the worst yields among constructs. Furthermore, the N- and C-termini of the analyzed proteins were most sensitive to this introduction of leucines; we posit that these domains are important for overall translocation and secretion, similarly to how transmembrane protein domains signal translocation topology [12]. One caveat of this experiment is the potential confounding factors of 1WA3's trimeric nature; the trimerization domain of the protein was not altered in this experiment. Also, this protein is naturally highly stable and fairly tolerant of mutation, which may result in unclear differences in overall secretion (if the protein can always be expressed). Another caveat is that this experiment was not highly quantitative, but more qualitative and taking advantage of normalizations in Western blot densitometry, and so worth revisiting. Again, this experiment is limited by throughput with respect to constructs examined: dozens of constructs for one protein, but nowhere near the thousands that would be ideal for refining such a model.

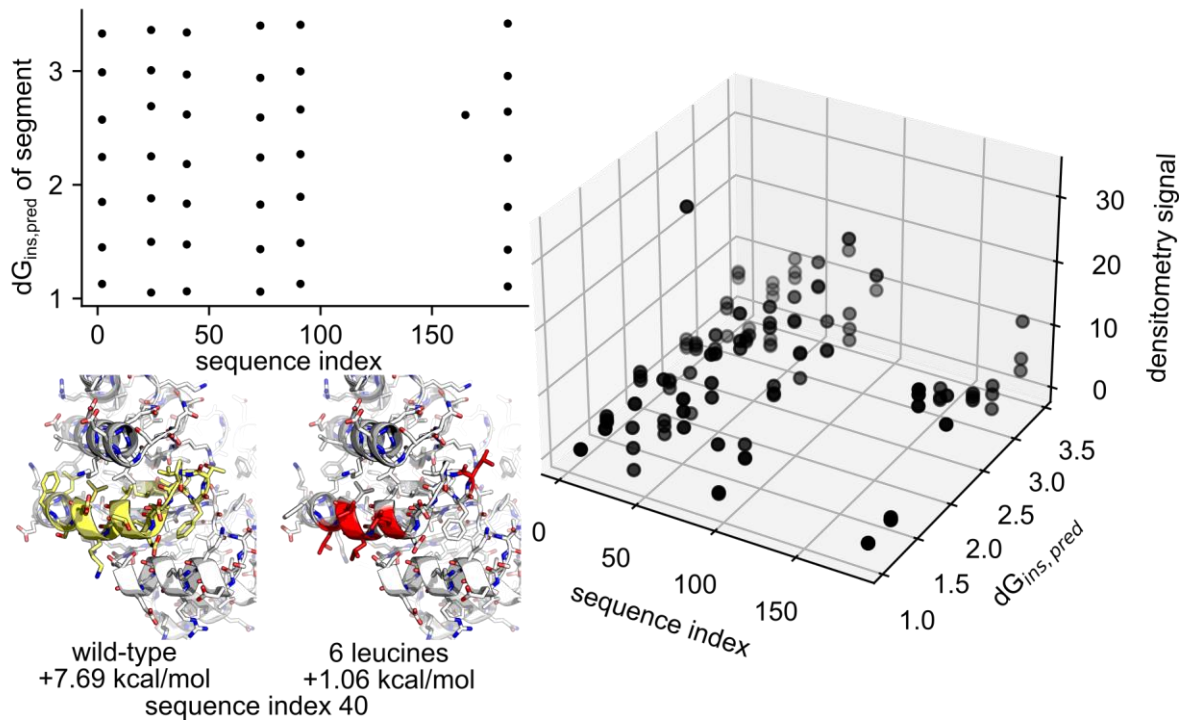


Figure 3.7. Schematic and results of introducing hydrophobic segments within 1WA3. (left top) The sequence indices and the $dG_{ins,pred}$ of the resultant mutant by adding different numbers of leucines at that index. The wild-type 1WA3 sequence has its lowest $dG_{ins,pred}$ at sequence index 165. (left bottom) representative example of a wild-type segment with no leucines added (yellow), and the lowest $dG_{ins,pred}$ segment analyzed for that segment (red). (right) Sequence index and $dG_{ins,pred}$ plotted against Western blot densitometry measurements, which are averages of three biological replicates.

The yeast surface display platform is attractive, but non-quantitative as well, for larger sets of secretion analysis. Though the platform differs from soluble secretion with some fairly large differences, it allows for the rapid testing of many constructs in parallel, and has been used for library-scale design and characterization [80]. Although none of these

previous studies have explicitly examined $dG_{ins,pred}$ and its relationship with secretion, especially because often non-expressing constructs are not further pursued, they are worth examining *post hoc* because they represent rich data sources for this phenomenon. With the help of Susana Vazquez Torres, we were able to examine 480 proteins for secretion, as well as understanding their $dG_{ins,pred}$ values and the effects of the Degreaser. For these proteins, those that were not Degreased had a $dG_{ins,pred}$ distribution comparable to those of other sets of designed proteins, with many having highly hydrophobic segments. These proteins displayed poorly on the surface of yeast, as represented by a semi-quantitative approach. The expression ratio, or proportion of yeast cells with a signal above a threshold, divided by the total number of yeast cells examined, gives a rough measure of yield. This ratio is also used heuristically to determine whether or not a protein is well-displayed on the surface of yeast, such as for binder design screening [66]. Other constructs that had higher $dG_{ins,pred}$, whether incidentally or after Degreaser application, could be displayed strongly. Some still failed to display, though again, this could be due to any number of factors of that protein, but overall, because a larger number of constructs can be analyzed, these can be discarded with experimental noise. This yeast display experiment strongly highlights the importance of eliminating segments of low $dG_{ins,pred}$ for proteins that are intended for display or eukaryotic secretion: very few proteins with $dG_{ins,pred}$ below 0 kcal/mol can be expressed. This moderate set of proteins analyzed can also reveal other sequence determinants of protein secretion, such as position of hydrophobic segment within the sequence, but these determinants are so far inconclusive. With the power to assess large numbers of sequences at once; however, richer data sets can be obtained to dig

more deeply into the phenomena that affect efficient display, and therefore develop protocols intended to guide design of efficiently displayed or secreted proteins.

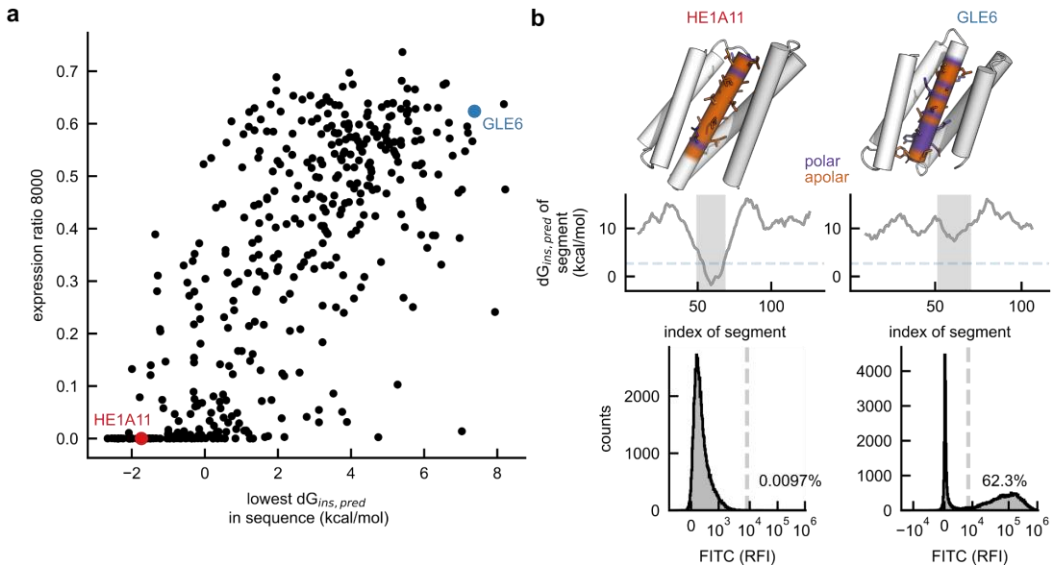


Figure 3.8. Yeast surface display expression ratio of designed binding proteins. (a) Scatterplot of 480 designed proteins and their surface expression. Expression ratio is the ratio of events per construct analyzed that have signal above 8,000 RFI, divided by total events, serving as a proxy for expression. (b) Representative examples of poorly displayed (left, HE1A11) and well-expressed (right, GLE6) proteins, their $dG_{ins,pred}$, and individual fluorescence distribution plots.

3.3 Discussion

A fair and unbiased way to quantify a large number of secreted proteins, designed or natural, is still difficult to formulate. Although methods exist that could be amenable to multiplex, high-throughput assays, the separation of the phenotype (secretion) from the genotype of a given construct makes the problem fundamentally challenging. With low-throughput quantification methods, such as Western blotting, we have demonstrated

that the Degreaser can strongly influence the secretion of previously-designed and characterized proteins, as well as exert this effect when applied during or post-design, before initial characterization. The Degreaser is safe to use for future applications where high secretion or surface expression is desirable, and acts on a wide variety of proteins. Viral glycoproteins, *de novo* designed monomeric proteins, oligomers, and nanoparticle assemblies can all be Degreased to boost their potential secretion. As with the computational aspect of formulating the Degreaser and its applications, however, there is still ample space to be explored with respect to engineering proteins for secretion. There must be other features within a protein sequence or structure that such model-guided protocols can capture that can further improve both the desired phenotype and augment the stability properties of designed proteins. This represents one of the advantages of the modularity of protein design, as well as of the Degreaser module itself: these steps can be taken a la carte or mixed and matched in any number of configurations, and so far there are not many significant conflicts among design methods. Ultimately, massive (tens to hundreds of thousands) scale protein design characterization must be enabled so that characterization of such broad design space is as routine as the design itself.

3.4 Materials and methods

Bacterial expression was carried out routinely. Briefly, BL21 or similar expression strains are cultured in TB or autoinduction medium with canonical expression vectors, namely pET29b(+), with 6x histidine tags at N- or C-termini, as determined by manual inspection of which terminus is more solvent exposed. Proteins are purified with Ni

affinity chromatography, and purified eluent is further resolved with size exclusion chromatography. For nanoparticle components, Superdex 200 is the most common resin used, while for assemblies Superose 6 is used. These purifications are carried out on AKTA FPLC instruments in Tris buffered saline (TBS), pH 8.0. For some applications, the NaCl concentration of the buffer is increased to 500 mM. For KWOCAs, 5% glycerol is added to the buffer for further preservation of stability and resistance to aggregation during freeze/thaw cycles.

Mammalian secretion is carried out via transient transfection. 1 microgram of purified expression vector plasmid, typically pCMV/R [81], with an IgK signal peptide fused to the construct of interest, as well as a myc tag and 6x histidine tag, is used per milliliter of cell culture, where cells are cultured at 3 million cells per milliliter. Clarified cell culture supernatant is mixed with Ni affinity resin, and purified according to manufacturer's instructions. Purified material is further resolved as with bacterial Ni eluents, with size exclusion chromatography.

Biochemical characterization of purified proteins is carried out with SDS-PAGE, DLS using an UNChained Labs UNcle or Wyatt DLS, and negative stain EM. nsEM is carried out on a Morgagni or Talos transmission electron microscope. Samples are stained with uranyl formate or nano-tungsten stain on glow-dsicharged carbon-coated copper mesh grids. LC/MS is carried out on an Agilent 6230B TOF on an AdvanceBio RP-Desalting column, and subsequently deconvoluted by way of Bioconfirm using a total entropy algorithm.

For yeast display, *Saccharomyces cerevisiae* EBY100 strain cultures were grown in C-Trp-Ura medium supplemented with 2% (w/v) glucose (CTUG). For induction of expression, yeast cells initially grown in CTUG were transferred to SGCAA medium supplemented with 0.2% (w/v) glucose and induced at 30 °C for 16–24 h. Cells were washed with PBSF (PBS with 1% (w/v) BSA) and labelled for 30 minutes at room temperature with FITC-conjugated anti-Myc at 10 µg/ml (Immunology Consultants Lab, CYMC-45F). After incubation time, cells were washed once more and resuspended in PBSF for cell sorting (Attune NxT Flow Cytometer, Thermo Fisher Scientific).

For secretion yield quantification, cells were centrifuged to separate medium from cells, and pelleted cells were resuspended in the same volume of removed medium in phosphate-buffered saline (PBS). All samples were then treated for 10 min at 37°C with 0.05% Triton-X 100 (Sigma) containing a 1:400 dilution of Benzonase endonuclease (EMD Millipore) to permeate membranes and prevent nucleic acid aggregation, facilitating quantitative gel loading. Internal myc tag protein standard was also added to a final concentration of 0.06 mg/mL. Treated samples were then diluted into 4× SDS loading buffer (200 mM Tris pH 6.8, 40% glycerol, 8% SDS, bromophenol blue, 4 mM DTT) and incubated at 95°C for 5 min. 14.3 µL of boiled samples were loaded onto Criterion 4-20% precast polyacrylamide gels (BIO-RAD). Precision Plus WesternC standards were included in each gel (BIO-RAD). Gels were run using BIO-RAD Criterion gel boxes and power supplies, then transferred using the Trans-Blot Turbo system onto 0.2 µm nitrocellulose membranes according to manufacturer instructions

(BIO-RAD). Transferred blots were blocked in 3% milk in wash buffer (10 mM Tris pH 8.0, 150 mM NaCl, 0.1% Tween-20) for 30 min, then incubated with a 1:20,000 dilution of mouse anti-myc tag antibody (9B11, Cell Signaling Technology) with agitation, either 75 min at room temperature or 16 h at 4°C. Blots were then washed three times with wash buffer and incubated 75 min at room temperature with a 1:10,000 dilution of goat anti-mouse HRP conjugated antibody (Cell Signaling Technology). After three washes with wash buffer, blots were developed with Clarity ECL substrates according to manufacturer directions on a Gel Doc XR+ Imager with Image Lab software (BIO-RAD).

Western blot images were analyzed using ImageJ/FIJI software for quantification. Calibration curves of known myc-tagged protein were used to establish a linear range (data not shown), and four points for each blot were included to allow absolute concentration determination. Three biological replicates (i.e., independent transfections) were included for each construct. For some constructs, the measured cellular level of protein was higher than the linear range of the calibration curve. However, for nearly all measurements, the secretion yield measurement was within linear range.

Chapter 4: Ongoing developments and future directions

4.1 Broader applications of the Degreaser

4.1.1 Application strategies for in silico maximization of transmembrane insertion potential

The aforementioned applications of the Degreaser are the most conservative approaches possible, allowing one or three mutations, and setting a relatively low score tolerance. Furthermore, only one application per set of protein was used. This led to fairly striking changes in the $dG_{ins,pred}$ distributions of the designs but retained their other phenotypes. When characterized, some of these proteins were still difficult to secrete, or expressed at low yields. Finally, for some of the individual cases examined, orthogonal and parallel redesign was necessary to reach the final, optimized secretable protein. Taken together, these data suggest that the Degreaser alone, or one application alone, may not be sufficient to fully optimize a given structure. Indeed, there are many such cases where, for example, in a repeat protein, if one repeat is found to be significantly hydrophobic, and that sequence is repeated, then the Degreaser will not change all repeats within the protein, instead choosing only one to redesign. Of course, the repeats can then be manually redesigned, but the intention of the protocol is to allow streamlined use. Therefore, we would like to be able to envision approaches where the Degreaser is seamlessly integrated into even complex design protocols.

The first approach that could further boost secretion of many of these proteins is multiple applications of the Degreaser. One such method would be as follows, following

the course of the KWOCA design already outlined in this work. For the KWOCAs, the Degreaser is added as a final step, after the interfaces between subunits are designed. However, for a future application, the Degreaser could be introduced in two or more steps. For example, the Degreaser can be the first Mover applied to the input scaffolds, already in their target geometry, before any interface design takes place. This enables the Degreaser to identify scaffolds that have hydrophobic segments, even before any other residues are changed to be more hydrophobic. This, in a sense, sanitizes the input scaffold so that more outputs are possible. Secondly, the Degreaser can be applied once after the nanoparticle interfaces are designed, mirroring the previous application. Then, the Degreaser can be applied again, perhaps with a scorefunction that more severely penalizes buried polar residues, so that any remaining hydrophobic segments are designed away, but not so severely that the designed interface is ruined. This type of iterative design is already built in to some features of Rosetta, such as FastDesign, which uses multiple rounds of MonteCarlo sampling with ramping repulsive energy terms so that the final design is an annealed product, rather than whatever happened to fit the design criteria initially [82,83]. This approach is likely to prevent any low $dG_{ins,pred}$ segments but does not address repeated segments within proteins, but the same designations used to generate repeat proteins can be used in conjunction with the Degreaser to allow each member of the repeat to mutate at the same time.

Another more involved approach would be the simulated annealing of $dG_{ins,pred}$ to some desired tolerance, such as setting a threshold of +3 kcal/mol, or in a way that if the $dG_{ins,pred}$ no longer increases by a certain amount per application (dG tolerance

convergence). This would allow *in silico* “evolution” of a particular design trajectory, because the sampled sequence space for the Degreaser could be much greater than one application, and *in silico* evolution for other purposes has already been established [84,85]. This approach is inherently more risky, because with each round of design, there is not only the chance of reverting to previous mutations, but the danger of deviating so far away from the original design that some phenotype or overall protein stability is lost. In theory, the Rosetta-enforced design metrics, or even just the imposition of the score function on a particular protein, should be enough to keep the designs from becoming unreasonable. This type of intensive redesign is most appropriate for when a small number of valuable initial protein sequences are already characterized, and the maintenance of those properties is critical for future applications. For cases where many proteins can be designed to fulfill a function, iteration may not be necessary. However, this computational iterative design would still have much faster optimization cycles than would expressing proteins, making variants of those proteins, and then expressing those variants.

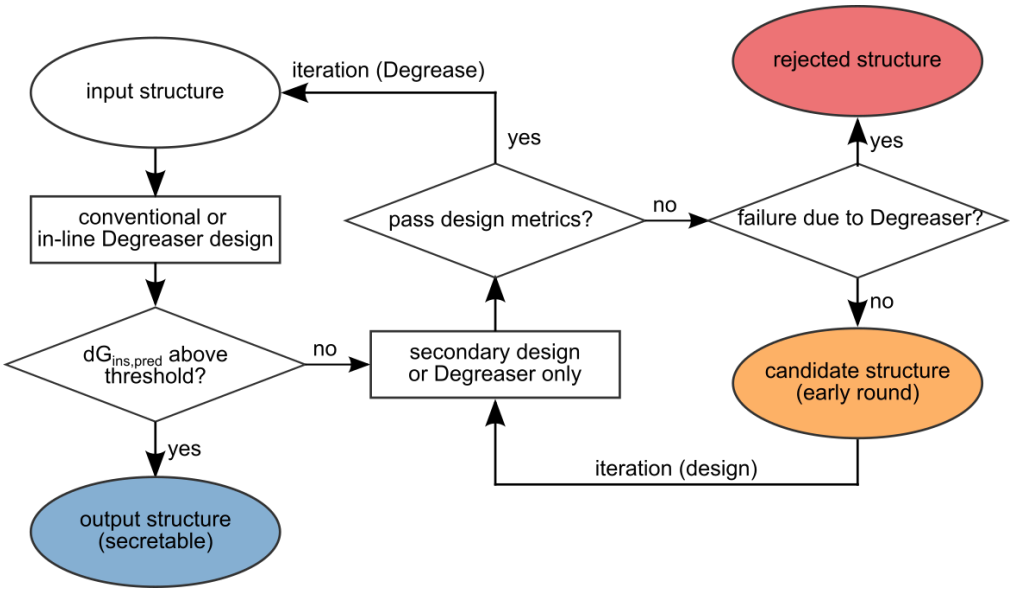


Figure 4.1. Graphical representation of Degreaser simulated annealing. Two key regions of computational examination will be the evaluation of design metrics with Degreaser application, and whether the Degreaser significantly perturbs those design metrics. If so, more care must be taken in order to preserve the original intention of the design protocol. This iteration can be performed until the $dG_{ins,pred}$ of any segment within the protein is above a certain threshold or until some other design criteria are met.

