



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:
Hilditch, Alex

Title:
Assembly and functionalisation of membraneless organelles from de novo designed proteins

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Assembly and functionalisation of membraneless organelles from *de novo* designed proteins

Alexander T. Hilditch



A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Life Sciences.

School of Biochemistry, University of Bristol
September 2023

Word count: 39204

Abstract

In recent years, protein design has undergone a revolution, giving rise to increasingly complex protein assemblies with atomistic accuracy. This has unlocked the potential to design sophisticated macromolecular complexes that can emulate biological systems. As one of the foundations of cellular organisation, the design of artificial organelles is a key target for synthetic biologists. Designed assemblies emulating bacterial microcompartments or fibrous scaffolds have been accessible for some time. A less-explored mechanism for protein assembly is by liquid-liquid phase separation. Despite its newfound relevance to cell biology, the design of protein assemblies that can reversibly de-mix is still in its infancy.

In this thesis, the bottom-up design of proteins for liquid-liquid phase separation is described. These proteins use *de novo* α -helical coiled coils as protein-protein interaction motifs, combined with unstructured linkers, to drive self-assembly in cells. Using rational coiled coil design principles, the interactions between these motifs are calibrated to drive dynamic liquid-liquid de-mixing, over arrested aggregation. Soft-matter biophysical techniques confirm the liquid-like properties of the *de novo* assemblies both *in vitro* and in *Escherichia coli* (*E. coli*). Moreover, the designer condensates can be functionalised with catalytic enzymes. Co-condensation of the two enzyme pathway for the production of indigo dye in the *de novo* organelles results in 6-fold more product than the comparable free enzymes.

In addition to phase separation, the formation of macromolecular fibres by the designed proteins is described. Analogous to droplet maturation, the designed proteins can form both fibres and de-mixed droplets in cells. Finally, following functionalisation in bacteria, phase separation of the designed helical proteins is investigated in mammalian cells. Here, the assemblies are re-designed for phase separation within the larger, and less crowded, eukaryotic cytoplasm. Taken together, these studies begin to highlight the power of bottom-up design for the construction and rationalisation of protein condensates.

Acknowledgements

The greatest acknowledgement is to my wife, Sofia, without whom I would be infinitely more lost. Thank you for your patience and support as we both worked on our PhDs together. I am incredibly proud to have done this with you.

Thank you to Dek, for hosting me and guiding me through my PhD. I feel like I have developed a lot personally, and as a scientist, in the past 4 years, and I hope to carry on the ideas that I've built here into my work in the future. Thanks also to Jen, for your excellent guidance as we attempted to learn about phase separation, and for bridging the gap between chemistry and physics.

Thank you to past and present members of the Max Planck Bristol Centre for Minimal Biology. We may be small, but it was a pleasure to work with all of you. Particular thanks go to Richard, who has been an excellent mentor and guide through academic life. Without you this project would have never gotten off the ground. I enjoyed our fun experiments, and building the lab together. Thanks also to Andrey, for being a partner in working on this wonderful thing called phase separation, and Bram, for the tireless discussions and ideas. Thanks also to Matt, Holly, and Dora, who made the lab what it is, and were always there to help.

Thanks also to members of the Woolfson group, both past and present. You've all created an amazing, hard-working atmosphere where I felt welcomed, and I will miss working with you. Special thanks to Will, for being the lab manager we didn't deserve, Katherine, for ideas and discussions, as well as reading and commenting on this thesis without complaint, and Jess, for teaching me to be a better scientist, and to love prosecco.

Special thanks also go to my family. To my mum Corinne, dad Paul, brother Ben, and sister Angela. Thank you all for your support and encouragement through this journey. It's been incredibly valuable, and I appreciate everything that you have all done for us. Although I never intended it, it's been fun to follow in your footsteps, and continue the tradition of studying for a PhD in Biochemistry.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed:

Date:

Table of Contents

Abstract	i
Acknowledgements	ii
Author's declaration	ii
Table of Contents	viii
List of Figures	ix
List of Tables	xii
Abbreviations	xiii
1 Introduction	1
1.1 The cellular organisation problem	1
1.1.1 Protein-based cellular organisation	2
1.2 Protein condensates - a new paradigm	4
1.3 Phase separation in biology	5
1.3.1 A diverse and widespread phenomenon	7
1.4 Phase separation in disease	9
1.4.1 Phase separation in neurodegeneration	9
1.4.2 Phase separation in cancer and infectious diseases	10
1.5 Molecular principles of phase separation	11
1.5.1 Multivalency and modular domains	11
1.5.2 Weak intermolecular interactions	13
1.6 Protein design meets biology	13

1.6.1	Rational protein design	14
1.6.2	Computational protein design	15
1.6.3	Coiled coil design	17
1.7	Design of assemblies for cellular organisation	20
1.7.1	Designed geometric protein assemblies	21
1.7.2	Designer protein condensates	22
1.8	Scope of this thesis	26
2	Materials & methods	28
2.1	Materials	28
2.1.1	Molecular biology	28
2.1.2	Oligonucleotides	29
2.1.3	Mammalian cell culture	29
2.1.4	Instrumentation	29
2.2	Recombinant protein expression	30
2.2.1	Molecular biology and DNA assembly	30
2.2.2	Protein expression and purification	31
2.2.3	TEV cleavage	32
2.3	Cell culture	33
2.3.1	Mammalian cell culture	33
2.4	Microscopy and protein detection	33
2.4.1	<i>E. coli</i> cell imaging	33
2.4.2	Mammalian cell imaging	34
2.4.3	Automated image analysis	34
2.4.4	TC-ReAsH II labelling	34
2.4.5	Western blotting	35
2.4.6	Transmission electron microscopy	36
2.5	Soft-matter biophysics	37
2.5.1	Cloud-point measurements	37
2.5.2	Fluorescence recovery after photobleaching	37
2.6	Enzymology measurements	38
2.6.1	Measurement of indigo production	38

2.7	Structural biophysics	39
2.7.1	Solid phase peptide synthesis	39
2.7.2	Peptide purification	39
2.7.3	Circular dichroism spectroscopy	39
2.7.4	X-ray fibre diffraction	40
2.8	Computational tools	40
3	Rational design of proteins for LLPS	42
3.1	Chapter introduction	42
3.2	Bottom-up design of <i>de novo</i> stickers and spacers	44
3.2.1	Coiled-coil stickers	45
3.2.2	Unstructured spacers	48
3.3	HERD proteins form subcellular protein condensates	51
3.3.1	Destabilisation of HERD proteins creates soluble condensates	52
3.3.2	HERD-2.2–GFP undergoes a temperature-dependent phase transition	54
3.3.3	The helical regions of HERD-2.2 are unstructured in solution	57
3.3.4	Helical interactions of HERD-2.2 are important for condensation in cells	60
3.3.5	AlphaFold2 predicts α -helical secondary structure for HERD-2.2	60
3.3.6	Changes to the linker do not significantly alter condensation	61
3.4	HERD-2.2–GFP undergoes LLPS <i>in vitro</i>	62
3.4.1	HERD-2.2–GFP droplets are reversible and dynamic	63
3.4.2	HERD-2.2–GFP still undergoes LLPS following TEV cleavage	64
3.4.3	The LLPS boundary of HERD-2.2–GFP can be mapped	65
3.5	HERD-2.2–GFP droplets recover rapidly after bleaching	68
3.5.1	FRAP <i>in vitro</i> gives rapid and complete recovery of fluorescence	69
3.5.2	HERD-2.2–GFP produces liquid-like condensates in cells	69
3.5.3	HERD-0–GFP produces immobile aggregates in <i>E. coli</i>	71
3.6	Chapter summary	71
4	Functionalisation of <i>de novo</i> MLOs	76
4.1	Chapter introduction	76

4.2	HERD-2.2 can colocalise multiple proteins in <i>E. coli</i>	77
4.3	Enzymatic production of indigo as an example pathway	78
4.3.1	HERD-2.2–TnaA but not FMO undergoes LLPS <i>in vitro</i>	80
4.3.2	HERD-2.2–TnaA and FMO are enriched in HERD-2.2–GFP droplets	81
4.4	Indigo production in HERD-2.2 condensates is sensitive to enzyme loading	84
4.5	HERD-2.2–GFP condensates accelerate pathway efficiency	86
4.5.1	HERD-0–GFP condensates restrict indigo production	88
4.6	Pathway efficiency correlates with material properties	88
4.6.1	HERD-2.2 condensates grown at 18 °C do not recover after bleaching	89
4.6.2	HERD-2.2 condensates grown at 18 °C inhibit enzymatic activity	89
4.7	HERD-2.2 re-design boosts FMO co-condensation and indigo production	91
4.8	Chapter summary	95
5	Investigation into fibre formation by HERD proteins	97
5.1	Chapter introduction	97
5.2	HERD-2.2 can form fibres as well as drive LLPS	98
5.2.1	HERD-2.2 forms protein fibres 13 nm in diameter	99
5.2.2	HERD-2.2-E forms fibres within the <i>E. coli</i> cytoplasm	101
5.3	HERD-2.2–GFP can be recruited to HERD fibres	103
5.3.1	Cells expressing HERD-2.2 fibres become elongated	103
5.4	CD spectroscopy of HERD-2.2 fibres	105
5.4.1	X-ray fibre diffraction indicates β secondary structure	107
5.5	HERD-2.2-E fibres can pattern LLPS droplets	108
5.6	Fibres are not nucleated from HERD-2.2–GFP droplets	108
5.7	Chapter summary	110
6	<i>De novo</i> proteins for LLPS in eukaryotes	112
6.1	Chapter introduction	112
6.2	HERD-2.2–GFP does not phase separate in HeLa cells	113
6.3	HERD re-design is required for LLPS in mammalian cells	115
6.3.1	Helical regions with tighter affinities	115
6.3.2	Additional stickers and spacers	116

6.3.3	Longer coiled coils for LLPS	116
6.4	LLPS and membrane localisation	117
6.4.1	HERD designs can be localised to the plasma membrane	120
6.5	Intramolecular interactions may inhibit phase separation	122
6.6	Chapter summary	122
7	Discussion and future work	125
7.1	Principal conclusions of this thesis	125
7.1.1	<i>De novo</i> design delivers proteins for LLPS	125
7.1.2	<i>De novo</i> droplets are highly efficient reaction crucibles	127
7.1.3	The interplay between LLPS and fibre formation remains elusive	129
7.1.4	Alternative designs are required for LLPS in mammalian cells	131
7.2	Possible directions for future research beyond this thesis	132
7.2.1	Globular proteins for LLPS	132
7.2.2	Predictive computational frameworks for phase separation	134
	Bibliography	137
8	Appendix	159

List of Figures

1.1	The crowded intracellular environment	2
1.2	Levels of cellular organisation	3
1.3	Protein-based cellular organisation	4
1.4	Liquid-liquid phase separation in biology	5
1.5	Phase diagram for LLPS	6
1.6	The rapid expansion of liquid-liquid phase separation (LLPS) as an organising principle in biology	8
1.7	Phase separation in disease	10
1.8	Varied interactions leading to phase separation	12
1.9	A gallery of structures developed by protein design	14
1.10	Steps in structural prediction of proteins using machine learning	16
1.11	Coiled coils in protein design	18
1.12	Engineered protein nanoparticles and fibres	22
1.13	Design and engineering of protein condensates	25
3.1	Workflow for the design and characterisation of proteins for LLPS	43
3.2	The stickers-and-spacers model of phase-separating proteins	44
3.3	Orthogonal coiled coils as peptide stickers	45
3.4	Theoretical pI distribution of the <i>E. coli</i> proteome	46
3.5	Linker effective solvation volume models	48
3.6	The designed linker, and disorder prediction of the coiled coils and linker by IUPred3	49
3.7	HERD protein architecture	50
3.8	HERD-0–GFP forms fluorescent subcellular condensates	51
3.9	Destabilisation of <i>de novo</i> coiled-coil stickers	52

3.10	Destabilised HERD proteins create soluble condensates	53
3.11	Severely disrupted helical regions produce fewer condensates	54
3.12	Automated image analysis of condensate formation in <i>E. coli</i>	55
3.13	HERD-2.2-GFP condensation in cells is temperature dependent	56
3.14	Fully mutated HERD constructs do not form condensates	57
3.15	The helical regions of HERD-2.2 are unstructured in solution	58
3.16	Helical interactions appear important for condensation of HERD-2.2-GFP	59
3.17	AlphaFold2 predicts α -helical secondary structure for HERD-2.2-GFP . .	61
3.18	Changes to the linker do not significantly alter condensation	62
3.19	HERD-2.2-GFP forms de-mixed droplets <i>in vitro</i>	63
3.20	HERD-2.2-GFP droplets are reversible and dynamic	64
3.21	HERD-2.2-GFP still undergoes LLPS following TEV cleavage	65
3.22	Measurements of the binodal phase boundary for HERD-2.2-GFP	66
3.23	HERD-2.2-GFP droplets recover rapidly after photobleaching <i>in vitro</i> . .	67
3.24	HERD-2.2-GFP forms liquid-like droplets in cells	68
3.25	HERD-0-GFP forms immobile aggregates in <i>E. coli</i>	70
3.26	Graphical representation of HERD designs 0 through 2.6	74
3.27	Graphical representation of HERD designs 2.7 through 3.4, and Ctrl1 and 2	75
4.1	HERD-2.2-GFP colocalisation of GFP and mCherry	78
4.2	HERD-2.2-GFP colocalisation of functional enzymes	79
4.3	Enzymatic production of indigo dye	80
4.4	HERD-2.2-TnaA de-mixed droplets <i>in vitro</i>	81
4.5	Enzymatic production of indigo in <i>de novo</i> condensates	82
4.6	HERD-2.2-TnaA and HERD-2.2-FMO can be co-condensed <i>in vitro</i> and in cells	83
4.7	HERD-2.2-GFP condensates are sensitive to enzyme loading	84
4.8	Indigo production is accelerated by condensation	85
4.9	HERD-2.2 labelled enzymes produce more than 2-fold more indigo	87
4.10	HERD-2.2-GFP condensates grown at 18 °C are not dynamic	89
4.11	HERD-2.2 condensates perform worse than free enzymes at 18 °C	90
4.12	HERD-2.2 charge re-design boosts FMO co-condensation	92

