Easy Synthesis of Complex Biomolecular Assemblies: Wheat Germ Cell-Free Protein Expression in Structural Biology

Marie-Laure Fogeron¹, Lauriane Lecoq¹, Laura Cole¹, Matthias Harbers²,³* and Anja Böckmann¹*

¹Molecular Microbiology and Structural Biochemistry, Labex Ecofect, UMR 5086 CNRS/Université de Lyon, Lyon, France, ²CellFree Sciences, Yokohama, Japan, ³RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Japan

Cell-free protein synthesis (CFPS) systems are gaining more importance as universal tools for basic research, applied sciences, and product development with new technologies emerging for their application. Huge progress was made in the field of synthetic biology using CFPS to develop new proteins for technical applications and therapy. Out of the available CFPS systems, wheat germ cell-free protein synthesis (WG-CFPS) merges the highest yields with the use of a eukaryotic ribosome, making it an excellent approach for the synthesis of complex eukaryotic proteins including, for example, protein complexes and membrane proteins. Separating the translation reaction from other cellular processes, CFPS offers a flexible means to adapt translation reactions to protein needs. There is a large demand for such potent, easy-to-use, rapid protein expression systems, which are optimally serving protein requirements to drive biochemical and structural biology research. We summarize here a general workflow for a wheat germ system providing examples from the literature, as well as applications used for our own studies in structural biology. With this review, we want to highlight the tremendous potential of the rapidly evolving and highly versatile CFPS systems, making them more widely used as common tools to recombinantly prepare particularly challenging recombinant eukaryotic proteins.

Keywords: cell-free protein expression, wheat germ, structural biology, NMR, labeling

INTRODUCTION

Efficient, easy-to-use, and rapid protein expression methods for protein analysis are in great demand for structural determination, biochemical research, and applications in synthetic biology, such as the design of new biological circuits or the development of new proteins for technical applications and therapies (Gregorio et al., 2019; Silverman et al., 2020). The rapid response to the recent COVID-19 pandemic shows how the scientific community is applying the latest technologies to study viral proteins and to make them available for structural analysis (Zhu et al., 2020b) and drug testing (Dai et al., 2020; Jin et al., 2020), the development of antibodies, or creation of new serological tests to monitor infection rates. In this context, cell-free protein synthesis (CFPS) was used to make versions of the SARS-CoV-2 N-protein using wheat germ cell-free protein synthesis (WG-CFPS) for use in serological testing (Matsuba et al., 2020; Yamaoka et al., 2020) and antibody development leading to tests for COVID-19 now available on the market to serve patients. Also, a variety of accessory and structural proteins have been synthesized and purified in milligram amounts using this approach.
expression of a marker protein, which will enable new concepts in the development of CFPS systems (Altincekic et al., 2021). CFPS platforms in general are versatile tools to address such needs (Rosenblum and Cooperman, 2014), building on previous work on pathogen-related research (Matsunaga et al., 2014; Yamaoka et al., 2016). These protein expression platforms can be customized to work on individual proteins or have been scaled for high-throughput protein expression for analysis and production on large scales. Hence, CFPS (recently also called TXTL for “transcription-translation”) is getting more attention these days with the development of new methods that try to make the best use of the unique features of an in vitro method rather than relying on established systems depending on a host cell (Zemella et al., 2015) for synthesis of recombinant proteins. The high potential of new CFPS systems was demonstrated by an E. coli system that is used for protein expression on an industrial scale (Zawada et al., 2011; Salehi et al., 2016; Hershewe et al., 2020). It was suggested that such systems could be used more in the future for the production of dedicated pharma proteins, for example, incorporating noncanonical amino acids (Hong et al., 2014; Quast et al., 2015; Wu et al., 2020), preparation of dedicated proteins that were produced under more defined conditions than possible in cell-based systems (Oza et al., 2015), or allowing for the “on-demand” production of protein therapeutics in the clinic (Mohr et al., 2016; Sullivan et al., 2016; Timm et al., 2016). The diverse features of CFPS systems promoted also their recent use in teaching (Stark et al., 2019), protein engineering (Kido et al., 2020), and synthetic biology (Tinafar et al., 2019), which holds great promises for studies on genetic networks or rapid prototyping (Karim et al., 2020) in metabolic engineering (Perez et al., 2016) as well as future drug development (Dondapati et al., 2020). Moreover, the in vitro reaction format of CFPS systems allows for full automation, miniaturization (Ayoubi-Joshaghani et al., 2020), and working with large sample numbers (Zhu et al., 2015). This advantage has been utilized in large-scale screening experiments (Khnouf et al., 2016; Kim et al., 2015), searches for malaria vaccine candidates (Kanoi et al., 2017; Morita et al., 2017; Kanoi et al., 2020), identifying interactions between E3 ligases and their substrates (Takahashi et al., 2016), building a protein array holding human Deubiquitinating Enzymes (DUBs) (Takahashi et al., 2020), or the development of protein array platforms (Romanov et al., 2014; Zarate and Galbraith, 2014; Morishita et al., 2019). Other promising developments make use of the stability of the reagents, where the extracts and buffers can be lyophilized for long-term storage at room temperature (Smith et al., 2014). This enabled the development of a paper-based diagnostic assay for the detection of Ebola (Pardee et al., 2014), a concept that could be extended to the development of more sensitive rapid tests for other infectious diseases suitable for use in developing countries or testing water quality (Jung et al., 2020) with a simple assay (Grawe et al., 2019; Thavarajah et al., 2020). It is a promising approach to combine DNA detection with the expression of a marker protein, which will enable new concepts for biosensor developments (Duyen et al., 2016; Ogawa et al., 2016; Zhang et al., 2020). For such applications, the translation system could also be miniaturized or used in a fluidic array device (Jackson et al., 2015) for automation and easy use.

Whether used in high-throughput or on individual proteins, CFPS systems can be optimized in ways not possible for cell-based systems. The open nature of an in vitro reaction allows for changes to the reaction environment to mimic better individual protein needs. This was demonstrated in many studies for the most commonly used commercial or self-made CFPS systems from E. coli using customized extract preparations on a large variety of proteins for different applications (Gregorio et al., 2019; Cole et al., 2020). Another well-established system on which we will focus here is based on wheat germ extracts (Roberts and Paterson, 1973; Madin et al., 2000). Eukaryotic ribosomes from plants are better adapted for protein folding during synthesis than prokaryotic ribosomes from E. coli extracts, notably when eukaryotic proteins are targeted. Besides those established CFPS systems (Rosenblum and Cooperman, 2014; Zemella et al., 2015; Dondapati et al., 2020), new systems were developed for rapid protein expression that better match the features of cell-based systems, for instance, using extracts from HeLa (Mikami et al., 2008) or Chinese Hamster Ovary (CHO) cells (Brodel et al., 2015; Thoring et al., 2016). Other advancing systems are based on extracts from Saccharomyces cerevisiae (Gan and Jewett, 2014), Pichia pastoris (Spice et al., 2020), tobacco BY-2 cells (Buntru et al., 2015), rice (Suzuki et al., 2020), or modified E. coli strains (Seki et al., 2009; Cole et al., 2020) to name a few. Our growing understanding of translation reactions and a deeper understanding of the cell extracts led to new protocols for extracts having improved activity (Borkowski et al., 2020; Contreras-Llano et al., 2020) or been engineered for specialized applications such as working better with noncanonical amino acids (Martin et al., 2018). All those modern CFPS systems have often been optimized for high protein yields and better cost performance, thus by far exceeding the abilities of the classical rabbit reticulocyte lysate system that is still widely used in protein labeling reactions and biochemical studies. Among the eukaryotic systems, high-performance wheat germ extracts have shown the highest protein expression activity (Perez et al., 2016), leading to the wide use of this system in research and applied sciences. Since the germ is in a dormant stage, it is an extraordinarily rich source for the protein factors and the ribosomes needed for rapidly performing protein synthesis from stored mRNAs during early germination (Sano et al., 2020).

In the context of structural biology, protocols developed for the preparation of highly active wheat germ extracts lead to a universal protein expression system (Sawasaki et al., 2002b) that is used in a variety of structural approaches, such as preparing stable-isotope-labeled samples for protein NMR (Lacabanne et al., 2019), making reference standards for mass spectrometry in proteomics (Singh et al., 2009; Takanori et al., 2017), or preparing samples for cryogenic electron microscopy (cryo-EM) (Novikova et al., 2018). Notably, sample amounts needed in structural biology have significantly diminished in the last years with the development of crystallization robots for X-ray studies, high-performance detectors in cryo-EM, and higher magnetic fields.
in NMR (Dobson, 2019). Particularly, in solid-state NMR, faster magic-angle spinning (MAS) recently reduced sample needs by a spectacular factor of 100 through proton detection under MAS frequencies exceeding 100 kHz (Agarwal et al., 2014; Bockmann et al., 2015; Lecoq et al., 2018; Lecoq et al., 2019; Wang et al., 2019), a milestone that enables investigation of submilligram amounts of sample. As solid-state NMR can typically target large protein assemblies such as viral capsids (Zhang et al., 2016; Wang et al., 2017; Quin et al., 2018), envelopes (David et al., 2018), microtubules (Guo et al., 2019), or membrane proteins (Ijirasko et al., 2020) and their assemblies (Ong et al., 2013; Kaur et al., 2015; Kaur et al., 2016; Kaur et al., 2018; Kaur et al., 2019), an in vitro protein synthesis system using a high-yielding eukaryotic ribosome is a central asset for such studies. This approach can generally be used to also produce proteins of pathogens that hijack the eukaryotic host cell machinery during infections making it a powerful tool for pathogen research.

Here, we review a typical workflow for using WG-CFPS and report our experiences about recombinantly preparing protein samples in this expression system. For all our experiments, we are using a WG-CFPS that had been developed in the Endo Lab at Ehime University (Sawasaki et al., 2002b). Endo and coworkers published the detailed protocol on how to prepare highly active wheat germ extracts by completely removing the endosperm in careful washing steps; the same protocol also describes how to utilize their wheat germ extracts in translation experiments (Takai et al., 2010). This protocol allows to establish extract preparation and CFPS in any reasonably equipped biochemistry laboratory; wheat germ extracts prepared according to the same procedure are also commercially available from CellFree Sciences (Japan). Figure 1 provides information on the basic steps for conducting protein expression experiments in this WG-CFPS. These conditions allow for direct expression of proteins in high-throughput experiments or also joint expression of several proteins in a single reaction, as shown, for example, for chromatin reconstruction experiments using premixed mRNAs for up to four core histones, three chromatin assembly factors, and histone H1 (Okimune et al., 2020). The expression of eight proteins in a single reaction is an impressive achievement not possible in most cell-based systems. However, the WG-CFPS can achieve this by simply adjusting the mRNA ratios in the translation reaction. The basic reaction conditions of the WG-CFPS can be adopted in many ways for more advanced applications further outlined in this review. The aim of this review is to give practical advice on how to plan and run such experiments and to highlight the extraordinary potential of the system, with a focus on (structural) studies on viral (membrane) proteins and the analysis of their assemblies.

EXPERIMENTAL DESIGN

The workflow of an experiment in a wheat germ system (WGS) is given in Figure 2; refer to the guide by Wingfield (2015) for an overview on the purification of recombinant proteins which provides some general introduction into protein expression, particularly in E. coli. For experimental design, information on
the nature of the protein is useful to better understand its requirements for production and purification. Next, the design of the cDNA template needed to produce the recombinant protein can include further considerations on working with fusion proteins such as constructs with affinity tags for detection and purification. The expression template determines conditions for the protein expression reactions and purification steps. For proteins requiring special conditions, the results of expression tests are analyzed and then optimized in iterative cycles. This may include different additives necessary to obtain soluble and correctly folded proteins. Finally, specific tests for samples later used for structural or functional analysis must be established to make sure that the protein is suitable for the intended use. This includes particularly biological and biophysical tests assessing whether the protein is correctly folded. We discuss these steps in the following sections in more detail.

**INFORMATION ON THE PROTEIN OF INTEREST**

Before starting experiments, information about the protein of interest needs to be collected. This includes its biological role, possible binding partners, physicochemical properties, protein modifications, known or predicted structure, domains, signal peptides, disordered zones, stability, solubility, or hydrophobicity. This information enables optimal design of the expression template before a cDNA encoding the protein is cloned into a suitable expression vector or prepared by polymerase chain reaction (PCR) and is also useful for planning the expression reaction. For instance, the reaction may need to be performed at lower temperatures to change folding kinetics or decreasing hydrophobic interactions and self-aggregation. Currently, gene synthesis in combination with new cloning methods like Gibson assembly of overlapping DNA molecules (Gibson et al., 2009) offers a very flexible means to quickly prepare expression templates based on publicly available sequence information, including the results of high-speed sequencing experiments. However, one should be careful in selecting expression templates only based on assembled contigs from sequencing reads. It is best to use fully annotated protein/gene sequences if there is not a special reason for utilizing experimental sequence data. Since the protein sequence is the only needed information, gene synthesis may also be used to optimize the expression template by analysis of the RNA structure, looking for certain sequence elements, and to optimize the codon use for expression in a given host (Gustafsson et al., 2012; Athey et al., 2017) where codon optimization for expression in WGS-CFPS is offered by most gene synthesis providers. Gene synthesis may further help with preparing templates for expression of fusion proteins having an affinity tag or creating artificial designer proteins such as protein standards for mass spectrometry (Takemori et al., 2017).