4.1.2 Inversion of Degreaser calculations for design of transmembrane proteins

The Hessa *et al.* model on which the Degreaser is based simply calculates $dG_{ins,pred}$, and thus the Degreaser can in theory be run in reverse: lowering the $dG_{ins,pred}$ of a given segment of a protein until it is predicted to partition favorably into membranes. Although transmembrane protein design is already well underway, with several suites of available Movers, score functions, and publications describing the design of transmembrane proteins, they are focused toward the generation of hydrophobic surfaces on the

exterior of a given protein so that the protein is stable within a membrane [47,86,87]. This typically results in hydrophobic sequence segments, but is not explicitly accounted for. For many membrane proteins, especially those in eukaryotic cells, the membrane-spanning domains are often single helices or are relatively small. Thus, for *de novo* designed proteins such as the HFuse scaffolds used to design KWOCAs, there is value in addition sequence design in order to favor not only the stable final conformation of a transmembrane protein, but also to help the insertion of that protein into the membrane from the Sec translocon.

As yet, the Degreaser is hard-coded to increase $dG_{ins,pred}$, but the flipping of an inequality in the protocol would allow for the decrease of $dG_{ins,pred}$. The added benefit of this type of design is that it is semi-exhaustive, and so positions that can be mutated to favor transmembrane insertion, if only a few are available to mutate, can be robustly identified. Also, if there are structural domains that are to be preserved, the Degreaser can take those into account. For example, if the helical bundle portion of an HFuse is to be redesigned to be a transmembrane domain, we could simultaneously stabilize its final conformation (via such features as RosettaMP) as well as ensure that a segment of that helical bundle is sufficiently hydrophobic. This may become more important as such features as hydrogen bonding networks are introduced into cores of proteins; interfaces and cores of future proteins need not be strictly hydrophobic. The final added benefit of this type of redesign rather than direct generation of transmembrane proteins is that it enables the addition of this transmembrane feature to an existing structure, assuming secondary structure elements allow it. Thus, it’s envisionable to have one protein that is

designed with another phenotype in mind, and have both a soluble and membrane-anchored version of the same protein. This hearkens to the ways that nature re-uses valuable protein domains: antibodies and B-cell surface receptors share an immunoglobulin for antigen recognition but antibodies are solubly secreted while the receptors are membrane-anchored by another domain [88]. To improve on nature's example, we could design a single functional protein with two forms, rather than require multiple polypeptides.

4.2 Applications and next generations of secreted protein nanoparticles

Before this work, the only available naturally-occurring nanoparticles that were adapted for scaffolding of viral glycoproteins were lumazine synthase and ferritin. Now, several secretable protein nanoparticles are available. The natural progression for King lab materials is to generate novel fusions of KWOCAs and of I3-01 to viral glycoproteins, express, and characterize them. Relative to ferritin, most KWOCAs do not secrete as robustly, but next generations of designed nanoparticles could. Even so, these proteins are ready for use: they are ready to be encoded genetically for, say, mRNA-based vaccines. These are currently in development by King lab member Grace Hendricks, who has demonstrated that secreted KWOCAs are readily suitable for genetic fusion to immunogen proteins, and can be encoded as mRNA and delivered to mice. These nanoparticle fusion retain immunogenicity whether they are expressed and purified, then injected, or delivered directly as mRNA. This marks the translation of the concepts that underlie the Degreaser into practical outcomes for vaccine design. Furthermore, KWOCAs can be used to understand how assembled proteins interact with biological

compartments. Finally, because these nanoparticles are modularly generated, they are customizable in a way that naturally-occurring proteins are not.

Because the design protocol used to generate the KWOCAs has been established, retracing its path but with knowledge of features such as flexibility within scaffolds enables more focused design that has potentially higher design success rate. Two-component, secretion-optimized nanoparticles can be designed, and scaffolds that led to off-target assemblies can be excluded. Other nanoparticle design protocols using diverse sets of scaffolds are also Degreaser-compatible, and so there is no effective limit on what proteins can be secreted. What does remain is choosing a strategy to express two components simultaneously and then simultaneously assessing protein assembly and secretion. Chelsea Shu has demonstrated that assemblies can be secreted, but there remains much optimization of expression conditions and many more designs to be screened.

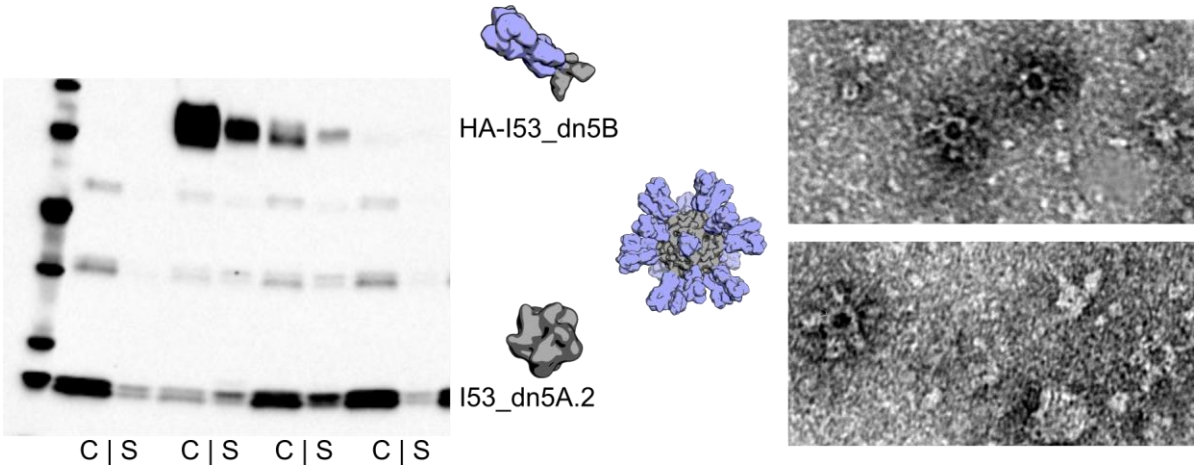


Figure 4.2. Secretion of a two-component nanoparticle, I53_dn5. (left) Optimization of co-transfection of I53_dn5A.2 and HA-I53_dn5B vector ratios for near-stoichiometric expression of the two components. (center) Depiction of each nanoparticle component, as well as the final assembly. (right) nsEM of unpurified cell culture supernatant from a co-transfection experiment showing presence of icosahedral assemblies with HA. Figure generated with data collected by Chelsea Shu.

4.3 *In vivo* assays for parallel, high-throughput characterization of protein secretion

As the ability to query many proteins at once increases, and as sequencing of large numbers of unique members of a given population becomes more facile, the critical limiting point of understanding a particular phenomenon is the development of an assay that can report on the underlying mechanisms of that phenomenon. Deep mutational scanning, as an example, has enabled the in-depth analysis of clinically-relevant proteins [89]. The data sets that can be generated and conclusions drawn from them vastly dwarf what could be possible with low-throughput methods. This type of rapid

generation of large numbers of variants is akin to that of *de novo* protein design, but computational protein design is not as limited in sequence space as is the mutation of natural proteins. Thus, for a phenotype such as eukaryotic secretion, or specifically translocation efficiency of a protein into the ER, an assay to determine the 'fitness' of a particular variant or unique design that is amenable to multiplexing would be highly desirable. Although similar types of assays exist for surface display, bulk quantification in culture supernatant, or within the cellular cytoplasm, unique considerations are necessary when determining protein translocation *in vivo* [90–93].

4.3.1 Fluorescence ratiometric assay

One of the difficulties of querying translocation or secretion is its intrinsic tie to overall protein expression. Furthermore, most multiplex methods do not distinguish among subcellular locales, such as organellar or surface membranes. Thus, an assay to determine a given protein's translocation efficiency into the ER during expression has to account for both the overall expression of the protein *and* the protein that translocates. Measurement of total protein expression is subject to stochastic noise on both the cellular and molecular levels, typically resulting in a distribution of measured expression yield. But, if the translocated yield within a given cell could be normalized to something that can count the total "expression strength" of that cell, then the final measurement can be less biased. Thus, a ratiometric assay that measures the ratio of two fluorescence signals can be used to reduce the noise and variance that can result from a particular proteins' behavior [94]. Fluorescent proteins are an attractive *in vivo* probe

when contrasted with organic or inorganic labels because they can be genetically fused to a queried protein. Flow cytometry methods are also amenable with fluorescent proteins, and when coupled with next-generation sequencing, allow for sorting and sequencing strategies that can simultaneously characterize tens to hundreds of thousands of constructs.

The assay in development takes advantage of a “self-cleaving peptide,” in quotation marks because they are not truly self-cleaving; rather, the sequence of the amino acids prevents the formation of a peptide bond as the protein is translating [95]. This enables multiple proteins to be expressed from the same mRNA transcript, allowing expression levels among proteins to be closer to stoichiometric. The separate proteins in this assay allow for simultaneous accumulation of a capture protein in the cell’s secretory pathway, a cytoplasmic normalization fluorescence signal, and a fluorescence signal for the queried protein. The fluorescent proteins chosen are mCherry and a superfolder variant of roGFP, a redox-sensitive green fluorescent protein [96]. Redox sensitivity has been used as an indicator of whether a protein exists in a more oxidizing environment, such as the ER, where disulfide bonds and enzymes dedicated to disulfide formation are prevalent, or in a more reducing environment, such as the cellular cytoplasm. With these components, we can detect two distinct signals that report on the status of a particular cell: (1) whether the cell is translating highly or lowly, based on the mCherry signal, and (2) whether the translated protein has been successfully translocated or if it remains in the cytoplasm. The capture of the protein is a SpyCatcher, which has been previously reported to be a highly robust biochemical and *in vivo* reagent for efficient

covalent isopeptide bond formation [97]. This SpyCatcher is then equipped with a C-terminal ER retention signal, so that the protein stays within the cell (with some potential leakage). Other assays have demonstrated that association with a protein that bears an ER retention signal is sufficient for capture of the unlabeled protein [93].

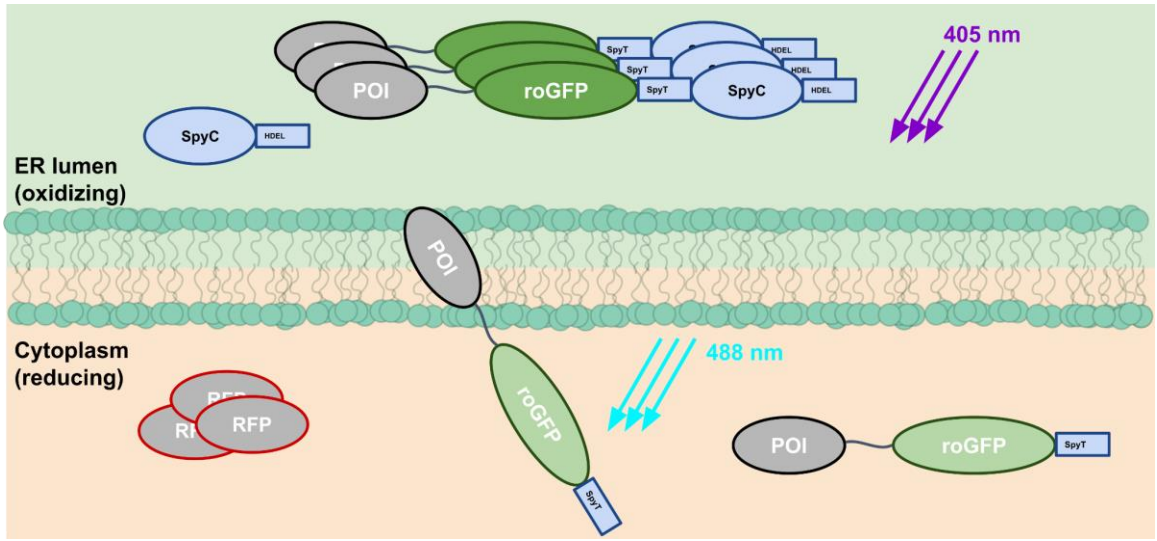


Figure 4.3. Schematic representation of fluorescence ratiometric assay for protein secretion. Bhaskar *et al*. describe the usage of a redox-sensitive GFP in cell-based assays [98], which we can leverage with a cytoplasmically-expressed RFP to normalize for total expression activity. Captured secretory protein and cytoplasmic protein can be simultaneously detected by taking a ratio of the emission signal from two different stimulation frequencies. Although this leads to differences in absolute signals because two different lasers and detectors would be used, controls would be run in parallel that establish baseline ratios for cytoplasmic or secreted protein. SpyT - SpyTag, SpyC - SpyCatcher, HDEL - ER retention signal, POI - protein of interest.

The proof-of-concept for this assay has been demonstrated to show expression of all of the components of the assay with no protein fused to the roGFP, which indicates that the secreted GFP can be successfully captured within the cell. Cells transfected with these plasmids are also fluorescent in both the red and green channels, and their flow cytometry signal can be interpreted as strong secretion of the GFP, which is somewhat to be expected, as sfGFP has been demonstrated previously to be robust with respect to soluble secretion [99]. Thus, the other controls necessary for the demonstration of the validity of this assay are membrane-spanning domains or proteins of varying topologies, such as N-in-C-out or N-out-C-in, as well as cytoplasmically targeted or known secretory proteins. With this assay, a large number of variants can be rapidly assessed in a multiplex format for high secretion efficiency as well as high secreted yield, because high ratios of green fluorescence to red fluorescence imply strong expression relative to mCherry. However, at present, there is not a particular pipeline in progress for the secretion of a large variety of proteins, but this platform should be soon available.

4.3.2 Fluorescence complementation assay

In parallel to a fluorescence ratiometric assay, a fluorescence complementation assay, which takes advantage of a split green fluorescent protein, is also in development. This assay has the advantage that only one fluorescent signal is measured, as opposed to normalization against another (although we lose some of the expression robustness information). Here, a non-fluorescent GFP1-10 is retained in the secretory pathway much as the capture reagent of the previously described assay. The protein of interest is then fused to GFP11, which completes the fluorescent protein in the endoplasmic

reticulum[100]. This leads to a steady-state fluorescent signal, and the association of the components of the split fluorescent protein is sufficient to delay the constitutive secretion of the target protein. Indeed, preliminary indications are that the protein of interest is secreted at low levels, but the cells expressing the entire reporter construct are fluorescent in the green channel, indicating that the semi-transient association of the two components of the GFP are sufficient to generate fluorescent signal.

The advantage of a fluorescence complementation assay is that there are fewer components to assess, and only one fluorescent channel to detect. The other is that the signal directly corresponds with total secreted yield, which may be sufficient for the determination of highly secreting proteins. Looking at a set of KWOCAs that are assayed by this fluorescence complementation assay, we see a positive correlation of the fluorescent signal obtained with the calculated secretion yield. There are other factors that could lead to different results by these two assays, such as what happens to each KWOCA when fused to fluorescent proteins, and whether retention of assembly influences secretion yield. Further investigation is merited before a fluorescence assay is suitable for large-scale studies, and, as before, a large-scale set of designs is necessary to examine.

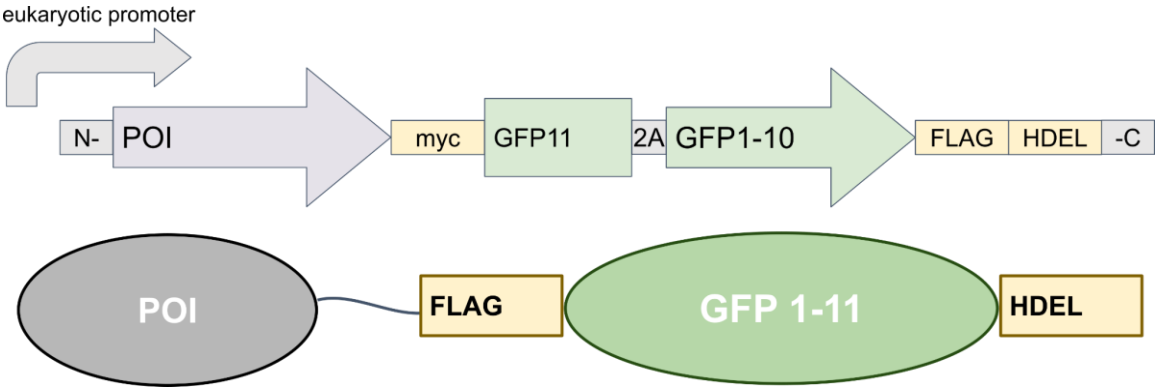


Figure 4.4. Schematic representation of fluorescence complementation assay. The expression of an ER-resident GFP 1-10 (GFP1-10, FLAG tag for detection, HDEL for retention), enables the generation of a fluorescent signal when the expressed protein of interest (POI) is successfully translocated into the ER. Feng *et al.* demonstrated that the 1-10 construct has very low baseline fluorescence signal [100]. This makes a split GFP reporter attractive because signal is directly correlated with overall secretion yield.

There remains an opportunity to assess the secretion of a wide variety of designed and natural proteins alike. Whether these proteins are initially recalcitrant to secretion, or need further design optimization before characterization, the Degreaser is easy to use for anyone familiar with protein design. The hope for protein design moving forward is a comprehensive suite of software that enables application-tailored generation of any phenotypes of interest, such as target binding, adoption of some particular geometry, or stimulation of a precise immune response. The Degreaser, then, plays its part as one portion of a larger pipeline that can be used to generate ready-to-secrete proteins. Coupled with an appropriate multiplex assay, generation of large sets of designed proteins that can be characterized in parallel can become routine. Insights gained from

previous large-scale studies are still used today, such as the miniproteins designed by Rocklin *et al.* in 2017, many of which serve as scaffolds for protein binders due to their robustness. There are many more protein phenotypes to be designed for eukaryotic expression, and this work represents one step in that direction.