4.13	HERD-4.1 based condensates improve indigo production 6-fold	93
4.14	<i>De novo</i> condensates can accelerate enzymology	94
5.1	Maturation of LLPS is due to its metastability	98
5.2	HERD-2.2 forms micron length fibres <i>in vitro</i>	99
5.3	HERD-2.2 forms fibres 13 nm in diameter	100
5.4	HERD-2.2-E forms fibres at a shifted pH	101
5.5	HERD-2.2-E forms fibres <i>in vitro</i> and in cells	102
5.6	HERD-2.2–GFP can be recruited to HERD fibres in cells	104
5.7	Cells expressing HERD-2.2 fibres are elongated	105
5.8	CD spectroscopy of HERD-2.2 fibres indicates secondary structure	106
5.9	X-ray fibre diffraction of HERD-2.2 fibres	107
5.10	HERD-2.2-E fibres can pattern LLPS droplets	109
5.11	TEV cleavage of GFP does not induce fibre formation	110
6.1	HERD-2.2–GFP does not phase separate in HeLa cells	113
6.2	HERD re-design to introduce additional valency or affinity	114
6.3	Neither HERD-0–GFP nor HERD-2.2-XL–GFP phase separate in HeLa cells	115
6.4	Longer coiled coils do not drive condensation	116
6.5	Expression of HERD-5.4 through 5.7–GFP in HeLa	118
6.6	LLPS and membrane interactions	119
6.7	HERD-2.2–GFP can be localised to the plasma membrane	121
6.8	Intramolecular interactions may be inhibiting phase separation	123
7.1	<i>De novo</i> design delivers solutions for LLPS	127
7.2	Redesign of proteins for LLPS in eukaryotes	130
7.3	Design of globular proteins for LLPS	133
7.4	Creation of a predictive computational framework for LLPS	135

List of Tables

2.1	Oligonucleotides	29
3.1	Theoretical pI of <i>de novo</i> coiled coils	47
8.1	Sequences of the designed HERD proteins used in Chapter 3	160
8.2	Sequences of the designed HERD proteins used in Chapter 4	161
8.3	Sequences of the designed HERD proteins used in Chapter 5	161
8.4	Sequences of the designed HERD proteins used in Chapter 6	162
8.5	Additional protein sequences used in this thesis	163

Abbreviations

C. elegans *Caenorhabditis elegans*

E. coli *Escherichia coli*

S. cerevisiae *Saccharomyces cerevisiae*

AI Artificial intelligence

ALS Amyotrophic lateral sclerosis

AMP Ampicillin

APS Ammonium persulfate

Bis-tris 2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol

BMC Bacterial microcompartment

BMV Bacterial microcompartment vertex

CASP Critical assessment of structure prediction

CD Circular dichroism

CLEM Correlative light-electron microscopy

CMP Chloramphenicol

CRISPR Clustered regularly interspaced short palindromic repeat

Csat Saturation concentration

DMF Dimethylformamide

DNA Deoxyribonucleic acid

dNTP Deoxynucleotide-triphosphate

EDT 1,2-ethanedithiol

FCS Fluorescence correlation spectroscopy

FKBP FK506 binding protein

FMO Flavin containing monooxygenase

FMRP Fragile X mental retardation protein

FRAP Fluorescence recovery after photobleaching

FRB FKBP–rapamycin binding

FTD Frontotemporal dementia

FUS Fused in sarcoma

GFP Green fluorescent protein

HERD Helical repeat domain

HisTrap HP Ni-Sepharose HisTrap High Performance

hnRNPA1 Heterogeneous nuclear ribonucleoprotein A1

hnRNPA2 Heterogeneous nuclear ribonucleoprotein A2

HPLC High pressure liquid chromatography

IDR Intrinsically disordered region

IMAC Immobilised metal ion affinity chromatography

IPTG Isopropyl β -D-1-thiogalactopyranoside

LB Lysogeny broth

LCST Lower critical saturation temperature

LLPS Liquid-liquid phase separation

LOVTRAP LOV2 trap and release of protein

MIA ModularImageAnalysis

MLO Membraneless organelle

MRE Mean residue ellipticity

mRNA Messenger RNA

NaCl Sodium chloride

NB Nuclear body

Ni-NTA Ni-nitrilotriacetic acid

NMP N-Methyl-2-pyrrolidone

NMR Nuclear magnetic resonance

PAE Predicted aligned error

PAGE Polyacrylamide gel electrophoresis

PBS Phosphate buffered saline

PCM Pericentriolar material

pCMV Cytomegalovirus vector

PCR Polymerase chain reaction

PDB Protein data bank

PEG Polyethylene glycol

pI Isoelectric point

pLDDT Predicted local distance difference test

PPI Protein-protein interaction

PRM Proline rich motif

PVDF Polyvinylidene fluoride

RNA Ribonucleic acid

rRNA Ribosomal RNA

SDS Sodium dodecyl sulfate

SH3 SRC Homology 3

SPOP Speckle-type BTB/POZ

SPPS Solid phase peptide synthesis

TAE Tris/acetate/EDTA

TC Tetra-cysteine

TCEP Tris(2-carboxyethyl)phosphine

TDP-43 TAR DNA binding protein of 43 kDa

TEM Transmission electron microscopy

TEMED N,N,N',N'-tetramethylethylenediamine

TEV Tobacco etch virus

TFA Trifluoroacetic acid

TFE Trifluoroethanol

TIM Triosephosphate isomerase

TnaA Tryptophanase

Tris-base 2-amino-2-(hydroxymethyl)-1,3-propanediol

tRNA Transfer RNA

UA Uranyl acetate

UCST Upper critical saturation temperature

WASP Wiskott–Aldrich syndrome protein

Chapter 1

Introduction

1.1 The cellular organisation problem

At any one time, the cellular lumen contains hundreds of thousands of different constituent molecules. A single cell of the budding yeast *Saccharomyces cerevisiae* contains approximately 42 million protein molecules from a proteome of nearly 6,000 unique proteins.^{1,2} Not only that, but the cell also encapsulates DNA, RNA, polysaccharides, lipids, and metabolites. In fact, overall, the cellular interior is occupied approximately 20 – 30% by volume by macromolecules.^{3,4} Put into context, if the entire human population (7.7 billion people) occupied a volume with a similar level of crowding, that volume would be only just over 1.6 km³. This packing of molecules into a dense biochemical environment is known as macromolecular crowding (Fig. 1.1).⁵ The question of how the cell is able to organise its components in such a crowded and complex environment is a longstanding problem in cell biology.^{6,7} Each of these components is only present in a relatively low concentration, but must be able to find and identify its interaction partners quickly. Further, the cell must be able to regulate its components to promote, or attenuate, their interactions. In order to overcome these problems, cellular organisms rely on complex systems of trafficking and compartmentalisation to organise their components.

Methods of cellular organisation can be broadly split into two categories: organisation using phospholipid bilayers (membranes), and protein-based organisation methods (Fig. 1.2). Membrane-bound organelles are the textbook example of cellular compartments, and indeed, the presence of membrane-bound organelles is so significant that it partially defines the three domains of life: prokaryotes and archaea contain no internal membrane-bound organelles, only a double phospholipid cell membrane creating an internal cytoplasm and surrounding periplasm; while eukaryotes contain internal membrane-bound organelles in addition to their double cell membrane.^{8–11} However, despite the diverse

and critical roles of membrane-bound organelles in eukaryotes, all 3 domains of life also use protein-based cellular organisation. These organelles and structures are just as varied in their assemblies and functions as the membrane-bound organelles. However, instead of using phospholipid bilayers, these assemblies rely solely on protein-protein interactions (PPIs) to self-assemble *in vivo*.

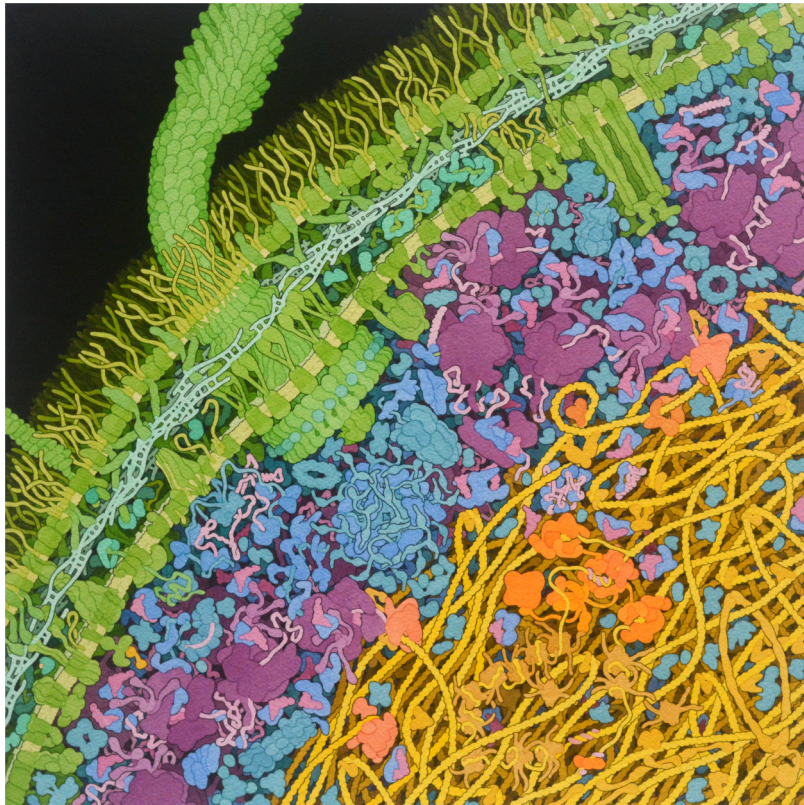


Figure 1.1: The crowded intracellular environment. Artwork depicting the crowded intracellular environment of an *E. coli* cell by David S. Goodsell, reproduced with permission from PDB-101, part of the RCSB Protein Data Bank.

1.1.1 Protein-based cellular organisation

There are several quintessential examples of protein-based cellular organisation in both eukaryotic and prokaryotic cells. One of these is the cytoskeleton (Fig. 1.3). Both eukaryotic and prokaryotic organisms share some form of a cytoskeleton, a system of fibres and tubes formed by protein oligomers.¹² The eukaryotic cytoskeleton is made up of 3 classes of protein assembly made by different protein oligomers. The largest are microtubules, hollow, cylindrical structures 25 nm in diameter, assembled from α and β tubulin heterodimers. The smallest are actin fibres, only 6 nm in diameter, formed by actin oligomers.¹³ Finally, there are additional intermediate filaments, typically around 10 nm in diameter and constructed from a variety of protein oligomers. More recently the cytoskeleton in prokaryotic cells has also begun to be understood, with structures

homologous to the eukaryotic tubules and filaments.^{14,15} These supramolecular cellular structures are responsible for a huge range of functions, including cell morphology, movement, division and intracellular transport. Without the cytoskeleton, the entire cellular apparatus, including its membrane-bound organelles, deoxyribonucleic acid (DNA), and morphology could not be maintained.

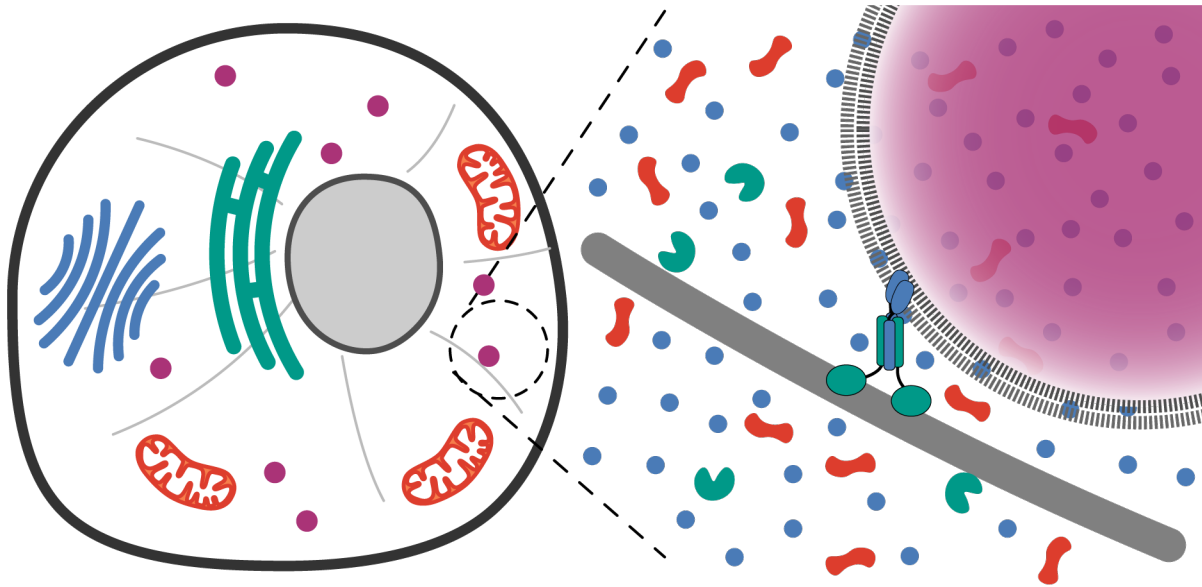


Figure 1.2: Levels of cellular organisation. Cells are organised on a macroscopic scale using both membrane-bound organelles, such as the nucleus, endoplasmic reticulum, golgi and mitochondria (left), and protein-based cellular structures, such as the cytoskeleton (right), where molecular motors traffic cellular components.

Other proteinacious compartments exist that more closely resemble membrane-bound organelles. One of these classes of compartment are bacterial microcompartments (BMCs). BMCs are structures found in prokaryotic organisms that fulfill some of the roles played by membrane-bound organelles in eukaryotic cells.¹⁶ However, instead of a semi-permeable lipid membrane, BMCs are formed entirely by rigid, polyhedral, protein scaffolds ranging from 100 – 400 nm in diameter.^{17,18} These protein scaffolds are created by highly conserved BMC proteins that form hexameric building blocks to create the facets of the polyhedron, while pentameric bacterial microcompartment vertex (BMV) proteins form the vertices.¹⁹ While BMC proteins and the polyhedral BMC shells are highly conserved, BMCs have diverse functions, participating in both anabolic and catabolic processes by encapsulating catalytic enzymes within their geometric shell.²⁰⁻²³

Both these types of cellular structure are formed by rigid, geometrically defined PPIs. Initially, this was believed to be the only way by which supramolecular protein structures could assemble. However, in addition to these rigid proteinacious scaffolds, there are a further class of protein-based cellular compartments that form by a very different mechanism. These organelles, known broadly as protein condensates, have been observed

for several decades, but are only now beginning to be understood as a key mechanism of cellular organisation.