For many proteins, matching cDNAs have already been prepared during large-scale cloning projects (Harbers, 2008), and those clones can be obtained from public depositories or distributors. Particularly for human genes, large-scale cDNA collections are available including the ready-to-use open reading frame clones from the Human Gene and Protein Database (Goshima et al., 2008) or the international ORFeome Collaboration (Collaboration, 2016). For example, most of the cDNA clones in those two collections have been used in a study to...
identify reference peptides for targeted proteomics on the human proteome (Matsumoto et al., 2017). Since those clones are provided in Gateway entry vectors, the cDNA inserts can be easily transferred onto other vector formats (Reece-Hoyes and Walhout, 2018). Those and other cDNA collections can be readily searched for genes of interest as a convenient way to find cDNA clones from distributors rather than requesting published materials from other researchers or starting from scratch preparing them by gene synthesis. In general, for inquiries on a given gene, the “Gene” database at NCBI (https://www.ncbi.nlm.nih.gov/gene/) is a very good starting point (Brown et al., 2015; Coordinators, 2017). This database holds information on reference sequences from RefSeqs, maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources. The information provided in Gene can be valuable to learn more about a certain gene, while the sequence information may be useful for domain analysis, using gene synthesis services, or to confirm the sequence of a cDNA clone after an ID check. Moreover, Gene provides links to worldwide resources including cDNA clone providers (go to “Gene LinkOut,” you may have to click on the + sign to see the entire list at the end of the web page). NCBI allows suppliers to link (“LinkOut”) products and services on the specified gene shown in the Gene output page to help researchers to find resources in the public domain. This service is best known for links from publishers in PubMed but can also be used in other NCBI databases (https://www.ncbi.nlm.nih.gov/projects/linkout/). Refer to the following link on NCBI for more information on how to find cDNA clones in the public domain: https://www.ncbi.nlm.nih.gov/genome/clone/finding_cdna.shtml. In addition to the Gene database, there are many other protein-focused databases, like UniProt (https://www.uniprot.org/) that offers important information on the protein and its annotation, families, domains, and isoforms. For annotated proteins, the UniProt section on “amino acid modifications” includes possible disulfide bonds (Feige and Hendershot, 2018), which are formed under oxidizing conditions and thus may require changes to the protein expression and handling as further outlined below. Disulfide bonds are important for protein folding and stability and are mostly found in extracellular, secreted, and periplasmic proteins. We describe in the next chapter protein analysis tools available in the public domain that can provide information for template design beyond the information that is already provided in UniProt.

### TEMPLATE DESIGN

Template design is the starting point for making a protein, and a careful analysis of the protein and its features helps to prepare the template. There are several tools freely available on the Internet with information on protein properties, domain structures, or folding (refer to http://molbiol-tools.ca/Protein_Chemistry.htm and Table 1 for links to some of these tools).

While many proteins can be expressed in the WGS as full-length proteins in their native form, it may also be of interest to work on isolated domains or with other protein fragments (Figure 3). For example, some protein domains can reduce translation efficiency and may be removed from the recombinant proteins such as leader peptides, if not particularly needed for working with microsomes (Brodel et al., 2015). Leader peptides can be rather hydrophobic and frequently prevent correct folding of proteins. In general, the sequences at the N-terminus of proteins can have a large impact on protein yields in recombinant expression experiments. Therefore, it can be helpful to modify the N-terminus to improve yields for poorly expressed protein, for example, using a systematic tag variation strategy in combination with CFPS (Haberstock et al., 2012); a similar effect was described for

| Table 1 | Protein analysis tools and selected databases. |
|---|---|---|
| Tool | Description | URL |
| Gene | Reference database and resources | https://www.ncbi.nlm.nih.gov/gene/ |
| UniProt | Protein sequence and functional information | http://www.uniprot.org/ |
| Wheat proteome | Reference on background protein analysis | https://www.wheatproteome.org/ |
| Protein Chemistry | Links to useful tools | http://molbiol-tools.ca/Protein_Chemistry.htm |
| ProtParam tool | Calculating physical and chemical parameters | https://web.expasy.org/protparam/ |
| Mfold | RNA folding | http://unafold.maastrichtuniversity.nl/ |
| JPred | Protein secondary structure prediction | http://www.ics.uci.edu/~jeremy/tmfinder/TM_Finder.html |
| Esprit | Alignment and secondary structure prediction | http://sable.cchmc.org |
| Proter | Visualization of proteoforms | https://sites.google.com/site/protter/start/ |
| Sable | Solvent accessibility | http://sable.cchmc.org |
| Scratch | Protein predictor on protein structures | http://scratch.proteomics.ics.uci.edu/ |
| InterPro | Protein classification and predicting domains | http://www.ebi.ac.uk/interpro/ |
| FFAS | Folding and function assignment | http://ffas.sanfordburnham.org/ffas/cgi/ffas.pl |
| CDTree | Protein domain hierarchy viewer and editor | https://www.ncbi.nlm.nih.gov/Structure/cdtree/cdtree.shtml |
| Cn3D | Macromolecular structure viewer | http://p4d-info.nig.ac.jp/vaposr/ |
| VaProS | Variation effect on Protein structure and function | http://bioinf.cs.ucl.ac.uk/vaposr/ |
| PSIPRED | Protein sequence analysis workbench | http://www.pondr.com/ |
| PONDR | Predictor of natural disordered regions | http://www.rcsb.org/pdb/home/home.do |
| Protein data bank | 3D structures of proteins | https://www.pymol.org/ |
| PyMOL | Molecular visualization system | http://tmfinder.research.sickkids.ca/cgi-bin/TMFinderForm.cgi |
| TM finder | Transmembrane region finder | |
N-terminal fusion with the GB1 domain (Michel and Wuthrich, 2012). Screening a library of 250,000 reporters led to the concept of a “short translational ramp” indicating that the amino acids in positions three and five impact protein yields (Verma et al., 2019).

Added sequences at the N- or C-terminus can have additional functions. The protein of interest may further be expressed with an affinity tag for later analysis and purification (see below). Other helpful tools such as Expasy ProtParam can help elucidate whether the primary sequence of the target protein is basic or acidic; which protein family it belongs to (e.g., UniProt; InterPro); whether it has hydrophobic stretches that need detergent or lipids to correctly fold (e.g., TM finder, for a more detailed review look at reference Punta et al., 2007); whether the protein has subdomains (e.g., Ipred, Scratch, and InterPro); whether it presents important functional motifs located directly at the N- or C-terminal, with which an affinity tag could interfere (e.g., FFAS; CDTree); which functions it could fulfill (e.g., UniProt); and what structures are predicted for folding (e.g., Proter). Corresponding websites for those tools are given in Table 1. Many of these tools are centralized on web portals, as, for example, on the Network Protein Sequence Analysis server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html). Other important questions relate to the presence of cysteines and potential disulfide bonds in the protein (e.g., annotations in UniProt). Because of the importance of disulfide bonds for protein folding, stability, or complex formation, different computational methods have been described for working with them (Sun et al., 2017; Gao et al., 2020).

However, we should point out that some cysteines do not form disulfide bonds in natively folded proteins. It can be important to prevent those cysteines from forming artificial disulfide bonds, since they can cause problems when proteins are refolding during later processing. Therefore, it may be better to mutate such cysteines to alanine or serine as done for the expression of the G protein-coupled neuropeptide Y receptor type 2, which had improved protein stability without any significant loss of functionality (Witte et al., 2013; Krug et al., 2020); refer to (Rawlings, 2018) for more information on membrane protein engineering. Another important example to note is zinc-binding motifs, which could require the use of additives like zinc ions or chaperones in the CFPS system.

For WG-CFPS, cDNA clones from various clone collections from different organisms (e.g., Arabidopsis, mouse, and human) have been used with good success, and it was commonly not necessary in those cases to do codon optimization. However, codon optimization is often used when the cDNA template was prepared by gene synthesis. This proved helpful when working on malaria-related projects, because Plasmodium falciparum uses very irregular A/T rich coding sequences (Arumugam et al., 2014). Codon optimization has been widely used, however, when expressing proteins in bacterial systems, which can have vastly different codon preferences as compared to higher organisms. It should be noted that codon usage has a direct influence on the elongation rate and thus regulates cotranslational folding (Yu et al., 2015). Theoretically, it could be useful to adjust the tRNA concentrations in a cell-free
translation reaction to mimic those from the organism from which the recombinant protein is originally derived or to redesign
the genetic code (Hibi et al., 2020). In principle, this could greatly assist correct protein folding; however, such experiments remain
difficult without a good method to prepare individual tRNAs
(Berg and Brandl, 2020) on a large scale. We expect more work in
this area because studies on cancer genomes have revealed
relevant synonymous mutations in tumors (Diederichs et al.,
2016) that could function by changing protein folding via
manipulating translation speed. Furthermore, it is interesting
to note that structural limitations of tRNAs for binding to
tRNA-modifying enzymes may have restricted the genetic code
to 20 amino acids (Saint-Leger et al., 2016) and thus defined
the chemical space for proteins. However, today, several approaches
are in use to extend the genetic code and the space of amino acids
that can be introduced into proteins (Arranz-Gibert et al., 2018).
In addition to naturally occurring proteins, the WGS has also
been successfully used to express artificial proteins such as the
preparation of stable isotope-labeled peptide libraries (Takemori
et al., 2016). Here, representative peptide sequences have been
concatenated and expressed together in artificial proteins
that were subjected to tryptic digestion to release individual peptides.
This technique is a very cost-effective way to produce many
different peptides when needed for quantitative protein mass
spectrometry and proteome analysis.

USE OF AFFINITY TAGS

When designing the expression template, considerations should
also be given to means of protein detection, purification, or
further modifications for the expression of fusion proteins. As
mentioned above, fusion tags can have different functions
including enhancing heterologous protein expression when
placed at the N-terminus (Haberstock et al., 2012; Ki and
Pack, 2020). This allowed, for instance, to increase yields of a
GPCR 5–38 times, resulting in sufficient protein amounts for
structural-functional studies (Lyukmanova et al., 2012b). Most
fusion proteins have added sequences encoding an affinity tag
that can be added at either end of the CDNA; small tags may also
be added by primer extension PCR. Ready-to-use commercial
expression vectors are available for the WGS or have been
described in the literature (Bardocz et al., 2008; Nagy et al.,
2020). While affinity tags can be especially useful in protein
purification, they also offer means for protein detection and
analysis (Kimble et al., 2013; Wood, 2014; Yadav et al., 2016).
Most affinity tags can be used in any protein expression system
and are commonly host independent, but the recently developed
AGIA-tag (Yano et al., 2016; Kido et al., 2020) and CP5-tag
(Takeda et al., 2017) systems have presently only been described
for the WG-CFPS. While epitope/antibody combinations allow for
short tags and high-affinity binding, larger tags may add undesired
functionalities to proteins of interest. It should also be noted that
antibody-based tag systems are often better for analytical purposes,
whereas they may be expensive for protein purification.

There is a preference for working with an affinity tag at the
C-terminus to make sure that only full-length proteins are
purified when translation is incomplete. Tags can be
eliminated after purification by insertion of an enzymatic
cleavage site for thrombin or Tobacco Etch Virus (TEV)
proteases (Waugh, 2011). Examples of other cleavage sites are
given in (Malhotra, 2009). Table 2 summarizes published affinity
tags that have been used in combination with the WGS.

The Histidine tag or short His-tag (Malhotra, 2009),
composed of 6 to 12 histidine residues, is very frequently used.
When fused at either the N- or C-terminal end, it allows for
a rapid, easy, and cost-effective purification on metal-chelating
resins with high binding capacity. Nickel and cobalt resins are
sensitive to reducing conditions such as those commonly used in
CFPS reactions, and hence crude reaction mixtures should be
diluted. Other commonly used purification tags like the
glutathione S-transferase (GST) tag (Malhotra, 2009) are larger
and can impart higher protein solubility when fused to the
N-terminus (Malhotra, 2009). With a length of 220 amino
acids (about 26 kDa in size), the GST-tag is large, which can
be helpful in pull-down assays, where the GST protein not only
facilitates the binding to a resin but also functions as a spacer to
better expose the fused protein used in binding assays. The main
drawback of the His-tag, and to a lesser extent of the GST-tag, is
the unspecific binding of endogenous proteins from the wheat
germ extract to metal-chelating and glutathione resins, which can
lead to significant contamination of the affinity-purified proteins
when working with proteins having low expression levels. This
limitation can be addressed when using extracts pretreated on a
nickel or glutathione resin for higher purity of His- or GST-
tagged proteins (Takai et al., 2010; Harbers, 2014). Such extracts
are commercially available (CellFree Sciences, Japan) for the His-
and GST-tags. Both tags can also be combined with other tags
where doubled-tagged proteins offer superior means to prepare
highly purified proteins. It should further be mentioned that,
against the His- and the GST-tag, commercial antibodies are
available that can be used in protein detection, which is very
handy to confirm protein expression when otherwise no
antibodies are available recognizing the target protein.