Appendix A - Sequences of proteins characterized in this work

Protein ID	full protein sequence
1E4C (N29S)	METDTLLLWVLLLWVPGSTGDGSGMERNKLARQIIDTCLEMTRLGLNQGTAGSVSVRYQDGMLITPTGIPYEKLT ESHIVFIDGNGKHDEGKLPSSEWRFHMAAYQSRPDANAVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNSIPCAP YATFGTRELSEHVALALKNRKATLLQHHGLIACEVNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIAVVLE KFKTYGLRIEEGSEQKLISEEDL
1WA3	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKEL SFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFP GEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGCTE GSEQKLISEEDL
2JFB	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARWNRKIIDALVAGAVKRLQEFGVKEENIIIETVPGSFEL PYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIE GKMHNHGEDWGAAAVEMATKFNGSEQKLISEEDL
2OBX	METDTLLLWVLLLWVPGSTGDGSNQHSHKDYETVRIAVVRARWHADIVDQCVSAFEAEMADIGGDRFAVDVFD VPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVIDGMMNVQLSTGVPVLSAVLTPHNYHDSAEH HRFFFEHFTVKGKEAARACVEILAAREKIAAGSEQKLISEEDL
5HRZ	METDTLLLWVLLLWVPGSTGDGSNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKKG EYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNAEAKQNLGNAKQKQGGSEQK LISEEDL
O43-38B N29S (O43-38 tetramer)	METDTLLLWVLLLWVPGSTGDGSGMERNKLARQIIDTCLEMTRLGLNQGTAGSVSVRYQDGMLITPTGIPYEKLT ESHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNSIPCAP AATFGTDELSMLVAVALLNRKATLLQHHGLIACEVNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIAVVLE KFKTFGLRIEEGSEQKLISEEDL
O43-38B A141E	METDTLLLWVLLLWVPGSTGDGSGMERNKLARQIIDTCLEMTRLGLNQGTAGSVSVRYQDGMLITPTGIPYEKLT ESHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNSIPCAP AATFGTDELSMLVEVALLNRKATLLQHHGLIACEVNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIAVVLE KFKTFGLRIEEGSEQKLISEEDL
O43-38B N146D	METDTLLLWVLLLWVPGSTGDGSGMERNKLARQIIDTCLEMTRLGLNQGTAGSVSVRYQDGMLITPTGIPYEKLT ESHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNSIPCAP AATFGTDELSMLVAVALLDRKATLLQHHGLIACEVNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIAVVLE KFKTFGLRIEEGSEQKLISEEDL
O43-38B R147N	METDTLLLWVLLLWVPGSTGDGSGMERNKLARQIIDTCLEMTRLGLNQGTAGSVSVRYQDGMLITPTGIPYEKLT ESHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNSIPCAP AATFGTDELSMLVAVALLNNKATLLQHHGLIACEVNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIAVVLE KFKTFGLRIEEGSEQKLISEEDL
O43-38B T150S	METDTLLLWVLLLWVPGSTGDGSGMERNKLARQIIDTCLEMTRLGLNQGTAGSVSVRYQDGMLITPTGIPYEKLT ESHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHCTAVSILNRSIPAIHYMIAAAGGNSIPCAP AATFGTDELSMLVAVALLNRKASLLQHHGLIACEVNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIAVVLE KFKTFGLRIEEGSEQKLISEEDL
DEO5	METDTLLLWVLLLWVPGSTGDGSGMNRNELARQIIDTMKEMTRLGLNQGTAGSVSVRYQDGMLITPIGIPYEKLT EDHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHSTAVSILNREIPAIHYMIAAAGGNSIPSAP AATFGTDELSMLVEVALLDRKATLLQHHGLIAVETNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIKTVLEK FKTFGLRIEEGSEQKLISEEDL
DEO5 E141A D146N	METDTLLLWVLLLWVPGSTGDGSGMNRNELARQIIDTMKEMTRLGLNQGTAGSVSVRYQDGMLITPIGIPYEKLT EDHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHSTAVSILNREIPAIHYMIAAAGGNSIPSAP AATFGTDELSMLVAVALLNRKATLLQHHGLIAVETNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIKTVLEK FKTFGLRIEEGSEQKLISEEDL
I3-01-wt	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLF

	PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCT EGSEQKLISEEDL
I3-01-K25D	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKKDALAVFLGGVHLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCT EGSEQKLISEEDL
I3-01-H35D	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVDLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCT EGSEQKLISEEDL
I3-01-F41Q	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITQTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCT EGSEQKLISEEDL
I3-01-F30Y	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKKKALAVYLGGVHLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCT EGSEQKLISEEDL
I3-01-L171Q- S177E-V180N	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVQAVGVGEALNKGTPVEVAEKAKAFVEKIRGCT EGSEQKLISEEDL
I3-01-H35D- L171Q-S177E- V180N	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVDLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVQAVGVGEALNKGTPVEVAEKAKAFVEKIRGCT EGSEQKLISEEDL
I3-01-negative- interior	METDTLLLWVLLLWVPGSTGDGSKMEELFKEHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQAREAVESGAEFIVSPHLDEEISQFAKEEGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVEAMKGPFPNVKFVPTGGVNLDNVAEWFEAGVLAVGVGSALVEGTPVEVAEKAKAFVEKIEGAT EGSEQKLISEEDL
I3-01-negative- interior-secmuts	METDTLLLWVLLLWVPGSTGDGSKMEELFKEHKIVAVLRANSVEEAKKKALAVFLGGVDLIEITFTVPDADTVIKEL SFLKEMGAIIGAGTVTSVEQAREAVESGAEFIVSPHLDEEISQFAKEEGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVEAMKGPFPNVKFVPTGGVNLDNVAEWFEAGVQAVGVGEALNEGTPVEVAEKAKAFVEKIEGAT EGSEQKLISEEDL
O43-38A (trimer)	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVVLALAKALAVFLGGVHLIEITFTVPDADTVIKEL SFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFP GEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGEPEEVREKAKAFVEKIRGCTE GSEQKLISEEDL
O43-38AV19D	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSDVLALAKALAVFLGGVHLIEITFTVPDADTVIKEL SFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFP GEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGEPEEVREKAKAFVEKIRGCTE GSEQKLISEEDL
O43-38AL21E	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVVEALAKALAVFLGGVHLIEITFTVPDADTVIKEL SFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFP GEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGEPEEVREKAKAFVEKIRGCTE GSEQKLISEEDL
O43-38AL31N	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVVLALAKALAVFNGGVHLIEITFTVPDADTVIKEL SFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFP GEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGEPEEVREKAKAFVEKIRGCTE GSEQKLISEEDL
I53-50 A.1PT1	METDTLLLWVLLLWVPGSTGDGSKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVPDADTVIKAL SVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHDILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCKWFKAGVLAVGVGKALVKGKPDEVREKAKKFVKKIRGCT EGSEQKLISEEDL

I53_dn5A	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARWNAEIILALVLGALKRLQEFGVKRENIIIETVPGSFEL PYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIE GKMHNHGEDWGAAAVEMATKFNGSEQKLISEEDL
dn5A W16E	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARENAEIILALVLGALKRLQEFGVKRENIIIETVPGSFELP YGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIEG KMHNHGEDWGAAAVEMATKFNGSEQKLISEEDL
dn5A L26E	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARWNAEIILALVEGALKRLQEFGVKRENIIIETVPGSFEL PYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIE GKMHNHGEDWGAAAVEMATKFNGSEQKLISEEDL
dn5A L29N	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARWNAEIILALVLGANKRLQEFGVKRENIIIETVPGSFEL PYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIE GKMHNHGEDWGAAAVEMATKFNGSEQKLISEEDL
dn5A K30E	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARWNAEIILALVLGALERLQEFGVKRENIIIETVPGSFEL PYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIE GKMHNHGEDWGAAAVEMATKFNGSEQKLISEEDL
dn5A.2	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARGNAEIILELVLGALKRLQEFGVKRENIIIETVPGSFELP YGSKLFVEKQKRLGKPLDAIIPIGVLIRGSTAHFDYIADSTTHQLMKLNFELGIPVIFGVLTTESDEQAEERAGTKAG NHGEDWGAAAVEMATKFNGSEQKLISEEDL
dn5A.2 G16W E23A	METDTLLLWVLLLWVPGSTGDGSGKYDGSKLRIGILHARWNAEIILALVLGALKRLQEFGVKRENIIIETVPGSFEL PYGSKLFVEKQKRLGKPLDAIIPIGVLIRGSTAHFDYIADSTTHQLMKLNFELGIPVIFGVLTTESDEQAEERAGTKA GNHGEDWGAAAVEMATKFNGSEQKLISEEDL
I53-50 B.4PT1	METDTLLLWVLLLWVPGSTGDGSNQHSHKDHETVRIAVVRARWHAEIVDACVSAFEAAMRDIGGDRFAVDVFDV PGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVINGMMNVQLNTGVPVLSAVLTPHNYDKSKAHT LLFLALFAVKGMEAARACVEILAAREKIAAGSEQKLISEEDL
I53_dn5B	METDTLLLWVLLLWVPGSTGDGSEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGR YREAIEYYQKALELDPNNAEAWYNLGNAYYERGEYEEAIEYYRKALRLDPNNADAMQNLLNAKMREEGSEQKLI SEEDL
BOLAS	MRKWEEIAERLREEFNINPEEAREAVEKAGGNEEEARRIVKKRLGGSEQKLISEEDLGGSTQTQEFDNEEEARKA EKELRKENRRVTVTQENGRWRVTWDGGSLEHHHHHH
NAC26-DEO5	METDTLLLWVLLLWVPGSTGDYKDESGMNRNELARQIIDTMKEMTRLGLNQGTAGSVSVRYQDGMLITPIGIPYE KLTEDHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHSTAVSILNREIPAIHYMIAAAGGNSIP SAPAATFGTDELSMLVEVALLDRKATLLQHHGLIAVETNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIKT VLEKFKTFGLRIEEGSQKLISEEDLGSGSGSGSGSGSHHHHHHSSSDYSDLQRVKQELLEEVKKELQKVKEEIIEA FVQELRKRGSLVPRGSGGVKLAGNSSLCPVSGWAIYSKDNSVRIGSKGDVFVIREPFISCSPLECRTFFLTQGAL LNDKHSNGTIKDRSPYRTLMSCPIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNGIITDTI KSWRNNILRTQESECACVNGSCFTVMTDGPSNGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSCYPDSSEITCV CRDNWHGSNRPWVSFNQNLEYQIGYICSGIFGDNPRPNDKTGSCGPVSSNGANGVKGFSYKYGNGVWIGRTK SISSRNGFEMIWDPNGWTGTDNNFSIKQDIVGINEWSGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKENTIWT SGSSISFCGVNSDTVGWSWPDGAELPFTIDK
NAC130-DEO5	METDTLLLWVLLLWVPGSTGDYKDESGMNRNELARQIIDTMKEMTRLGLNQGTAGSVSVRYQDGMLITPIGIPYE KLTEDHIVFIDGNGKHEEGKLPQSEWRFHMAAYKARPDANAVVHNHAVHSTAVSILNREIPAIHYMIAAAGGNSIP SAPAATFGTDELSMLVEVALLDRKATLLQHHGLIAVETNLEKALWLAHEVEVLAQLYLTTLAITDPVPVLSDEEIKT VLEKFKTFGLRIEEGSQKLISEEDLGSGSGSGSGSGSHHHHHHSSSDYSDLQRVKQELLEEVKKELQKVKEEIIEA FVQELRKRGSLVPRGSGGSGWAPLSKDNSVRIGSKGDVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKD RSPYRTLMSVPIGSVPSPYNARFESIAWSASACHDGINWLTIGITGPDNGAVAILKYNGIITDTIKSWRNNILRTQES ECACVNGSCFTVMTDGPSNGQASYKIFRIEKGKIVKSVEMNAPNYHYEECSCYPDSSEITCVCRDNWHGSNRP WVSFNQNLEYQIGYICSGIFGDNPRPNDKTGSCGPVSSNGANGVKGFSFKYGNGVWIGRTKSISSRNGFEMIWD PNGWTGTDNNFSIKQDIVGINEWSGYSGSFVMHPELTGLDCIVPCFWVELIRGRPKENTIWTSGSSISFCGVNSD TTGWSWPDGAELPFTIDK
hMPV-F-wt-50A	MSWKVVIIFSLLITPQHGLKESYLEESCSTITEGYLSVLRTGWYTNVFTLEVGDVENLTCSDGPSLIKTELDLTKSAL RELKTVSADQLAREEQIENPRQSRFVLGAIALGVCTAAAVTCGVAIAKTIRLESEVTAIKNALKTTNEAVSTLGNGV RVLAFAVRELKDFVSKNLTRALNKNKCDIDDLKMAVSFSQFNRRFLNVVRQFSDNAGITPAISLDLMTDAELARAV SNMPTSAGQIKLMLENRAMVRRKGFGILIGVYGSSVIYMVQLPIFGVIDTPCWIVKAAPSCSGKKGNYACLLREDQ GWYCQNAGSTVYYPNEKDCETRGDHVFCDTACGINVAEQSKECNINISTTNYPCKVSTGRHPISMVALSPLGAL

	VACYKGVSCSIGSNRVGIIKQLNKGCSYITNQDADTVTIDNTVYCLSKVEGEQHVIKGRPVSSSFDPIKFPEDQFN VALDQVFENIENSQALVDQSNRILSSAEKGNTGSGGSGSGSGEKAAKAEEAARKMEELFKKHKIVAVLRANSVEE AIEKAVAVFAGGVHLIEITFTVPDADTVIKALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKE KGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAV GVGSALVKGTPDEVREKAKAFVEKIRGCTELEHHHHHH
dghMPV-F-50A	MSWKVVIIFSLLITPQHGLKESYLEESCSTITEGYLSVLRTGWYTNVFTLEVGDVENLTCADGPSLIKTELDLTKSAL RELRTVSADQLAREEQIENPRRRRFVLGDIALGRAEAAARTAGVAIAKTIRLESEVTAIKNALKKTNEAVSTLGNGV RVLATAVRELKDFVSKNLTRAINKNKCDIPDLKMAVSFSQFNRRFLNVVRQFSDNAGITPAISLDLMTDAELARAV SNMPTSAGQIKLMLENRAMVRRKGFGILIGVYGSSVIYMVQLPIFGVIDTPCWIVKAAPSCSEKKGNYACLLREDQ GWYCQNAGSTVYYPNEKDCETRGDHVFCDTAAGINVAEQSKECNINISTTNYPCKVSTGRHPISMVALSPLGAL VACYKGVSCSIGSNRVGIIKQLNKGCSYITNQDADTVTIDNTVYQLSKVEGEQHVIKGRPVSSSFDPVKFPEDQFN VALDQVFESIENSQALVDQSNRILSSAEKGNTSGRENLYFQGGGGSGYIPEAPRDGQAYVRKDGEWVLLSTFLG SGGSGSGSGGSGSGEKAAKAEEAARKMEELFKKHKIVAVLRANSVEEAIEKAVAVFAGGVHLIEITFTVPDADTVI KALSVLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTIL KLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIR GCTELEHHHHHH
FR-CONTROL	MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGS KAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMG WEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERA EGRHSTGGMDELYKEQKLISEEDLGSGEGRGSLLTCGDVEENPGPMETDTLLLWVLLLWVPGSTGDEFGGSGG SDYKDDDDKVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFISTTGKLPVPWPTLVTTLSY GVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKL EYNYNCHNVYIMADKQKNGIKVNFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTCSALSKDPNEKR DHMVLLEFVTAAGITLGMDELYKGSGESGSGVPTIVMVDAYKRYKGGSATNFSLLKQAGDVEENPGPMETDTLL LWVLLLWVPGSTGDYDYDVPDYAGGSEGGSGVTTLSGLSGEQGPSGDMTTEEDSATHIKFSKRDEDGRELAGA TMELRDSSGKTISTWISDGHVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGEATKGDAHTGSG ESGSGHDEL
FC-CONTROL	METDTLLLWVLLLWVPGSTGDMEFKTEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVA ATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKT WEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKH MNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKE FLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASG RQTVDEALKDAQTGGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLW VLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWI LVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPA DGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTE LKHSMGSDYKDDDDKHDEL
FC-KWOCA4	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEVEKKAREVAKEAVELAS LLRSETAIRVAQAILEAAEAAKRAAEQGKTEVAKLALKVLEEAIELAKEKRSEEALKVVLEIARAALAAAQAAEEGFT DVAKMALEVLERAIELAKDDRSEEALKEVLEIARAALAAAQLAKKGRDDEARKILMKLRIRITLRKLEESLRELRRIL EELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELAGSEQKLISEEDLGGSTELNFKEWQKA FTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVD FDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQ FEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGK RYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA18	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERALKAAL EAAEQALREGDEDAFKCAVELLEQALEARKKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLR LACEAARKKNSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAEL LKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAVGSEQKLISEE DLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMA SLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQ AAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPN DKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA46	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIEEARQAIKEAERALR EGDPRFAEMAVRIALALVRMLERLARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQAR KTGSTEVLIEAARLAIEVARVAFKVGSPETAKEAVRTALELVKELIQQALKTGSDEVLERAAELAKEVARVAKEVGD PRAARKADMVAKIADTLRELMESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQAA

	LASLAGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGST GDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFH QYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLT AADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYK DDDDKHDEL
FC-KWOCA47	METDTLLLWVLLLWVPGSTGDMEFGSADELRAVAELQRLNIELARKLLEAVARLQELNIDLVRKTSELTDEKTIRE EIRKVKEESKRIVEEAEEEIRRAKEDSKRIVTEALRRAREQIREKWEELEERAKRAETPEEALRAAEEMVKLIEELIR IAEMLQRAGLKEEAEDVLREATELIKRATELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVL REATELIKRVARLLLAIALLADTPELAKRAAELLKRLIELLKEIAKLLEEEGNEDEAEKVKEEAKELEELVRWLEEQIR GGSWSGGGSEQKLISEEDLGGSLEHHHHHHGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGD VEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYE ELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEG SHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMA ANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA51	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAK EALEQAKREGDEIARLCAKMLEILAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRL ACSAAARQDSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELL KALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAVGSEQKLISEED LGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMAS LPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQA AMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPND KTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA58	METDTLLLWVLLLWVPGSTGDMEFGSDECERLETEVMKAAKELMKLATQSTDKEVRKIAWEVADQLLRLAEEAC RSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQEAVK LAEQAKDERVIEVALEMARLLIELAQEACRRNDEEALRRASEIVKRVQELIKEAEKATDEEEIERLLRKAAIDITLAQ LEISLKELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKMLAELAGSWSGGGSEQKLISE EDLGGSLEHHHHHHGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLW VLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWI LVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPA DGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTE LKHSMGSDYKDDDDKHDEL
FC-KWOCA67	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIAKEISRIAEESKKAIE TLARLADKMTDENQVDTAIELIAKIAREAIKRIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLAELHRTDTFMAKA IEAIAELAKEAIKAIADLAKKHTTEEFMARAISAIAELAKKAIEAIWRLASLHKTDEFMDKAAEAIAELAEEAIRAIRELA KKHTTEEFVRKAESAVREISKKAKDAIRKLADAMRDPTAREKAKKLEIKVELAEALAELAVALLKLKLSLDELERSL EELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLDINAQLVGSEQKLISEEDLGGSTELNFKEWQKAFTDMMAT NFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQ GTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASL TVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTAR TTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA70	METDTLLLWVLLLWVPGSTGDMEFGSDESVDLAVKLAEALRKEAEELIKKARKTGDPELLRKALEALEKAVKLVE DAIKRNPDNDEAVETAVRLARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVKTAVE LAKELEKVARELLERARKTGDDELLKLAKRALEVARRAVELALKSRPDAEEARRVYIRLTEMELEISLTELRKILEEL KEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELAGSWSGGGSEQKLISEEDLGGSLEHHHHH HGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDM VSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYL PYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAAD WCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDD DKHDEL
FC-KWOCA43	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDEEVWKAVIDAIELMKEAR ELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNKIAVIVAVLLARELKKVAEELQERAKKTGDPELLKLALR ALEVAVRAVELAIKSNPDNDEAVETAVRLAEELAKVAKELIERAKKTGDADLLRLAKRAIEVARRAVELAKKSRPDA ERADEAYKRLKELEREIRELLRKMLTEALRKLEKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIV EVLSIIAEVLKINAALVGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLW VLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWI LVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPA

	DGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA100	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEETIEMARQLIKEAERALREGDPEEARMAVEMALAAVRILERQARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAKEAVRTALELVKELIQQALKTGSDEVLERAAELAKEVARVAKEVGDPRAARKADMVAKIADTLRELMESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQAALASLAGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA101	METDTLLLWVLLLWVPGSTGDMEFGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDEIARLCAEMLEILAEEARRKKDSEEAEAVYWAARATLAALEALEQAKREGDEDARRCAEELLRLACSAAARQDSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELLKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAVGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL
FC-KWOCA102	METDTLLLWVLLLWVPGSTGDMEFGSTEEKIEEARQSIKEAERSLREGNPEKALDAVARALSLVNELERLARKTGSTEVLIEAARLAIEVARVALKVGSPEMAQLAVELALRLVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAARTALELVEELERQARKTGSEEVLERAARLAEEVARVAEEIGDPELARKAMKVAIRLTEELLKKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISADNQRALARLAGSWSGGGSEQKLISEEDLGGSLEHHHHHHHGSEQKLISEEDLGGSTELNFKEWQKAFTDMMATNFSLLKQAGDVEENPGPMETDTLLLWVLLLWVPGSTGDMVSKGEEDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVMGTGFPADGPVMTNTLTAADWCMSKKTYPNDKTIISTFKWSYTTVNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSMGSDYKDDDDKHDEL

1wa3-mono-clean_2_1.128.pdb	KMLLLFKLHLIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_2_1.450.pdb	KMELLFKKLLIVAVLRALSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_2_1.849.pdb	KMLLLFLKHLIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_2_2.244.pdb	KMLELFLKHLIVAVLLANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_2	KMLLLFKKLKIVAVLRALSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC

_2.573.pdb	
1wa3-mono-clean_2_2.989.pdb	KMLELFKLHKIVAVLLALSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_2_3.330.pdb	KMLLLFKKHKIVAVLLALSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_24_3.007.pdb	KMEELFKKHKIVAVLRANSVEEALLKALAVFEGGVHLIEILFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_24_3.361.pdb	KMEELFKKHKIVAVLRANSVEEALLKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_25_1.052.pdb	KMEELFKKHKIVAVLRANSVEEAKLKALAVFLGGVHLIEITFLVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_25_1.498.pdb	KMEELFKKHKIVAVLRANSVEEAKLKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_25_1.881.pdb	KMEELFKKHKIVAVLRANSVEEAKLLALAVFEGGVHLIEILFLVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_25_2.250.pdb	KMEELFKKHKIVAVLRANSVEEAKLLALAVFEGGVHLIEITFLVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_25_2.691.pdb	KMEELFKKHKIVAVLRANSVEEAKLKALAVFEGGVHLIEILFLVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC

1wa3-mono-clean_40_1.063.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFLVPDADLVILLLLFLKEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_40_1.475.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFLVPLADTVILELLFLLEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_40_1.834.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPLADLVILELSFLLEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_40_2.182.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFLVPDADLVILLLSFLKEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_40_2.618.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADLVILLLSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_40_2.970.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDALTVIKLLSFLLEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_40_3.339.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDALLVIKELLFLKEKGAIIGAGTVTSVEQCRKAVESG AEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_73_1.060.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVLQCRLAVLLG ALFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLD NVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_73_1.434.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVLLCRKAVLSG ALFIVSPLLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLD NVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_73_1.826.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVELCRLAVLSG ALFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLD NVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC

1wa3-mono-clean_73_2.240.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVLQCLKAVLSGALFIVSPLLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_73_2.592.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVLQCRLAVESGALFIVSPLLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_73_2.941.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVLLCRKAVELGALFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_73_3.400.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVLLCRKAVLSGAEFIVSPLLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_91_1.129.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLLLEISQFCLLLGVFLMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_91_1.489.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDELILLFCLLKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_91_1.894.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLLEEILLFCLLKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_91_2.269.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLDLEISLFCLELGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_91_2.662.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLLEEISLFCLLKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_91_2.998.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVESGAEFIVSPHLLLEILQFCKLLGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC

1wa3-mono-clean_91_3.407.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDLLILQFCLEKGVFLMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNL DNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC
1wa3-mono-clean_184_1.106.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGLPDEVLLLALAFVLKILGC
1wa3-mono-clean_184_1.429.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPLEVRLLALAFVLKILGC
1wa3-mono-clean_184_1.805.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPLEVRLLALAFVLKIRGC
1wa3-mono-clean_184_2.235.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPLLVRLLAKAFVLKILGC
1wa3-mono-clean_184_2.643.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGLPLEVRLLALAFVEKILGC
1wa3-mono-clean_184_2.957.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGLPDEVLLLALAFVEKILGC
1wa3-mono-clean_184_3.418.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGLPDEVLLLAKAFVLKILGC
1wa3-mono-clean.pdb	KMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCRKAVES GAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVN LDNVCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGC

KWOCA	Symmetry_HF_Protocol_DesignNumber	full protein sequence (*E. coli*)	full protein sequence (Human)
0	I3_HF_02 (I3_HF_OG_00)	MSDEVDRRVLELAIKASRATDKEEVIEIVKELAELAKQSTDSRLVERIVT LLALVAIDATDKELVIYIVKILAELAKQSTDSELVKKIVEMLAQVARFATDK ELVEYIARILLELAKQADDATLVAFIAEMLAEVRKEAKDKELKEKIDEILK	METDTLLLWVLLLWVPGSTGDGSSDEVDRRVLELAIKASRATDKEEVIE IVKELAELAKQSTDSRLVERIVTLLALVAIDATDKELVIYIVKILAELAKQST DSELVKKIVEMLAQVARFATDKELVEYIARILLELAKQADDATLVAFIAE

		ELAKITLKALEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQEISAANQKALALLGLEHHHHHH	MLAEVRKEAKDKELKEKIDEILKELAKITLKALEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQEISAANQKALALLGGSEQKLISEEDL
1	I3_HF_OG_01	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIAKEISRIAEESKKRIEELARDADNAKDEEWVDRAIERIAKLAREAIKRIFELALQLKSEEFMARAISAIAELAKKAIEAIYRLADNHKTDTFMAKAIEAIAELAKEAIKAIAWLALAHDTEEFMARAISAIAELAKKAIEAIYRLAKNHRTDEFMRKAAEAIAELAKEAIKAIKDLAKLHTTREFKAKAESAIREIVKKAIEAIKRLAELMKDIKARLFAALLILDLMLLLLKVSLEKLEESLEELEKNPSEDALVENNRINVENNKIIVEVLRIIAEVLKLNAKAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIAKEISRIAEESKKRIEELARDADNAKDEEWVDRAIERIAKLAREAIKRIFELALQLKSEEFMARAISAIAELAKKAIEAIYRLADNHKTDTFMAKAIEAIAELAKEAIKAIAWLALAHDTEEFMARAISAIAELAKKAIEAIYRLAKNHRTDEFMRKAAEAIAELAKEAIKAIKDLAKLHTTREFKAKAESAIREIVKKAIEAIKRLAELMKDIKARLFAALLILDLMLLLLKVSLEKLEESLEELEKNPSEDALVENNRINVENNKIIVEVLRIIAEVLKLNAKAV
2	I3_HF_OG_02	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTRQKEQLDEVLKEIWRLATEAMKLMTDWKEAIKIAEEALRAMEMLKRAVEKVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRAAEEAMKLMRDEEEAAKIAKKALEAIKALAEAVELIKDKKQIETLLELVKLLIETAELEARKMSDREKAARIIERALEKIRDLAKIASTLAELKKSLDELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQRMLAKLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTRQKEQLDEVLKEIWRLATEAMKLMTDWKEAIKIAEEALRAMEMLKRAVEKVTDNEVIEKLLEVVKEIIRLAEEAMKLMRDEEEAAKIAKAIKMLARAVEEVTDNEVIEKLLEVVKEIIRAAEEAMKLMRDEEEAAKIAKKALEAIKALAEAVELIKDKKQIETLLELVKLLIETAELEARKMSDREKAARIIERALEKIRDLAKIASTLAELKKSLDELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQRMLAKLA
3	I3_HF_OG_03	MPFEVLLKAKESLERAREASERGDEEEFRKAAEKALKLARLLVIAAKMAGIPELVRAAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPKLVEEAAKVALKVAELAARNGDLEVFKKAAKSAKEVAKRLLEVATKRGDASLAAKAMEIMIELALLAVKVLQDLNIELARKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAKEISREVAEAARGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSPFEVLLKAKESLERAREASERGDEEEFRKAAEKALKLARLLVIAAKMAGIPELVRAAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPKLVEEAAKVALKVAELAARNGDLEVFKKAAKSAKEVAKRLLEVATKRGDASLAAKAMEIMIELALLAVKVLQDLNIELARKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAKEISREVAEAARGSWSGGGSEQKLISEEDLGGSLEHHHHHH
4	I3_HF_OG_04	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEVEKKAREVAKEAVELASLLRSETAIRVAQAILEAAEAAKRAAEQGKTEVAKLALKVLEEAIELAKEKRSEEALKVVLEIARAALAAAQAAEEGFTDVAKMALEVLERAIELAKDDRSEEALKEVLEIARAALAAAQLAKKGRDDEARKILMKLRIRITLRKLEESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEVEKKAREVAKEAVELASLLRSETAIRVAQAILEAAEAAKRAAEQGKTEVAKLALKVLEEAIELAKEKRSEEALKVVLEIARAALAAAQAAEEGFTDVAKMALEVLERAIELAKDDRSEEALKEVLEIARAALAAAQLAKKGRDDEARKILMKLRIRITLRKLEESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA
5	I3_HF_OG_05	PREDLEEAKERVKEIEKLIIKAELLQEQGNKEEAEKVLREAREQIREVTTELWIIARDSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEKIAENARKPELALRAAELLVRLIELLIEIAKLLREQGNREEAEKVLREAKELIKRVKELLEKIARLASTPEESKRAKKLILKLALLLLDMALLELKESLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKLNAELV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPREDLEEAKERVKEIEKLIIKAELLQEQGNKEEAEKVLREAREQIREVTTELWIIARDSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEKIAENARKPELALRAAELLVRLIELLIEIAKLLREQGNREEAEKVLREAKELIKRVKELLEKIARLASTPEESKRAKKLILKLALLLLDMALLELKESLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKLNAELV
6	I3_HF_OG_06	MSEEEKERIMRIAFEAMKSGTEESLRQAIEDVAQLAKKSQDSEVLEFAIRVIQLIAEASGSEEALRQAIRAVAEIAKEAQDSRVLREAIDVIRRIAEESGSEEALRQAKRAVTEIAKEADDPKVKEFAKRVILKLALLKLRVALQKLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAEVLEINAQLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSSEEEKERIMRIAFEAMKSGTEESLRQAIEDVAQLAKKSQDSEVLEFAIRVIQLIAEASGSEEALRQAIRAVAEIAKEAQDSRVLREAIDVIRRIAEESGSEEALRQAKRAVTEIAKEADDPKVKEFAKRVILKLALLKLRVALQKLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAEVLEINAQLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
7	I3_HF_OG_07	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVENRVRDLAWKAKVAKDKEEVRLIVLRLKLLADLSTDSELVNEIVKQLAEVAKEATDKELVIYIVIKILAELAKQSTDSELVNEIVKQLAEVAKEATDQELVKRIAEILLKLAKQSTDEKLVEEIEKQLREVEKWARDEKLQRDIRLILALLRLLVALLRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLQLNAEAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVENRVRDLAWKAKVAKDKEEVRLIVLRLKLLADLSTDSELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDQELVKRIAEILLKLAKQSTDEKLVEEIEKQLREVEKWARDEKLQRDIRLILALLRLLVALLRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLQLNAEAV
8	I3_HF_OG_08	MSEEEQERIRRILMEARLFGDEEALREAIIRVAKLAIDSQDSEVLEEAIRVILRIAKESGSEEALRLAIIAVAMIAKAAQDSEVLKEAIRVIKRIAEESGSADALALAAQAVAEIANEARDPEVLEFALRVALELLEQALRILKKSLDELERSLEELEKNPSEEALVENNRLNVENNKIIVAVLKIIAAVLKINAMLVGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSSEEEQERIRRILMEARLFGDEEALREAIIRVAKLAIDSQDSEVLEEAIRVILRIAKESGSEEALRLAIIAVAMIAKAAQDSEVLKEAIRVIKRIAEESGSADALALAAQAVAEIANEARDPEVLEFALRVALELLEQALRILKKSLDELERSLEELEKNPSEEALVENNRLNVENNKIIVAVLKIIAAVLKINAMLVGSWSGGGSEQKLISEEDLGGSLEHHHHHH
9	I3_HF_OG_09	MDECERLLTEVFKAAEELMKLARQSTDWEVRKIAMEVARQLIQLAEEACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQELVKLAEKATDEKVIKMALEVARDLIKLAQEACRRNNEEALRLASEAVKDVQELVKKATKATDERKIELLLALAQLAAALAQLRASLRELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINADLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDECERLLTEVFKAAEELMKLARQSTDWEVRKIAMEVARQLIQLAEEACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQELVKLAEKATDEKVIKMALEVARDLIKLAQEACRRNNEEALRLASEAVKDVQELVKKATKATDERKIELLLALAQLAAALAQLRASLRELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINADLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
10	I3_HF_OG_10	MSMQLRAVAMLQDLNIELARLLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAKRMIDAAKKASDAIRLAAEAMELAEELRQRRPNLQAADDARKLAKQAMEAVKLAVELLKEHPGSDIADLCIKAAREAAEAASKAAELALRHPNSQAARDAIKLASLAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDSQAARDAIKLASQAAKAVKIACFLAQEHPNADIAKKCIKAASEAAERASKAAELAQRHPDDELARDIIRLASQFAEEVIERCLRALGGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSSMQLRAVAMLQDLNIELARLLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAKRMIDAAKKASDAIRLAAEAMELAEELRQRRPNLQAADDARKLAKQAMEAVKLAVELLKEHPGSDIADLCIKAAREAAEAASKAAELALRHPNSQAARDAIKLASLAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDSQAARDAIKLASQAAKAVKIACFLAQEHPNADIAKKCIKAASEAAERASKAAELAQRHPDDELARDIIRLASQFAEEVIERCLRALGGSWSGGGSEQKLISEEDLGGSLEHHHHHH
11	I3_HF_OG_11	MDEEECEMLAKLVAKLVEALKRAGVSEDEIAEIVAEQISKVIEWLKRKGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVAEQISKVIEWLKRKGSSYEVICECVARIVAEIVEALKRSGTSEEEIAEIVARVLREVMRTLWESGSSAEVIYECLKRALEEIREALKRAGVSEDEIRRIKEYVDRVALRAALEILTEA	METDTLLLWVLLLWVPGSTGDGSDEEECEMLAKLVAKLVEALKRAGVSEDEIAEIVAEQISKVIEWLKRKGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEEEIAEIVARVLREVMRTLWESGSSAEVIYECLKRALEEIREALKRAGVSED

		LKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLDVIAMVLKLNADLAGSWSGGGSEQKLISEEDLGGS	EIRRIKEYVDRVALRAALEILTEALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLDVIAMVLKLNADLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
12	I3_HF_OG_12	MDEEELERFERIFREVRSPDPENIREAVRKAEELLRENPSLVAEALLILAIEAAVRAPDPEAIREAVRAALELLRENPSLVAEALLILAIEAAVRAPDPRAIREAVRAALELMLENPSEEAKELLRRAIESAKKAPDPRAQTLAEIAETLLRLLKSLDELRRILEELKEMLERLERNPDKDVIVKVLKVIVKAIEASVENQRISAENQAMLALLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDEEELERFERIFREVRSPDPENIREAVRKAEELLRENPSLVAEALLILAIEAAVRAPDPEAIREAVRAALELLRENPSEEAEELLRRAIEAAVRAPDPRAIREAVRAALELMLENPSEEAKELLRRAIESAKKAPDPRAQTLAEIAETLLRLLKSLDELRRILEELKEMLERLERNPDKDVIVKVLKVIVKAIEASVENQRISAENQAMLALLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
13	I3_HF_OG_13	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEKWIRRAKEVAKEVEKVAQRAEEEGNPRLRDAAKALRLAVEMAIVVAKLLGNPELVEWVARAARVAAEVIKVAIEAEKAGNRDLFRAALELVRAVIEAIAIAVALGDPEAVEAAARLAKVAAEVIKVAIKAEEGGFRELFRKALELVRKAIEAMEKAIKEGDARKIEKVAEEARERAEEIKKAMKDKARETLKRLEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKVLAAIA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEKWIRRAKEVAKEVEKVAQRAEEEGNPRLRDAAKALRLAVEMAIVVAKLLGNPELVEWVARAARVAAEVIKVAIEAEKAGNRDLFRAALELVRAVIEAIAIAVALGDPEAVEAAARLAKVAAEVIKVAIKAEEGGFRELFRKALELVRKAIEAMEKAIKEGDARKIEKVAEEARERAEEIKKAMKDKARETLKRLEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKVLAAIA
14	I3_HF_OG_14	MGEKVTAWQLAEEAKETAEKVGDPRLEALAEQAAQEGDSEKAKAILLAAEAARVAKEVGDPELIKLAEEAADRGDSQKAKAILLAAEAARVAKEAGMPDLIRLALMAARAGGDAAAAAILAAADAAREAKEEGDPRKMELARALAAVAALQALNLALALKLLEAVARLQELNIDLVRKTSELRAEEAIRLEIKKVKEESKRIVEEAEEEIRRAAKMSSEAAQLAGGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSGEKVTAWQLAEEAKETAEKVGDPRLEALAEQAAQEGDSEKAKAILLAAEAARVAKEVGDPELIKLAEEAADRGDSQKAKAILLAAEAARVAKEAGMPDLIRLALMAARAGGDAAAAAILAAADAAREAKEEGDPRKMELARALAAVAALQALNLALALKLLEAVARLQELNIDLVRKTSELRAEEAIRLEIKKVKEESKRIVEEAEEEIRRAAKMSSEAAQLAGGSWSGGGSEQKLISEEDLGGSLEHHHHHH
15	I3_HF_OG_15	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVETRVFELAVEAKRAKDKETVDAIVKRLAWLAKLSTDSELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDQELVKRIAEILLKLAKQSTDEKLVEEIEKQLREVEKAAKDEKLQRDIRLILALLRLLVALLRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLQLNAEAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVETRVFELAVEAKRAKDKETVDAIVKRLAWLAKLSTDSELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDQELVKRIAEILLKLAKQSTDEKLVEEIEKQLREVEKAAKDEKLQRDIRLILALLRLLVALLRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLQLNAEAV
16	I3_HF_OG_16	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDERDKQWEEVMRLALELAKKATDKELRKEIMKCLILATELAMRSTNDELIKQILEVAKLAFELASKATDEELIKEILKCCQLAFEAASRADDDELIKLLLTVAKLAFEVASNAKDEEFIKEVLKIAQLAFEAINRATDKRRQKILATLAILLASLMKLKQILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDERDKQWEEVMRLALELAKKATDKELRKEIMKCLILATELAMRSTNDELIKQILEVAKLAFELASKATDEELIKEILKCCQLAFEAASRADDDELIKLLLTVAKLAFEVASNAKDEEFIKEVLKIAQLAFEAINRATDKRRQKILATLAILLASLMKLKQILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA
17	I3_HF_OG_17	MPRQLLEEAKERVETIERLIFTAELLQLKGNKEEAEKVLREAREQIREVTLILEAIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREAEELIKRVLELLEKIAENADTEELARRAAELIKRLIELLKEIAKLLEEAGKKDEAEKVKEKAKEMKERVDILETLIELERSLRELRRILEELKEMLERLERNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALARLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSPRQLLEEAKERVETIERLIFTAELLQLKGNKEEAEKVLREAREQIREVTLILEAIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREAEELIKRVLELLEKIAENADTEELARRAAELIKRLIELLKEIAKLLEEAGKKDEAEKVKEKAKEMKERVDILETLIELERSLRELRRILEELKEMLERLERNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALARLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
18	I3_HF_OG_18	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERALKAALEAAEQALREGDEDAFKCAVELLEQALEARKKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRLACEAARKKNSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELLKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERALKAALEAAEQALREGDEDAFKCAVELLEQALEARKKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRLACEAARKKNSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELLKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAV
19	I3_HF_OG_19	MADELKAVSRLQRLNIELARKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAKEISRKITKEAAKRAADEIVREAARRAGLDSSEAKELADRLIKEVAENAERGYDDAEALLAAAEAFKRVAEAAKRAGIVSTEALELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGMASSEALRLAIEIIRFIVKVAQREGNDISEAARQAAEWFRKIATTAKRGGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSADELKAVSRLQRLNIELARKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAKEISRKITKEAAKRAADEIVREAARRAGLDSSEAKELADRLIKEVAENAERGYDDAEALLAAAEAFKRVAEAAKRAGIVSTEALELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGMASSEALRLAIEIIRFIVKVAQREGNDISEAARQAAEWFRKIATTAKRGGSWSGGGSEQKLISEEDLGGSLEHHHHHH
20	I3_HF_OG_20	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTDEERLELERVAWEAKKAAREGDTDRVRWALLEAKLIADQSGTTEAVKLALEVVARVAIDAARRGNTDAVREALEVALEIARASGTTEAVKLALEVVARVAIEAARRGNSDAAREALRVAKEIARESGTEEARRLAKEVIKRVADEAEKAGAKKVAILATLALLKISLDELREILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTDEERLELERVAWEAKKAAREGDTDRVRWALLEAKLIADQSGTTEAVKLALEVVARVAIDAARRGNTDAVREALEVALEIARASGTTEAVKLALEVVARVAIEAARRGNSDAAREALRVAKEIARESGTEEARRLAKEVIKRVADEAEKAGAKKVAILATLALLKISLDELREILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA
21	I3_HF_OG_21	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVENRVRKLAWEAKVAKDKETVRLIVLELKLLADLSTDSELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDKELVDWIAKILLDLARQSTDKELVKEIEKQLDEVKKWARDEELQRRIELILALLRLLVALLELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLMLNADAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVENRVRKLAWEAKVAKDKETVRLIVLELKLLADLSTDSELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDKELVDWIAKILLDLARQSTDKELVKEIEKQLDEVKKWARDEELQRRIELILALLRLLVALLELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLMLNADAV
22	I3_HF_OG_22	MPEEALKRVEKLVWKALELLIKAKLRGSEEDLEKALRTAEEAAREAKVVLAMAEREGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEELVRAVELAEKQGDPKVALRAVELVVRVAELLMRIAKESGSEEAKERAAKVAAEAARLAKRVAELAAKAGDKDVAKKALELALRALEIMLEILRDMLRKLKESLEELKKNPSEDALVRNNELIVKVLEIIVMVLEAIAAVLKINALLVGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSPEEALKRVEKLVWKALELLIKAKLRGSEEDLEKALRTAEEAAREAKVVLAMAEREGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEELVRAVELAEKQGDPKVALRAVELVVRVAELLMRIAKESGSEEAKERAAKVAAEAARLAKRVAELAAKAGDKDVAKKALELALRALEIMLEILRDMLRKLKESLEELKKNPSEDALVRNN