1.2 Protein condensates - a new paradigm

The first cellular protein condensates were observed as far back as the 1830s. This compartment was later identified as the nucleolus, the dense site of ribosomal RNA (rRNA) synthesis in the nucleus.²⁴ Since then, a huge number of diverse cellular condensates have been identified in a host of different contexts.²⁵ For a long time, it was unclear how these cellular compartments were structured. These assemblies were not membrane-bound organelles – they were formed entirely of proteins or nucleic acids, and gained the denomination membraneless organelles (MLOs).²⁶ They were observed to be highly diverse in their structures, molecular compositions, and functions, but appeared to have some consistent, unexpected, physical characteristics, such as the ability for two such compartments to fuse.²⁷

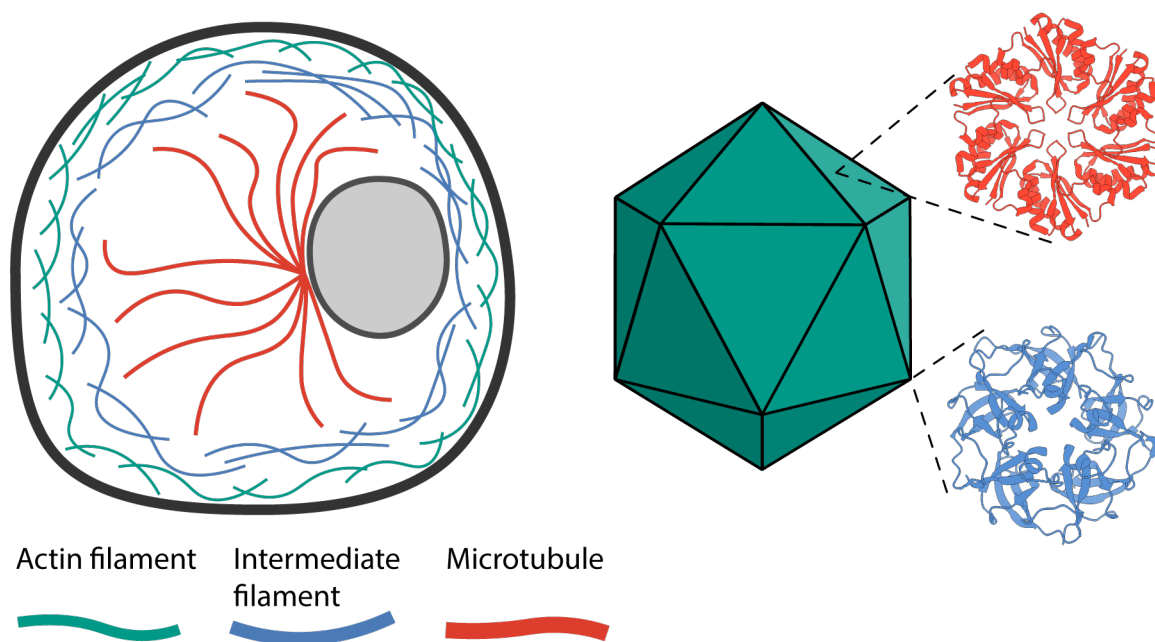


Figure 1.3: Protein-based cellular organisation. Examples of protein-based cellular organisation. The cytoskeleton (left) is composed primarily of three types of proteinaceous filaments, actin filaments, intermediate filaments, and microtubules. BMCs (right) are composed of hexameric (red) and pentameric (blue) protein oligomers that self-assemble to form polyhedral protein assemblies.

The watershed study on protein condensates came in 2009, with the discovery that P granules, germ granules from *Caenorhabditis elegans* (*C. elegans*) embryos, behaved like liquid droplets within the cytoplasm.²⁸ These ribonucleoprotein granules, formed by ribonucleic acid (RNA) and RNA binding proteins, were observed to form spherical droplets that were

responsive to shear forces, could dissolve and re-condense, and fused together like liquid droplets. Further, fluorescently labelled protein within the droplets was observed to be motile, with a viscosity approximately 1000 times greater than water. These observations were in contrast to the ideas that organelles within cells were either membrane-bound, or formed by a static proteinaceous assembly. This discovery opened up the door to the idea that organelles in living cells could be formed by a physical process called liquid-liquid phase separation (LLPS).²⁹

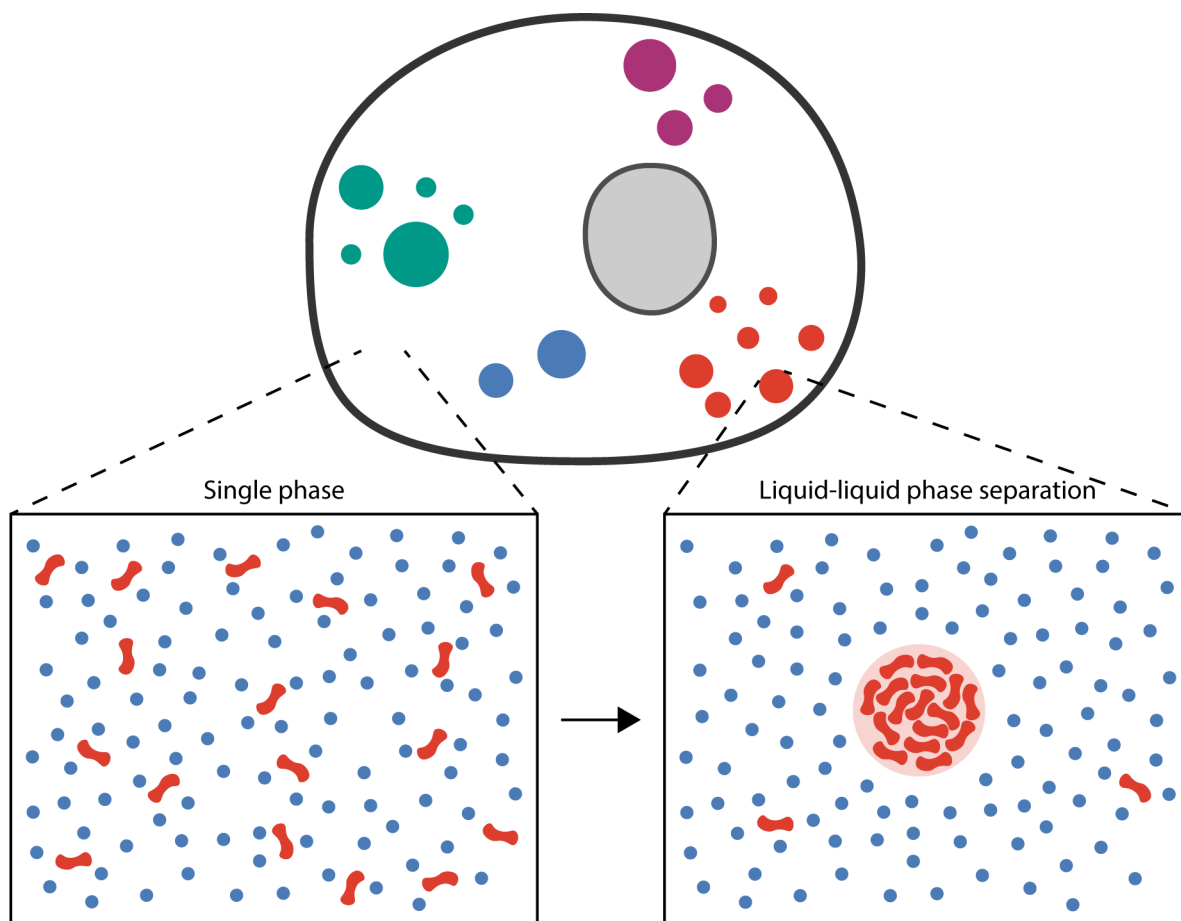


Figure 1.4: Liquid-liquid phase separation in biology. Proteins can undergo a de-mixing process called liquid-liquid phase separation.

1.3 Phase separation in biology

Phase separation is the process where a solution will reversibly de-mix into two or more distinct phases (Fig. 1.4). In the case of LLPS, it results in the formation of two liquid phases: a dense phase that is enriched in macromolecules, and a non-enriched dilute phase. In the *C. elegans* germ cells, the P granules are the dense phase formed by RNA and RNA binding proteins, while the surrounding cytoplasm forms the dilute phase.

LLPS, and phase separation in general, are physical principles that have been recognised in polymer chemistry for much longer than they have been identified in biology.³⁰ Phase separation is a phenomenon that can be observed frequently in everyday life: in a vinaigrette the oil and vinegar will not remain mixed, and will de-mix back into two distinct liquid phases. Similarly, in a macromolecular context, LLPS is the process where macromolecules in a solution reversibly de-mix to form two liquid phases: a dense phase and a dilute phase.²⁹ However, unlike mixtures of oil and water, both the de-mixed biomacromolecules and the aqueous solvent of the cell are usually formed by relatively polar molecules, so the driving forces of de-mixing are perhaps less intuitive, but these interactions can still be described by the entropy and molecular interactions of the system during mixing (Fig. 1.5).³¹

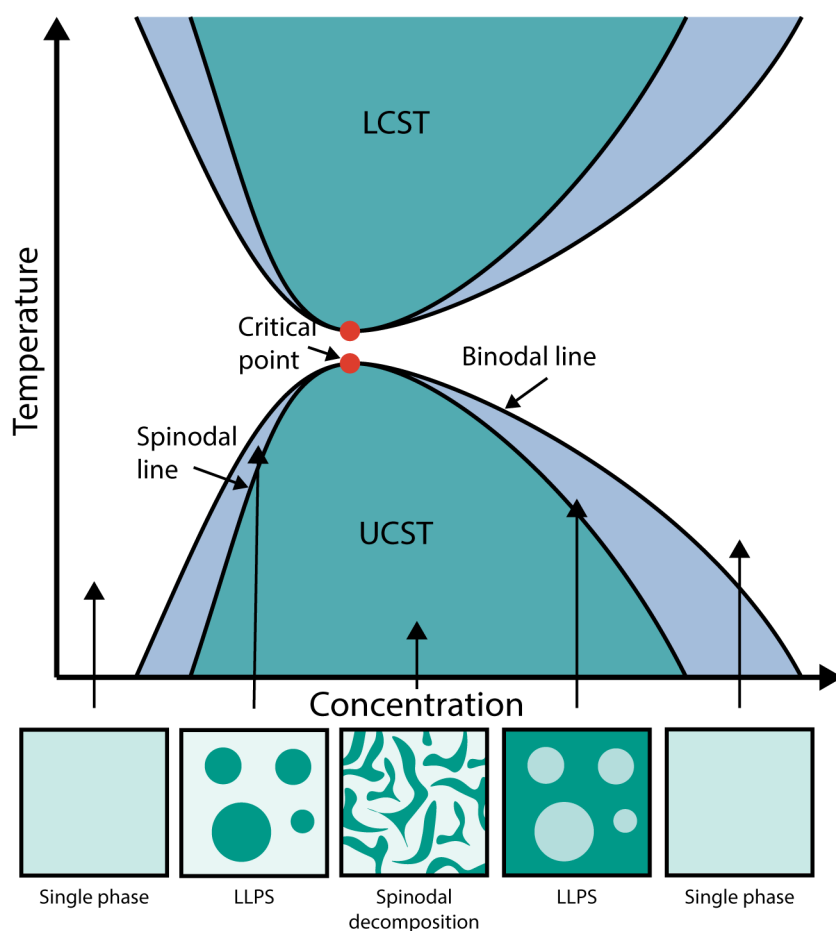


Figure 1.5: Phase diagram for LLPS. The blue shaded area indicates the nucleation limited region for LLPS for both enthalpically and entropically driven phase separation, limited by the binodal line. The green shaded area indicates the region limited by the spinodal line, giving spinodal decomposition. The maximum and minimum points for an upper critical saturation temperature (UCST) and lower critical saturation temperature (LCST) respectively are the critical point (red circle), where all 3 phases are in equilibrium.

Several complex and descriptive models for phase separation have been developed, beginning from models developed by Maurice L. Huggins and Paul J. Flory in 1942, known as

the Flory-Huggins solution theory.^{32,33} These models describe how de-mixing occurs as a combination of entropic and enthalpic interactions: LLPS can be enthalpically driven, where de-mixing occurs because of pairwise interactions between protein molecules that overcome the entropic penalty of de-mixing, or LLPS can be entropically driven, where entropic forces drive condensation by differential solubility effects between protein and solvent. These behaviours give rise to different dependencies on temperature, due to the inverse dependence of entropic and enthalpic forces on temperature. Enthalpically driven phase separation is dispelled as the temperature is increased, giving a upper critical saturation temperature (UCST), while entropically driven phase separation is dispelled as the temperature is reduced, giving a lower critical saturation temperature (LCST). There are also further phase transitions possible beyond LLPS. Pushing further into the de-mixed region can induce irreversible spinodal decomposition.

In cells, these dense regions of protein may initially appear similar to inclusion bodies, regions of misfolded or aggregated protein in both bacteria and eukaryotic cells. However, LLPS is distinct from protein aggregation or misfolding.²⁶ Phase separation is characterised by a sudden de-mixing event when certain physical thresholds are reached.³⁴ Further, LLPS is a reversible process, while protein aggregates are typically characterised by proteins being trapped in an irreversible non-native or misfolded state.³⁵ Such aggregates can still have structures and functions, but are frequently toxic to the organism. Condensates meanwhile, while potentially highly viscous, are characterised by dynamic protein diffusion within the dense phase.³⁶ These compartments are also frequently transient, and can be recycled depending on the cellular requirements.

1.3.1 A diverse and widespread phenomenon

Since the discovery of phase separation as a mechanism for cellular organisation, there has been a surge in the number of cellular structures that are described as protein condensates (Fig. 1.6). These condensates have diverse structures and are implicated in a huge range of biological functions. In the nucleus alone there are dozens of so called nuclear bodies (NBs), including the nucleolus, Cajal bodies, nuclear speckles, and paraspeckles.³⁷⁻⁴⁰ These NBs have the characteristic properties of phase separated compartments: they form spherical de-mixed structures, not encircled by a membrane, which can dissolve and fuse. The roles of these structures in the nucleus is still unclear, but they have been implicated with discriminating between actively transcribed DNA and repressed heterochromatin.

Alongside stress granules and RNA granules, there are a host of other protein condensate bodies that have been implicated to involve phase separation. However, unlike stress granules, which are quite widely accepted to assemble by some form of phase separation,

some cellular bodies have disputed mechanisms of self-assembly. One such body is the centrosome, the main site of microtubule organisation in eukaryotic cells, organising both the cytoskeletal network and the mitotic spindles during cell division. At the core of the centrosome are the centrioles, two barrel shaped structures formed by an inner cartwheel structure surrounded by nine-fold symmetric microtubules.⁴¹ Surrounding the centrioles is the pericentriolar material (PCM), containing hundreds of different protein components.