As alternative approaches, the FLAG (Einhuw and
Jungbauer, 2001) and Streptag II (Schmidt and Skerra, 2007)
tags do often better remove background contaminations than
possible for His- and GST-tagged proteins (Nagy et al., 2020).
The FLAG-tag is an octapeptide (DYKDDDDK) that is recognized by
a specific antibody which allows for sensitive protein detection.
Three combined FLAG-tags were described for working with a
WGS to achieve even higher binding affinity (Novikova et al.,
2018). While a tagged protein can be eluted by competition with
the peptide or by endonuclease cleavage, binding capacity of the
resin is low, making affinity purification quite expensive and less
attractive for large-scale routine protein production, but the
FLAG-tag is an excellent tool for binding assays to study
protein complexes. Here, the FLAG-tag has been used in
PerkinElmer AlphaScreen assays (Nemoto et al., 2018) in
combination with biotinylated proteins made in the WGS in
the presence of added biotin and the biotin ligase BirA (Matsuoka
et al., 2010). In contrast, the Streptag II (Schmidt and Skerra,
2007) allows for lower cost affinity purification, with high purity
levels reached already after a single purification step. It uses a
TABLE 2 | Summary of tags whose use has been described for protein expression with the WGS.

| Tag | Main types of application | Position | References | Advantages | Drawbacks |
|-----|--------------------------|----------|------------|------------|-----------|
| Single tag His<sub>6</sub> | Translation setup and protein detection | N-ter | (Mahotra, 2009; Vinarov et al., 2006a; Aoki et al., 2009; Revathi et al., 2010; Takahashi et al., 2013) | Cost effective, easy, and fast purification process | WGE endogenous proteins bind to the affinity support |
|  | Purification by affinity chromatography | C-ter | (Mahotra, 2009; Vinarov et al., 2006a; Aoki et al., 2009; Revathi et al., 2010) | Affinity support endless reusable | Need for EDTA-free buffers |
| GST | Translation setup at small scale | N-ter | (Los et al., 2008; Bardoczky et al., 2008; Nagy et al., 2020) | Solubility enhancement | WGE endogenous proteins bind to the affinity support |
|  | Purification by affinity chromatography | | | Low impact on protein folding when fused at the N-terminus | Higher purification cost than for His-tag |
|  | Protein detection by Western blot | | | Elution by enzymatic cleavage or competition | Degradation of GST under reductive conditions |
| HaloTag | Pull-down assays | N-ter | (Bardoczky et al., 2008; Ramadan et al., 2015; Novikova et al., 2018; Nagy et al., 2020) | Affinity support reusable up to five times | Large size of the tag (34 kDa) |
|  | Purification by affinity chromatography | | | Higher binding capacity of the affinity support | Affinity support not reusable |
|  | Protein detection by Western blot | | | Highly specific interaction to the affinity support | High purification cost |
|  | Pull-down assays | | | No binding of WGE endogenous proteins to the affinity support | |
| FLAG | Functional analysis | N-ter | (Einhauer and Jungbauer, 2001; Bardoczky et al., 2008; Ramadan et al., 2015; Novikova et al., 2018; Nagy et al., 2020) | No binding of WGE endogenous proteins to the affinity support | |
|  | Structural analysis by cryo-EM and cryo-lithography | | | Cost-effective and easy purification process | Lower binding capacity of the affinity support |
|  | Protein detection by Western blot | C-ter | (Novikova et al., 2018) | Elution by competition under native conditions | Higher purification cost than for other tags |
|  | Pull-down assays | | | Affinity support reusable up to five times |
| Strep-tag II | Purification by affinity chromatography | N-ter | (Schmidt and Scerra, 2007; David et al., 2019) | No binding of WGE endogenous proteins to the affinity support | Slightly lower binding capacity of the affinity support than for His- and GST-tags |
|  | Pull-down assays | C-ter | (Foger et al., 2015a; Foger et al., 2015b; Li et al., 2016; Fogeron et al., 2017a; Minkoff et al., 2017; David et al., 2019) | Cost-effective and easy purification process | Higher purification cost than for His- and GST-tags |
| Dual tag | Pull-down assays | N-ter | (Khan et al., 2006; Bardoczky et al., 2008; Nagy et al., 2020) | Improved binding capacity | |
| Double-His<sub>6</sub> | Purification by affinity chromatography | | | Increased detectability | |
| GST-His<sub>6</sub> | Pull-down assays | N-ter/C-ter | (Bardoczky et al., 2008; Nagy et al., 2020) | Solubility enhancement through GST | |
|  | Purification by affinity chromatography | | | Efficient purification through GST | |
|  | Protein detection by Western blot | | | Elimination of GST by TEV cleavage | |
| GST-AviTag | Pull-down assays | N-ter/C-ter | (Cull and Schatz, 2000; Bardoczky et al., 2008; Nagy et al., 2020) | Solubility enhancement through GST | |
|  | Purification by affinity chromatography | | | Efficient purification through GST | |
| His-Flag | Pull-down assays | N-ter (Takahashi et al., 2012) | (Continued on following page) | Efficient purification through His-tag | |
|  | In vitro binding assay (AlphaScreen) | | | Highly sensitive detection through FLAG-tag | |

Frontiers in Molecular Biosciences | www.frontiersin.org 8 March 2021 | Volume 8 | Article 639587
minimal peptide sequence (WSHPQFEK) with a high affinity to native streptavidin or an engineered streptavidin having even higher affinity for the tag (Strep-Tactin (Maertens et al., 2013)). Affinity purification using the Strep-tag II is rapid and easy to setup. The Strep-tag II is very suitable for routinely making proteins in the WGS. When even higher affinity is needed during purification, a Twin-Strep-tag (Schmidt et al., 2013) is available that can also be used with WGS (Fogeron et al., 2016; Boukadida et al., 2018; Jirasko et al., 2020).

The use of the HaloTag, a 297 amino acid peptide derived from a bacterial haloalkane dehalogenase (Los et al., 2008), has been recently described for the WGS (Nagy et al., 2020). Because the HaloTag forms a highly specific covalent bond with its synthetic ligand, this tag is of particular interest for pull-down assays, allowing for more stringent buffer conditions and washing steps (Los et al., 2008). Although affinity purification is possible, elution of proteins having a HaloTag must be performed by enzymatic cleavage (Nagy et al., 2020). This makes the method quite expensive and thus not very suitable for large-scale production.

In our hands, for NMR sample preparation, the Strep-tag II is so far the best choice since it combines high purity of the protein of interest with yields compatible with structural biology (Fogeron et al., 2015a; Fogeron et al., 2015b; Li et al., 2016; Fogeron et al., 2017a; Minkoff et al., 2017).

**EXPRESSION TEMPLATES**

For routine protein expression, working with a dedicated expression vector is the best choice, although CFPS can also be done with linear DNA templates. Several vectors are available for use with the WGS, as outlined in (Bardocz et al., 2008; Nagy et al., 2020) and references therein. We have always relied on vectors having the E01 enhancer (Kamura et al., 2005) to drive cap-independent translation (available from CellFree Sciences, Japan), but there are more expression vectors for WGS available from commercial providers (e.g., pIVEX Wheat Germ Vector Sets, biotrecrabbit, Germany) and depositories (PSI: Biology-Materials Repository (PSI: Biology-MR)) sometimes using other initiation sites (Sawasaki et al., 2002b; Bardocz et al., 2008). Commonly, the gene of interest should be inserted as near as possible to the E01 sequence to get better expression. For CFPS systems, the vectors commonly have promoters for an RNA polymerase like the SP6 or T7 RNA polymerases from bacteriophages, which catalyze the synthesis of RNA in a 5′−3′ direction (Mcclister and Raskin, 1993). In addition, a ribosomal binding site or translation enhancer is required to enable efficient protein expression. For the WGS, as well as for other eukaryotic systems, it is important to use a cap-independent translational initiation sequence like the E01 enhancer (Kamura et al., 2005) to avoid cumbersome steps for in vitro capping of the RNA transcripts. Alternatively, Internal Ribosome Entry Sites have been successfully used in various CFPS systems (Mikami et al., 2008; Anastasina et al., 2014; Hodgman and Jewett, 2014; Quast et al., 2016). This includes a Species Independent Translation Initiation Sequence that could be applied to prepare an expression vector for use in different CFPS systems, thus avoiding the need to clone into multiple expression vectors (Gagoski et al., 2015). Other translational enhancers in the 3′ untranslated region of the template have been described in the literature (Fan et al., 2012) that could potentially further improve protein expression (Ogawa et al., 2014). However, such elements are not commonly used in today’s expression systems. It was further reported that some noncoding antisense RNAs can stimulate the translation of a matching sense RNA. This observation led to developing synthetic long noncoding RNAs named SINEUPs to enhance protein translation in vivo or in vitro (Zucchelli et al., 2015a; Zucchelli et al., 2015b). To date, no examples for the successful use of this method in a WGS were published to our knowledge, although this biological principle may also exist in plants.

CFPS experiments can readily utilize linear DNA templates instead of circular vectors. While it is very convenient for many applications to directly prepare a template by the PCR, it should be noted that circular DNA templates are more stable and commonly provide better protein yields than linear DNA templates. Linear expression templates can be directly made by PCR methods (Schinn et al., 2016) without cloning experiments and thus allow for rapid expression screening. Different PCR protocols have been developed to add regularity sequences at the 5′ and 3′ ends of the coding region using overlap-extension PCR. In consecutive PCR reactions, a promoter to drive RNA expression and an enhancer to induce protein synthesis are added at the 5′ end; while working with the T7 RNA
polymerase, a terminator sequence has to be added at the 3′ end. In addition to the regulatory sequences, the PCR primers can also be used to add short sequences encoding an affinity tag at either end. Caution is required when working with linear DNA in CFPS systems, because some extracts have an exonuclease activity that will damage or even entirely destroy a linear DNA template (Schinn et al., 2016). This problem was addressed by different approaches to protect or to extend the noncoding regions of the linear DNA templates. One elegant approach circularizes the PCR products before use in the expression reaction (Wu et al., 2007). However, the method uses the endogenous DNA ligase activity in E. coli S30 extracts, and only a quarter of the PCR products can be protected in this way. Other approaches to protect linear DNA templates have been described in the patent literature (Heindl et al., 2002). One easy-to-implement option is the use of biotinylated primers during PCR and later addition of streptavidin to the protein expression reaction to block exonucleases attacking the template from the ends. The same concept had been recently used when adding a DNA-binding protein to linear templates having matching binding sites at the ends (Zhu et al., 2020a). This approach showed good template protection when working with an E. coli CFPS system, though it is less effective than circularizing the PCR product. The standard “Split-PCR” protocol commonly used in combination with the WGS uses an extended 3′ overhang to better protect the linear DNA templates (Sawasaki et al., 2002b). Uncoupling of transcription and translation reactions, as described below, may further help to avoid DNA degradation by exonuclease activities within the cell extracts used only in the translation reaction.

Regardless of the approach taken, we advise analyzing expression templates before use for having the correct sequence and all necessary elements for successful expression. It is our common routine to confirm the sequence of new expression vectors. We further recommend analysis of vector DNA on an agarose gel and determining the OD_{260/280} ratio to assure the purity of the DNA preparation. CFPS reactions are sensitive to the quality of DNA templates. If uncertain or unforeseen problems occur, often a phenol/chloroform extraction of the circular or even linear DNA templates can be immensely helpful to resolve problems with expression.

**TEST EXPRESSION**

Once the expression vectors or linear templates are available, all templates are then individually tested for expression of the target protein. Besides the templates, different wheat germ extracts can be compared for their properties foremost on the achieved protein yields. Although no clear data have been published, different extract preparations may lead to variations in posttranslational modifications during expression. Variations between extract preparations may be better controlled when using commercial reagents, with wheat germ extracts commercially available from different providers; alternatively,

---

**FIGURE 4** | Protein analysis. (A) Typical flowchart for protein analysis after small-scale expression test. Parameters to be considered are highlighted in blue. (B) Small-scale expression test of the nonstructural protein 2 (NS2) from hepatitis C virus (HCV). This membrane protein was produced in the absence or presence of various detergents at a 0.1% concentration (w/v). Samples were analyzed by SDS-PAGE followed by Coomassie blue staining (upper panels) and Western blotting with an antibody against the Strep-tag II fused at the C-terminus of NS2 (lower panels). CFS, total cell-free sample; pellet, pellet obtained after centrifugation of CFS; SN-beads, supernatant obtained after centrifugation of CFS and incubated with Strep-Tactin magnetic beads to capture Strep-tag II-tagged NS2 protein; −, negative control (no NS2); +, positive control (NS2 expressed in the absence of detergent). The black arrowheads indicate NS2, adapted from Fogeron et al. (2015a). (C) SDS-PAGE analysis followed by Coomassie blue staining of the different steps from the affinity purification of the NS2 membrane protein produced directly in a solubilized form in the presence of MNG-3, adapted from Fogeron et al. (2015b).
home-made wheat germ extracts can be used (Takai et al., 2010; Fogeron et al., 2015a).

The user of a cell-free translation reaction must choose between coupled or uncoupled reactions. In coupled reactions, transcription and translation are performed in a single reaction step, allowing for an easier setup and shorter overall time requirements. In uncoupled or linked reactions, the mRNA is prepared beforehand and then added to the wheat germ extract for the translation step. With modern protocols, the mRNA can be used directly after transcription without any prior purification (Takai et al., 2010). Although coupled reactions have been described for the WGS (Stueber et al., 1984), uncoupled reactions are usually preferred (Sawasaki et al., 2002a; Endo and Sawasaki, 2004, 2006; Takai et al., 2010). Uncoupling indeed allows for more flexibility to work under optimal reaction conditions (e.g., temperature), to use additives in the translation reaction without interfering with transcription, or to better identify and solve problems when they occur. These advantages clearly counterbalance the fact that uncoupled reactions might be more time-consuming. Note that both coupled and uncoupled reactions can be applied to the different reaction formats described in the Large-Scale Protein Production section.

Some proteins may require testing of different reaction conditions, which can be done in parallel, as shown in Figure 4 for added detergents. If there is an uncertainty on which regions of a protein could give best yields, PCR-based template generation can be used to test the expression of multiple protein fragments before cloning them into an expression vector (Novikova et al., 2018). Similarly, different affinity tags have been tested in this way to see their effect on protein expression (Haberstock et al., 2012; Kralicek, 2014).

Quick expression tests are preferably done in small batch reactions by adding a labeled-lysine-charged tRNA (FluoroTech™, Promega, United States) to reaction mixtures. The fluorescently labeled lysine is randomly incorporated at AAA codons into the synthesized protein during the translation reaction, thus allowing for easy detection of proteins (Zhao et al., 2010; Novikova et al., 2018). After completion of the translation reaction, the labeled protein can be directly detected by SDS-PAGE using a laser-based fluorescent gel scanner; we recommend digesting the remaining labeled tRNA by RNase A treatment before loading onto the gel. In an optimized expression system, the newly synthesized protein from the added template should be visible on SDS-PAGE as there is no background expression in the WGS. While the labeling reaction is providing information on whether the protein can be made from an expression template, it is good to also perform a regular cell-free protein expression experiment without the fluorescent label to further test protein yields, solubility, purification methods, and possibly protein function (Fogeron et al., 2017a), as it is unclear whether in certain cases the randomly incorporated labeled lysine could interfere with protein functions. The expression test experiments should be further extended if the protein of interest requires disulfide bonds, certain cofactors, the addition of metal ions or is, for example, a membrane protein with expected low solubility. We will provide below more information on additives that could be tested to improve protein expression and quality.