			ELIVKVLEIIVMVLEAIAAVLKINALLVGSWSGGGSEQKLISEEDLGGSLEHHHHHH
23	I3_HF_OG_23	MDECERLRTEVMKAAKELMKLAEQSTDKEVRKIAWEVANQLWRLAEEACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVAKALIRLAQLACRLNIEEALKTASEIVKLVQELVKEAEKATDENEIRKLLIRAKLLIALAELEAALRLLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINAEAVGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDECERLRTEVMKAAKELMKLAEQSTDKEVRKIAWEVANQLWRLAEEACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVAKALIRLAQLACRLNIEEALKTASEIVKLVQELVKEAEKATDENEIRKLLIRAKLLIALAELEAALRLLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINAEAVGSWSGGGSEQKLISEEDLGGSLEHHHHHH
24	I3_HF_OG_24	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSEEVNERVKQLAEKAKKATDKEEVRQIVMELFQLAKQSTDSELVNEIVKQLAEVAKEATDKELVIFIVLVLAFLASLSTDSELVNEIVKQLAEVAKEATDKELVEYILRILLSLARLSTDKELVREIKKQMTEVKKELKDDEARELAERILLALALIELELALRELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAEVLKLNARLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSEEVNERVKQLAEKAKKATDKEEVRQIVMELFQLAKQSTDSELVNEIVKQLAEVAKEATDKELVIFIVLVLAFLASLSTDSELVNEIVKQLAEVAKEATDKELVEYILRILLSLARLSTDKELVREIKKQMTEVKKELKDDEARELAERILLALALIELELALRELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAEVLKLNARLA
25	I3_HF_OG_25	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGNKEEAEKVLREAREQIREVFRELVTIAVNADTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTLLLEAIAAAAETPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEATKVLREAEELIDRVLELLIEIALNADTDELAKRALKLILDLIKLLREIAKLLEEAGLEDEAKRVKRKADLMAVAALQLLNLLLALKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAARKSQEIVQVAL	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGNKEEAEKVLREAREQIREVFRELVTIAVNADTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTLLLEAIAAAAETPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEATKVLREAEELIDRVLELLIEIALNADTDELAKRALKLILDLIKLLREIAKLLEEAGLEDEAKRVKRKADLMAVAALQLLNLLLALKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAARKSQEIVQVAL
26	I3_HF_OG_26	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIAKEISRIAEESKKLIEELARIAANMQDERKVDEAIELIAEIARAIAEIKRIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLASMHNTDTFMAKAIEAIAELAKEAIKAIADLAKKHTTEEFMARAISAIAELAKKAIEAIYQLADRHTTDEFMRKAIEAIAELAKEAIRAIADLAKNLRTEEFIRKAESAIRELAKKAEEAIKRLADLLTDPTAREKAIEARTEIELMARLMNALLELITALLRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLMINAKLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIAKEISRIAEESKKLIEELARIAANMQDERKVDEAIELIAEIARAIAEIKRIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLASMHNTDTFMAKAIEAIAELAKEAIKAIADLAKKHTTEEFMARAISAIAELAKKAIEAIYQLADRHTTDEFMRKAIEAIAELAKEAIRAIADLAKNLRTEEFIRKAESAIRELAKKAEEAIKRLADLLTDPTAREKAIEARTEIELMARLMNALLELITALLRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLMINAKLV
27	O3_HF_OG_01	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDQEVWEAVRRAWELLLEARELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNLAAVWAAVELARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVKTAAKLARELMKVARELMERAKKTGDVELLKLAKRALEVAKRAVELAKKSRPDMEEAEELRKRIAELERKVRELIKKMLEKALRELEKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLEINAALA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDQEVWEAVRRAWELLLEARELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNLAAVWAAVELARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVKTAAKLARELMKVARELMERAKKTGDVELLKLAKRALEVAKRAVELAKKSRPDMEEAEELRKRIAELERKVRELIKKMLEKALRELEKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLEINAALA
28	O3_HF_OG_02	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEELKKVLETVRTLSELAKRSDDPEEATLIAMVVILLALRAVKEDPSTDALRAVLEAVRLASEVAKRVTDPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRLASEVAKRVTDPKKALKIAELVIRLAEEAVKEDDSEEAQRAVEEAERLAREVAKRVRDKKESIRMTLLVLEMSLTRLRVILERLKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQKISAENQKALAELA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEELKKVLETVRTLSELAKRSDDPEEATLIAMVVILLALRAVKEDPSTDALRAVLEAVRLASEVAKRVTDPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRLASEVAKRVTDPKKALKIAELVIRLAEEAVKEDDSEEAQRAVEEAERLAREVAKRVRDKKESIRMTLLVLEMSLTRLRVILERLKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQKISAENQKALAELA
29	O3_HF_OG_03	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRKAVELAIEIARWLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIMIAVVLGDPRLVEAVARAAKVAAEVIKVAIDAEKAGAREMFRRALELVREVIEAIEEAVIEGDPERVERVARKATKEALDIALKLLEMLLQRLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSMIAKVLKLNAKLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRKAVELAIEIARWLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIMIAVVLGDPRLVEAVARAAKVAAEVIKVAIDAEKAGAREMFRRALELVREVIEAIEEAVIEGDPERVERVARKATKEALDIALKLLEMLLQRLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSMIAKVLKLNAKLV
30	O3_HF_OG_04	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTRQKEQLDEVLEWIEDLAKMARLLMTDEEEAKKIQEEAERAKEMLRRAVEKVTDNEVIEKLLEVVKEIIRLALQAMLKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRLAEEAMKLMRNEEEAAKIAKMALEAIKLLAEAVELITDKEQIEQLLRKVKELIRWAEEVARKMSDREKAAMIAAIALTLIALEASLASLKAILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISARNQEMLARLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTRQKEQLDEVLEWIEDLAKMARLLMTDEEEAKKIQEEAERAKEMLRRAVEKVTDNEVIEKLLEVVKEIIRLALQAMLKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRLAEEAMKLMRNEEEAAKIAKMALEAIKLLAEAVELITDKEQIEQLLRKVKELIRWAEEVARKMSDREKAAMIAAIALTLIALEASLASLKAILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISARNQEMLARLA
31	O3_HF_OG_05	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRKAVEIAITIAKLLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVILAILIAVILGDPRLVEAVARAAKVAADAIKVAIEAEKMGLREVFRDVRAVMKGDPEEVERVAREATKRAIEATLIILERSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISALNQKMLAQLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRKAVEIAITIAKLLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVILAILIAVILGDPRLVEAVARAAKVAADAIKVAIEAEKMGLREVFRDVRAVMKGDPEEVERVAREATKRAIEATLIILERSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISALNQKMLAQLA
32	O3_HF_OG_06	MALEKDRRALEALKRAQEAEKKGDVEEAVRAAQEAVRAAKESGASWILRLVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVDVAAKAAQVAAEAAKQAGDKDMLEKVAKVAEQIAKAAEKEGDKKVSIDATRIALEASLAALEIILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQKISAKNQKALAELAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSALEKDRRALEALKRAQEAEKKGDVEEAVRAAQEAVRAAKESGASWILRLVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVDVAAKAAQVAAEAAKQAGDKDMLEKVAKVAEQIAKAAEKEGDKKVSIDATRIALEASLAALEIILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQKISAKNQKALAELAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
33	O3_HF_OG_07	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEELKKVLETVRTLCELAKKSDDPDTARLIAAVVIKLALRAVKEDPSTDALRAVLEAVRCACEVAKR	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEELKKVLETVRTLCELAKKSDDPDTARLIAAVVIKLALRAVKED

		VTDPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAELVIRLAKEAVKEDLSEEASRAAEEAARCAREVAKRVTDPEKSKRLKKLVLDLLLALLEAALRLLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAKVLKANAEAV	PSTDALRAVLEAVRCACEVAKRVTDPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAELVIRLAKEAVKEDLSEEASRAAEEAARCAREVAKRVTDPEKSKRLKKLVLDLLLALLEAALRLLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAKVLKANAEAV
34	O3_HF_OG_08	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDKDIEEAVKWLAEWAKRATDKETVIEIVKWLAELAKKSTDSELVNEIVKQLAEVAKEATDRELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDRSLVARIALILADLAKQSTDKELRKEIEKQMEEVRKEATDEILRTLLELIRSLIRLLMILEDLKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQEISARNQAALAALA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDKDIEEAVKWLAEWAKRATDKETVIEIVKWLAELAKKSTDSELVNEIVKQLAEVAKEATDRELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDRSLVARIALILADLAKQSTDKELRKEIEKQMEEVRKEATDEILRTLLELIRSLIRLLMILEDLKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQEISARNQAALAALA
35	O3_HF_OG_09	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTKEDARSTCEKAARKAAESNDEEVAKDAAINCAKVAMEAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLIVALAAGMPTEEAARSFCEAAAKAAAEAGDARVAKIAEKACREVARQAGMPEKDADRAFKEAMKQAIEETLKRLEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAKNQAALAALA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTKEDARSTCEKAARKAAESNDEEVAKDAAINCAKVAMEAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLIVALAAGMPTEEAARSFCEAAAKAAAEAGDARVAKIAEKACREVARQAGMPEKDADRAFKEAMKQAIEETLKRLEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAKNQAALAALA
36	O3_HF_OG_10	MHHHHHHGGSEQKLISEEDLSGGGSWSGSEEVMKKAIEVIKEAKELAKELDSEEAKKVVERITEAAWAAVSAAIAGKTEVAKLALKVLEEAIELAKEKRSEEALKVVLEIARAALAAAQAAEEGFTEIAKLALKVLEDAIEMAKKKRSEEALKWVLEIAREALKAAQLAKKGDAERAKLILALLELTIALLELRLMLKRLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKINARAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSEEVMKKAIEVIKEAKELAKELDSEEAKKVVERITEAAWAAVSAAIAGKTEVAKLALKVLEEAIELAKEKRSEEALKVVLEIARAALAAAQAAEEGFTEIAKLALKVLEDAIEMAKKKRSEEALKWVLEIAREALKAAQLAKKGDAERAKLILALLELTIALLELRLMLKRLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKINARAV
37	O3_HF_OG_11	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRAVDEAIEIARWLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIMIAVVLGDPRLVEAVARAAKVAAEVIKVAIDAEKAGAREMFRRALELVREVIEAIEEAVIEGDPERVERVARKATKEALDIALKLLEMLLQRLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSMIAKVLKLNAKLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRAVDEAIEIARWLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIMIAVVLGDPRLVEAVARAAKVAAEVIKVAIDAEKAGAREMFRRALELVREVIEAIEEAVIEGDPERVERVARKATKEALDIALKLLEMLLQRLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSMIAKVLKLNAKLV
38	O3_HF_OG_12	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSEEVRERVTKLAMKAAVAKDKEEVIEIVKELAELAKQSTDSELVNKIVEALAAVAKAATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKAATDKELVEYIARILAELAKQSTDKELRKEIMKQLEEVAKEAKDKRLRLDILLVLLEIALELLREMLRELKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLKANARAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSEEVRERVTKLAMKAAVAKDKEEVIEIVKELAELAKQSTDSELVNKIVEALAAVAKAATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKAATDKELVEYIARILAELAKQSTDKELRKEIMKQLEEVAKEAKDKRLRLDILLVLLEIALELLREMLRELKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLKANARAV
39	O3_HF_OG_13	MPETFEAIARAIEVAREVEKVAQRAEEEGNPDLRDSAKELARAVDEAIEEAKKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIEEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGARDLFRLALELVRAVIEAIEFAVKLGDPEMVERAARIAKTAAELIKRAIRAKKEGDKDQEREAKKRVTRLIIELTLMVLKASLDLLRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVDNQRVSADNQKMLAELAGSWSGGGSEQKLISEDLGGS	METDTLLLWVLLLWVPGSTGDGSPETFEAIARAIEVAREVEKVAQRAEEEGNPDLRDSAKELARAVDEAIEEAKKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIEEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGARDLFRLALELVRAVIEAIEFAVKLGDPEMVERAARIAKTAAELIKRAIRAKKEGDKDQEREAKKRVTRLIIELTLMVLKASLDLLRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVDNQRVSADNQKMLAELAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
40	O3_HF_OG_14	MALRLLEKLLRELWKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLIMNAVAVAALAARVAKDKEEVIRLAKVMAELAKLADDSELVNAIVEALAEVAKVATDKELVIYLVKILAELAKQSTDSELVNEIVKQLAEVAKEATDNELVWEIVRILDELAKQSTDSELVNEIVKQLEEVAKEATDSMLVLAIKLVLENLKQQGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSALRLLEKLLRELWKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIARVLIMNAVAVAALAARVAKDKEEVIRLAKVMAELAKLADDSELVNAIVEALAEVAKVATDKELVIYLVKILAELAKQSTDSELVNEIVKQLAEVAKEATDNELVWEIVRILDELAKQSTDSELVNEIVKQLEEVAKEATDSMLVLAIKLVLENLKQQGSWSGGGSEQKLISEEDLGGSLEHHHHHH
41	O3_HF_OG_15	MALEKARRALEALKRAQEAEKKGDVEEAVRAAQEAVRAAKESGMSEILRLVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVDVAAKAAQVAAEAAKQAGDKDMLEKVAKVAEQIAKAAEKEGDKKVSIDATRIALEASLAALEIILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQKISAKNQKALAELAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSALEKARRALEALKRAQEAEKKGDVEEAVRAAQEAVRAAKESGMSEILRLVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVDVAAKAAQVAAEAAKQAGDKDMLEKVAKVAEQIAKAAEKEGDKKVSIDATRIALEASLAALEIILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQKISAKNQKALAELAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
42	O3_HF_OG_16	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEELKKVLEWVRKLCEDAKKSRDPDTARLIAAVVILLALEAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAELVIRLAKEAVEENLSEEARRAAEEAERCAREVAKRVTDPEKSKRLKRLVLDLALALLKAALELLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLRLNAEAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEELKKVLEWVRKLCEDAKKSRDPDTARLIAAVVILLALEAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAELVIRLAKEAVEENLSEEARRAAEEAERCAREVAKRVTDPEKSKRLKRLVLDLALALLKAALELLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLRLNAEAV
43	O3_HF_OG_17	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDEEVWKAVIDAIELMKEARELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNKIAVIVAVLLARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVETAVRLAEELAKVAKELIERAKKTGDADLLRLAKRAIEVARRAVELAKKSRPDAERADEAYKRLKELEREIRELLRKMLTEALRKLEKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAEVLKINAALV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDEEVWKAVIDAIELMKEARELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNKIAVIVAVLLARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVETAVRLAEELAKVAKELIERAKKTGDADLLRLAKRAIEVARRAVELAKKSRPDAERADEAYKRLKELEREIRELLRKMLTEALRKLEKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAEVLKINAALV
44	O3_HF_OG_18	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSEEVNERVKQLAEKAIKARDKEEVIEIVKELAELAKQSTDSELVNEIVRALALVAIAARDKELVIYIVKILAELAKQSTDSELVNEIVKALAEVAKAADDKELVDYIVKILLELAKQSTDK	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSEEVNERVKQLAEKAIKARDKEEVIEIVKELAELAKQSTDSELVNEIVRALALVAIAARDKELVIYIVKILAELAKQSTDSELVNEIVKALAEVAKAADDKELVDYIVKILLELAKQSTDKKLVEEIIKQLEEVRKAAKDKELQERIK