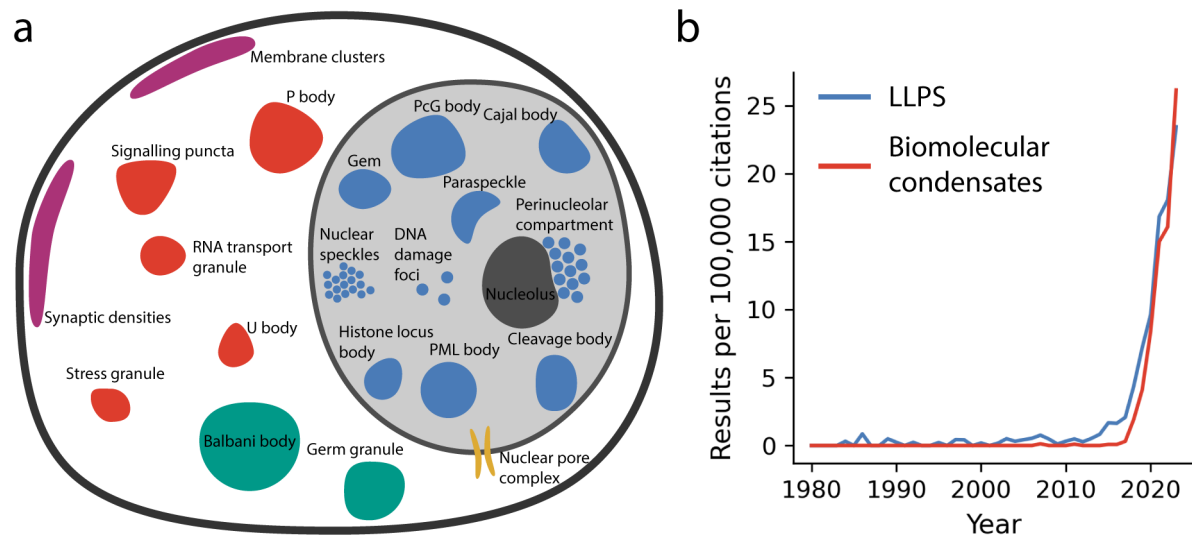


Figure 1.6: The rapid expansion of LLPS as an organising principle in biology. **a**, A selection of some of the cellular compartments identified to-date as assembling as a result of protein phase separation or condensation. Condensates are coloured according to their associated roles. Germ granules are coloured in green. Stress granules and RNA granules are coloured in red. Synaptic densities and membrane clusters are coloured in purple. Nuclear condensates are coloured in blue, with the nucleolus in dark grey, and the nuclear pore complex is coloured in yellow. Figure adapted from Banani *et al.* (2017) with permission from Springer Nature.³⁴ **b**, Proportional number of results (results per 100,000 citations) for LLPS (blue) or Biomolecular condensates (red) in PubMed by year from 1980 to 2023. Data collected using PubMed by Year (<https://esperr.github.io/pubmed-by-year/>) on 14th July 2023.

The overall structure of the PCM has been the subject of intense investigation. None of the the scaffolding proteins that make up the PCM contain motifs that are typically associated with LLPS, and electron microscopy indicates that the PCM is instead formed by a fibrous matrix with large pores, allowing it to be permeable to cytoplasmic macromolecules, similarly to a phase separated droplet.⁴² However, further studies have demonstrated that PCMs assembled *in vitro* can form spherical liquid-like droplets that functionally resemble centrosomes, concentrating α/β -tubulin dimers and nucleating microtubules.⁴³ Indeed, difficulties in determining whether cellular compartments are formed by phase separation or other phenomena is a recurring theme in this burgeoning field.⁴⁴

1.4 Phase separation in disease

Overall, LLPS has become a seemingly ubiquitous feature in cell biology in a remarkably short period of time. It follows therefore that there has been considerable focus on the roles of aberrant phase separation in disease.⁴⁵ Scaffold proteins that form membraneless organelles are susceptible to mutations, like other forms of protein assembly, causing altered or reduced functionality (Fig. 1.7).⁴⁶ In addition however, due to the sensitivity of protein condensates to their local environment, membraneless organelles are also particularly susceptible to changes in physical and chemical conditions.⁴⁷ This means that unstable cellular conditions, such as can occur in aging tissues, can also lead to aberrant phase separation.^{35,48,49}

1.4.1 Phase separation in neurodegeneration

Abnormal or deficient phase separation is most frequently implicated in neurodegenerative disorders.⁵⁰ Neuronal cells are considered to be particularly vulnerable to aberrant phase separation.⁵¹ Neurons are post-mitotic cells, and therefore do not undergo cell division, which is a mechanism by which many cells are able to clear or reduce the accumulation of toxic aggregates.⁵² Further, despite their size and being unable to undergo cell division, neurons are extremely plastic cells, with the capacity to undergo large morphological and molecular changes to react to stimulus.⁵³ These changes require that neurons transport and store pools of messenger RNA (mRNA) for local translation, as well as using intricate layers of genetic regulatory networks.^{54,55} Some of these networks have been directly implicated to function using phase separation, such as for the fragile X mental retardation protein (FMRP) transcriptional repressor.⁵⁶⁻⁵⁸

Aberrant phase separation has been closely implicated in both of the neurodegenerative disorders amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD).⁵⁹⁻⁶¹ ALS-causing mutations in several RNA binding proteins (fused in sarcoma (FUS), TAR DNA binding protein of 43 kDa (TDP-43), heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1), and heterogeneous nuclear ribonucleoprotein A2 (hnRNPA2)) have been identified to drive accumulation in stress granules.⁶²⁻⁶⁸ Further, several of the proteins found in these stress granules have also been found in the pathological FTD and TDP-43 aggregates identified in ALS and FTD neuronal tissues.⁶⁹⁻⁷¹ ALS-causing mutations have also been implicated to alter the morphology and material properties of stress granules, with such mutations suggested to accelerate the transition from a liquid-like organelle to a solid-like aggregate.⁷²⁻⁷⁷

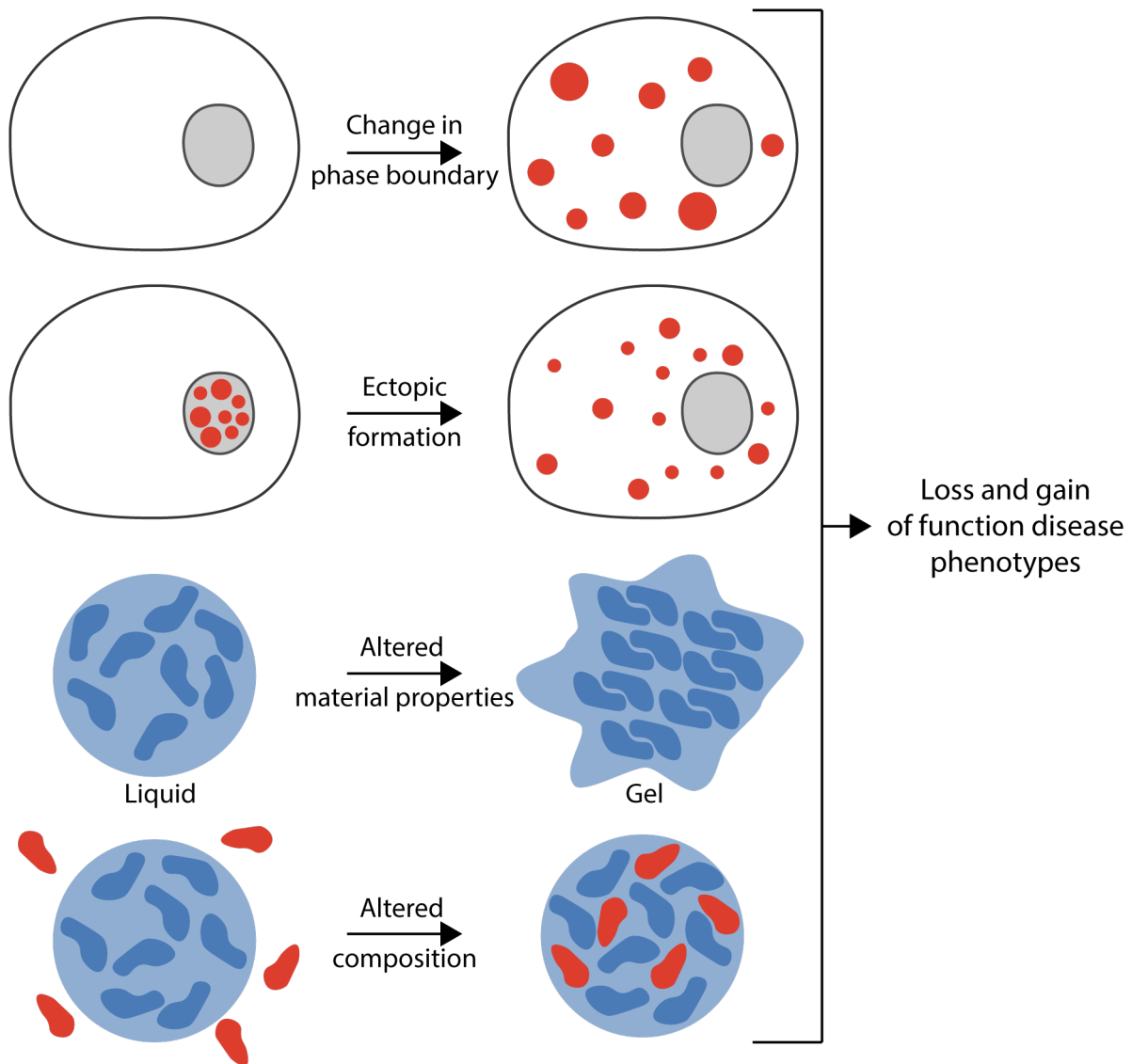


Figure 1.7: Phase separation in disease. Concepts of how defects or changes in phase separation can lead to disease phenotypes, such as by changes to protein assembly, localisation, or regulation leading to altered phase separation characteristics. Adapted from Alberti and Dorman (2019) with permission from Springer Nature.⁴⁶

1.4.2 Phase separation in cancer and infectious diseases

In addition to neurological disorders, phase separation has also been implicated to play roles in a range of other diseases. In cancers, aberrant phase separation is suspected to play a role in disrupting normal cell signalling homeostasis, and aid cancerous cells in entering a hyper-proliferative state.^{78,79} Direct evidence linking phase separation defects and cancers is so far limited, but one study has identified that cancerous mutations in the tumor suppressor protein speckle-type BTB/POZ (SPOP) lead to specific phase separation defects.⁸⁰ In infectious diseases, evidence suggests that viral replication centres, compartments within infected cells that promote the formation of new viral particles and aid in the evasion of the immune response, are formed by phase separation.^{80,81} Stud-

ies also suggest that phase transitions are used by bacteria and fungi in order to enter a dormant state, and survive unfavourable conditions.⁸¹ In this state, portions of their cytoplasm can harden and reversibly solidify to avoid stress, and can allow pathogenic bacteria to avoid cytostatic drugs.⁸² Overall, the evidence that phase separation and phase transitions play key roles in biology and disease is continuing to mount. These discoveries have re-enforced the need to understand the principles behind phase transitions in a biological context, and possibly reveal new opportunities for therapeutic intervention.

1.5 Molecular principles of phase separation

Understanding the molecular principles underlying phase separation and aberrant phase transitions is likely to be key to the treatment of pathological defects. This could aid in understanding how mutations within phase-separating proteins alter their structure or function, and lead the way to treating phase separation defects. Models of polymer solutions like the Flory-Huggins solution theory describe how de-mixing occurs for an ideal polymer solution, but these rarely translate well to complex proteins with anisotropic features. One of the challenges in studying biological phase separation is understanding the specific molecular interactions that drive phase separation.⁸³ As the number of proteins that have been identified to phase separate continues to increase, the diversity of their molecular structures and interactions is remarkable (Fig. 1.8).⁸⁴ This makes it challenging to identify a consistent molecular architecture that drives phase separation, particularly in a predictive manner.⁸⁵ However, some features are beginning to emerge as characteristic molecular markers of phase-separating proteins.

1.5.1 Multivalency and modular domains

The most consistent feature of phase-separating proteins is multivalency: phase separation is driven by a large number of protein-protein or protein-nucleic acid interactions.⁸⁶ Individually, each of these interactions is characteristically weak and short-ranged, but it is the net attractive interactions of these multivalent proteins that overcome the penalties of de-mixing.⁸⁷ In proteins, multivalency can occur in a number of different ways: well-folded globular proteins can be multivalent due to the presence of multiple interaction patches on their surface.⁸⁸ Alternatively, individual folded protein domains can be concatenated together by linkers to create a linear multivalent assembly, known as a modular domain protein. Moreover, largely or entirely unstructured protein sequences, known as intrinsically disordered regions (IDRs), can form the basis of multivalent PPI domains.^{89,90}

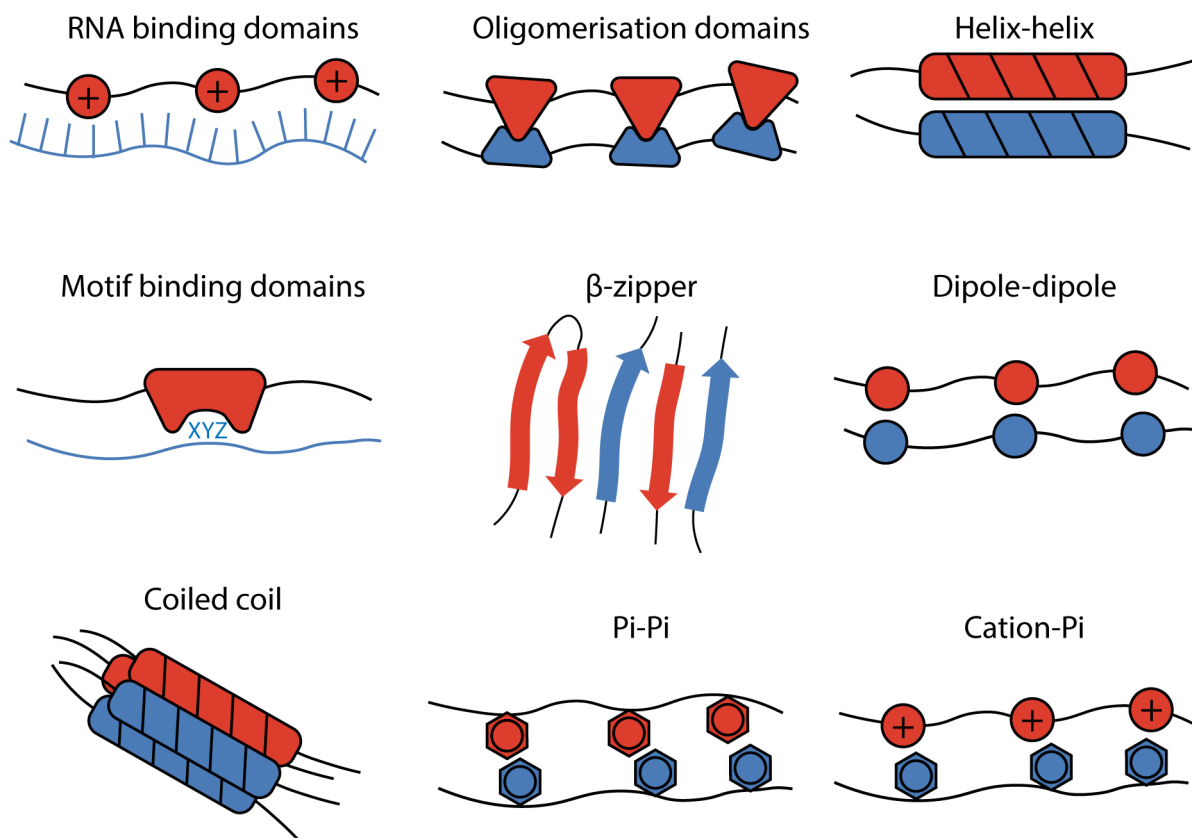


Figure 1.8: Varied interactions leading to phase separation. Schematic of a number of intermolecular interactions that have been characterised as leading to or implicated in protein phase separation. Figure adapted from Boenyaems *et al.* (2018) with permission from Elsevier.²⁵

All of these forms of multivalency have been identified to play roles in protein phase separation in different contexts. One of the first examples of an engineered phase-separating protein used the modular protein SRC Homology 3 (SH3) domain and its binding partner proline rich motif (PRM).⁹¹ Linear repeats of these folded protein domains formed complexes that readily liquid-liquid phase separated. Further, increasing the number of SH3 and PRM repeats increased the strength of their interactions, eventually leading to gelation.