**PROTEIN ANALYSIS**

SDS-PAGE analysis effectively assesses the expression and solubility of the protein within translation reactions, where it can be helpful to compare to a negative control expression reaction lacking the template or using an empty expression vector. While staining the proteins in the gel might be sufficient for protein detection, it can be advantageous to detect the protein of interest by Western blotting using a suitable antibody, which can also be directed against an affinity tag. Treating samples with benzonase, an endonuclease degrading DNA and RNA independently of their shape allows removing nucleic acids from the translation reaction. As indicated in Figure 4, both the full reaction mixture, the supernatant and pellet after centrifugation of the crude reaction mixture (e.g., at 20,000 g for 30 min) are analyzed to assess protein expression and solubility. The protein in the supernatant fraction can, for better visibility on the gel, be enriched using magnetic beads which can capture a tagged protein via the tag. Magnetic beads are fast, easy, and very convenient to use; they offer a higher binding capacity than standard chromatography resins and allow for efficient automation. Another fraction to be analyzed is the remaining supernatant of the binding assay, to confirm that the tag had worked properly.

SDS-PAGE of the full reaction already reveals if synthesis was successful. The protein in this fraction can best be seen by Western blotting, since there are many contaminating proteins present in the crude reaction mixture. When insoluble, the pellet fraction will be enriched in the protein target, which can typically be the case for membrane proteins or nucleic-acid-binding proteins like transcription factors. If this is the case, the protein can sometimes be seen using Coomassie staining, since there are few insoluble proteins present in wheat germ extracts. Otherwise, it should be identified using Western blotting for more reliable detection. The soluble fraction concentrated on beads will show the protein when soluble and if it attaches correctly to the beads via its tag. If the tag is inaccessible, the protein will remain in the soluble fraction. Confirming binding of the tag can help design the subsequent purification steps. Most structure determination techniques require soluble proteins; still, solid-state NMR and cryo-EM can be applied to proteins which are localized to the pellet fraction, due to either their size or aggregation state. One should however mention that while pellets formed by autoassembling proteins, or RNA-interacting proteins, can be correctly folded, membrane proteins found in the pellet after expression in absence of a detergent are likely misfolded. In the latter case, solubilization is an asset, as membrane reconstitution can then be done subsequently. Soluble expression can often be induced by additives (see below); data analysis will be carried out in a similar way to assess protein synthesis, solubility, and binding to the magnetic beads as proxy for purification. In structural genomics studies, at this point, one can distinguish if a protein will be directed to
analyses using a soluble protein, like solution-state NMR or X-ray crystallography, or if it will need approaches that can target insoluble proteins, as solid-state NMR and cryo-EM.

SDS-PAGE also allows confirming expected protein size, as well as stability, and the presence of degradation products (which should be largely absent, since wheat germ extracts have no significant protease activity). If a protein is expressed in a soluble state and attaches to the magnetic beads, one can proceed to column purification via standard protocols for the used tag, and fractions can be analyzed using SDS-PAGE. The purified protein can be used for the first biophysical characterization using mass spectrometry, which allows confirming its identity, as well as detecting possible posttranslational modifications. The latter, typically phosphorylation, was shown to be possible in the WGS and took place on sites identified in vivo (David et al., 2019). Acetylation was also observed (unpublished). An enzymatic test on the translation reaction can often determine already whether the protein is functional or not; precautions must be taken on background activities in the wheat germ extract, however. Also, when autoassemblies are expected to form, electron microscopy analysis allows for their direct observation, often also in the crude reaction (David et al., 2018; Wang et al., 2019). In addition, the secondary structure of the protein of interest can be investigated by circular dichroism (Kelly et al., 2005).

## TROUBLESHOOTING POOR EXPRESSION

When no expressed protein can be detected, the reaction conditions and template design should be checked for possible errors. Negative results are often about poor detection or the inability to see the overexpressed protein over the background of proteins from the extract. Also, often proteins may not show on SDS-PAGE at the expected molecular weight. Both problems can be addressed by Western blotting. Using the FluoroTect™ labeling method described above, we have seen only very few cases where no protein could be detected after expression in the WGS, as the method is very sensitive (note that the free label will run at the front of the gel which may interfere with very small proteins and thus may be better removed before SDS-PAGE analysis).

Further troubleshooting should consider the following points working with uncoupled reactions to better understand potential problems during transcription and translation: 1. Confirm the expression template was made correctly and no mistakes have been made during template design and preparation. 2. Confirm the DNA quality of the template on an agarose gel and measuring the OD_{260/280}. 3. Confirm the RNA quality using agarose gel or capillary electrophoresis; CFPS reactions must be done under RNase-free conditions. 4. Confirm reagent quality by working with a positive control known to work well in the expression system.

Regarding the template design, an N-terminal tag can have an impact on the secondary structure of RNA, and thus on protein expression where, for example, hairpin loops tend to repress translation. RNA secondary structures can be analyzed using the Mfold software (Zuker, 2003) (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form). In case of low protein yields, changes to the N-terminus could also be considered as, for example, shown for making Growth Hormone Secretagogue Receptor in a CFPS (Pacull et al., 2020). Further, for optimal purity of the DNA template, a phenol/chloroform extraction is recommended. A sign for an efficient transcription reaction is the appearance of a white magnesium pyrophosphate precipitate. Agarose gel electrophoresis allows verifying the expected size of the RNA. When working with circular DNA and an Sp6 promoter, the RNA can form a ladder as the polymerase may run several times around the vector. While wheat germ extracts can be stored for an extended time at ~80°C, they are overly sensitive to freeze/thawing cycles, or any storage at higher temperatures. We advise using a positive control like, for example, expressing an easy-to-detect Green Fluorescent Protein (GFP) to confirm the performance of wheat germ extracts and other reagents. Moreover, commercial buffers are preferred since they are less error prone.

## OPTIMIZING EXPRESSION REACTION CONDITIONS

### Expression Conditions

The expression yield is an important parameter, especially for structural studies which require higher amounts of protein. The temperature during translation reactions can have an impact on both expression yield and protein folding. When protein yields are not satisfying, it is thus worth testing different temperatures for protein synthesis within a range from 4 to 25°C; wheat germ extracts lose activity above 25°C. The lower the temperature is, the longer the translation reaction must be.

### Table 3: Maximal concentration for added ions.

| Ion     | Salt | Maximal concentration (µM) |
|---------|------|----------------------------|
| Mn^{2+} | Chloride | 100                       |
| Mg^{2+} | Chloride | 100                       |
| Ca^{2+} | Chloride | 100                       |
| Cu^{2+} | Chloride | 100                       |
| Cd^{2+} | Chloride | 100                       |
| Co^{2+} | Chloride | 100                       |
| Fe^{2+} | Chloride | 10                        |
| Ni^{2+} | Chloride | 100                       |
| Zn^{2+} | Chloride | 10                        |

Data obtained for expression of GFP using wheat germ extract WPRO7240 (CellFree Sciences, Japan) on a bilayer format. aCFPS systems are critically dependent on the Mg concentration. bAbout 25% reduction of protein yield when using 100 µM ferrous chloride.
**Additives**

*In vitro* reactions allow for the addition of factors that may be required for optimal protein folding, function, or solubility. Most common examples are the addition of isotope-labeled amino acids (Makino et al., 2010), metal ions like zinc (Okada et al., 2009) and iron (Samuel et al., 2015), redox reagents supporting the formation of disulfide bonds (Saaranen and Ruddock, 2019), or detergents and lipids (Sachse et al., 2014). Additionally, chaperones may be used to support or modulate folding. However, one should test beforehand whether additives do not interfere with protein synthesis. Note that wheat germ extracts commonly contain some lipids and metal cofactors that may assist already protein expression (Goren et al., 2009).

Additives, such as detergents, lipids, or chaperones may also help to improve protein yields. We have experienced that adding detergent to protein synthesis could improve their expression level and purity.

**Ions**

Besides zinc and iron ions, other ions can be used in the WGS. Refer to Table 3 for different ions that had been tested for use in the WGS (unpublished data provided by F. Tanabe and R. Morishita). In the table, we give the highest ion concentrations that can be used in the translation reactions without inhibiting synthesis in separated transcription and translation reactions. Goren and Fox (Goren and Fox, 2008) described the preparation of a functional human stearoyl-CoA desaturase complex by coexpression in the WGS, which requires nonheme iron for its catalytic function. Because the wheat germ extract lacked the necessary amount of iron ions and heme, ascorbate stabilized $\text{Fe}^{2+}$ was subsequently added to their proteoliposome preparation to activate the complex. They also provide information on an elemental analysis of a wheat germ extract. As another example, the yeast (m^2G10) methyltransferase (a Trm11 and Trm112 complex) was prepared using the WGS for coexpression and complex formation (Okada et al., 2009). Since Trm112 contains two zinc fingers, the authors showed that the system could be used in the presence of up to 20 µM added ZnCl$_2$ without reducing protein yields.

**Detergents and lipids**

Detergents and lipids are of special interest for working with membrane proteins, which are today the most important drug targets for therapy (Hopkins and Groom, 2002; Arinaminpathy et al., 2009). However, membrane proteins are notoriously difficult to express in living cells since they are often toxic and may depend on the lipid composition of membranes (Harayama and Riezman, 2018). This makes CFPS a highly valuable alternative, where the use of CFPS systems for the preparation of G protein-coupled receptors for structural investigations was recently reviewed (Kögl et al., 2019). Three dedicated protocols were established for their expression in CFPS systems: 1) the precipitate mode, 2) working in the presence of detergents, or 3) working in the presence of lipids. In the first mode, protein precipitates form during synthesis and can be afterward efficiently solubilized with a detergent (Klammt et al., 2004). Although there is evidence that detergent solubilization of membrane protein precipitates produced in the *E. coli* CFPS systems could result in functionally folded proteins (Klammt et al., 2004; Sansuk et al., 2008), it was also shown that such a process could lead to inactive proteins (Klammt et al., 2005; Klammt et al., 2007). Examples for solubilization and refolding after expression in *E. coli* CFPS have been published for the Growth Hormone Secretagogue Receptor (Pacull et al., 2020) and Neuropeptide Y2 Receptor (Krug et al., 2020). As lipids are not fully removed during wheat germ extract preparation, they may bind to proteins (Schwarz et al., 2008), which could explain why membrane proteins expressed in the precipitate mode are sometimes partially soluble.

An interesting alternative is the production of membrane proteins in the presence of detergents. While ionic detergents often denature proteins, nonionic and zwitterionic detergents are mild for membrane protein solubilization and in many cases preserve protein folding. Above the critical micelle concentration (CMC), detergents in aqueous solutions spontaneously form micellar structures (Garavito and Ferguson-Miller, 2001; Seddon et al., 2004). The CMC is influenced by pH, ionic strength, temperature, and the presence of protein, lipid, and other detergent molecules. Membrane protein expression in the presence of detergent leads to the formation of proteomicelles. Detergents are available instantly at the ribosomes, eliminating problems encountered regarding the transport to membranes and translocation processes of synthesized proteins (Schwarz et al., 2008). Importantly, not all detergents are compatible with CFPS systems. Their use in *E. coli* lysates has been broadly reported (Berrier et al., 2004; Elbaz et al., 2004; Ishihara et al., 2005; Klammt et al., 2005; Schwarz et al., 2008; Deniaud et al., 2010; Miot and Betton, 2011), suggesting that mild detergents with low CMC values allow for optimal solubilization yields without interfering with expression yields. However, some detergents affected protein expression levels in the WGS (Genji et al., 2010). Table 4 summarizes detergents whose use was described for the WGS. Detergent concentration can also impact both protein expression and solubilization levels. Alternatives to traditional detergents, such as the linear carbohydrate-based polymer NVoy (Guild et al., 2011) and peptide surfactants (Periasamy et al., 2013), have been described for use in the WGS as well. In addition, the use of fluorinorated compounds (Park et al., 2007; Park et al., 2011; Blesnec et al., 2012) and amphipols (Popot, 2010) has been reported for *E. coli* based systems and supports direct membrane protein reconstitution into membranes (Nagy et al., 2001; Park et al., 2007). Note however that there is only one commercially available amphipol compatible with CFPS (NAPol) (Popot, 2010; Park et al., 2011). To be analyzed in a native-like environment, membrane proteins expressed in the presence of traditional detergents or alternative surfactants can be reconstituted in lipids after purification (Seddon et al., 2004; Fogeron et al., 2016; Lacabanne et al., 2017; Jirasko et al., 2020), which has been described as the most successful approach for membrane protein insertion into membranes (Bayburt and Sligar, 2010). As protein loss needs to be minimized, fast lipid reconstitution
| TABLE 4 | Summary of detergents whose use has been described to produce membrane proteins with the WGS. |
|-------------------------------|-------------------------------------------------|-----------------------|-------------------------|---------------------------------|---------------------------------|
| Detergents                     | Proteins                                         | Yield  | Solubility | References                     | Applications/comments            |
| Anionic                        | Cholate                                          | 3,7,12-Trihydroxy-5-cholan-24-oic acid             | ▼                      | (Nozawa et al., 2007)          |                                  |
|                               | Insect odorant receptor subunits                  |         | –                 | (Carraher et al., 2013)       | Protein expression              |
|                               | Nonstructural proteins from HCV                  | ▼                      | –                 | (Fogeron et al., 2015a)       |                                  |
| Deoxycholate                   | 3,12-Dihydroxy-5-cholan-24-oic acid              | Insect odorant receptor subunits                  | ▼                      | (Carraher et al., 2013)       |                                  |
| N-Lauryl sarcosine             | Sodium dodeyl sulfate                            | Insect odorant receptor subunits                  | ▼                      | (Carraher et al., 2013)       | Not suitable for functional or structural studies |
| Nonionic                       | CHAPS                                            | Steroid derivative (3-(3-cholamidopropyl(dimethylammonio)-1-propansulfonat) | ▼                      | (Nozawa et al., 2007)          |                                  |
|                               | Bacteriorhodopsin                                | ▼                      | –                 | (Genji et al., 2010)          | Protein expression              |
|                               | CD36                                            | Nonstructural proteins from HCV                  | ▼                      | (Fogeron et al., 2015a)       |                                  |
|                               | EDG3                                            | EDG3; GPR84                                       | ▼                      | (Fogeron et al., 2015a)       |                                  |
|                               | DPC                                              | Monochain phosphocoline (dodecyl-phosphocholine)  | ▼                      | (Fogeron et al., 2015a)       | Protein expression              |
|                               | LDAO                                             | Lauryl dimethyl amide oxide                       | ▼                      | (Fogeron et al., 2015a)       | Protein expression              |
|                               | LPPG                                             | 1-Palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] | ▼                      | (Fogeron et al., 2015a)       |                                  |
|                               | Zwittergent                                      | Steroid derivative (3-[(3-cholamidopropyl(dimethylammonio)-1-propansulfonat]) | ▼                      | (Carraher et al., 2013)       | Insertion of purified proteins into preformed liposomes |
|                               | Brij-35                                          | AtPPT1                                           | ▼                      | (Nozawa et al., 2007)          |                                  |
|                               | Brij-58                                          | AtPPT1                                           | ▼                      | (Nozawa et al., 2007)          |                                  |
|                               | Brij-78                                          | AtPPT1                                           | ▼                      | (Nozawa et al., 2007)          |                                  |
|                               | Digitonin                                        | AtPPT1                                           | ▼                      | (Nozawa et al., 2007)          | Protein not functional |