		KLVEEIIKQLEEVRKAAKDKELQERIKKVIIDLTLIILELSLRDLRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA	KVIIDLTLIILELSLRDLRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELA
45	O3_HF_OG_19	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDEEVLEAVFRAEELKREAQELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNEMAVWVAVALARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVKTAAKLADELRKVARELIERAKKTGDFELLRLALRAVEVAKRAVELAKKSRPDMEEAEELKKRLDEVEREIRELLKRMLEKALKELELKALRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINALAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDEEVLEAVFRAEELKREAQELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNEMAVWVAVALARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVKTAAKLADELRKVARELIERAKKTGDFELLRLALRAVEVAKRAVELAKKSRPDMEEAEELKKRLDEVEREIRELLKRMLEKALKELELKALRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINALAV
46	T3_HF_OG_01	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIEEARQAIKEAERALREGDPRFAEMAVRIALALVRMLERLARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAKEAVRTALELVKELIQQALKTGSDEVLERAAELAKEVARVAKEVGDPRAARKADMVAKIADTLRELMESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQAALASLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIEEARQAIKEAERALREGDPRFAEMAVRIALALVRMLERLARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAKEAVRTALELVKELIQQALKTGSDEVLERAAELAKEVARVAKEVGDPRAARKADMVAKIADTLRELMESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQAALASLA
47	T3_HF_OG_02	MADELRAVAELQRLNIELARKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAKEDSKRIVTEALRRAREQIREKWEELEERAKRAETPEEALRAAEEMVKLIEELIRIAEMLQRAGLKEEAEDVLRETELIKRATELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVARLLLAIALLADTPELAKRAAELLKRLIELLKEIAKLLEEEGNEDEAEKVKEEAKELEELVRWLEEQIRGGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSADELRAVAELQRLNIELARKLLEAVARLQELNIDLVRKTSELTDEKTIREEIRKVKEESKRIVEEAEEEIRRAKEDSKRIVTEALRRAREQIREKWEELEERAKRAETPEEALRAAEEMVKLIEELIRIAEMLQRAGLKEEAEDVLRETELIKRATELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVARLLLAIALLADTPELAKRAAELLKRLIELLKEIAKLLEEEGNEDEAEKVKEEAKELEELVRWLEEQIRGGSWSGGGSEQKLISEEDLGGSLEHHHHHH
48	T3_HF_OG_03	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDERQKRRELVRILAELLAMTATDEELIKEIKKCAQLAEELASRSTNDELIKQILEVAQLAFKLAFSAEDEELIKEILKCCQLAFELASRSTNDELIKQILEVAQLAFKLAFSAEDEELIKEILKCCQLAFELASRSTNDELIKQILEVAKLAFELAATASSEELIKAILAACQLAFLAASLATNDELIKKILEDAKRAMERASKATDEKEIMKILLDAALKALEVLLRLLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRMIVIVLSAIAIVLKINALAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDERQKRRELVRILAELLAMTATDEELIKEIKKCAQLAEELASRSTNDELIKQILEVAQLAFKLAFSAEDEELIKEILKCCQLAFELASRSTNDELIKQILEVAKLAFELAATASSEELIKAILAACQLAFLAASLATNDELIKKILEDAKRAMERASKATDEKEIMKILLDAALKALEVLLRLLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRMIVIVLSAIAIVLKINALAV
49	T3_HF_OG_04	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGRKVDAAAVLMEARSQIREVTRELEEIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNTQLAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEATKVLREAEELIERVFELLKKIAENSDTPELAKRAEELIERLIELLEEIAKLLEEAGRRKEALRVLLKALELLLRLLKKSLDELERSLEELEKNPSEDALVENNRLNVKNNRIIVKVLEMIAKVLKMNAKAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGRKVDAAAVLMEARSQIREVTRELEEIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNTQLAEKVLREAEELIERVFELLKKIAENSDTPELAKRAEELIERLIELLEEIAKLLEEAGRRKEALRVLLKALELLLRLLKKSLDELERSLEELEKNPSEDALVENNRLNVKNNRIIVKVLEMIAKVLKMNAKAV
50	T3_HF_OG_05	MHHHHHHGGSEQKLISEEDLSGGGSWSGSYEDEAERKAWEVAAKVLRLVLSGTSEDEIAEEVAREISEVIRTLKESGSSYEVIAEIVARIVAAIVLALKLSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAKIVSYVMSVVLKTTLESGSSFEVIREILRRIEEEIKEALKRAGVSEEEIKRIILKVALAVLKAALDFLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLKAIVFVLELIARVLKINADLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSYEDEAERKAWEVAAKVLRLVLSGTSEDEIAEEVAREISEVIRTLKESGSSYEVIAEIVARIVAAIVLALKLSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAKIVSYVMSVVLKTTLESGSSSFEVIREILRRIEEEIKEALKRAGVSEEEIKRIILKVALAVLKAALDFLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLKAIVFVLELIARVLKINADLV
51	T3_HF_OG_06	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDEIARLCAKMLEILAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRLACSAAARQDSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELLKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDEIARLCAKMLEILAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRLACSAAARQDSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELLKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAV
52	T3_HF_OG_07	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDERQKKREDVRILAELLARTATDEELIKEIKKCAQLAEELASRSTNDELIKQILEVAELAFRLAFSASDEELIKEILKCCQLAFELASRSTNDELIKQILEVAKLAFELAATASSEELIKAILAACQLAFLAASLATNDELIKKILEDAKRAMERASKATDEKEIMKILLDAALKALEVLLRLLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRMIVIVLSAIAIVLKINALAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDERQKKREDVRILAELLARTATDEELIKEIKKCAQLAEELASRSTNDELIKQILEVAELAFRLAFSASDEELIKEILKCCQLAFELASRSTNDELIKQILEVAKLAFELAATASSEELIKAILAACQLAFLAASLATNDELIKKILEDAKRAMERASKATDEKEIMKILLDAALKALEVLLRLLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRMIVIVLSAIAIVLKINALAV
53	T3_HF_OG_08	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDLERALAFFVLEEKKKGRNPEEAKKEAKKLKKKLKKSAGSSDLLTALAKFVLEEVRKGRNPEEAVKEAIKLAEKLKRSAGSSSLLEALAKFVLDAVRKGRNPEEAVKRAIELAEALKRAADSSELEEKLADRVLELVLKGMDAAKAVLDAALKALEKLLRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVKVLEAIAAVLRVNALAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDLERALAFFVLEEKKKGRNPEEAKKEAKKLKKKLKKSAGSSDLLTALAKFVLEEVRKGRNPEEAVKEAIKLAEKLKRSAGSSSLLEALAKFVLDAVRKGRNPEEAVKRAIELAEALKRAADSSELEEKLADRVLELVLKGMDAAKAVLDAALKALEKLLRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVKVLEAIAAVLRVNALAV
54	T3_HF_OG_09	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDELARAVAEMLRILAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVAEELRLAAEAAASQDSEQARAVYAAARAVLLALEALEAAKRVGDEEARRIAEELLRKAEEAARKQDPELARAVAEKAELLKALMELAKALRELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKLNAKAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDELARAVAEMLRILAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVAEELRLAAEAAASQDSEQARAVYAAARAVLLALEALEAAKRVGDEEARRIAEELLRKAEEAARKQDPELARAVAEKAELLKALMELAKALRELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKLNAKAV
55	T3_HF_OG_10	MGEEALLWLLDLLKKALEELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLEIIAKVLMMNAAILAVKADDKNVREIVRDAAELAARSTNSEVQRLAIRAAILAALSTDSTVLLIVKLALELAKQSTNEEVIKLALKAAVLAAESTDLEVLATVLLALARATKSTDEEEIKEELRKAVEEAEGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSGEEALLWLLDLLKKALEELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLEIIAKVLMMNAAILAVKADDKNVREIVRDAAELAARSTNSEVQRLAIRAAILAALSTDSTVLLIVKLALELAKQSTNEEVIKLALKAAVLAAESTDLEVLATVLLALARATKSTDEEEIKEELRKAVEEAEGSWSGGGSEQKLISEEDLGGSLEHHHHHH
56	T3_HF_OG_11	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIEEARQSIKEAERSLREGNPEKALKAVARALALVAELEELARKTGSTEVLIEAARLAIEVARVAL	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIEEARQSIKEAERSLREGNPEKALKAVARALALVAELEELA

		KVGSPAMAQLAVELALRLVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAARTALELVEELERQARKTGSEEVLERAARLAEEVARVAEEIGDPELARKAMKVAIRLTEELLKKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISADNQRALARLA	RKTGSTEVLIEAARLAIEVARVALKVGSPAMAQLAVELALRLVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAARTALELVEELERQARKTGSEEVLERAARLAEEVARVAEEIGDPELARKAMKVAIRLTEELLKKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISADNQRALARLA
57	I3_HF_ND_01	MHHHHHHGGSEQKLISEEDLSGGGSWSGSAEAAELAMRAATEALELCKRSTDEELCKELFKLAARLMELAERYPDSEAAKLALKAALEAIELCKQSTDEELCEKLVRLAQALIELAKRFPDAEAAKLALKAALLAIRLCKDADAEDAAEALVAMALKAIELAKRRPRDEQAKEELRILAKLIKLLEELARALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIARVLMANAAAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSAEAAELAMRAATEALELCKRSTDEELCKELFKLAARLMELAERYPDSEAAKLALKAALEAIELCKQSTDEELCEKLVRLAQALIELAKRFPDAEAAKLALKAALLAIRLCKDADAEDAAEALVAMALKAIELAKRRPRDEQAKEELRILAKLIKLLEELARALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIARVLMANAAAV
58	I3_HF_ND_02	MDECERLETEVMKAAKELMKLATQSTDKEVRKIAWEVADQLLRLAEEACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQEAVKLAEQAKDERVIEVALEMARLLIELAQEACRRNDEEALRRASEIVKRVQELIKEAEKATDEEEIERLLRKAAIDITLAQLEISLKELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKMLAELAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDECERLETEVMKAAKELMKLATQSTDKEVRKIAWEVADQLLRLAEEACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQEAVKLAEQAKDERVIEVALEMARLLIELAQEACRRNDEEALRRASEIVKRVQELIKEAEKATDEEEIERLLRKAAIDITLAQLEISLKELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKMLAELAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
59	I3_HF_ND_03	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERALRAAIEALEQADREGDEDAKRCAIELIRQAWEAWEKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRLACEAARKKNSELARAVYEAARAVLAALEALELAKRAGDEDARRCAEELLRRACEAARKNDPELARAVYEKARELAKRLRLKEALRELELKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKLNAKLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERALRAAIEALEQADREGDEDAKRCAIELIRQAWEAWEKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRLACEAARKKNSELARAVYEAARAVLAALEALELAKRAGDEDARRCAEELLRRACEAARKNDPELARAVYEKARELAKRLRLKEALRELELKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKLNAKLV
60	I3_HF_ND_04	MIMSKIAETAKRLADSLRELRRILEELKEMLERLEKRPDKKVIVDVLKVIVKAIEASVENQRISASNQAALALAIAAEAVKEIEEDIDRARKLKDEGNKEEAEKVLRKAREKIREVRDALDAIAKGAGTPDIALKAAELLVRLIKLLIEIAKLLQDAGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATKMIIRVAQLLVKIAKNSDEPELAKRAAELLKRLIELLKEIAKLLEEEGNEDEAEKVKEIAKILEEAVRELEERIIGGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSIMSKIAETAKRLADSLRELRRILEELKEMLERLEKRPDKKVIVDVLKVIVKAIEASVENQRISASNQAALALAIAAEAVKEIEEDIDRARKLKDEGNKEEAEKVLRKAREKIREVRDALDAIAKGAGTPDIALKAAELLVRLIKLLIEIAKLLQDAGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATKMIIRVAQLLVKIAKNSDEPELAKRAAELLKRLIELLKEIAKLLEEEGNEDEAEKVKEIAKILEEAVRELEERIIGGSWSGGGSEQKLISEEDLGGSLEHHHHH
61	I3_HF_ND_05	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDEKFAEAIAKAVAEMVEDLKRSGTSEDEIAEIVAKFISLIIKFLKEAGSSYEVIAEIVARIVAEIVEALKRSGTNEDEIAKIVARVISEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAKIAAKVIRVVMETLRESGSSEEVIKEIVTRILEEIREALKRAGVSETEQREIETKIRRVVKAAETMLRLEKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKVLLDLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDEKFAEAIAKAVAEMVEDLKRSGTSEDEIAEIVAKFISLIIKFLKEAGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAKIVARVISEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAKIAAKVIRVVMETLRESGSSEEVIKEIVTRILEEIREALKRAGVSETEQREIETKIRRVVKAAETMLRLEKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKVLLDLA
62	I3_HF_ND_06	MDEELCELLAKTVAKIVELLKRKGVSEDEIAEIVAQIISAIIELLKRMGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEEEIAEIVARVLREVMRTLWESGSSAEVIYECLKRALEEIREALKRAGVSEDEIRRIKEYVDRVALRAALEILTEALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLDVIAMVLKLNADLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDEELCELLAKTVAKIVELLKRKGVSEDEIAEIVAQIISAIIELLKRMGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEEEIAEIVARVLREVMRTLWESGSSAEVIYECLKRALEEIREALKRAGVSEDEIRRIKEYVDRVALRAALEILTEALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLDVIAMVLKLNADLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
63	I3_HF_ND_07	MDEEICEMIAKLVALLVEALKRAGVSEDEIAEIVAQQISEVIRWLKEWGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEEEIAEIVARVLREVMRTLWESGSSAEVIYECLKRALEEIREALKRAGVSEDEIRRIKEYVDRVALRAALEILTEALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLDVIAMVLKLNADLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDEEICEMIAKLVALLVEALKRAGVSEDEIAEIVAQQISEVIRWLKEWGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEEEIAEIVARVLREVMRTLWESGSSAEVIYECLKRALEEIREALKRAGVSEDEIRRIKEYVDRVALRAALEILTEALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLDVIAMVLKLNADLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
64	I3_HF_ND_08	MAELAIEMARQSIREAERSLLEGNPEKAREDVRRALELVRLLEKIARREGSTEVLIEAARLAIEVARVALWVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFEVGSPETAREAARTALELVEELDRQAEKTGSKEVLERAARLAKEVARVAKEIGDPELARKADEVAERLDIKRTLLDLEDSLRELRRILEELKRALEMLEKLPDKEMIRDVLKVIVKAIEASVENQRISAENQKALARLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSAELAIEMARQSIREAERSLLEGNPEKAREDVRRALELVRLLEKIARREGSTEVLIEAARLAIEVARVALWVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFEVGSPETAREAARTALELVEELDRQAEKTGSKEVLERAARLAKEVARVAKEIGDPELARKADEVAERLDIKRTLLDLEDSLRELRRILEELKRALEMLEKLPDKEMIRDVLKVIVKAIEASVENQRISAENQKALARLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
65	I3_HF_ND_09	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVTERVAELAAEAVRATDKEEVIEIVKELAELAKQSTDSELVNFIVRALAAVAIAAQDKELVIYIVKILAELAKQSTDSELVNEIVKALAEVAKAATDKELVKYIVDILLELAKQADDATLVAKIAEQLAEVREEAKDKELQERIDRVLKKLIEITLRALEESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVRNQQISAANQKALALLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSKEVTERVAELAAEAVRATDKEEVIEIVKELAELAKQSTDSELVNFIVRALAAVAIAAQDKELVIYIVKILAELAKQSTDSELVNEIVKALAEVAKAATDKELVKYIVDILLELAKQADDATLVAKIAEQLAEVREEAKDKELQERIDRVLKKLIEITLRALEESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVRNQQISAANQKALALLA
67	I3_HF_ND_11	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIAKEISRIAEESKKAIETLARLADKMTDENQVDTAIELIAKIAREAIKRIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLAELHRTDTFMAKAIEAIAELAKEAIKAIADLAKKHTTEEFMARAISAIAELAKKAIEAIWRLASLHKTDEFMDKAAEAIAELAEEAIRAIRELAKKHTTEEFVRKAESAVREISKKAKDAIRKLADAMRDPTAREKAKKLEIKVELAEALAELAVALLKLKLSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLDINAQLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIAKEISRIAEESKKAIETLARLADKMTDENQVDTAIELIAKIAREAIKRIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLAELHRTDTFMAKAIEAIAELAKEAIKAIADLAKKHTTEEFVRKAESAVREISKKAKDAIRKLADAMRDPTAREKAKKLEIKVELAEALAELAVALLKLKLS

			LDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLDINAQLV
69	O3_HF_ND_01	MSSEEAERIARILEEVWSPDPENIREAVRKAEELLRENPSRQAEELLREAIEAAVRAPDPEAIREAVRAAEELLRENPSTEAEELLRRAIEAAVRAPDPEAIREAVRAARELFKENPSEEAEELLKRAADSALKAPDPRAIREALEALLELLEAALRRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVKVLEIIARVLKANARLVGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSSSEEAERIARILEEVWSPDPENIREAVRKAEELLRENPSRQAEELLREAIEAAVRAPDPEAIREAVRAAEELLRENPSTEAEELLRRAIEAAVRAPDPEAIREAVRAARELFKENPSEEAEELLKRAADSALKAPDPRAIREALEALLELLEAALRRLKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVKVLEIIARVLKANARLVGSWSGGGSEQKLISEEDLGGSLEHHHHHH
70	O3_HF_ND_02	MDESVDLAVKLAEALRKEAEELIKKARKTGDPELLRKALEALEKAVKLVEDAIKRNPDNDEAVETAVRLARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVKTATAVELAKELEKVARELLERARKTGDDELLKLAKRALEVARRAVELALKSRPDAEEARRVYIRLTEMELEISLTELRKILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDESVDLAVKLAEALRKEAEELIKKARKTGDPELLRKALEALEKAVKLVEDAIKRNPDNDEAVETAVRLARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNDEAVKTAVELAKELEKVARELLERARKTGDDELLKLAKRALEVARRAVELALKSRPDAEEARRVYIRLTEMELEISLTELRKILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAELAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
71	O3_HF_ND_03	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDRELKQLAEVLNEIQRLAEEARKLMTDEEEAKKIQEEAERAKTMLARAVWAVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRLAEEAMKLMRNEEEAAKIAKKALEAIKALAEAVELMKDKEDIETLLRKVKELIRKAEEEARKMSDREEAAKIAERALELIKKLAKIAATLARLKRSLEELRRILEELKEMLERLERNPDKAVIVRVLKVIVKAIEASVENQRISAENQKALAELA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDRELKQLAEVLNEIQRLAEEARKLMTDEEEAKKIQEEAERAKTMLARAVWAVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRLAEEAMKLMRNEEEAAKIAKKALEAIKALAEAVELMKDKEDIETLLRKVKELIRKAEEEARKMSDREEAAKIAERALELIKKLAKIAATLARLKRSLEELRRILEELKEMLERLERNPDKAVIVRVLKVIVKAIEASVENQRISAENQKALAELA
72	O3_HF_ND_04	MCDALARAASYIIRWVIKENPEYSEKVARIASVIVKAIIEGYPNGCDCVAKAASSIIRAVIEKNPNYDKVVSDVASAIIRAIAEGLPGGADCISKAAESIRRAVEEKMPNDSEAEESVTRAIIKATEEALKRSLEELKRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVRNQALSAANQEMLARLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSCDALARAASYIIRWVIKENPEYSEKVARIASVIVKAIIEGYPNGCDCVAKAASSIIRAVIEKNPNYDKVVSDVASAIIRAIAEGLPGGADCISKAAESIRRAVEEKMPNDSEAEESVTRAIIKATEEALKRSLEELKRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVRNQALSAANQEMLARLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
74	O3_HF_ND_06	MHHHHHHGGSEQKLISEEDLSGGGSWSGSGRELKQLAEVLEEIQRLAEEARKLMTDEEEAKKIQEEAERAKRMLASAVWAVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRAAEEAMKLMRNEEEAAKIAKKALEAIKALAEAVEAIKNKEEIERLLTLVKELIRKAEEEARRMSDREKAAEIIERALEKIKKLAKLAKALADLEKALRELKKSLDELERSLEELERNPSEAALVENNRLNVENNKIIVEVLRIIAEVLKINAELA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSGRELKQLAEVLEEIQRLAEEARKLMTDEEEAKKIQEEAERAKRMLASAVWAVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRAAEEAMKLMRNEEEAAKIAKKALEAIKALAEAVEAIKNKEEIERLLTLVKELIRKAEEEARRMSDREKAAEIIERALEKIKKLAKLAKALADLEKALRELKKSLDELERSLEELERNPSEAALVENNRLNVENNKIIVEVLRIIAEVLKINAELA
77	T3_HF_ND_02	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGDRIRATAVLMEARAQIEEVTRELEEIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGQTQSAEDVLRQATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEATKVLREAEELIERVFELLKKIAENSDTPELAKRAEELIERLIELLEEIAKLLEEAGRRKEALRVLLKALELLLRLLKKSLDELERSLEELEKNPSEDALVENNRLNVKNNRIIVKVLEMIAKVLKMNAKAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGDRIRATAVLMEARAQIEEVTRELEEIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGQTQSAEDVLRQATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEATKVLREAEELIERVFELLKKIAENSDTPELAKRAEELIERLIELLEEIAKLLEEAGRRKEALRVLLKALELLLRLLKKSLDELERSLEELEKNPSEDALVENNRLNVKNNRIIVKVLEMIAKVLKMNAKAV
78	T3_HF_ND_03	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDERARAVAEMLEIAAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVAELLLRAAELLAALMELRKALEELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLELNMKAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDERARAVAEMLEIAAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVAELLLRLAASAAAEQDSEKAKAVYEAARAVLAALRALELAKRLGDEEARRIAEELLRKAEEAARKNDPELARAVRLRAELLAALMELRKALEELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLELNMKAV
81	T3_HF_ND_06	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIEEARQSIKEAERSLREGNPEKALDAVMRALSLVLELERLARKTGSTEVLIEAARLAIEVARVALKVGSPSAADLAVKVALRLVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAARTALELVEELERQARKTGSEEVLERAARLAEEVARVAEEIGDPELARKAMKVAIRLTEELLKKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISADNQRALARLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEKIEEARQSIKEAERSLREGNPEKALDAVMRALSLVLELERLARKTGSTEVLIEAARLAIEVARVALKVGSPSAADLAVKVALRLVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAARTALELVEELERQARKTGSEEVLERAARLAEEVARVAEEIGDPELARKAMKVAIRLTEELLKKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISADNQRALARLA
82	T3_HF_ND_07	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDKDAEAVAELLRLLAEKARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVAEELLRLAAEAAASQDSELAKAVYAAARAVLAALKALEAAKRVGDEEARREAEELLRKAEEAARKNDPELARAVFERAELLRALMELAKALRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINAKLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDKDAEAVAELLRLLAEKARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVAEELLRLAAEAAASQDSELAKAVYAAARAVLAALKALEAAKRVGDEEARREAEELLRKAEEAARKNDPELARAVFERAELLRALMELAKALRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINAKLA
83	I3_HF_DG_01	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDEDFAEAIAKAVARMVETLKRSGTSEDEIAEIVAKFISLIIKFLKEAGSSYEVIAEIVARIVAEIVEALKRSGTNEDEIAKIVARVISEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAKIAAKVIRVVMETLRESGSSEEVIKEIVTRILEEIREALKRAGVSETEQREIETKIRRVVKAAETMLRLEKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKVLLDLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDEDFAEAIAKAVARMVETLKRSGTSEDEIAEIVAKFISLIIKFLKEAGSSYEVIAEIVARIVAEIVEALKRSGTNEDEIAKIVARVISEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAKIAAKVIRVVMETLRESGSSEEVIKEIVTRILEEIREALKRAGVSETEQREIETKIRRVVKAAETMLRLEKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKVLLDLA
84	I3_HF_DG_02	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPRLLLEEAKERVETIERLIFTAKALQAKGNKEEAEKVLREAREQIREVTLILELIAKDSDTPELATRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREAEELIKRVLELLEKIAENADTEELARRAAELIKRLIELLKEIAKLLEEAGKKDEAEKVKEKA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPRLLLEEAKERVETIERLIFTAKALQAKGNKEEAEKVLREAREQIREVTLILELIAKDSDTPELATRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREAEELIKRVLELLEKIAENADTEELARRAAELIKRLIEL

		KEMKERVDILETLIELERSLRELRRILEELKEMLERLERNPDKDVIVKVLK VIVKAIEASVENQRISAENQKALARLA	LKEIAKLLEEAGKKDEAEKVKEKAKEMKERVDILETLIELERSLRELRRIL EELKEMLERLERNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALARL A
85	I3_HF_DG_03	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTRQKEQLDEVLKEIWRLA TEAMKLMTDWKEAIKIAKEALEAMEMLKRAVEKVTDNEVIEKLLEVVKEI IRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEV VKEIIRAAEEAMKLMRDEEEAAKIAKKALEAIKALAEAVELIKDKKQIETLL ELVKLLIETAELEARKMSDREKAARIIERALEKIRDLAKIASTLAELKKSLD ELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQ RMLAKLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGS WSGSTRQKEQLDEVLKEIWRLATEAMKLMTDWKEAIKIAKEALEAMEM LKRAVEKVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALE AIKMLARAVEEVTDNEVIEKLLEVVKEIIRAAEEAMKLMRDEEEAAKIAK KALEAIKALAEAVELIKDKKQIETLLELVKLLIETAELEARKMSDREKAARI IERALEKIRDLAKIASTLAELKKSLDELRRILEELKEMLERLEKNPDKDVIV KVLKVIVKAIEASVENQRISAENQRMLAKLA
86	I3_HF_DG_04	MDEEECEMLAKLVALLVEALKRAGVSEDEIAKIVAEFISEVIRWLKEKGS SYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSY EVICECVARIVAEIVEALKRSGTSEEEIAEIVARVLREVMRTLWESGSSA EVIYECLKRALEEIREALKRAGVSEDEIRRIKEYVDRVALRAALEILTEAL KELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLDVIAMV LKLNADLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDEEECEMLAKLVALLVEALKRAGVS EDEIAKIVAEFISEVIRWLKEKGSSYEVICECVARIVAEIVEALKRSGTSE DEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEEE IAEIVARVLREVMRTLWESGSSAEVIYECLKRALEEIREALKRAGVSEDE IRRIKEYVDRVALRAALEILTEALKELKKSLDELERSLEELEKNPSEDALV ENNRLNVENNKIIVEVLDVIAMVLKLNADLAGSWSGGGSEQKLISEEDL GGSLEHHHHHH
87	I3_HF_DG_05	MTEVEKKAREVAKEAVELASLLRSEEAIKVAQAILEAAEAAKRAAEQGK TEVAKLALKVLEEAIELAKEDRSEEALKVVLEIARAALAAAQAAEEGFTD VAKMALEVLERAIELAKDERSEEALKEVLEIARAALAAAQLAKKGRDDE ARKILMKLRIRITLRKLEESLRELRRILEELKEMLERLEKNPDKDVIVKVL KVIVKAIEASVENQRISAENQKALAELAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSTEVEKKAREVAKEAVELASLLRSEE AIKVAQAILEAAEAAKRAAEQGKTEVAKLALKVLEEAIELAKEDRSEEAL KVVLEIARAALAAAQAAEEGFTDVAKMALEVLERAIELAKDERSEEALK EVLEIARAALAAAQLAKKGRDDEARKILMKLRIRITLRKLEESLRELRRIL EELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQKALAEL AGSWSGGGSEQKLISEEDLGGSLEHHHHHH
89	I3_HF_DG_07	MDECKRLREEVMKAAKELMKLASQSTDDEVREIAMRVAEQLWRLAEE ACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRL AQEACRSNDDECLRLASEVVKAVQELVKLAEQAKDEDVIRVALKVAEA LIRLAQKACRLNDEKALREASEVVKFAQEAVKEAEKATDEEEIKKLLKK VLLEIALADLSIALEELKKSLDELERSLEELEKNPSEDALVENNRLNVEN NKIIVEVLRIIAEVLKINAEAVGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSDECKRLREEVMKAAKELMKLASQS TDDEVREIAMRVAEQLWRLAEEACRSNSDECLRLASEVVKAVQELVKL AEQATDEEVIRVALEVARELIRLAQEACRSNDDECLRLASEVVKAVQEL VKLAEQAKDEDVIRVALKVAEALIRLAQKACRLNDEKALREASEVVKFA QEAVKEAEKATDEEEIKKLLKKVLLEIALADLSIALEELKKSLDELERSLE ELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINAEAVGSWSGGG SEQKLISEEDLGGSLEHHHHHH
90	I3_HF_DG_08	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSEEEYKRLVEVAIRAEA AREGNTDEVREQLQRALEIARESGTTEAVKAALEVVANVAIEAARRGN TRAVKEALEVALEIARESGTTEAVKLALEVVARVAKDAAERGNDDAVK EALMVAIEIAKESGTEEALELAKEVIKRVAIVAEAAGAERAKKAAELARA LAELLRALIELKRSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEV LRIIAEVLMINAKLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGS WSGSSEEEYKRLVEVAIRAEAAREGNTDEVREQLQRALEIARESGTT EAVKAALEVVANVAIEAARRGNTRAVKEALEVALEIARESGTTEAVKLA LEVVARVAKDAAERGNDDAVKEALMVAIEIAKESGTEEALELAKEVIKR VAIVAEAAGAERAKKAAELARALAELLRALIELKRSLDELERSLEELEKN PSEDALVENNRLNVENNKIIVEVLRIIAEVLMINAKLV
91	I3_HF_DG_09	MSYDERARKAVKRYVKEEGGSEEEAEREAEKVREEIRKKASDKYLIQA AAAVVAYVIELGGSPDEAVKLAEAVVRAIKAAADDSYLEQAAAAVVAFV IRKGGSPMEAVIKAKEVVDRIKEAADSREATRKAARMVATVIQAGGSPE EAVKDAKKLVDLLRALRELEKALRELKKSLDELERSLEELEKNPSEDAL VENNRLNVENNKIIVEVLRIIAEVLKINAELVGSWSGGGSEQKLISEEDL GGS	METDTLLLWVLLLWVPGSTGDGSSYDERARKAVKRYVKEEGGSEEEA EREAEKVREEIRKKASDKYLIQAAAAVVAYVIELGGSPDEAVKLAEAVV RAIKAAADDSYLEQAAAAVVAFVIRKGGSPMEAVIKAKEVVDRIKEAAD SREATRKAARMVATVIQAGGSPEEAVKDAKKLVDLLRALRELEKALRE LKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKI NAELVGSWSGGGSEQKLISEEDLGGSLEHHHHHH
92	O3_HF_DG_01	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVE KVAQRAEEEGNPDLRDSAKELRKAVELAILIAKMLGNPELVEWVARAA KVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAINIAVVLGDPRLVEAV ARAAKVAAEVIKVAILAEKMGAREMFRKALELVRKVIEAIETAVIEGDPE KVERVAREATKEALKILLWLLEKLLRELKKSLDELERSLEELEKNPSEDA LVENNRLNVENNKIIVKVLEMIARVLKMNAKAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGS WSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRKAV ELAILIAKMLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALE LVRAVIEAINIAVVLGDPRLVEAVARAAKVAAEVIKVAILAEKMGAREMF RKALELVRKVIEAIETAVIEGDPEKVERVAREATKEALKILLWLLEKLLRE LKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVKVLEMIARVLK MNAKAV
93	O3_HF_DG_02	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPELEEWIRRAKEVAKEVE KVAQRAEEEGNPDLRDSAKELRKAVELAIEIARWLGNPELVEWVARAA KVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAINIAVVLGDPRLVEAV ARAAKVAAEVIKVAIDAEKAGAREMFRRALELVREVIEAIEEAVIEGDPE RVERVARKATKEALDIALKLLEMLLQRLREMLRKLKESLEELKKNPSED ALVRNNELIVEVLRVIVEVLSMIAKVLKLNAKLV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGS WSGSPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRKAV ELAIEIARWLGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAAL ELVRAVIEAINIAVVLGDPRLVEAVARAAKVAAEVIKVAIDAEKAGAREM FRRALELVREVIEAIEEAVIEGDPERVERVARKATKEALDIALKLLEMLLQ RLREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSMIAKVL KLNAKLV
94	O3_HF_DG_03	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDEEVVDAVIRALRLMLEA RELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNELAVIVAVRL ARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDND KAVETAVRLAEELAKVAKELIERAKKTGDADLLRLAKRAIEVARRAVELA KKSRPDAERADEAYKRLKELEREIRELLRKMLTEALRKLEKALQELREM LRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAEVLKINAALV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGS WSGSDEEVVDAVIRALRLMLEARELIKKARKTGDPELLRKALEALEEAV RAVEEAIKRNPDNELAVIVAVRLARELKKVAEELQERAKKTGDPELLKL ALRALEVAVRAVELAIKSNPDNDKAVETAVRLAEELAKVAKELIERAKKT GDADLLRLAKRAIEVARRAVELAKKSRPDAERADEAYKRLKELEREIRE LLRKMLTEALRKLEKALQELREMLRKLKESLEELKKNPSEDALVRNNEL IVEVLRVIVEVLSIIAEVLKINAALV
95	O3_HF_DG_04	MALEKDRRALEALRRAQEAEKKGDVEEAVRAAQEAVRAAKESGASWI LRLVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDNDVLRKVA EQALRIAKEAEKQGNVDVAAKAAQVAAEAAKQAGDKDMLEKVAKVAE QIAKAAEKEGDKKVSIDATRIALEASLAALEIILEELKEMLERLEKNPDKD VIVKVLKVIVKAIEASVKNQKISAKNQKALAELAGSWSGGGSEQKLISEE DLGGS	METDTLLLWVLLLWVPGSTGDGSALEKDRRALEALRRAQEAEKKGDV EEAVRAAQEAVRAAKESGASWILRLVAEQALRIAKEAEKQGNVEVAVK AARVAVEAAKQAGDNDVLRKVAEQALRIAKEAEKQGNVDVAAKAAQV AAEAAKQAGDKDMLEKVAKVAEQIAKAAEKEGDKKVSIDATRIALEASL AALEIILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVKNQKISAKNQ KALAELAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
96	O3_HF_DG_05	MHHHHHHGGSEQKLISEEDLSGGGSWSGSGRELKQLAEVLEEIQRLA EEARKLMTDEEEAKKIQEEAERAKRMLASAVWAVTDNEVIEKLLEVVK EIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLE VVKEIIRAAEEAMKLMRNEEEAAKIAKKALEAIKALAEAVEAIKNKEEIER	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGS WSGSGRELKQLAEVLEEIQRLAEEARKLMTDEEEAKKIQEEAERAKRM LASAVWAVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALE AIKMLARAVEEVTDNEVIEKLLEVVKEIIRAAEEAMKLMRNEEEAAKIAK