However, by far the most common multivalent architecture driving phase separation, are IDRs.^{92,93} These domains do not form a stable structure, instead forming highly flexible and dynamic linear protein chains.⁹⁴ Despite their lack of secondary structure, IDRs have widespread roles in biology, often as a result of their unique flexibility. IDRs are found in a large number of biomolecular condensates, most notably in those that also condense nucleic acids, and these unstructured proteins have also been identified to phase separate *in vitro*.^{95,96} IDRs typically have a low sequence complexity, with repeats of a limited palette of amino acids, frequently enriched in serine, asparagine, glutamine, phenylalanine, glycine, and tyrosine residues. In phase separation, individual residues or short repeats of residues in IDRs make up interaction motifs that drive de-mixing.⁹⁷

1.5.2 Weak intermolecular interactions

In addition to multivalency, another molecular principle applied to phase separation is the requirement for relatively weak or transient intermolecular interactions.⁸⁶ Strong, irreversible, PPIs can lead to forms of protein assembly with larger energy minima, such as precipitation, gelation, or crystallisation.^{25,98} Instead, to create a de-mixed phase with dynamic, liquid-like properties, the intermolecular interactions must be able to re-arrange on very short timescales. This necessitates that interactions between proteins that undergo LLPS are either weak, or rapidly reversible.^{99–101}

The requirement for weak or transient PPIs is reflected in the types of molecular interaction that are believed to be important for phase separation. Weak molecular interactions such as π - π stacking, cation- π interactions, and electrostatic interactions have all been heavily implicated in phase-separating proteins, particularly in IDRs.^{83,85} However, there is also an extraordinary diversity of interaction motifs and contacts implicated in phase separation: from π - π interactions and RNA binding domains, to β -zipper and coiled coil motifs, suggesting that there is a flexibility around phase separation and the molecular interactions that drive it.^{25,83}

1.6 Protein design meets biology

As the molecular mechanisms surrounding phase separation begin to be understood, the potential for these systems to be engineered or even designed from the bottom-up has increased dramatically. This can not only contribute to our understanding of how membraneless organelles can assemble *in vivo*, but also towards developing phase-separating systems not seen in nature. This brings protein assembly and phase separation into the realm of synthetic biology, and specifically, protein design.

Protein design is the bottom-up creation of defined 3D assemblies from a primary amino acid sequence. The goal of protein design is distinct from that of protein engineering.¹⁰² Protein engineering modifies defined protein scaffolds to introduce new or modified functionalities. This has the advantage that it enables the accessible re-designing of proteins either rationally or through screening methods such as directed evolution.¹⁰³ In *de novo* protein design, by contrast, the entire conformation and structure of the protein is part of the design target.¹⁰⁴ This can require a greater level of understanding of protein folding, and consideration of all elements of a protein's sequence. As a reward for these challenges however, the entirely bottom-up design of proteins presents the potential to create new structures and functions unexplored in nature, and prove our understanding of the sequence-to-structure relationship.¹⁰⁵ In the context of designing supramolecular

cellular assemblies, this could deliver artificial cytoskeletons, or organelles that function orthogonally to the endogenous cellular machinery. From its inception in the early 1980s, with a series of poorly soluble β -sheet and β -sandwich peptides, protein design has developed rapidly, and become a tractable tool for the design of a huge diversity of protein structures and architectures.^{106,107} As the field has matured, two distinct approaches to protein design have developed: rational or minimal protein design, and computational protein design (Fig. 1.9).¹⁰⁸

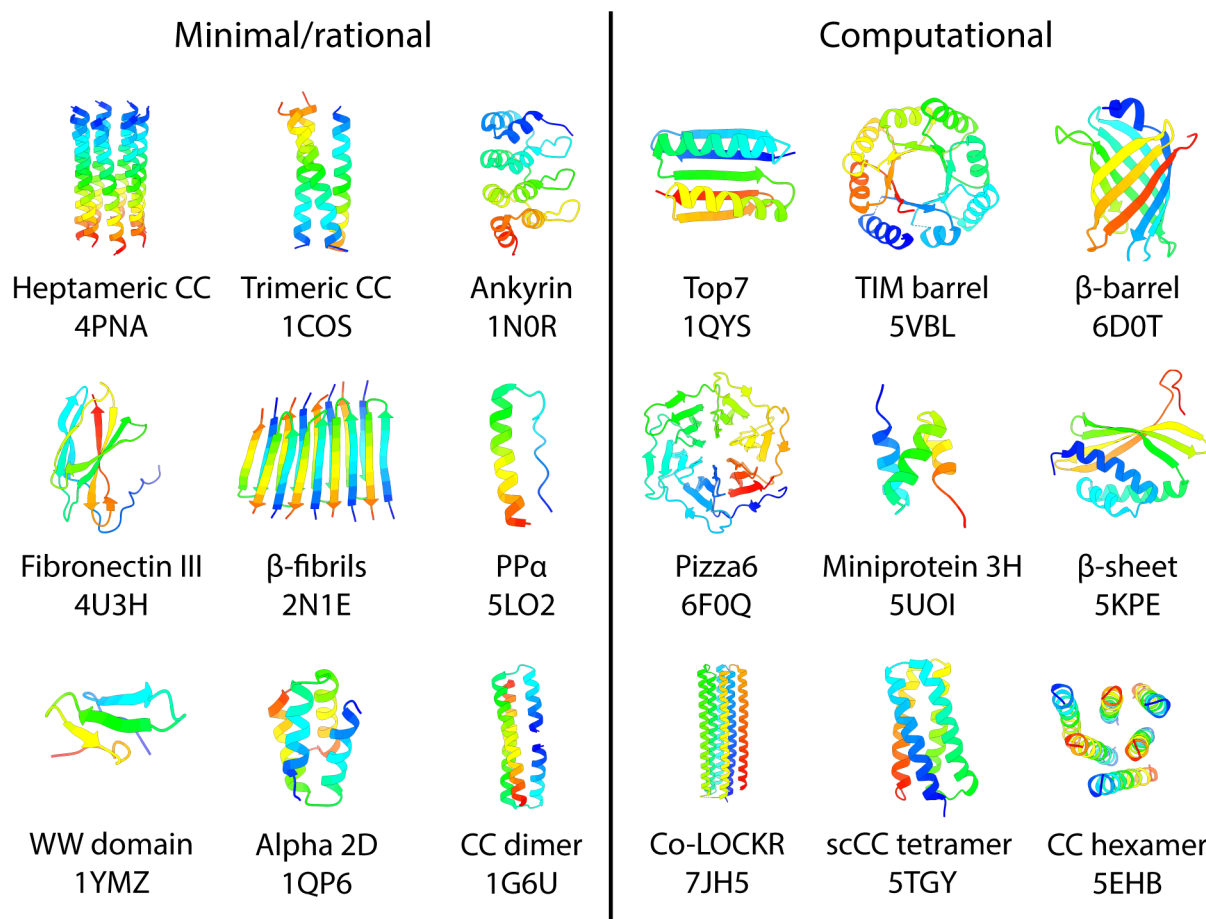


Figure 1.9: A gallery of structures developed by protein design. Structures are divided broadly into those designed by minimal or rational protein design techniques, and those developed by computational design techniques. Below the structures are their common names and 4 figure PDB IDs. Coiled coil is abbreviated CC and scCC denotes a single chain coiled coil. This gallery is by no means exhaustive and may be subject to personal bias, but is designed to show the versatility and diversity of structures that can be generated by protein design methods. Figure adapted from Woolfson (2021) with permission from Elsevier.¹⁰²

1.6.1 Rational protein design

The effort to understand the *why* of protein folding has been the focus of the field of rational protein design.¹⁰² Rational protein design uses logical and understandable sequence

rules and repeats to generate a protein fold from a pattern of amino acids.^{109,110} In entirely minimal protein design, these sequence repeats may be patterns of hydrophobic and polar amino acids (hpphppp) used to create amphipathic α -helices.¹⁰⁶ However, rational protein design can go further and parameterise more complex sequence-to-structure relationships through specific patterns of amino acids.¹¹¹ Rational protein design methods promise to deliver the human understanding behind protein folding, through the creation of simple rules for why a pattern of amino acids has a tendency to fold into a certain structure.

In the past 50 years, rational protein design has built a steadily increasing understanding of why proteins fold into differing secondary, tertiary, and quaternary structures. Many of the first *de novo* proteins were designed rationally, and used recent advances in solid phase peptide synthesis (SPPS) to use peptides as scaffolds for testing protein folding. In 1985 DeGrado and Lear created some of the first rationally structured *de novo* peptides, with peptides using alternating polar and hydrophobic residues (LKLKLLKL) folding into a β secondary structure. However, when the pattern of hydrophobic and polar residues was changed to match the rotation of an α -helix (LKKLLKL), the peptides now self assembled into tetrameric α -helical bundles.¹⁰⁷ This was some of the first evidence that chemical principles of amphipathic peptides could be related directly to their secondary structure in solution. Since then, minimal and rational models of protein folding have delivered a range of increasingly complex protein structures, from β -barrels and β -propellers, to rubredoxin folds, triosephosphate isomerase (TIM) barrels, and coiled coils.^{112–116}

This approach differs significantly from the use of computational design tools, where the rationalisation of why certain combinations or patterns of amino acids produce a stable protein fold is lost through the use of complex physical models, or more recently, machine learning techniques.

1.6.2 Computational protein design

In the preceding two years, computational protein design has undergone a revolution.¹¹⁷ In 2021, at the 14th meeting of the critical assessment of structure prediction (CASP), AlphaFold2 was released.¹¹⁸ Created by researchers at DeepMind, AlphaFold2 showed a remarkable and highly consistent ability to predict a protein's 3D structure from its primary sequence. When tested on previously unreleased protein structures from the community, AlphaFold2 correctly predicted their structure with an overall root mean square deviation across C α atoms of less than a 1 Å. Such accurate predictions were due to the revolutionary use of neural networks in structure prediction (Fig. 1.10).¹¹⁹

The results are structural predictions that far outperform any previous entries to the

CASP meetings, and likely one of the greatest scientific revolutions of the century.^{120,121} The release of AlphaFold2 allows biologists to rapidly generate models of protein structures that have never been solved before. While the direct evaluation of these models should be taken with caution, AlphaFold2 has proven itself to be an excellent tool for generating new hypotheses, fitting models to low resolution structural data, or rapidly testing experimental ideas *in silico*. All of which come with a remarkably low barrier to entry.¹²²

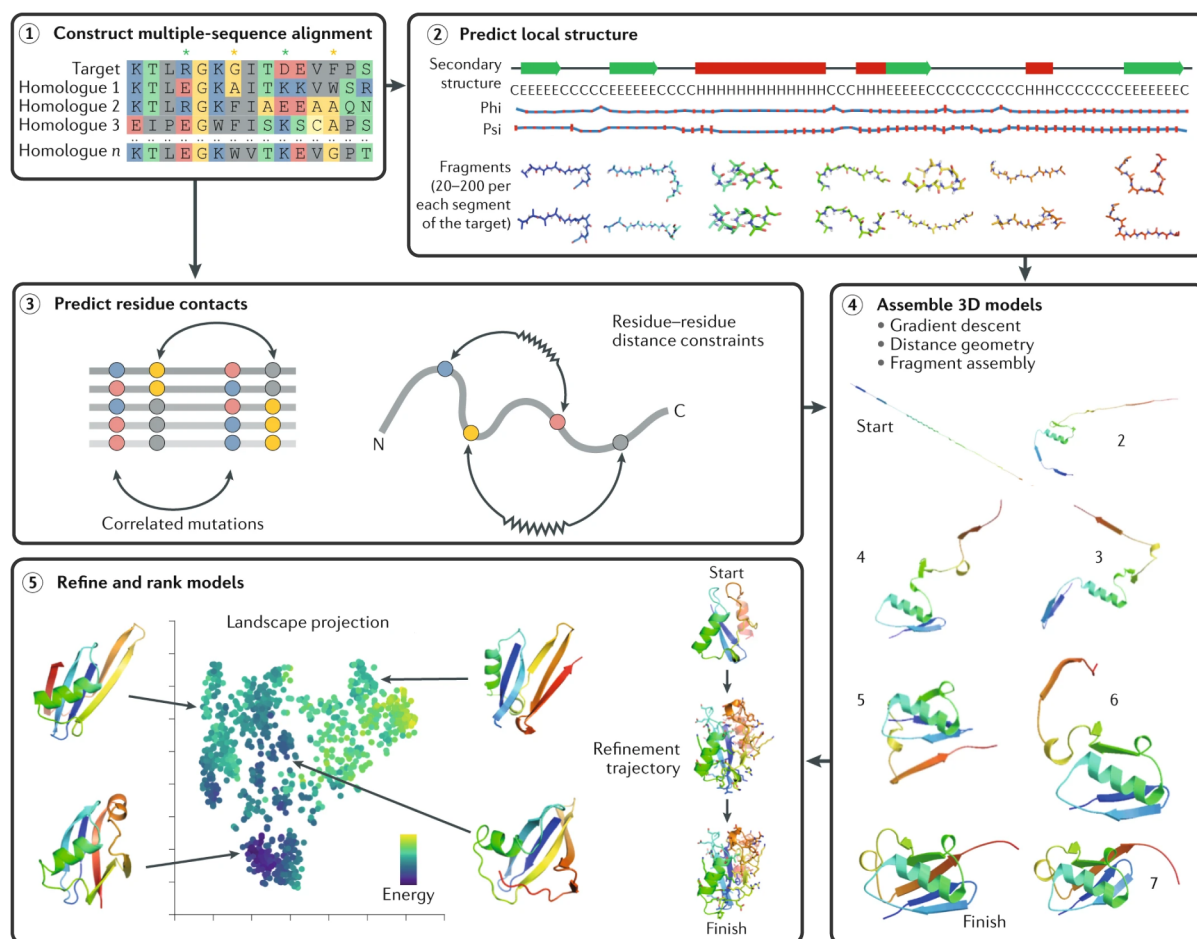


Figure 1.10: Steps in structural prediction of proteins using machine learning. A schematic of the key steps in structural prediction of a protein without a template. Target sequences are initially passed through a multiple sequence alignment, before local areas of structure are generated from homologous fragments, and incorporate elements of physical parameters such as torsion angles. The sequence alignment also feeds into predicted residue contacts, where residues are predicted to be in close proximity by assessing correlated mutations and distance constraints. These predictions then feed into model building that are refined to minimise energy and create the final energy landscape projections. Figure adapted from Kuhlman and Bradley (2019) with permission from Springer Nature.¹¹⁷