(Continued on following page)
### Table 4 (Continued) Summary of detergents whose use has been described to produce membrane proteins with the WGS.

| Detergents | Proteins | Yield | Solubility | References | Applications/comments |
|------------|----------|-------|------------|------------|-----------------------|
| CrdS (Agrobacterium curdlan synthase) | Insect odorant receptor subunits | ▲ | ▲ | (Periasamy et al., 2013) | Insertion of purified proteins into preformed liposomes |
| DM | Alkyl glucoside (n-decyl-D-maltoside) | Bacteriorhodopsin | ▽ | ▲ | (Chae et al., 2010a) | Protein expression |
| DDM | Alkyl glucoside (n-dodecyl-D-maltoside) | APPT1 | ▲ | ▲ | (Nozawa et al., 2007) | Protein expression |
| MNG-3 | Lauryl maltose neopentyl glycol | Bacteriorhodopsin | ▲ | ▲ | (Chae et al., 2010a) | Protein expression |
| Nonidet P-40 | Polyethylene glycol derivative | APPT1 | ▲ | ▲ | (Periasamy et al., 2013) | Protein expression |
| β-OG | Alkyl glucoside (n-octyl-D-glucopyranoside) | Insect odorant receptor subunits | – | ▲ | (Carraher et al., 2013) | Insertion of purified proteins into preformed liposomes |
| Triton X-100 | Polyethylene glycol derivative | APPT1 | ▲ | ▲ | (Nozawa et al., 2007) | Protein expression |
| Triton X-114 | Polyethylene glycol derivative | APPT1 | ▲ | ▲ | (Nozawa et al., 2007) | Protein expression |
| Tween-20 | Polyoxyethylene alkyl-ether | APPT1 | ▲ | ▲ | (Nozawa et al., 2007) | Protein expression |
| Tween-40 | Polyoxyethylene alkyl-ether | APPT1 | ▲ | ▲ | (Nozawa et al., 2007) | Protein expression |
| Tween-60 | Polyoxyethylene alkyl-ether | APPT1 | ▲ | ▲ | (Nozawa et al., 2007) | Protein expression |
| Tween-80 | Polyoxyethylene alkyl-ether | CrdS (Agrobacterium curdlan synthase) | ▲ | ▲ | (Periasamy et al., 2013) | Protein expression |
| Mixtures | Fos-choline FC-12 or FC-14 CHAPS | Bacteriorhodopsin | ▲ | ▲ | (Genji et al., 2010; Nozawa and Tozawa, 2014b) | Counteracting the inhibitory effect of detergent |
without the need for extensive protein handling is an asset. This is possible by using, instead of lengthy dialysis for detergent removal (Althoff et al., 2012; Lacabanne et al., 2017), complexation of the detergents with cyclodextrin (Degrip et al., 1998). Proteoliposomes can then simply be separated by centrifugation for further analysis (Jirasko et al., 2020).

Another option is to express the proteins in the presence of lipids, as membrane proteins can cotranslationally incorporate into the lipid bilayer to form directly proteoliposomes (Nozawa et al., 2007). Different protocols have been developed in this context (Shadiac et al., 2013; Zhou and Takeda, 2020). Although the endoplasmic reticulum (ER) is removed during wheat germ extract preparation, mimicking the natural membrane environment is indeed possible with the addition of lipids to the translation reaction. CFPS systems tolerate relatively high concentrations of lipids and lipid mixtures, and even slightly beneficial effects have been observed on the expression efficiency (Klammt et al., 2004). Most often lipids are used in the WGS in the form of liposomes (Akbarzadeh et al., 2013), which are artificial spherical vesicles formed by lipid bilayers from either synthetic lipids or biological lipid extracts (Akbarzadeh et al., 2013). Insertion of membrane proteins into liposomes leads to the direct formation of proteoliposomes which are generally isolated by ultracentrifugation on density gradients (Nozawa et al., 2011; Periasamy et al., 2013). Such proteoliposomes can be easily purified and have been used in studies on membrane proteins (Banerjee and Datta, 1983; Rigaud, 2002; Wang and Tonggu, 2015). Lipid type and composition are highly important to ensure cotranslational insertion (Periasamy et al., 2013), and screening biologically relevant lipids instead of using commercially available lipids might be a better choice. Examples of membrane proteins produced in the presence of liposomes using a WGS are summarized in Table 5. This approach is in theory very attractive, but not all proteins can be integrated into liposomes, some requiring a more complex lipid environment and others depending on the translocon machinery (Sachse et al., 2014). The absence of the translocon can, however, be problematic for the topology of multispanning membrane proteins. Also, when low lipid-to-protein ratios are crucial, it might be that these ratios cannot be reached using lipid addition, since spontaneous insertion might not be quantitative.

Another possibility is the addition of microsomes, which are membranous vesicles obtained from the ER often from dog pancreas (Jackson and Blobel, 1977), oocytes (Kobilka, 1990; Lyford and Rosenberg, 1999), or oviduct cells (Rosenberg and East, 1992). Canine pancreatic microsomal membranes are commercially available (Promega, United States) and allow for signal peptide cleavage, membrane insertion, translocation, and core glycosylation according to the maker. In the presence of microsomes, membrane proteins having a signal peptide are translocated through the translocon of the ER membrane and then can be glycosylated within the lumen of the membranes (Dobberstein and Blobel, 1977; Katz et al., 1977; Lingappa et al., 1978). Since the protein synthesis machinery is present only outside the vesicles, a prevalent inside-out orientation of membrane proteins can be expected (Schwarz et al., 2008). There are, however, few reports on this approach using the WGS (Dobberstein and Blobel, 1977; Jackson and Blobel, 1977; Katz et al., 1977; Lingappa et al., 1978), mainly because of low expression yields making this approach only suitable for functional protein analyses.

Alternatively, cotranslation insertion in a synthetic membrane of block copolymer vesicles has been described for the CXCR4 GPCR (De Hoog et al., 2014). Other alternatives to liposomes and biological membrane vesicles are bicelles and nanodiscs (Ritchie et al., 2009; Bayburt and Sligar, 2010; Dürr et al., 2012; Lyukmanova et al., 2012a; Sachse et al., 2014). The diameter of nanodiscs ranges from 10 to 20 nm, depending on the length and type of the membrane scaffold protein (Sachse et al., 2014). During synthesis, membrane proteins are incorporated into nanodiscs in a passive manner and can later on be extracted from them in a native functional form (Ranaghan et al., 2011). A major advantage is that nanodiscs keep membrane proteins soluble in a detergent-free environment, possibly yielding monodisperse and homogenous samples (Borch and Hamann, 2009; Henrich et al., 2015; Danmaliki and Hwang, 2020). A tag fused to the membrane scaffold protein allows moreover for a simple purification procedure (Bayburt and Sligar, 2010). Recent examples of membrane proteins produced in the presence of nanodiscs using the WGS are summarized in Table 5. This includes, for example, the synthesis of the G protein Signaling 1 (ArGS1) protein from Arabidopsis thaliana (Li et al., 2016). A major drawback of nanodiscs in solid-state NMR is that they might result in low signal-to-noise ratios due to high lipid-to-protein ratios (Jirasko et al., 2020). For solution NMR, optimized nanodiscs have been developed with smaller diameters (Hagn et al., 2013). Nanodiscs can be unstable, and therefore polymer-enhanced versions have been described to extend the use of this promising platform for studies on membrane proteins (Chen et al., 2020).

To summarize, there are different alternatives to produce membrane proteins in a native form. The most suitable one depends mainly on the nature of the protein and the final application. For our purposes in NMR sample preparation, the expression directly in a detergent-solubilized form, followed by affinity purification and lipid reconstitution, has given the most convincing results, since it also allowed the selection of a lipid-to-protein ratio which minimizes the amount of lipids in NMR rotors (Lacabanne et al., 2019; Jirasko et al., 2020). In one special case, membrane envelopes of the duck hepatitis B virus were autoassembled when using the WGS in the presence of mild detergents, likely using lipids present in the wheat germ extract (David et al., 2018), which made reconstitution dispensible.

Chaperones

Molecular chaperones are important protein factors often needed for correct conformational folding of proteins. A recombinant E. coli CFPS system, the PURE system, was used to systematically test the impact of chaperones on the solubility of ~800 proteins (Niwa et al., 2012) without interference of other proteins from a cell extract, showing their importance to improve protein quality. The eukaryotic translation machinery is thought to have been optimized through evolution to support cotranslational protein folding (Endo and Sawasaki, 2006). Newly synthesized proteins in the WGS can indeed be
TABLE 5 | Summary of lipids whose use has been described for the production of membrane proteins with the WGS.

| Lipid composition | Proteins | References | Applications/comments |
|-------------------|----------|------------|-----------------------|
| **Anionic lipids** | **DOPG** | CrdS (Agrobacterium curdlan synthase) | (Periasamy et al., 2013) | No protein expression |
| **DOTAP** | CrdS (Agrobacterium curdlan synthase) | (Periasamy et al., 2013) | No protein expression |
| **Asolectin** | A/PPT1 (Arabidopsis thaliana phosphoenolpyruvate/ phosphate translocator 1) and 40 other membrane proteins | | Functional analysis |
| **DMPC** | A/DTC (Arabidopsis thaliana dicarboxylate/ tricarboxylate carrier) | (Nozawa and Tazawa, 2014a) | Isolation on density gradient |
| **Cationic lipids** | Cytochrome b₃₈ | (Nomura et al., 2008) | Transport activity |
| **EYPC** | Cytochrome b₃₈ | (Nomura et al., 2008) | Transport activity |
| **POPC** | CrdS (Agrobacterium curdlan synthase) | (Periasamy et al., 2013) | Functional analysis |
| **POPE** | CrdS (Agrobacterium curdlan synthase) | (Periasamy et al., 2013) | Functional analysis |
| **Lipid mixtures** | Soybean total extract, containing 20% lecithin | | Functional and structural analysis by NMR |
| **DMPC/cholesterol (70/30, mol/mol)** | TbsLS4 (Trypanosoma brucei siphingolipid synthase 4) | (Sevova et al., 2010) | Enzymatic specificity analysis |
| **POPC/CL/POPS/POPA (54/24/16/4, mol/mol)** | A/RGS1 (Arabidopsis thaliana regulator of G protein Signaling 1) | (Li et al., 2016) | Functional and structural characterization |
| **E. coli lipids** | CrdS (Agrobacterium curdlan synthase) | (Periasamy et al., 2013) | Functional and structural characterization |
| **Asolectin liposomes** | Cytochrome b₃₈ | (Nomura et al., 2008) | Functional and structural characterization |
| **Other mixtures** | Human stearoyl-CoA desaturase complex | (Goren and Fox, 2008) (Goren et al., 2009) | Functional and structural analysis by NMR |
| **Nanodiscs** | TblSLS4 (Trypanosoma brucei siphingolipid synthase 4) | (Sevova et al., 2010) | Enzymatic specificity analysis |
| **MSP1D1** | A/RGS1 (Arabidopsis thaliana regulator of G protein Signaling 1) | (Ji et al., 2016) (Fogeron et al., 2015a; Fogeron et al., 2015b; Li et al., 2016; Fogeron et al., 2017a; Minkoff et al., 2017) | Functional and structural characterization |

Liposomes. DOPG, 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol); DOTAP, 1,2-dioleoyl-3-trimethylammonium-propane; DMPC, 1,2-dimyristoyl-sn-glycero-3-phosphocholine; DOPC, 1,2-dioleoyl-sn-glycero-3-phosphocholine; EYPC, egg yolk phosphatidylcholine; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; POPE, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine; CL, 19,39-bis[1,2-dioleoyl-sn-glycero-3-phospho]-sn-glycerol; POPS, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoserine; POPA, 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-1′-glycerol.