		LLTLVKELIRKAEEEARRMSDREKAAEIIERALEKIKKLAKLAKALADLEKALRELKKSLDELERSLEELERNPSEAALVENNRLNVENNKIIVEVLRIIAEVLKINAELA	KALEAIKALAEAVEAIKNKEEIERLLTLVKELIRKAEEEARRMSDREKAAEIIERALEKIKKLAKLAKALADLEKALRELKKSLDELERSLEELERNPSEAALVENNRLNVENNKIIVEVLRIIAEVLKINAELA
97	O3_HF_DG_06	MHHHHHHGGSEQKLISEEDLSGGGSWSGSDRELKQLAEVLNEIQRLAEEARKLMTDEEEAKKIQEEAERAKRMLASAVWAVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRLAEEAMKLMRNEEEAAKIAKKALEAIKALAEAVELMKDKEDIETLLRKVKELIRKAEEEARKMSDREEAAKIAERALELIKKLAKIAATLARLKRSLEELRRILEELKEMLERLERNPDKATIVRVLKVIVKAIEASVENQRISAENQKALAELA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSDRELKQLAEVLNEIQRLAEEARKLMTDEEEAKKIQEEAERAKRMLASAVWAVTDNEVIEKLLEVVKEIIRLAEEAMKLMRNEEEAAKIAKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRLAEEAMKLMRNEEEAAKIAKKALEAIKALAEAVELMKDKEDIETLLRKVKELIRKAEEEARKMSDREEAAKIAERALELIKKLAKIAATLARLKRSLEELRRILEELKEMLERLERNPDKATIVRVLKVIVKAIEASVENQRISAENQKALAELA
98	O3_HF_DG_07	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEWIELARRLIKLAEELLRAGDPEAARAAVEMALKAVRTLEKLARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAVRTALELVEELERQAKKTGSKEVLKRAAELAREVARVAREVGDPEAARKAEEVARRLDLLLKLMEALRELEKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKINAKLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEEWIELARRLIKLAEELLRAGDPEAARAAVEMALKAVRTLEKLARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAVRTALELVEELERQAKKTGSKEVLKRAAELAREVARVAREVGDPEAARKAEEVARRLDLLLKLMEALRELEKALQELREMLRKLKESLEELKKNPSEDALVRNNELIVEVLRVIVEVLSIIAKVLKINAKLA
99	O3_HF_DG_08	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTKEDARSTCEKAARKAAESNDEEVAKRAAIECARVAMEAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLIVARAAGMPTEEAARSFCEAAAKAAAEAGDARVAKIAEKACREVARQAGMPEKDADRAFKEAMKQAIEETLKRLEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAKNQAALAALA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTKEDARSTCEKAARKAAESNDEEVAKRAAIECARVAMEAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLIVARAAGMPTEEAARSFCEAAAKAAAEAGDARVAKIAEKACREVARQAGMPEKDADRAFKEAMKQAIEETLKRLEDSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAKNQAALAALA
100	T3_HF_DG_01	MHHHHHHGGSEQKLISEEDLSGGGSWSGSTEETIEMARQLIKEAERALREGDPEEARMAVEMALAAVRILERQARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAKEAVRTALELVKELIQQALKTGSDEVLERAAELAKEVARVAKEVGDPRAARKADMVAKIADTLRELMESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQAALASLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSTEETIEMARQLIKEAERALREGDPEEARMAVEMALAAVRILERQARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAKEAVRTALELVKELIQQALKTGSDEVLERAAELAKEVARVAKEVGDPRAARKADMVAKIADTLRELMESLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQAALASLA
101	T3_HF_DG_02	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDEIARLCAEMLEILAEEARRKKDSEEAEAVYWAARATLAALEALEQAKREGDEDARRCAEELLRLACSAAARQDSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELLKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSDEEEAREWAERAEEAAKEALEQAKREGDEIARLCAEMLEILAEEARRKKDSEEAEAVYWAARATLAALEALEQAKREGDEDARRCAEELLRLACSAAARQDSEQARAVYEAARAVLAALRALEAAKRAGMEEARKEAEELLRRACEAARKQDPELARAVRDKAELLKALADLFKALKELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLRINARAV
102	T3_HF_DG_03	MTEEKIEEARQSIKEAERSLREGNPEKALDAVARALSLVNELERLARKTGSTEVLIEAARLAIEVARVALKVGSPEMAQLAVELALRLVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAARTALELVEELERQARKTGSEEVLERAARLAEEVARVAEEIGDPELARKAMKVAIRLTEELLKKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISADNQRALARLAGSWSGGGSEQKLISEEDLGGS	METDTLLLWVLLLWVPGSTGDGSTEEKIEEARQSIKEAERSLREGNPEKALDAVARALSLVNELERLARKTGSTEVLIEAARLAIEVARVALKVGSPEMAQLAVELALRLVQELERQARKTGSTEVLIEAARLAIEVARVAFKVGSPETAREAARTALELVEELERQARKTGSEEVLERAARLAEEVARVAEEIGDPELARKAMKVAIRLTEELLKKSLRELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISADNQRALARLAGSWSGGGSEQKLISEEDLGGSLEHHHHHH
103	T3_HF_DG_04	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSEKEKVEELAQRIREQLPDTRLALMAQALANLANALDDSEALKVVYLALRIVQQLPDTELAREALELALDAAKSTDSKALEVVKLALRIVQLLPDTEDAREALELAKEAVKSTDEEERKKVKIKLKLLEALAELEKALRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINAKLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSEKEKVEELAQRIREQLPDTRLALMAQALANLANALDDSEALKVVYLALRIVQQLPDTELAREALELALDAAKSTDSKALEVVKLALRIVQLLPDTEDAREALELAKEAVKSTDEEERKKVKIKLKLLEALAELEKALRELKKSLDELERSLEELEKNPSEDALVENNRLNVENNKIIVEVLRIIAEVLKINAKLA
104	T3_HF_DG_05	MHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGNRVDATAVLMEARSQIREVTRELEEIAKNSDTPELATRAAELLVRLIKLLIEIAKLLQEQGQTQSAEDVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEATKVLREAEELIERVFELLKKIAENSDTPELAKRAEELIERVFELLKKIAENSDTPELAKRAEELIERLIELLEEIAKLLEEAGRRKEALRVLLKALELLLRLLKKSLDELERSLEELEKNPSEDALVENNRLNVKNNRIIVKVLEMIAKVLKMNAKAV	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSPRERLEEAKERVEEIRELIDKARKLQEQGNRVDATAVLMEARSQIREVTRELEEIAKNSDTPELATRAAELLVRLIKLLIEIAKLLQEQGQTQSAEDVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEATKVLREAEELIERVFELLKKIAENSDTPELAKRAEELIERVFELLKKIAENSDTPELAKRAEELIERLIELLEEIAKLLEEAGRRKEALRVLLKALELLLRLLKKSLDELERSLEELEKNPSEDALVENNRLNVKNNRIIVKVLEMIAKVLKMNAKAV
105	T3_HF_DG_06	MHHHHHHGGSEQKLISEEDLSGGGSWSGSSEVMKKADEVIKEAKELAKELDSEEAKKVVERITEAALAAVKAAIAGKTEVAKLALKVLEEAIELAKENRDEEALKVVLEIARAALAAAQAAEEGFTEIAKEALRVLEEAIEFAKKNRDEEALKIVLDIARMALEAAQRAKKGDDERAKILLKIAETLLRLLESLLELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQRMLALLA	METDTLLLWVLLLWVPGSTGDGSHHHHHHGGSEQKLISEEDLSGGGSWSGSSEVMKKADEVIKEAKELAKELDSEEAKKVVERITEAALAAVKAAIAGKTEVAKLALKVLEEAIELAKENRDEEALKIVLDIARMALEAAQRAKKGDDERAKILLKIAETLLRLLESLLELRRILEELKEMLERLEKNPDKDVIVKVLKVIVKAIEASVENQRISAENQRMLALLA

References

1. Uhlén M, Karlsson MJ, Hober A, Svensson A-S, Scheffel J, Kotol D, et al. The human secretome. Sci Signal. 2019;12. doi:10.1126/scisignal.aaz0274
2. Farhan H, Rabouille C. Signalling to and from the secretory pathway. J Cell Sci. 2011;124: 171–180.
3. Sun Z, Brodsky JL. Protein quality control in the secretory pathway. J Cell Biol. 2019;218: 3171–3187.
4. Fisher AC, Haitjema CH, Guarino C, Çelik E, Endicott CE, Reading CA, et al. Production of secretory and extracellular N-linked glycoproteins in Escherichia coli. Appl Environ Microbiol. 2011;77: 871–881.
5. Kightlinger W, Duncker KE, Ramesh A, Thames AH, Natarajan A, Stark JC, et al. A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. Nat Commun. 2019;10: 5404.
6. Thomas G. Furin at the cutting edge: From protein traffic to embryogenesis and disease. Nat Rev Mol Cell Biol. 2002;3: 753–766.
7. Braun E, Sauter D. Furin-mediated protein processing in infectious diseases and cancer. Clin Transl Immunology. 2019;8: e1073.
8. Rapoport TA. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature. 2007;450: 663–669.
9. Janda CY, Li J, Oubridge C, Hernández H, Robinson CV, Nagai K. Recognition of a signal peptide by the signal recognition particle. Nature. 2010;465: 507–510.
10. Denks K, Vogt A, Sachelaru I, Petriman N-A, Kudva R, Koch H-G. The Sec translocon mediated protein transport in prokaryotes and eukaryotes. Mol Membr

Biol. 2014;31: 58–84.

11. Stalder D, Gershlick DC. Direct trafficking pathways from the Golgi apparatus to the plasma membrane. Semin Cell Dev Biol. 2020;107: 112–125.
12. Higy M, Junne T, Spiess M. Topogenesis of membrane proteins at the endoplasmic reticulum. Biochemistry. 2004;43: 12716–12722.
13. Braakman I, Hebert DN. Protein Folding in the Endoplasmic Reticulum. Comprehensive Biotechnology, Second Edition. 2011;1: 217–227.
14. von Heijne G. Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem. 1983;133: 17–21.
15. Hegde RS, Bernstein HD. The surprising complexity of signal sequences. Trends Biochem Sci. 2006;31: 563–571.
16. Harada Y, Li H, Li H, Lennarz WJ. Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. Proc Natl Acad Sci U S A. 2009;106: 6945–6949.
17. Ruddock LW, Molinari M. N-glycan processing in ER quality control. J Cell Sci. 2006;119: 4373–4380.
18. Walter P, Ron D. The unfolded protein response: from stress pathway to homeostatic regulation. Science. 2011;334: 1081–1086.
19. Ushioda R, Hoseki J, Nagata K. Glycosylation-independent ERAD pathway serves as a backup system under ER stress. Mol Biol Cell. 2013;24: 3155–3163.
20. Nakatsukasa K, Brodsky JL. The recognition and retrotranslocation of misfolded proteins from the endoplasmic reticulum. Traffic. 2008;9: 861–870.
21. Byun H, Gou Y, Zook A, Lozano MM, Dudley JP. ERAD and how viruses exploit it.

Front Microbiol. 2014;5: 330.

22. Hartmann E, Sommer T, Prehn S, Görlich D, Jentsch S, Rapoport TA. Evolutionary conservation of components of the protein translocation complex. Nature. 1994;367: 654–657.
23. Robson A, Collinson I. The structure of the Sec complex and the problem of protein translocation. EMBO Rep. 2006;7: 1099–1103.
24. White SH. Membrane protein insertion: the biology-physics nexus. J Gen Physiol. 2007;129: 363–369.
25. Junne T, Kocik L, Spiess M. The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. Mol Biol Cell. 2010;21: 1662–1670.
26. Tian P, Andricioaei I. Size, motion, and function of the SecY translocon revealed by molecular dynamics simulations with virtual probes. Biophys J. 2006;90: 2718–2730.
27. Zhang B, Miller TF 3rd. Long-timescale dynamics and regulation of Sec-facilitated protein translocation. Cell Rep. 2012;2: 927–937.
28. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol. 1996;3: 842–848.
29. Radzicka A, Wolfenden R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. Biochemistry. 1988;27: 1664–1670.
30. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature.

2005;433: 377–381.

31. Moon CP, Fleming KG. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. Proc Natl Acad Sci U S A. 2011;108: 10174–10177.
32. Jones DT, Taylor WR, Thornton JM. A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry. 1994;33: 3038–3049.
33. Pasquier C, Hamodrakas SJ. An hierarchical artificial neural network system for the classification of transmembrane proteins. Protein Eng. 1999;12: 631–634.
34. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305: 567–580.
35. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022. p. 2022.04.08.487609. doi:10.1101/2022.04.08.487609
36. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature. 2007;450: 1026–1030.
37. Elazar A, Weinstein J, Biran I, Fridman Y, Bibi E, Fleishman SJ. Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. Elife. 2016;5. doi:10.7554/eLife.12125
38. Elazar A, Chandler NJ, Davey AS, Weinstein JY, Nguyen JV, Trenker R, et al. De

novo-designed transmembrane domains tune engineered receptor functions. Elife. 2022;11. doi:10.7554/eLife.75660

39. Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016;537: 320–327.
40. Brunette TJ, Parmeggiani F, Huang P-S, Bhabha G, Ekiert DC, Tsutakawa SE, et al. Exploring the repeat protein universe through computational protein design. Nature. 2015;528: 580–584.
41. Brunette TJ, Bick MJ, Hansen JM, Chow CM, Kollman JM, Baker D. Modular repeat protein sculpting using rigid helical junctions. Proc Natl Acad Sci U S A. 2020;117: 8870–8875.
42. Rocklin GJ, Chidyausiku TM, Goreshnik I, Ford A, Houliston S, Lemak A, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. Science. 2017;357: 168–175.
43. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. 2011;487: 545–574.
44. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science. 2016;352: 680–687.
45. Fallas JA, Ueda G, Sheffler W, Nguyen V, McNamara DE, Sankaran B, et al. Computational design of self-assembling cyclic protein homo-oligomers. Nat Chem. 2017;9: 353–360.
46. Chevalier A, Silva D-A, Rocklin GJ, Hicks DR, Vergara R, Murapa P, et al.

Massively parallel de novo protein design for targeted therapeutics. Nature. 2017;550: 74.

47. Lu P, Min D, DiMaio F, Wei KY, Vahey MD, Boyken SE, et al. Accurate computational design of multipass transmembrane proteins. Science. 2018;359: 1042–1046.
48. Ueda G, Antanasijevic A, Fallas JA, Sheffler W, Copps J, Ellis D, et al. Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. Elife. 2020;9. doi:10.7554/eLife.57659
49. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, et al. Design of a hyperstable 60-subunit protein icosahedron. Nature. 2016;535: 136–139.
50. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. Science. 2016;353: 389–394.
51. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, et al. Accurate design of co-assembling multi-component protein nanomaterials. Nature. 2014;510: 103–108.
52. Marcandalli J, Fiala B, Ols S, Perotti M, de van der Schueren W, Snijder J, et al. Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. Cell. 2019;176: 1420–1431.e17.
53. Walls AC, Fiala B, Schäfer A, Wrenn S, Pham MN, Murphy M, et al. Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. Cell. 2020;183: 1367–1382.e17.

54. Antanasijevic A, Ueda G, Brouwer PJM, Copps J, Huang D, Allen JD, et al. Structural and functional evaluation of de novo-designed, two-component nanoparticle carriers for HIV Env trimer immunogens. PLoS Pathog. 2020;16: e1008665.
55. Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, et al. Evolution of a designed protein assembly encapsulating its own RNA genome. Nature. 2017;552: 415–420.
56. Baycin-Hizal D, Tabb DL, Chaerkady R, Chen L, Lewis NE, Nagarajan H, et al. Proteomic analysis of Chinese hamster ovary cells. J Proteome Res. 2012;11: 5265–5276.
57. Rupp O, MacDonald ML, Li S, Dhiman H, Polson S, Griep S, et al. A reference genome of the Chinese hamster based on a hybrid assembly strategy. Biotechnol Bioeng. 2018;115: 2087–2100.
58. Kuo C-C, Chiang AW, Shamie I, Samoudi M, Gutierrez JM, Lewis NE. The emerging role of systems biology for engineering protein production in CHO cells. Curr Opin Biotechnol. 2018;51: 64–69.
59. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One. 2011;6: e20161.
60. Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nat Methods. 2020;17: 665–680.
61. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial

polypeptide libraries. Nat Biotechnol. 1997;15: 553–557.

62. Hsia Y, Mout R, Sheffler W, Edman NI, Vulovic I, Park Y-J, et al. Design of multi-scale protein complexes by hierarchical building block fusion. Nat Commun. 2021;12: 2294.
63. Black SD, Mould DR. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. Anal Biochem. 1991;193: 72–82.
64. Lauer TM, Agrawal NJ, Chennamsetty N, Egodage K, Helk B, Trout BL. Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. J Pharm Sci. 2012;101: 102–115.
65. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.
66. Cao L, Coventry B, Goreshnik I, Huang B, Sheffler W, Park JS, et al. Design of protein-binding proteins from the target structure alone. Nature. 2022;605: 551–560.
67. Hessa T, Reithinger JH, von Heijne G, Kim H. Analysis of transmembrane helix integration in the endoplasmic reticulum in S. cerevisiae. J Mol Biol. 2009;386: 1222–1228.
68. Malm M, Saghaleyni R, Lundqvist M, Giudici M, Chotteau V, Field R, et al. Evolution from adherent to suspension: systems biology of HEK293 cell line development. Sci Rep. 2020;10: 18996.
69. Kanekiyo M, Wei CJ, Yassine HM, McTamney PM, Boyington JC, Whittle JRR, et al. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1

antibodies. Nature. 2013;499: 102–106.

70. Georgiev IS, Joyce MG, Chen RE, Leung K, McKee K, Druz A, et al. Two-Component Ferritin Nanoparticles for Multimerization of Diverse Trimeric Antigens. ACS Infectious Diseases. 2018;4: acsinfecdis.7b00192.
71. Kober L, Zehe C, Bode J. Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol Bioeng. 2013;110: 1164–1173.
72. Taylor SC, Posch A. The design of a quantitative western blot experiment. Biomed Res Int. 2014;2014: 361590.
73. Pillai-Kastoori L, Schutz-Geschwender AR, Harford JA. A systematic approach to quantitative Western blot analysis. Anal Biochem. 2020;593: 113608.
74. Dalvie NC, Rodriguez-Aponte SA, Hartwell BL, Tostanoski LH, Biedermann AM, Crowell LE, et al. Engineered SARS-CoV-2 receptor binding domain improves manufacturability in yeast and immunogenicity in mice. Proc Natl Acad Sci U S A. 2021;118. doi:10.1073/pnas.2106845118
75. Malhi H, Homad LJ, Wan Y-H, Poudel B, Fiala B, Borst AJ, et al. Immunization with a self-assembling nanoparticle vaccine displaying EBV gH/gL protects humanized mice against lethal viral challenge. Cell Rep Med. 2022;3: 100658.
76. White JM, Delos SE, Brecher M, Schornberg K. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit Rev Biochem Mol Biol. 2008;43: 189–219.
77. Apellániz B, Huarte N, Largo E, Nieva JL. The three lives of viral fusion peptides. Chem Phys Lipids. 2014;181: 40–55.
78. Battles MB, Más V, Olmedillas E, Cano O, Vázquez M, Rodríguez L, et al.

Structure and immunogenicity of pre-fusion-stabilized human metapneumovirus F glycoprotein. Nat Commun. 2017;8: 1528.

79. Jardine J, Julien J-P, Menis S, Ota T, Kalyuzhniy O, McGuire A, et al. Rational HIV immunogen design to target specific germline B cell receptors. Science. 2013;340: 711–716.
80. Cherf GM, Cochran JR. Applications of Yeast Surface Display for Protein Engineering. Methods Mol Biol. 2015;1319: 155–175.
81. Isomura H, Stinski MF. The human cytomegalovirus major immediate-early enhancer determines the efficiency of immediate-early gene transcription and viral replication in permissive cells at low multiplicity of infection. J Virol. 2003;77: 3602–3614.
82. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol. 1997;268: 209–225.
83. Loshbaugh AL, Kortemme T. Comparison of Rosetta flexible-backbone computational protein design methods on binding interactions. Proteins. 2020;88: 206–226.
84. Macalino SJY, Basith S, Clavio NAB, Chang H, Kang S, Choi S. Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery. Molecules. 2018;23. doi:10.3390/molecules23081963
85. Wang Y, Chen J, Kang Z. In Silico Protein Design Promotes the Rapid Evolution of Industrial Enzymes. Biochemistry. 2019;58: 1451–1453.
86. Alford RF, Koehler Leman J, Weitzner BD, Duran AM, Tilley DC, Elazar A, et al. An

Integrated Framework Advancing Membrane Protein Modeling and Design. PLoS Comput Biol. 2015;11: e1004398.

87. Koehler Leman J, Lyskov S, Bonneau R. Computing structure-based lipid accessibility of membrane proteins with mp_lipid_acc in RosettaMP. BMC Bioinformatics. 2017;18: 115.
88. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. B Cells and Antibodies. Garland Science; 2002.
89. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014;11: 801–807.
90. Duellman T, Burnett J, Yang J. Quantitation of secreted proteins using mCherry fusion constructs and a fluorescent microplate reader. Anal Biochem. 2015;473: 34–40.
91. Lombardelli L, Logiodice F, Kullolli O, Piccinni M-P. Evaluation of Secreted Cytokines by Multiplex Bead-Based Assay (X MAP Technology, Luminex). Methods Mol Biol. 2021;2285: 121–130.
92. Lu H, Yu S, Qin F, Ning W, Ma X, Tian K, et al. A secretion-based dual fluorescence assay for high-throughput screening of alcohol dehydrogenases. Biotechnol Bioeng. 2021;118: 1624–1635.
93. Praznik A, Fink T, Franko N, Lonzarić J, Benčina M, Jerala N, et al. Regulation of protein secretion through chemical regulation of endoplasmic reticulum retention signal cleavage. Nat Commun. 2022;13: 1323.
94. Yu H-B, Li M, Wang W-P, Wang X-L. High throughput screening technologies for ion channels. Acta Pharmacol Sin. 2016;37: 34–43.

95. Liu Z, Chen O, Wall JBJ, Zheng M, Zhou Y, Wang L, et al. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep. 2017;7: 2193.
96. Lohman JR, Remington SJ. Development of a family of redox-sensitive green fluorescent protein indicators for use in relatively oxidizing subcellular environments. Biochemistry. 2008;47: 8678–8688.
97. Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, et al. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc Natl Acad Sci U S A. 2012;109: E690–7.
98. Bhaskar A, Chawla M, Mehta M, Parikh P, Chandra P, Bhave D, et al. Reengineering Redox Sensitive GFP to Measure Mycothiol Redox Potential of Mycobacterium tuberculosis during Infection. PLoS Pathog. 2014;10. doi:10.1371/journal.ppat.1003902
99. Aronson DE, Costantini LM, Snapp EL. Superfolder GFP is fluorescent in oxidizing environments when targeted via the Sec translocon. Traffic. 2011;12: 543–548.
100. Feng S, Sekine S, Pessino V, Li H, Leonetti MD, Huang B. Improved split fluorescent proteins for endogenous protein labeling. Nat Commun. 2017;8: 370.