The use of neural networks for protein structure prediction has not only revolutionised structural biology, but has had enormous implications for protein design.¹²³ Until recently, computational protein design used arrays of physical functions that attempted to minimise a protein's energy.¹²⁴⁻¹²⁶ These packages compute elements like van der Waals

forces, steric repulsion, burying of hydrophobic residues and solvation of polar residues, hydrogen bonding effects, electrostatic potentials, and scoring for residue backbone and side chain torsion angles and rotamers.^{127,128} Now, neural-network-based design tools have been developed that show much greater proficiencies at generating soluble protein assemblies that match their design targets, while also being several orders of magnitude faster and computationally cheaper.¹²⁹ In the last year, the advance of artificial intelligence (AI) methods in computational protein design has only increased.¹³⁰ Recently, several new techniques taking advantage of methods in AI image generation such as DALL-E have been generated. These techniques use a process called diffusion to sample a range of protein backbone conformations by a de-noising process. Several tools have been released in the past year to generate protein backbones for a variety of structures and applications.^{131,132}

As the role of AI and machine learning in protein design is likely to only increase in the future, it is worth considering the lost "human" element in protein design. Generative models take almost no input from the user in their generation of new protein backbones, and both these tools and other machine learning tools tell us very little about why proteins fold. Indeed, these tools can be considered very much a black box in their weighting and decision making. This means that while they may serve as incredibly useful and accurate tools for protein design, in much the same way as AlphaFold2 is a useful tool for structural prediction, they have not solved the protein folding problem. Understanding the rules and requirements for why a protein sequence will fold into a predictable structure will require a much clearer understanding of protein chemistry, with increased emphasis on the rationalisation of protein folding.

1.6.3 Coiled coil design

Both rational and computational protein design have been highly successful in delivering solutions and understanding for synthetic biology. In the future, they will almost certainly continue to work synergistically to develop new protein folds and architectures, as they have already done so. Indeed, one of the best examples of this marriage is in coiled coil design. Coiled coils represent a special case in protein folding, and an example of one of the great success stories of both rational and computational protein design.^{111,133} Coiled coils are a highly regular quaternary protein structure created by two or more α helices, that wrap around each other in a left-handed super-helix.¹³⁴ The reason for this special quaternary structure comes from the unique geometry of the α helix, which is precisely defined in order to maximise backbone hydrogen bonding along its length, while minimising steric constraints. As a result, α helices have exactly 3.6 residues per turn of the helix, giving a rise per residue of 1.5 Å (Fig. 1.11a).

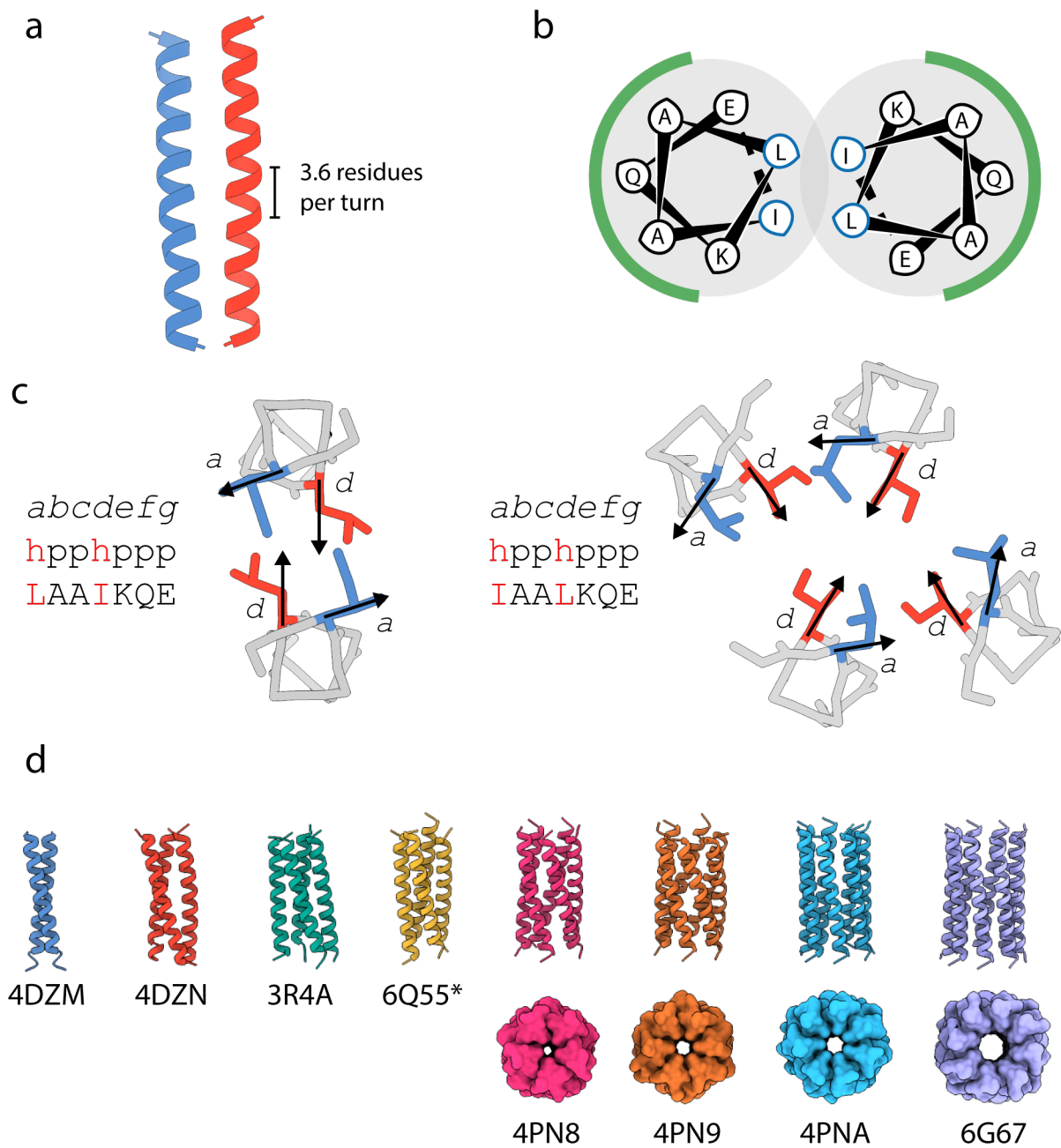


Figure 1.11: Coiled coils in protein design. **a**, A model of a dimeric coiled coil, with the characteristic 3.6 residue per turn geometry. **b**, A helical wheel representation of a dimeric coiled coil, with the hydrophobic interface residues (leucine and isoleucine) highlighted in blue, and the solvent exposed exterior to the coiled coil highlighted with a green semi-circle. **c**, Interfacial angles and heptad repeats for dimeric (left) and tetrameric (right) coiled coils. The change in position of the hydrophobic leucine and isoleucine from the *a* and *d* positions is highlighted in red on the sequence, and on the models *a* positions are coloured in blue, and *d* positions in red. **d**, A selection of some of the coiled coil assemblies that have been designed *de novo*, with oligomeric states ranging from dimer to octamer. Surface representations of the coiled coils from pentamer to octamer are shown side on, to visualise the solvent accessible cavity. PDB codes are given below each structure. * denotes that 6Q55 is an anti-parallel tetramer, while all other structures are parallel assemblies.

When two or more α helices come together to form an interface, they interact via a seam of residues along their length. Due to the 3.6 residues per turn rotation of the helix, the interacting residues will be spaced 3 and 4 residues apart in a 7 residue (heptad) repeat. These heptad repeats are denoted *abcdefg*, with interacting residues occupying the *a* and *d* positions (Fig. 1.11b). This creates an average spacing of 3.5 residues, as close as possible, but not exactly identical, to the 3.6 residues per turn of the α helix. The coiled-coil structure comes from this mismatch between the 3.5 residue repeat in interacting residues, and the 3.6 residues per turn of the α helix. For the two seams of interacting residues to perfectly match, the helices must coil around each other in a left-handed super-helix.^{135,136}

From this highly regular secondary structure come highly specific quaternary interactions. Helical nets showing the 3-4 residue spacing within the helices highlight the specific packing of interacting residues within the coiled coil. Along the interacting seam of residues, *a* and *d* residue side chains from one helix interdigitate with diamond shaped holes between residues on the other helix. These interactions, specific to coiled coils, were denoted knobs-into-holes packing by Francis Crick in 1953, along with predictions of coiled coil structure that would later be experimentally determined to be highly accurate.^{137,138} Because of this regularity in sequence, and the ability to rationally and parametrically predict their structure, coiled coils have been studied by protein designers for several decades, and our understanding of their folding and assembly goes far beyond the mathematical analysis available to Crick in the 1950s. We now understand the chemistry of coiled coils perhaps better than any other protein fold.¹³⁹ Initially, all of the interacting residues between helices (*a* and *d* residues) are assumed to be hydrophobic. This has the result that when two helices interact, the hydrophobic residues are buried, producing a positive entropic effect. Further, because the knobs-into-holes packing within this hydrophobic core is so compact, its quaternary structure is highly sensitive to the identities of these hydrophobic residues.¹⁴⁰

In a dimeric coiled coil, the side chains of residues at the *d* position are angled directly towards the centre of the coiled coil, into the hole formed by residues on the opposing helix (Fig. 1.11c).¹⁴¹ This direct packing is known as perpendicular packing, and due to the compact interface, overwhelmingly prefers non- β -branched leucine residues. In contrast, the side chains of residues at the *a* position are angled slightly away from the centre of the coiled coil, parallel to the knob on the opposing helix. As this residues side chain is less intimately packed in the core of the coiled coil, it can more readily accommodate β -branched or non- β -branched residues, with a slight preference for β -branched isoleucine.¹⁴²

If the hydrophobic interface is enlarged to accommodate a third helix, forming a trimeric coiled coil, the packing angles of the *a* and *d* residues rotate slightly to face the centre

of the hydrophobic core. The resulting packing is now more similar between the a and d residues, both now somewhere between parallel and perpendicular, known as acute packing. As the positions are equivalent, there is no strong preference for β -branched or non- β -branched residues, and hence helices with isoleucine at both the a and d position will prefer to form trimeric coiled coils.^{116,142} If the interface is enlarged even further to create a tetrameric coiled coil, the interfacial angles are essentially the inverse of the dimeric coiled coil.¹⁴³ Now the side chains of residues at the a position are pointing directly towards the centre of their respective holes, in perpendicular packing, and the residues at the d position are pointed parallel to their holes. This creates a preference for non- β -branched leucine at a , and β -branched or non- β -branched residues at d .

Natural coiled coils typically form either dimers, trimers, or tetramers, but through protein design oligomeric states above those found in nature have been made accessible.¹⁴¹ By expanding the hydrophobic interface to include the residues flanking the a and d positions (the e and g residues), the coiled-coil interface is widened enough to accommodate oligomeric states above tetramers. Through a process of rational design and physical computational modelling, new rules for the assembly of pentameric, hexameric, and heptameric coiled coils have been developed (Fig. 1.11d). These include rules for the e and g positions, where progressively smaller residues at these flanking positions drive the assembly of larger oligomeric states, up to a nonameric coiled coil with glycine at the e and g positions.

Overall, coiled coils have been a triumph of rational and computational protein design, with robust rules in place for designing coiled coils far beyond even those found in nature.^{144–147} As protein design too has matured, work has progressed beyond the understanding of their chemistry and physics, to the potential applications in biology.

1.7 Design of assemblies for cellular organisation

As the complexity of protein design increases, the field comes closer to the bottom-up design of systems that rival biological components. Protein designers can create new protein folds both rationally and computationally, but also create proteins for expanded functionality in living cells.^{108,148–150} Methodologies have been created to design proteins for protein and ligand binding, membrane protein assembly, protein assembly *in vivo*, and increasingly, proteins that incorporate conformational changes or switch-like behaviours.^{104,146,151–160} One of the key challenges remaining is the design of proteins that mimic the supramolecular cellular components found in nature, such as the cytoskeleton, microcompartments, or even entire organelles.

1.7.1 Designed geometric protein assemblies

Some of the first designed protein assemblies were created to emulate the natural rigid cellular complexes such as the cytoskeleton and BMCs. It was quickly realised that the design of large protein assemblies requires precise control over protein geometry. As the protein assembly increases in size, small deviations in interfacial angles can have large implications for the geometry of the overall assembly.¹⁶¹ As a result, most designed protein nanoparticles or scaffolds have stringent requirements for protein and interface rigidity, as well as symmetry.^{162–165}

Due to these requirements, it is perhaps not surprising that the first designed supramolecular assemblies designed were linear assemblies. Both designed β -sheet proteins and α -helical coiled coils have been used to create fibrous materials *in vitro*.¹⁶⁶ β -sheet assemblies typically take advantage of the tendency of β -sheets to assemble laterally, much like naturally forming amyloid-like fibres.^{167,168} Helical assemblies meanwhile use the rationality of coiled coil design to create helices that have "sticky ends" for linear assembly, or by creating heteromeric coiled coils that self-assemble into fibres by domain swapping.^{169–171} All of these designs however will essentially assemble indefinitely as long as their components are available.