Nanodiscs. MSP1D1, membrane scaffold protein 1D1; MSP1E3D1, membrane scaffold protein 1E3D1; TOCL, 1',3'-bis[1,2-dioleoyl-sn-glycero-3-phospho]-1'-sn-glycerol.

stabilized by eukaryotic chaperones promoting folding. This is best documented for the formation of disulfide bonds. Since translation reaction buffers commonly contain the reducing agent dithiothreitol (DTT), the production of disulfide bond-containing proteins is a delicate issue. However, lowering DTT concentration commonly leads to decreased expression yields,
although disulfide bond formation was demonstrated in a DTT-deficient reaction in the presence of a protein disulfide isomerase (PDI) (Kawasaki et al., 2003). The addition of chaperones that catalyze disulfide bond formation to the translation reaction therefore might be necessary for correct folding (Saaranen and Ruddock, 2019). As an example, the cooperative role of quiescin sulfhydryl oxidase (QSOX) and PDI was investigated for the generation of native pairings in unfolded reduced proteins (Rancy and Thorpe, 2008; Gad et al., 2013). Interestingly, glutathione-based redox buffers are not needed for efficient folding using a combination of QSOX and PDI (Rancy and Thorpe, 2008). If used, the ratio of reduced and oxidized glutathione might have to be optimized depending on the protein of interest. The use of a combination of a PDI and ErO1α system (Inaba et al., 2010; Shergalis et al., 2020) was established by CellFree Sciences for the synthesis of an anti-AGIA-IgG antibody Fab fragment exhibiting antigen-binding capacity after protein synthesis (https://www.cfsciences.com/eg/resources/application-note/519-note-9). In addition to using exogenous chaperones, pretreatment of E. coli extracts with the alkylating agent iodoacetamide (IAA) which covalently blocks the free sulfhydryl groups of enzymes has been reported to support oxidative folding (Kim and Swartz, 2003; Yin and Swartz, 2004). This approach could be considered for the WGS as well to prevent aggregation and to improve both solubility and activity, yielding more of the disulfide bond-containing proteins. Fine-tuning of redox conditions for disulfide bond formation has been described for human and mouse prion-like Doppel proteins and mouse interleukin-22 (Michel and Wuthrich, 2012) and has also proven to be efficient for disulfide bond formation in virus-like particles (Bundy and Swartz, 2011), all these proteins being produced using E. coli lysates. This approach was further shown to support the formation of intramolecular disulfide bonds during the synthesis of antibody fragments using an insect cell lysate (Stech et al., 2014). Although there are few data available about this approach for the WGS (Rancy and Thorpe, 2008; Gad et al., 2013), it could definitely be of interest when disulfide bond formation is desired. Since disulfide bonds are commonly found under oxidizing conditions, further considerations should be given to the storage and processing of proteins carrying this modification. Several methods have been published for the analysis of disulfide bonds including Ellman’s reagent [5,5′-dithiobis-(2-nitrobenzoic acid) or DTNB] used to quantify the concentration of thiol groups (Ellman, 1959; Winther and Thorpe, 2014), mass spectrometry (Lakbub et al., 2018), NMR (Denisov et al., 2019; Wiedemann et al., 2020), or simpler methods using PEG-maleimide to modify free thiol groups. Proteins may also be analyzed under nonreducing and reducing conditions by SDS-PAGE (Braakman et al., 2017). However, posttranslational formation of disulfide bonds in CFPS experiments or later during processing of proteins has its own challenges as well as not every cysteine may form correct bonds. As a first indication for correct folding, we suggest looking for an increase in protein solubility and, where possible, to confirm protein activity in a functional test or take an NMR spectrum.

The chaperone function of Ric-8 proteins was shown to be required for proper folding of heterodimeric G proteins (Chan et al., 2013). More recently, it was also shown that coexpression of J-domain containing chaperone proteins with potassium channels is essential for their folding, stabilization, and tetrameric assembly (Li et al., 2017). Another example of the correlation between cofactor binding and protein folding was demonstrated for the Flavin Mono Nucleotide- (FMN-) binding protein (Abe et al., 2004). The WGS allows for coexpression of two or more proteins in the same translation reaction, where the expression of binding partners may assist proper expression. Direct preparation of protein complexes in cotranslation experiments will open new ways to make use of chaperones to assist the production of functional proteins.

LARGE-SCALE PROTEIN PRODUCTION

Cell-free protein expression systems can use different reaction formats, and the choice of the best suited reaction format depends on the application. This can include functional and structural investigations for research or clinical purposes, small-scale assays, high-throughput screening or large-scale production including industrial use for diagnostic or therapeutic applications. Commonly, protein quantities for structural and functional studies or antigen production are in the range of milligram amounts, while much larger quantities could be desired for industrial applications. All those needs have been achieved by different CFPS. Therefore, the main parameters to be considered include the size and nature of the protein of interest, ease of implementation, productivity, reaction time, ability to scale, and the cost of the platform. To meet those needs, WG-CFPS reactions can be performed in either one-compartment or two-compartment reactions, and final yields between 1 and 20 mg of GFP per mL wheat germ extract can be achieved using high-quality extracts (Harbers, 2014).

The batch mode (Kawasaki et al., 2003) is a one-compartment reaction in which all reagents are mixed in a single container and is thereby the least complicated. The system works, however, only for a few hours, mainly due to the accumulation of inhibitory byproducts in the single reaction compartment (Schwarz et al., 2008), and the amount of the synthesized protein is usually not sufficient for structural investigations (Sawasaki et al., 2002a). The batch mode is ideal for small-scale high-throughput expression screening experiments (Sawasaki et al., 2002a; Schwarz et al., 2008). One alternative to the regular batch reaction format is the so-called repeat-batch or discontinuous batch mode (Harbers, 2014). After incubation, the batch reaction is concentrated by centrifugation step; then, fresh reaction buffer is added to provide new substrates. Multiple concentration cycles can thus be performed, leading to higher protein yields than the batch mode (Harbers, 2014). An automated discontinuous batch system was described for the production of soluble Galdieria sulphuraria protein DCN1, leading to a yield higher than 2 mg/mL in the reaction mixture allowing for its structure determination by X-ray crystallography from a 10 mL reaction (Beebe et al., 2011). This approach has also been applied to the production of membrane proteins in the presence of detergents or lipids (Beebe et al., 2011).
The continuous-exchange cell-free (CECF) system is a two-compartment setup in which the cell-free protein expression reaction is separated from the feeding buffer by a semipermeable membrane (Katzen et al., 2005). The cell-free expression reaction takes place in the dialysis device, and the dialysis buffer containing fresh substrates can diffuse in, while byproducts passively diffuse out (Figures 5A, B). Mini- and maxi-CECF reactors, as well as further CECF reactor designs, have been described in detail (Schneider et al., 2009). Interestingly, a microfluidic platform was also designed for CECF, allowing for reduced reaction volumes and simultaneous expression of up to 96 proteins (Jackson et al., 2014). Making proteins for NMR use, we standardly use 500 μL and 3 mL commercial dialysis cassettes and the CECF system (David et al., 2018; Wang et al., 2019) (Figures 5A, B). When larger protein amounts are needed, larger dialysis cassettes can be used, or more generally, CECF reactions can be run in parallel as described in (Aoki et al., 2009).

The continuous-flow cell-free (CFCF) system is another two-compartment setup like CECF. First described by Spirin (Spirin et al., 1988; Spirin, 2004), the CFCF system provides, through the use of a pump, the automated and continuous supply of substrates into the reaction chamber and the removal of byproducts which are pushed out through an ultrafiltration membrane, simultaneously retaining the protein of interest (Endo et al., 1992). The translation reaction can thus proceed for more than two days, which is more than ten times longer than the batch mode, and can yield up to several milligrams of protein (Spirin et al., 1988; Morita et al., 2003). This approach has been reported to be not suitable for high molecular weight proteins over 300 kDa (Spirin, 2004). Continuous reaction formats such as CECF and CFCF are nonetheless attractive for industrial protein production, and automated systems have been optimized in this direction (Vinarov et al., 2006a; Aoki et al., 2009; Revathi et al., 2010). The bilayer method is a simplified and less expensive version of CECF and CFCF (Sawasaki et al., 2002a). In contrast to the CECF and CFCF systems, the two compartments are not separated by a semipermeable membrane, and the total reaction has thus to be harvested for further analysis (Figure 5C). This method allows for the synthesis of protein amounts compatible with functional and structural analyses. The substrate buffer is overlaid onto the translation mixture, forming two separate layers through their different density, thus allowing for a diffusion-controlled translation process (Sawasaki et al., 2002a; Takai et al., 2010). The bilayer method can be fully automated for large-scale and efficient screening (Endo and Sawasaki, 2004; Vinarov et al., 2006b). Moreover, its flexible format permits screening different additives for the translation reaction, such as detergents or lipids for the expression of membrane proteins in a soluble form (Harbers, 2014), and can be readily scaled up from 96-well plates to 6-well plates (Figure 5D). Over 13,000 human cDNA clones have already been tested for protein expression using this method (Goshima et al., 2008). Bilayer expression is easier to handle and less expensive than the CECF and CFCF modes but much more efficient than batch reactions. While yields are about three times higher in the dialysis mode, the cost is proportionally higher in the case of protein labeling because of the larger buffer volume. In addition, solubility could be an issue in the dialysis mode since proteins are more concentrated.

For much higher throughput, protein synthesis using microfluidics approaches was also described for the WGS and has been used for the kinetic analysis of transcription factor-DNA interactions (Geertz et al., 2012) and to perform 96 dialysis reactions in parallel. The microfluidics approach can improve protein expression, offering much higher yields as compared to batch reactions (Jackson et al., 2014). These methods have the potential to become more important in the context of biomedical and diagnostic approaches as well as applications in systems biology looking at many proteins at the same time (Ayoubi-Joshaghani et al., 2020).
In the context of NMR sample preparation, we mainly use the bilayer method for small-scale expression tests and screening of additives in 96-well plates (Fogeron et al., 2015b). We have also implemented the mini-CECF reactor (Schneider et al., 2009) in our laboratory for samples that need to be more concentrated for analysis (Figures 5E,F, unpublished). When sample concentration is not an issue, the bilayer method is definitely the method of choice when working at a small scale and to reduce the volumes needed for translation reactions and feeding buffer or while adding stable isotope-labeled amino acids. For larger-scale production of labeled NMR samples, we usually perform the translation reaction either using the bilayer method in 6-well plates followed by affinity purification (Fogeron et al., 2015b) or in the CECF mode using dialysis cassettes followed by isolation of the protein on a density gradient in the case of proteoliposomes, capsids, or viral envelope assemblies (David et al., 2018; Wang et al., 2019). In our hands, proteins in this setup typically yield between 0.2 and 1 mg protein per mL wheat germ extract used. The cost for this screening is around 300 € per reaction using 1 mL wheat germ extract (including purification), where the triply labeled amino acids only represent about 120 €. For 1 mg of protein, sufficient for solution and solid-state NMR experiments (0.7 mm rotor), this results in a cost of 600–1,200 € for a $^2$H/$^{13}$C/$^{15}$N labeled sample. When using commercial extracts, the cost of the extract must be added to this. The production of home-made wheat germ extracts can be done at negligible cost when only reagents are considered. Our lab routinely produces around 100 mL wheat germ extract per year, with eye sorting of the wheat germs done on a single day every two months by the entire group (<10 people). The extract is then prepared in two days by one person (Fogeron et al., 2015a). In bacterial expression, a triply $^2$H/$^{13}$C/$^{15}$N labeled protein preparation costs around 1700 €/L culture; the sample cost then depends on the yields which can be achieved but for complex systems are often not above 1–5 mg/L. Costs for triply labeled samples are thus rather similar, until proteins express with high yield in bacteria. $^{13}$C/$^{15}$N labeling in WG-CFPS is not much less costly than triple labeling and reduces the cost of amino acids by only a factor of two, whereas in bacterial expression, the factor is nearly ten. Therefore, it makes most sense to use CFPS for complex proteins where bacterial expression fails and in cases where deuteriation is of high importance.

APPLICATION TO STRUCTURAL BIOLOGY

While CFPS is important for a variety of applications, we will give some examples from structural biology (Figure 6), where most experiments are done on recombinant proteins. While in many cases E. coli expression gives satisfactory results with respect to yield, it does not produce well-folded proteins in all cases. From experience, expression in cells corresponding to the origin of the protein (e.g., often mammalian cells) would be the most adapted approach with respect to correct folding; though, yield is often prohibitively low. CFPS, notably using eukaryotic systems, is a good compromise for mammalian proteins, providing sufficient yield and in many cases correct folding. The eukaryotic ribosomes are also ensuring a slower synthesis as compared to bacterial systems, which in turn promotes cotranslational folding. Furthermore, CFPS-generated proteins are easy to purify,
which is important for crystallographic studies. The WGS can easily provide around 50 μg of protein needed for cryo-EM and even more than the minimum of around 1 mg needed for X-ray crystallography and NMR. The following section illustrates some examples of how WGS was successfully applied in the past years for structural biology approaches, including solution and solid-state NMR, cryo-EM, and X-ray crystallography.

**NMR**

WGS is particularly attractive for its ability to produce complex eukaryotic (membrane) proteins for NMR and importantly combines this with the major advantage other CFPSs have, that is, the efficient and specific isotopic labeling required for NMR protein studies. Indeed, only the cell-free synthesized protein is isotopically labeled during expression (Morita et al., 2003); therefore, even if some remaining contaminants are present in the sample, they will be invisible in the NMR spectra. Amino acid selective labeling, easily implemented in CFPS by simply adding the desired amino acids into the reaction mixture, can also be used in WGS (Morita et al., 2004; Kohno and Endo, 2007; Tonelli et al., 2011; Fogeron et al., 2015a; Jirasko et al., 2020). Selective labeling results in significantly reduced NMR spectral complexity and enables application to higher molecular weight systems (>50 kDa) (Tugarnov et al., 2006). In the case of WGS, 15N-selective labeling was shown to be efficient for most amino acids, except for Ala, Glu, and Asp (Morita et al., 2004). For these residues, the addition of inhibitors for transaminases and glutamine synthase during protein synthesis is required to avoid scrambling (transfer of isotope labels between amino acids) (Morita et al., 2004). This method was first demonstrated on the RNA-binding protein RbpA1 and yeast ubiquitin (Morita et al., 2004) and then successfully applied to produce specific labeling schemes in β2-microglobulin (Kameda et al., 2009), a structural component of amyloid fibrils. The specific labeling scheme was crucial in the structural characterization of the refolding intermediate of β2-microglobulin and enabled to reveal the regions important for amyloidogenicity (Kameda et al., 2009). Another example is the 15N-Val selectively labeled yeast ubiquitin, where the four valine residues could be observed in the 1H–15N HSQC spectrum (Kohno, 2010).

For NMR studies, it is sometimes necessary to deuterate the protein, except for the amide protons, to improve the spectral linewidth. In the WGS, deuterated amino acids can be used, while the expression is done in H2O to obtain labeling. Hence, there is no need for a posteriori proton back exchange as in cell-based expression, since amide protonation is achieved directly during synthesis, avoiding a denaturation and refolding step, which can compromise the native fold of proteins. The usefulness of this approach was reported in the case of HBV capsids, where 20% of amide protons from the hydrophobic core are missing when bacterial expression is used (Lecoq et al., 2019), while they are present in samples prepared by WG-CFPS (Lecoq et al., 2019). It has been shown that metabolic scrambling cannot be avoided, and proton back exchange can occur on CH groups of Gly, Ala, Asp, Glu, Gln, and Lys (Tonelli et al., 2011). This problem can be alleviated by using similar transaminase inhibitors (Morita et al., 2004; Tonelli et al., 2011). The WGS and NMR were also applied to NS5A from the hepatitis C virus (HCV), where it revealed phosphorylation sites on the protein (Badillo et al., 2017).

With advances in MAS solid-state NMR, membrane proteins can be studied in lipids (reviewed in (Ladizhansky, 2017)). Recently, MAS frequencies exceeding 100 kHz have allowed structural investigation of submilligram amounts of protein (Agarwal et al., 2014; Andreas et al., 2016), as typically can be produced using WGS. This has enabled studies of HCV NS4B (Figure 7A) (Fogeron et al., 2016; Jirasko et al., 2020) and NS5A. NMR could show that the NS5A dimer in lipids presented a different orientation than in crystals and a model could be forwarded which proposes a binding mode for the directly acting antiviral Daclatasvir (Jirasko et al., 2020) (Figure 7B). It was also shown that high-quality NMR spectra can be obtained on HBV capsids (Figure 7C) and envelope proteins (the latter from the duck virus variant) proteins (Figure 7D) (David et al., 2018; Wang et al., 2019) which are spontaneously self-assembled in the WGS (Lingappa et al., 2005; David et al., 2018; Wang et al., 2019). For the HBV capsid, the combination of WGS and solid-state NMR allowed studying the effect of capsid assembly modulators at the exit from the ribosome (Wang et al., 2019).

Solution NMR applications of the WGS were advanced in the field of structural genomics by Markley and coworkers, who used it to screen 238 eukaryotic hypothetical proteins from *A. thaliana* and human genomes (Vinarov et al., 2004). Nearly half of these proteins were found to be soluble, and 40% yielded 1H–15N HSQC spectra indicative of folded proteins. Several solution NMR structures were solved of WG-CF-synthesized proteins, including the *A. thaliana* protein At3g01050.1 (Figure 7E) (Vinarov et al., 2004). A detailed comparison of *E. coli* expression with the WG-CFPS revealed that solubility and folding reached a higher success rate in the WGS (Tyler et al., 2005). The implementation of a high-throughput cell-free translation platform at the Center for Eukaryotic Structural Genomics enabled the use of WGS for fast screening and solution NMR structural determination (Makino et al., 2010; Makino et al., 2014; Vinarov and Markley, 2014).

Another recent work reports WG-CFPS of virtually all SARS-CoV-2 accessory proteins and M and E structural proteins, which showed that most of them can be produced and purified in soluble form and in milligram amounts (Altincelik et al., 2021).1

**X-Ray Crystallography and Cryo-EM**

The first X-ray structure of a protein expressed in the WGS was solved in 2007 (Miyazono et al., 2007; Watanabe et al., 2010), on the cytotoxic R.PabI (Figure 7F), one of the 4 bp cutter restriction enzymes, which are highly toxic to the cells. Recent literature shows further examples (Novikova et al., 2018), including the hexameric assembly of the dioxygen-concentrating mechanism protein (CCMK) (Novikova et al., 2018). Also, diffracting crystals of the glutamine synthetase from *O. tauri* complex were obtained from a crystallization screen using WGS (Novikova et al., 2018). Further negative-stain EM data provide indications that autoassemblies of larger superstructures are commonly observed when working with WGS, without showing aggregation nor disordered complexes, suggesting highly homogeneous samples compatible with high-resolution studies (Novikova et al., 2018). Furthermore, the EM envelope of the pyridoxal 5'-phosphate synthase-like subunit
(PDX1.2) from *A. thaliana* (Figure 7G) was solved (Novikova et al., 2018). As WGS application for cryo-EM is now just emerging, we expect a much wider use of WGS for preparation of samples for cryo-EM studies.

**DISCUSSION**

Today, CFPS is a mature technology that can serve most needs in protein expression. In recent years, the technology has made interesting contributions to the new fields of synthetic biology and the development of minimal cells. In contrast to *in vivo* expression systems, CFPS reactions enable manipulation of the reaction conditions in accordance with the requirements of a given protein. They can also be implemented in fully automated protein production to obtain larger yields or throughput and are thus a suitable approach to rapidly screen many mutations for changes to protein function and later production on a larger scale to obtain structural data. A former disadvantage of WG-CFPS used for structural studies was the lower yield when compared to bacterial cell-free or cell-based expression. However, structural studies currently need less protein, which makes these lower

**FIGURE 7** Examples of structural studies on proteins expressed in WGS using NMR, X-ray crystallography, and cryo-EM. (A) Solid-state NMR spectrum of the HCV membrane protein NS4B reconstituted into DMPC lipids (Jirasko et al., 2020). (B) Dimer orientation in lipids of the HCV helix anchor and domain 1 (AHD1) of the NSSA protein as determined by solid-state NMR (Jirasko et al., 2020). (C) Solid-state NMR spectra of the hepatitis B virus capsid (Wang et al., 2018) and of (D) the subviral particles made of duck HBV small envelope protein (DHBs S) (David et al., 2018). The three spectra have been recorded at 110 kHz MAS on an 850 MHz spectrometer. Both HBV capsids and subviral particles were autoassembled during cell-free synthesis; their negative-staining electron microscopy images are shown inside the corresponding spectrum. (E) 20 conformers obtained by solution NMR of At3g01050.1 protein (Vinarov et al., 2004) (PDB 1se9, figure prepared using PyMol (https://pymol.org/2/)). (F) Structure of restriction endonuclease PabI obtained by X-ray crystallography (Miyazono et al., 2007; Watanabe et al., 2010) (PDB 2dvy). (G) 3D cryo-EM reconstruction of PDX1.2 complex at 15 Å resolution (Novikova et al., 2018). Figures were adapted with permission from Jirasko et al. (2020) for panel A, from Wang et al. (2019) for panel C, and from David et al. (2018) for panel D and reprinted with permission from Jirasko et al. (2020) for panel D, from Miyazono et al. (2007), Watanabe et al. (2010) for panel F, and from Novikova et al. (2018) for panel G.
yields less of a barrier. Therefore, WGS today is fully compatible with high-resolution structural biology methods such as cryo-EM, X-ray crystallography, solution NMR, and solid-state NMR. The unique possibilities of this system for studies on cytotoxic proteins (Miyazono et al., 2007), complex membrane proteins (Fogeron et al., 2016; Badillo et al., 2017; David et al., 2018), or molecular assemblies in a native-close state shall thus create opportunities for structural approaches to study complex and difficult proteins. The examples given in this review demonstrate that CFPS production can be a powerful alternative to cell-based methods and thus could enable entirely new applications.

With recent developments in bioinformatics and methods of applied synthetic biology, we foresee rapid progress for new approaches to optimize protein expression using CFPS systems. These approaches will take advantage of computational protein design (Gustafsson et al., 2012), rapid template generation by gene synthesis, and working with linear DNA templates. Such linear templates can be used in fully automated and highly parallel testing of different template/protein designs and translation reaction conditions using robotic or microfluidic devices to perform effective expression tests yielding optimized proteins and conditions for their functional synthesis in a truly short timeframe (Ayoubi-Joshaghani et al., 2020; Borkowski et al., 2020). Conditions can subsequently be scaled to produce the protein for further use. This progress will largely be driven by the unique features of CFPS systems and the freedom they offer to adopt the system as this is already done for rational biodesigns (Laohakanakorn, 2020). Another important aspect for future developments will be the production of proteins having posttranslational modifications, where approaches have already been discussed for tyrosine and serine phosphorylation, lysine acetylation, and lysine methylation (Venkat et al., 2019). Similarly, we expect more progress on protein glycosylation and engineering glycosylation reactions in CFPS (Hershew et al., 2020; Kightlinger et al., 2020), a key aspect for making biologicals for treatment. With this outlook, we expect remarkably interesting development’s for new methods and applications that will involve a CFPS step as an essential part of the assay system. We hope that the WGS will contribute to these developments as it proved to be one of the most effective eukaryotic CFPS systems.

AUTHOR CONTRIBUTIONS

MLF, LL, MH, and AB wrote the manuscript, with input from I.C.

FUNDING

This research was supported by the CNRS (CNRS-Momentum 2018) and the French Agence Nationale de Recherches sur le Sida et les Hépatites virales (ANRS, ECTZ71388, and ECTZ100488).

ACKNOWLEDGMENTS

We are grateful to F. Tanabe and R. Morishita for sharing unpublished results and to E. Beebe, K. Kami, R. Morishita, and B.H. Meier for critically reading and advising on the manuscript.

REFERENCES

Abe, M., Hori, H., Nakanishi, T., Arisaka, F., Ogasawara, T., Sawasaki, T., et al. (2004). Application of cell-free translation systems to studies of cofactor binding proteins. Nucleic Acids Symp. Ser. 13, 143–144. doi:10.1039/nasi48.1.143

Aceti, D. J., Bingman, C. A., Wrobel, R. L., Frederick, R. O., Makino, S., Nichols, K. W., et al. (2015). Expression platforms for producing eukaryotic proteins: a comparison of E. coli cell-based and wheat germ cell-free synthesis, affinity and solubility tags, and cloning strategies. J. Struct. Funct. Genomics 16, 67–80. doi:10.1016/j.sfg.2015.01-0198-1

Agarwal, V., Penzel, S., Szekely, K., Cadalbert, R., Testori, E., Oss, A., et al. (2014). De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy. Angew. Chem. Int. Ed. Engl. 53, 12233–12236. doi:10.1002/anie.201405730

Akbarzadeh, A., Rezaei-Sadabad, R., Davaran, S., Joo, S. W., Zarghani, N., Hanifehpour, Y., et al. (2015). Liposome: classification, preparation, and applications. Nanoscale Res. Lett. 10, 102. doi:10.1186/1556-276X-10-102

Althoff, T., Davies, K. M., Schulze, S., Joos, F., and Kühlbrandt, W. (2012). GRecon: a method for the lipid reconstitution of membrane proteins. Angew. Chem. Int. Ed. Engl. 51, 8343–8347. doi:10.1002/anie.2012005730

Aly, K. A., Beebe, E. T., Chan, C. H., Goren, M. A., Sepúlveda, C., Makino, S., et al. (2012). Cell-free production of integral membrane aspartic acid proteases reveals zinc-dependent methyltransferase activity of the Pseudomonas aeruginosa pre-protein peptidase PpID. MicrobiologyOpen 2, 94–104. doi:10.1002/mbo3.51

Anastasina, M., Terenin, I., Butcher, S. J., and Kainov, D. E. (2014). A technique to increase protein yield in a rabbit reticulocyte lysate translation system. Biotechniques 56, 36–39. doi:10.2144/000114125

Andreas, L. B., Jaudzems, K., Stanek, J., Lalli, D., Bertarello, A., Le Marchand, T., et al. (2016). Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. Proc. Natl. Acad. Sci. U.S.A. 113, 9187–9192. doi:10.1073/pnas.1602248113

Aoki, M., Matsuda, T., Tomo, Y., Miyata, Y., Inoue, M., Kigawa, T., et al. (2009). Automated system for high-throughput protein production using the dialysis cell-free method. Protein Expr. Purif. 68, 128–136. doi:10.1016/j.pep.2009.07.017

Arumugam, T. U., Ito, D., Sawasaki, T., Ikeda, Y., Hiasa, Y., et al. (2014). The ligand binding ability of dopamine D1 receptors synthesized using a wheat germ cell-free protein synthesis system with liposomes. Eur. J. Pharmacol. 745, 117–122. doi:10.1016/j.ejphar.2014.10.011

Arrinamipathy, Y., Khurana, E., Engelman, D. M., and Gerstein, M. B. (2009). Computational analysis of membrane proteins: the largest class of drug targets. Drug Discov. Today 14, 1130–1135. doi:10.1016/j.drudis.2009.08.006

Arranz-Gibert, P., Vanderschuren, K., and Isaacs, F. J. (2018). Next-generation genetic code expansion. Curr. Opin. Chem. Biol. 46, 203–211. doi:10.1016/j.cbpa.2018.07.020

Arumugam, T. U., Ito, D., Takashima, E., Tachibana, M., Ishino, T., Torii, M., et al. (2014). Application of wheat germ cell-free protein expression system for novel malaria vaccine candidate discovery. Expert Rev. Vaccin. 13, 75–85. doi:10.1586/14760854.2014.861777

Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L. V., Katneni, U., et al. (2017). A new and updated resource for codon usage tables. BMC Biol. 18, 391. doi:10.1186/s12862-017-1793-7

Ayoubi-Joshaghani, M. H., Dianat-Moghadam, H., Seidi, K., Jahanban-Esfahalan, A., Zare, P., and Jahanban-Esfahalan, R. (2020). Cell-free protein synthesis: the transition from batch reactions to minimal cells and microfluidic devices. Biotechnol. Bioeng. 117, 1204–1229. doi:10.1002/bit.27248
using detergents and thioredoxin-fusion vectors. *Protein Expr. Purif.* 41, 27–37. doi:10.1016/j.pep.2005.01.013

Jackson, K., Jin, S., and Fan, Z. H. (2015). Optimization of a miniaturized fluid array device for cell-free protein synthesis. *Biotechnol. Bioeng.* 112, 2459–2467. doi:10.1002/bit.25668

Jackson, K., Khnouf, R., and Fan, Z. H. (2014). Cell-free protein synthesis in microfluidic 96-well plates. *Methods Mol. Biol.* 1118, 157–168. doi:10.1007/978-1-62703-782-2_10

Jackson, R. C., and Blobel, G. (1977). Post-translational cleavage of presecretory proteins with an extract of rough microsomes from dog pancreas containing signal peptidase activity. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5598–5602. doi:10.1073/pnas.74.12.5598