The first discrete protein assemblies designed were protein nanocages.¹⁶¹ It was discovered that the combination of two oligomerising proteins of differing symmetries into a single polypeptide could self-assemble using the natural interfaces of the component proteins. Further, it was identified that if the fusion protein was created with the correct orientation, geometry, and was sufficiently rigid in its construction, the protein could self-assemble into discrete nanocages.¹⁶¹ In this pioneering work, dimeric and trimeric oligomerising domains combined using a rigid helical linker created tetrahedral protein cages with a radius of approximately 9 nm. Since this discovery, a number of groups have continued to apply protein design principles to nanocages and nanoparticles, increasingly using computational protein design to create new interfaces of defined geometries and symmetries, for complex assemblies (Fig. 1.12a).^{172,173} Now, these nanoparticles are being functionalised to leverage the effect of oligomerisation and assembly on protein-protein interactions and binding to create neutralising antibody platforms and vehicles for protein delivery.^{174–176}

Indeed, an increasing focus of protein design is not only the construction of supramolecular protein assemblies, but also their functionalisation. Both fibrous and discrete protein assemblies have been functionalised by recruitment of client proteins to the designed scaffold (Fig. 1.12b).^{152,178} Engineered protein filaments have been decorated with alcohol dehydrogenase to produce ethanol in *Escherichia coli* (*E. coli*), and protein nanoparticles have been used to encapsulate nucleic acids and small molecules for delivery.^{152,162}

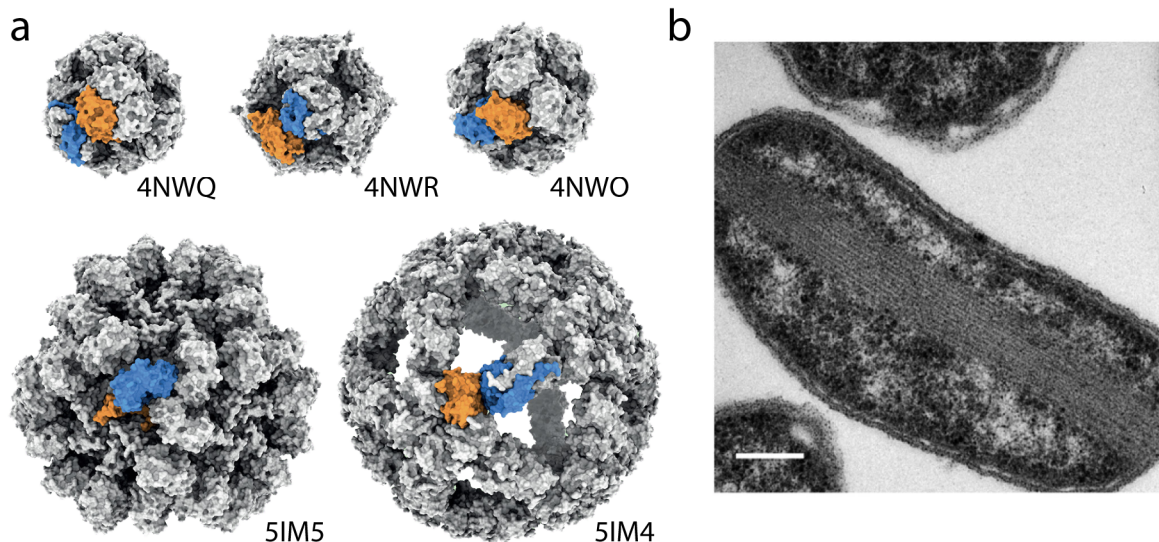


Figure 1.12: Engineered protein nanoparticles and fibres. Protein design and engineering has been leveraged to create discrete protein nanoparticles (a) and supramolecular protein fibres that self-assemble in living cells (b). Panels adapted from Edwardson *et al.* (2022) and Lee *et al.* (2018) with permission from the American Chemical Society and Springer Nature.^{152,177}

Central to all of these protein scaffolds however, is that they require strictly defined geometric interfaces to self-assemble correctly, and small changes to these assemblies can change their structure significantly. The recent discovery and exploration of proteins that can self-assemble into largely unstructured compartments by LLPS has presented an alternative route to protein self-assembly. Instead of designing geometrically defined, rigid scaffolds, protein designers and engineers have recently begun exploring the potential for constructing synthetic MLOs using natural and engineered proteins.

1.7.2 Designer protein condensates

The potential utility of LLPS in biology meant that it very quickly caught the attention of synthetic biologists. This not only presents the opportunity to design condensates from the bottom-up, but also circumvents many of the challenges and restrictions associated with designing a rigid, geometric protein assembly (Fig. 1.13).¹⁷⁹ These designer condensates typically incorporate elements from natural phase-separating proteins to drive protein condensation. Several such systems have now been generated using elements from a variety of endogenous condensates, including FUS and TDP-43 from ribonucleoprotein granules, LAF-1 from P granules, as well as some recombinant proteins such as spider silk and elastin.^{180–184}

These disordered motifs have proven highly capable of scaffolding engineered MLOs, and robust systems for phase separation have been achieved relatively readily through fusion of these extant IDRs to client proteins. In some cases, duplication of the phase-separating

domain is used to lower the critical saturation concentration, and drive phase separation more readily.^{185,186} The challenge for the field has been to build increasing levels of control and functionality into the designed condensates that have been created.¹⁸⁷ Controlled assembly and disassembly of MLOs has been introduced by several different triggers. Irreversible disassembly has been triggered by proteolytic digestion of the scaffold IDR, while highly specific and reversible phase separation has been engineered by the elegant Corelet and OptoDroplet systems.^{180,188} These systems use the FUS IDR as a scaffold component combined with photoactivatable oligomerisation domains.¹⁸² The increase in valency on photoactivation of the oligomerisation domains is sufficient to induce phase separation, and is reversible on deactivation of the photoactivatable domain. To expand the repertoire of synthetic triggers still further, an alternative system instead focuses on condensate disassembly. One system inducibly dissolves condensates by recruitment of a disruptive ligand called a C-BLOCK.¹⁸⁹ Recruitment of C-BLOCKS to FUS condensates inducibly dissolves MLOs in mammalian cells. These systems, and other works, demonstrate that phase separation is a process that can be rationally controlled by protein design concepts, analogous to more conventional protein assemblies.^{185,187}

In addition to its controllability, phase separation has gained significant interest due to its potential utility. The ability to inducibly sequester client proteins into MLOs offers several clear potential applications. An elegant study first demonstrated that protein condensates can be used to sequester natural proteins, and as a result directly control endogenous cellular processes.¹⁹⁰ This demonstrated several controllable and reversible methods of recruiting endogenous proteins to synthetic protein condensates, including by fusion of coiled coils, DNA oligonucleotides, and inducible PPIs to both the endogenous protein and a phase-separating RGG domain from LAF-1. Sequestration of natural proteins led to a functional knock-down of the targeted protein, with predictable impacts on cell biology.¹⁹⁰ Further work on the control of cell biology using synthetic condensates used variations of photoactivatable phase separation scaffolds, combined with clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 proteins, to drive controllable LLPS of nuclear proteins at specified genomic loci for targeted chromatin restructuring. In prokaryotes, researchers used the oligomeric PopZ protein from *Caulobacter crescentus*, and DivIVA from *Bacillus subtilis*, to create polar condensates in *E. coli* that induce cell asymmetry and asymmetrical division in the typically symmetrical bacteria.^{191,192}

Some of the most impressive demonstrations of the utility of protein condensates have come from the development of synthetic systems that act orthogonally to endogenous cellular processes. One approach used photo-inducible phase separation to direct metabolic flux through a branched metabolic pathway. Here, compartmentalisation of enzymes for deoxyviolacein biosynthesis was used to direct metabolic flux through either linear or branched metabolic pathways, using phase separation for post-translational control of

biosynthesis.¹⁹³ Another approach used the protective capacity of protein sequestration within condensates to protect the host cell from toxic intermediates. A recent study measured the production of a sesquiterpene, α -farnesene in *E. coli* within RGG condensates.¹⁹⁴ They demonstrated that *E. coli* expressing the enzymes Ldi and LspA within phase separated compartments demonstrated improved production of α -farnesene, which they attributed to alleviated cytotoxicity of Ldi-catalysed products. However, perhaps the most impressive example of engineered protein condensates comes from the use of biomolecular condensates to create organelles for orthogonal translation.¹⁹⁵ This enables replacement of amber stop codons with non-canonical amino acids only in mRNAs recruited to the engineered condensates, while the endogenous translation machinery operates simultaneously. These condensates have perhaps the greatest versatility demonstrated to-date, recruiting mRNAs, transfer RNAs (tRNAs), and ribosomes selectively to the synthetic organelles, but also demonstrates a powerful use case for phase separation in synthetic biology. This system has been developed further, by creating multiple orthogonal membraneless organelles using membrane localisation as a second layer of localised assembly, for simultaneous incorporation of several non-canonical amino acids within a single cell.^{196,197}

Despite these advances, several challenges remain in our capacity to engineer biomolecular phase separation - chief among them, the bottom-up design of scaffolds to induce phase separation. While several groups have demonstrated that it is possible to introduce new functionality into biomolecular condensates and control their assembly or disassembly, the vast majority of these systems rely on natural protein sequences as scaffolds to drive phase separation. In addition, these sequences are predominantly extensively intrinsically disordered, making elucidating the precise molecular interactions involved in phase separation challenging. However, some pioneering attempts have been made to design proteins for phase separation from the bottom-up. Artificial intrinsically disordered proteins have been developed using a minimal 8 residue repeat from Rec-1 resilin, a protein that naturally undergoes phase separation.¹⁹⁸ This created between 20 – 80 repeats of this octa-residue repeat, and includes rational mutations to understand the impact of changes to its residue composition on protein LLPS.

Instead of looking at natural sequences, other work has examined the construction of condensates using synthetic protein oligomerisation domains.¹⁹⁹ Combinations of dimeric and tetrameric oligomerisation domains produced synthetic condensates in *Saccharomyces cerevisiae* (*S. cerevisiae*). The relative concentrations of these two components, and their affinities, was shown to determine the phase boundary within cells. This work is noteworthy because it reduces the disordered component of the synthetic proteins almost entirely, building a system with quantifiable and structurally defined PPIs.

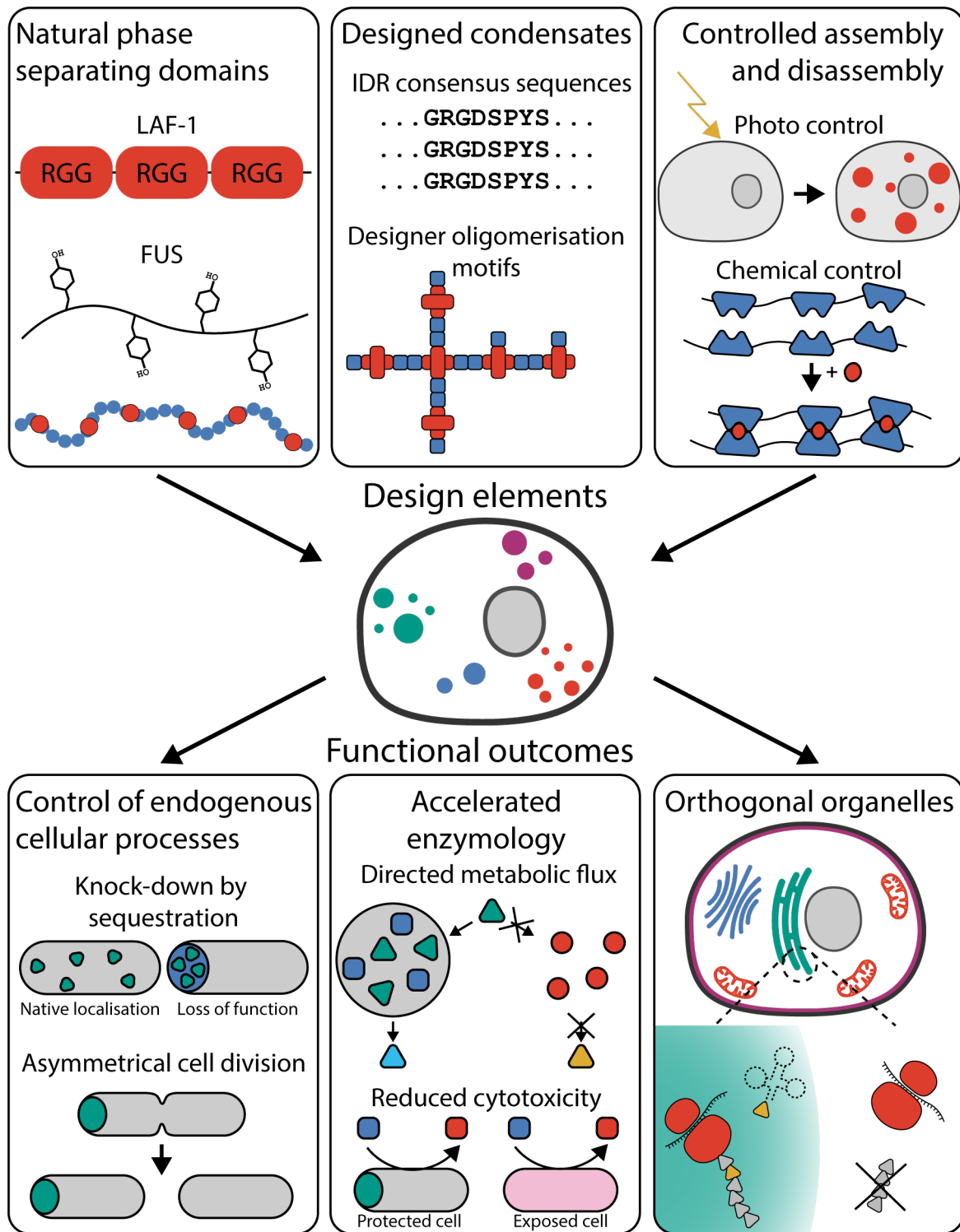


Figure 1.13: Design and engineering of protein condensates. Upper: designed elements that have been introduced into synthetic protein condensates, including engineering of natural phase-separating domains, the design of condensates from a bottom-up approach using natural consensus sequences and entirely artificial protein domains, and mechanisms for controlled assembly and disassembly of protein condensates. Lower: functional outcomes of the engineered condensates, including knock-downs of endogenous proteins, induction of asymmetry into cell division, accelerated enzymology by direction of metabolic flux or protection from cytotoxic intermediates, and the creation of orthogonal organelles for amber stop codon suppression.

Further work has described the design of other self-assembling condensates, also using two oligomerising domains.^{200,201} In addition to condensation, these assemblies use the well-characterised FK506 binding protein (FKBP) and FKBP–rapamycin binding (FRB) rapamycin dependent PPI interface, so that proteins can be selectively sequestered on addition of a small molecule trigger. Cargo release was engineered using the photoinducible LOV2 trap and release of protein (LOVTRAP) system, which uses the photosensitive interaction between LOV2 and Zdk1. On illumination with blue light the interaction is disrupted, and cargo released from the synthetic condensates.