Jarecki, B. W., Makino, S., Beebe, E. T., Fox, B. G., and Chanda, B. (2013). Function of Shaker potassium channels produced by cell-free translation upon injection into Xenopus oocytes. *Sci. Rep.* 3, 1040. doi:10.1038/srep01040

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro rhodopsin 2 and its H30A mutant. *J. Struct. Biol.* 206, 55–65. doi:10.1016/j.jsb.2018.06.001

Kawasaki, T., Gouda, M. D., Sawasaki, T., Takai, K., and Endo, Y. (2003). Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. *Eur. J. Biochem.* 270, 4780–4786. doi:10.1046/j.1432-1323.2003.03880.x

Kelly, S. M., Jess, T. J., and Price, N. C. (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta* 1751, 119–139. doi:10.1016/j.bbapap.2005.06.005

Khan, F., He, M., and Taussig, M. J. (2006). Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on nitrocellulose acid surfaces. *Anal. Chem.* 78, 3072–3079. doi:10.1021/ac060184l

Khnouf, R., Olivero, D., Jin, S., Coleman, M. A., and Fan, Z. H. (2010). Cell-free expression of soluble and membrane proteins in an array device for drug screening. *Anal. Chem.* 82, 7021–7026. doi:10.1021/ac1015479

Ki, M. R., and Pack, S. P. (2020). Fusion tags to enhance heterologous protein expression. *Appl. Microbiol. Biotechnol.* 104, 2411–2425. doi:10.1007/s00253-020-10402-8

Kido, K., Yamanaoka, S., Nakano, S., Motani, K., Shinohara, S., Nozawa, A., et al. (2020). AirID, a novel proximity biotinylation enzyme, for analysis of protein-protein interactions. *Elife* 9, 11355. doi:10.7554/eLife.54993

Kißlinger, W., Waltel, K. F., Delia, M. P., and Jess, T. J. (2020). Synthetic glycochemistry: parts, systems, and applications. *ACS synth.* 9, 1534–1562. doi:10.1021/acssynbio.0c00120

Kim, D. M., and Swartz, J. R. (2003). Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of *Escherichia coli*. *Biotechnol. Bioeng.* 85, 122–129. doi:10.1002/bit.10865

Kim, H. C., Kim, K. S., Kang, T. J., Choi, J. H., Song, J. J., Choi, Y. H., et al. (2015). Implementing bacterial acid resistance into cell-free protein synthesis for buffer-free expression and screening of enzymes. *Biotechnol. Bioeng.* 112, 2630–2635. doi:10.1002/bit.25671

Kimple, M. E., Brill, A. L., and Pasker, R. L. (2013). Overview of affinity tags for protein purification. *Curr. Protoc. Protein Sci.* 73, 9. doi:10.1002/0471140864.ps09967

Klammt, C., Lohr, F., Schäfer, B., Haase, W., Dötsch, V., Rüterjans, H., et al. (2004). High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* 271, 568–580. doi:10.1111/j.1432-1323.2003.03959.x

Klammt, C., Srivastava, A., Elfner, N., Junge, F., Meyermann, M., Schwarz, D., et al. (2007). Functional analysis of cell-free-produced human endothelin B receptor reveals transmembrane segment 1 as an essential area for ET-1 binding and homodimer formation. *FEBS J.* 274, 3257–3269. doi:10.1111/j.1745-6649.2007.0585x.x

Klammt, C., Schwarz, D., Fendler, K., Haase, W., Dötsch, V., and Bernhard, F. (2005). Evaluation of detergents for the soluble expression of a-helical and B-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J.* 272, 6024–6038. doi:10.1111/j.1745-6648.2005.05020.x

Koblikka, B. K. (1990). The role of cytotoxic and membrane factors in processing of the human beta-2 adrenergic receptor following translocation and glycosylation in a cell-free system. *J. Biol. Chem.* 265, 7610–7618. doi:10.1074/jbc.265.12.39158-6

Kögler, L. M., Stichel, J., and Beck-Sickinger, A. G. (2019). Structural investigations of cell-free expressed G protein-coupled receptors. *Biotechnol. Bioeng.* 401, 97–116. doi:10.1002/bit.10865

Kohn, T., and Endo, Y. (2007). Production of protein for nuclear magnetic resonance study using the wheat germ cell-free system. *Methods Mol. Biol.* 375, 257–272. doi:10.1007/978-1-59745-388-2_13

Kohn, T. (2010). NMR assignment method for amide signals with cell-free protein synthesis system. *Methods Mol. Biol.* 607, 113–126. doi:10.1007/978-1-60327-331-2_11

Kralick, A. (2014). A cell-free expression screen to identify fusion tags for improved protein expression. *Methods Mol. Biol.* 1118, 35–54. doi:10.1007/978-1-62703-782-2_3

Krug, U., Gloge, A., Schmidt, P., Becker-Baldus, J., Bernhard, F., Kaiser, A., et al. (2020). The conformational equilibrium of the neproptide Y2 receptor in bilayer membranes. *Angew. Chem. Int. Ed. Engl.* 59, 23854–23861. doi:10.1002/anie.202006075
protein to lipid bilayers. FEBS Lett. 501, 115–120. doi:10.1016/S0014-5793(01)02627-8

Nagy, S. K., Kállai, B. M., András, J., and Mészáros, T. (2020). A novel family of expression vectors with multiple affinity tags for wheat germ cell-free protein expression, Berlin: Springer, 1–9.

Nemoto, K., Kagawa, M., Nozawa, A., Hasegawa, Y., Hayashi, M., Imai, K., et al. (2018). Identification of new abasic acid receptor agonists using a wheat cell-free based drug screening system. Sci. Rep. 8, 4268. doi:10.1038/s41598-018-22538-9

Niwa, T., Nanamiya, H., Miyata, T., Linka, N., Endo, Y., Weber, A. P., et al. (2010). Production and partial purification of membrane proteins using a liposome-supplemented wheat cell-free translation system. BMC Biotechnol. 11, 35. doi:10.1186/1472-6750-11-35

Nozawa, A., and Tozawa, Y. (2014a). Incorporation of adenine nucleotide transporter, Ant1p, into liposomes facilitates ATP translocation and activation of encapsulated luciferase. J. Biosci. Bioeng. 118, 130–133. doi:10.1016/j.jbiosc.2014.02.001

Nozawa, A., Nanamiya, H., Miyata, T., Linka, N., Endo, Y., Weber, A. P., et al. (2007). A cell-free translation and proteoliposome reconstitution system for functional analysis of plant solute transporters. Plant Cell Physiol. 48, 1815–1820. doi:10.1093/pcp/pcm150

Ong, Y. S., Lakatos, A., Becker-Baldus, J., Pos, K. M., and Glaubitz, C. (2013). Reconstitution of Drosophila and human chromatins by wheat germ cell-free translation. BMC Biotechnol. 20, 62. doi:10.1186/1472-6879-13-6655-6

Ong, Y. S., Lakatos, A., Becker-Baldus, J., Pos, K. M., and Glaubitz, C. (2013). Detecting substrates bound to the secondary multidrug efflux pump EmrE by DNP-enhanced solid-state NMR. J. Am. Chem. Soc. 135, 15754–15762. doi:10.1021/ja402605s

Oza, J. P., Aerni, H. R., Pirman, N. L., Barber, K. W., Ter Haar, C. M., Rogulina, S., et al. (2018). Proteoliposome reconstitution in nanodiscs as photoactive monomers. Biochemistry 47, 12047–12056. doi:10.1021/acs.biochem.8b01604x

Rawlings, A. E. (2018). Membrane protein engineering to the rescue. Biochem. Soc. Trans. 46, 1541–1549. doi:10.1042/BST20181040

Rhee-Hoyes, J. S., and Walhout, A. J. M. (2018). Gateway recombinatorial cloning. Cold Spring Harb. Protoc. 12, 114. Revathi, B. R., K., V. A., Hinnebusch, A. G., Bhuyan, A. K., Ksr, S., Sudhakar, A., et al. (2010). Trends in wheat germ cell free protein expression system with an emphasis on up-scaling and industrial application. Indian J. Sci. Technol. 3, 349–354.

Rigaud, J. L. (2002). Membrane proteins: functional and structural studies using reconstituted proteoliposomes and 2-D crystals. Braz. J. Med. Biol. Res. 35, 753–766. doi:10.1590/S0100-879X2002000700001

Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolneric, K. J., Atkins, W. M., et al. (2009). Chapter 11 - reconstitution of membrane proteins in phospholipid bilayer nanodiscs. Meth Enzymol. 464, 211–231. doi:10.1016/S0070-0252(06)67974-3

Romano, V., Davidoff, S. N., Miles, A. R., Grainger, D. W., Gale, B. K., and Brooks, B. D. (2014). A critical comparison of protein microarray fabrication technologies. Analyst 139, 1303–1326. doi:10.1039/c3an06177g

Rosenberg, R. L., and East, J. E. (1992). Cell-free expression of functional Shaker potassium channels. Nature 360, 166–169. doi:10.1038/360166a0

Rosenblum, G., and Cooperman, B. S. (2014). Engine out of the chassis: cell-free protein synthesis and its uses. FEBS Lett. 588, 261–268. doi:10.1016/j.febslet.2013.10.016

Playing with new surfactants. N. Biotechnol. 28, 255–261. doi:10.1016/j.nbt.2010.08.008

Peri, I. G., Stark, J. C., and Jewett, M. C. (2016). Cell-free synthetic biology: engineering beyond the cell. Cold Spring Harb Perspect. Biol. 8, 121. doi:10.1101/cshperspect.a023853

Periasamy, A., Shadiac, N., Amalraj, A., Garajov, S., Nagarajan, Y., Waters, S., et al. (2013). Cell-free protein synthesis of membrane (1,3)-β-glucan (curdlan) synthase: co- translational insertion in liposomes and reconstitution in nanodiscs. Biochim. Biophys. Acta 1828, 743–757. doi:10.1016/j.bbamem.2012.10.003

Popot, J. L. (2010). Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solutions. Annu. Rev. Biochem. 79, 737–775. doi:10.1146/annurev.biochem.052208.114057

Punta, M., Forrest, R. L., Bigelow, H., Kennyts, A., Liu, J., and Rost, B. (2007). Membrane protein prediction methods. Methods 41, 460–474. doi:10.1016/j.ymeth.2006.07.026

Quast, R. B., Mursiek, D., Hoffmeister, C., Sonnabend, A., and Kubick, S. (2015). Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis. FEBS Lett. 589, 1703–1712. doi:10.1016/j.febslet.2015.04.041

Quast, R. B., Sonnabend, A., Stech, M., Wustenhagen, D. A., and Kubick, S. (2016). High-yield cell-free synthesis of human EGFR by IRES-mediated protein translation in a continuous exchange cell-free reaction format. Sci. Rep. 6, 30399. doi:10.1038/srep30399

Quinn, C. M., Wang, M., Fritz, M. P., Runge, B., Ahn, J., Xu, C., et al. (2018). Dynamic regulation of HIV-1 capsid interaction with the restriction factor TRIM5α identified by magic-angle spinning NMR and molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A. 115, 11519–11524. doi:10.1073/pnas.1807961115

Rigaud, J. L. (2002). Membrane proteins: functional and structural studies using reconstituted proteoliposomes and 2-D crystals. Braz. J. Med. Biol. Res. 35, 753–766. doi:10.1590/S0100-879X2002000700001

Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerkics, K. J., Atkins, W. M., et al. (2009). Chapter 11 - reconstitution of membrane proteins in phospholipid bilayer nanodiscs. Meth Enzymol. 464, 211–231. doi:10.1016/S0070-0252(06)67974-3

Romano, V., Davidoff, S. N., Miles, A. R., Grainger, D. W., Gale, B. K., and Brooks, B. D. (2014). A critical comparison of protein microarray fabrication technologies. Analyst 139, 1303–1326. doi:10.1039/c3an06177g

Rosenberg, R. L., and East, J. E. (1992). Cell-free expression of functional Shaker potassium channels. Nature 360, 166–169. doi:10.1038/360166a0

Rosenblum, G., and Cooperman, B. S. (2014). Engine out of the chassis: cell-free protein synthesis and its uses. FEBS Lett. 588, 261–268. doi:10.1016/j.febslet.2013.10.016

Rosenblum, G., and Cooperman, B. S. (2014). Engine out of the chassis: cell-free protein synthesis and its uses. FEBS Lett. 588, 261–268. doi:10.1016/j.febslet.2013.10.016

Rosenblum, G., and Cooperman, B. S. (2014). Engine out of the chassis: cell-free protein synthesis and its uses. FEBS Lett. 588, 261–268. doi:10.1016/j.febslet.2013.10.016
Zhu, B., Gan, R., Cabezas, M. D., Kojima, T., Nicol, R., Jewett, M. C., et al. (2020). Increasing cell-free gene expression yields from linear templates in Escherichia coli and Vibrio natriegens extracts by using DNA-binding proteins. *Biotechnol. Bioeng.* 12, 157. doi:10.1101/2020.07.22.214221

Zhu, G., Zhu, C., Zhu, Y., and Sun, F. (2020). Minireview of progress in the structural study of SARS-CoV-2 proteins. *Curr. Res. Microb. Sci.* 1, 53–61. doi:10.1016/j.crmicr.2020.06.003

Zucchelli, S., Cotella, D., Takahashi, H., Carrieri, C., Cimatti, L., Fasolo, F., et al. (2015a). SINEUPs: a new class of natural and synthetic antisense long non-coding RNAs that activate translation. *RNA Biol.* 12, 771–779. doi:10.1080/15476286.2015.1060395

Zucchelli, S., Fasolo, F., Russo, R., Cimatti, L., Patrucco, L., Takahashi, H., et al. (2015b). SINEUPs are modular antisense long non-coding RNAs that increase synthesis of target proteins in cells. *Front Cell Neurosci.* 9, 174. doi:10.3389/fncel.2015.00174

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi:10.1093/nar/gkg595

**Conflict of Interest:** MH declares competing financial interests as an employee of CellFree Sciences Co., Ltd., Japan. The other authors do not have any competing interests.

Copyright © 2021 Fogeron, Lecoq, Cole, Harbers and Böckmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.