Overall, our ability to engineer and even design phase-separating proteins is growing rapidly. Physical interrogation of LLPS has shed light on the molecular mechanisms of how these phase transitions occur in biology, while chemical dissection of PPIs provides new rationalisation of the interactions involved. These advances in our understanding have made phase separation accessible to synthetic biologists for the creation of artificial MLOs.^{202–204} However, significant questions still remain. Namely, how can we design a protein condensate so that it behaves like a liquid, rather than a diffusion-limited gel or arrested aggregate? Further, would it be possible to make such a protein from the bottom-up, using principles of protein design to create the interaction interfaces? And could such a *de novo* condensate confer similar benefits on catalysis as natural IDR-based systems? These questions are still relatively unexplored, and will form the premise of this thesis.

1.8 Scope of this thesis

Biomolecular condensates formed by protein phase separation have critical roles in cell biology, and the utility of phase separation in synthetic biology has been increasingly realised in recent years. Current state-of-the-art protein condensates re-purpose extant IDRs to scaffold LLPS, but more recently the potential for protein design to offer solutions for protein phase separation has been realised. The aim of this thesis is to use concepts in rational protein design to create a protein for LLPS from the bottom-up. This can offer insights into how the rational design of PPIs can drive the formation of different condensate materials, and the implications on catalysis within these designer MLOs. Chapter 3 describes the design and characterisation of a *de novo* polypeptide for LLPS *in vitro* and in *E. coli*, using rationally designed coiled coils as PPI motifs. Chapter 4 applies these designed MLOs to the colocalisation of a multi-enzyme cascade in cells, producing a clear improvement on pathway efficiency due to condensation. Chapter 5 investigates the formation of ordered fibres of the designed protein *in vitro* and in cells, and the interplay between LLPS and fibre formation. Chapter 6 applies the developed empirical design rules for LLPS to attempt to port the designer protein from prokaryotes

into eukaryotic cells. Together, these studies present new insights into the bottom-up design and functionalisation of *de novo* proteins for LLPS.

Chapter 2

Materials & methods

2.1 Materials

2.1.1 Molecular biology

Restriction enzymes, DNA polymerase, purple gel loading dye, DNA and protein electrophoresis size markers were supplied by New England Biolabs. DNA ligase and ProTEV Plus tobacco etch virus (TEV) protease were supplied by Promega. Agar, lysogeny broth (LB), agarose, glycine, sodium chloride (NaCl), polyethylene glycol (PEG) 3350, 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris-base), 2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol (Bis-tris), β -mercaptoethanol, ampicillin, chloramphenicol, isopropyl β -D-1-thiogalactopyranoside (IPTG), D-arabinose, N-Methyl-2-pyrrolidone (NMP), paraformaldehyde, phosphate buffered saline (PBS) tablets, sodium dodecyl sulfate (SDS), N,N,N',N'-tetramethylethylenediamine (TEMED), methanol, BugBuster lysis buffer, urea, tris(2-carboxyethyl)phosphine (TCEP) and imidazole were purchased from Sigma Aldrich. QIAprep miniprep and midiprep kits, and QIAquick DNA gel extraction kits were purchased from Qiagen. SybrSafe stain, SimplyBlue SafeStain, 40% Bis-acrylamide/acrylamide, ammonium persulfate (APS), and SuperSignal Pico chemiluminescence substrate was purchased from Thermo Fisher Scientific. cOmplete protease inhibitor tablets were supplied by Roche. Ni-Sepharose HisTrap High Performance (HisTrap HP) columns, and HiLoad 16/600 Superdex 75 pg and 200 pg size exclusion columns were supplied by Cytiva. ProLong Gold Antifade Mountant, Power Blotter 1-Step Transfer Buffer, and TC-ReAsH II was purchased from Invitrogen. Synthesised gene fragments ordered as double stranded DNA blocks and DNA oligonucleotides were supplied by Integrated DNA Technologies. DNA sequencing was performed by Source Bioscience. Plasmids were provided as gifts from other researchers. Mouse anti-polyhistidine antibody (H1029) was purchased from Sigma Aldrich, horseradish

peroxidase coupled secondary anti-mouse (J1430) was purchased from Invitrogen, and polyvinylidene fluoride (PVDF) western blotting membranes were purchased from GE Healthcare.

2.1.2 Oligonucleotides

Name	Sequence (5' – 3')	Length	%GC	T _m (°C)
HERD-2.2-GFP CMV fwd	CACGCGGCCCGCATGGGC AGTCATCACCACCATCA	34	65	61
HERD-2.2-GFP CMV rev	CACGGATCCTTATTTGT AAAGTTCGTCCATGCC TAATGTAATGC	44	41	61
HERD-0-GFP CMV fwd	GGTTTCCCTCTAGAAAT AATTTTGTTTAACTTT AA	35	26	56
HERD-0-GFP CMV rev	CACGGATCCTTATTTGT AAAGTTCGTCCATGCC TAATGTAATGC	44	41	61
HERD-2.2-XL-GFP CMV fwd	CACGCGGCCCGCATGG GCAGTCATCACCACC ATCA	34	65	61
HERD-2.2-XL-GFP CMV rev	CACGGATCCTTATTTG TAAAGTTCGTCCATGC CTAATGTAATGC	44	41	61

Table 2.1: Oligonucleotides used for sub-cloning of designed HERD constructs. GC content and melting temperature (T_m) calculated using SnapGene.

2.1.3 Mammalian cell culture

Dulbecco's Modified Eagle's Medium (DMEM), PBS for cell culture, and foetal calf serum were purchased from Gibco Invitrogen. Effectine transfection reagent was purchased from Qiagen. The antibiotics penicillin and streptomycin were supplied by PAA.

2.1.4 Instrumentation

Confocal fluorescence microscopy and fluorescence recovery after photobleaching were performed on a Leica SP8 confocal using a 65 mW Ar laser (488 nm) and 20 mW solid state laser (561 nm), with a 63x 1.4 numerical aperture oil immersion objective lens. Sonication was performed using a Sonics VibraCell. Western blots were imaged with a

G:Box Chemi-XT4 chemiluminescent imager (SynGene). An Äkta Pure 25L (Cytiva) was used for protein purification. Western blots were transferred using a Power Blotter XL (Invitrogen), and imaged G:Box Chemi-XT4 chemiluminescent imager (SynGene). Cloud-point measurements used a Perkin Elmer Lambda 35 UV/Vis spectrophotometer with a recirculating water bath and Peltier to control the sample temperature. Measurements of indigo production used a Perkin Elmer Lambda 25 UV/Vis spectrophotometer. Transmission electron microscopy was performed on a Tecnai 12 - FEI 120 kV BioTwin Spirit. Circular dichroism spectroscopy measurements were performed on a JASCO J-810 spectropolarimeter with a Peltier temperature controlled cuvette holder.

2.2 Recombinant protein expression

2.2.1 Molecular biology and DNA assembly

Synthetic DNA fragments for HERD protein construction were ordered as double stranded DNA blocks from Integrated DNA Technologies. For expression in *E. coli*, HERD sequences were subcloned with XbaI/NdeI restriction sites into the pET28a(+) derivative pDIC vector generously provided by Dr Matthew Lee. DNA was codon optimised for expression in *E. coli* K12 and to enable DNA synthesis of the repetitive sequences. HERD proteins include an *N*-terminal His-TEV tag to enable affinity purification by immobilised Nickel resins. DNA sequences for the client proteins mEmerald, mCherry, alcohol dehydrogenase, glucose-6-phosphatase, tryptophanase A, and flavin containing monooxygenase were synthesised as DNA fragments and subcloned *C*-terminally to the HERD protein using NdeI/SacI restriction sites, with an additional (GS)₃ linker.

Plasmids were assembled by restriction digest and ligation. Restriction digest reactions were assembled with 1 µg DNA, 10x digestion buffer (1 µl), and 0.5 µl each restriction enzyme made up to a volume of 10 µl with nuclease free water and incubated at 37 °C for 1 hour. Reaction products were analysed by agarose gel electrophoresis using 1% agarose in tris/acetate/EDTA (TAE) running buffer supplemented with SybrSafe DNA stain, run at 60 V for 45 minutes. Digested DNA fragments were gel extracted using QIAQuick gel extraction kit following the manufacturer's recommended protocol. Purified fragments were assembled by ligation using T4 DNA ligase. 100 ng of the insert fragment was incubated at 16 °C for 2 hours with 33 ng of the vector backbone fragment, 2x rapid ligation buffer (5 µl), and 1 µl T4 DNA ligase in nuclease free water to a total volume of 10 µl.

Ligated assemblies were transformed into chemically competent DH5α *E. coli*. 50 µl of ultracompetent DH5α were defrosted on ice and incubated with DNA (typically 3 µl for

a ligation reaction, and 0.5 μ l for an intact vector) for 30 minutes on ice. Reactions were heat shocked at 42 °C for 30 seconds, before chilling on ice for 1 minute. 200 μ l of LB was added and cells incubated at 37 °C for 1 hour, before spreading the cell suspension on LB-agar plates containing the appropriate antibiotics (ampicillin (AMP) - 100 μ g/ml, chloramphenicol (CMP) - 25 μ g/ml) and incubating at 37 °C overnight. A single colony was picked from the transformation plate and inoculated into 5 ml LB with antibiotics, before incubating overnight at 37 °C shaking at 200 rpm. Plasmid DNA was extracted from the overnight cultures using the QIAprep spin miniprep or midi-prep (for transfection of mammalian cells) following the manufacturer's recommended protocols. All plasmid assemblies were verified by test restriction digest, followed by DNA sequencing by Source Bioscience.

For mammalian expression, HERD-GFP sequences were subcloned into the commercial cytomegalovirus vector pCMV. The restriction sites NotI/BamHI were introduced upstream and downstream of the open reading frame by polymerase chain reaction (PCR) and used for subcloning into pCMV. PCR was performed using Q5 high fidelity DNA polymerase from New England Biosciences. 50 μ l reactions were assembled with 100 ng of template DNA, 5x polymerase reaction buffer (10 μ l), 10 mM deoxynucleotide-triphosphates (dNTPs), 10 μ M each forward and reverse primer, and 0.5 μ l Q5 DNA polymerase made up 50 μ l with nuclease free water. PCR cycles used an initial 98 °C, 30 s denaturation step, followed by 30 cycles of 98 °C (10 s), 55 °C – 72 °C (oligonucleotide annealing temperature dependent, 30 s), and 72 °C (30 s per kb of product length). Final extension was provided by a 2 minute 72 °C incubation before incubating at 4 °C. The PCR reaction was verified by agarose gel electrophoresis and the product purified by gel extraction, before assembly into pCMV by restriction digest and ligation.

2.2.2 Protein expression and purification

For protein expression, chemically competent BL21*(DE3) *E. coli* were transformed with the vector of interest as described. A single colony from the transformation plate was inoculated into a 5 ml LB overnight culture supplemented with AMP or CMP, and grown overnight (37 °C, 200 rpm). Expression cultures (50 ml – 1 l) supplemented with AMP or CMP were inoculated 1:100 with overnight culture and grown (37 °C, 200 rpm) to OD₆₀₀ = 0.4 – 0.6. Protein expression was induced with 400 μ M IPTG/0% – 0.2% D-arabinose and the culture grown at the chosen expression temperature (18 °C – 37 °C, 200 rpm) for up to 18 hours.

For protein purification, cells were collected by centrifugation (3400 xg, 20 minutes) and the pellet resuspended in 40 ml of resuspension buffer (20 mM Tris pH 7.5, 50 mM imidazole, 500 mM NaCl, 2 M urea, plus 1 tablet cComplete protease inhibitor cocktail). Cells

were lysed on ice by sonication for 15 minutes (5 s on, 2 s off, 75% amplitude). Following complete lysis, the suspension was centrifuged (16000 xg, 20 minutes). The supernatant was retained and filtered using 0.2 μ m syringe filters to clarify. The clarified supernatant was then added to a 5 ml HisTrap HP immobilised metal ion affinity chromatography (IMAC) column pre-equilibrated in Ni-nitrilotriacetic acid (Ni-NTA) binding buffer (20 mM Tris pH 7.5, 500 mM NaCl, 50 mM imidazole, 2 M urea). After sample application, the column was washed in binding buffer for 4 column volumes. Then the bound protein was eluted using a gradient of NiNTA elution buffer (0 – 100%) (20 mM Tris pH 7.5, 500 mM NaCl, 2 M urea, 500 mM imidazole). Following, the elution fractions were further purified using a HiLoad Superdex 200 pg size exclusion column at 1 ml/min flow rate (20 mM Tris pH 7.5, 2 M urea). The resulting fractions were analysed by SDS-polyacrylamide gel electrophoresis (PAGE). 6 μ l of sample was added to 6 μ l of prestained loading dye and boiled at 95 °C for 10 minutes. Then, 8 μ l of the fully denatured solution was loaded on a 10 – 12% acrylamide/bis-acrylamide (29:1) gel, with 8 μ l of broad-range pre-stained protein standard (New England Biolabs) used as a molecular weight ladder, and run at 200 V for 1 hour. The gel was then stained with SimplyBlue SafeStain Coomassie G-250 total protein stain for 1 hour, and destained in water. Fractions were pooled and dialysed into 20 mM Tris pH 7.5 using a 26/10 desalting column run at 5 ml/min. Desalted fractions were pooled and concentrated using 10 kDa molecular weight cut-off Amicon spin columns, and snap frozen in liquid nitrogen for storage at -70 °C. Protein concentration was measured by absorbance at 280 nm using a Nanodrop and calculated into molar concentrations using the molecular weight and molar extinction coefficient from the primary amino acid sequence.

2.2.3 TEV cleavage

TEV cleavage of HERD-2.2-GFP was performed using ProTEV Plus (Promega). The reaction was set up with 18 mg of HERD-2.2-GFP, 0.5 mM EDTA, 1 mM DTT, and 200 units of ProTEV Plus, to create a total volume of 12 ml. The reaction was incubated at 30 °C overnight, shaking at 50 rpm. Following cleavage, the protein was purified from the tag a HisTrap HP column and collecting the unbound flow-through. Complete cleavage and purification was confirmed by running samples on SDS-PAGE as described above, and total protein staining with SimplyBlue SafeStain.

2.3 Cell culture

2.3.1 Mammalian cell culture

HeLa cells were cultured in Dulbecco's Modified Eagle's Medium with high glucose, 10% foetal calf serum and 5% penicillin/streptomycin. Cells were grown at 37 °C and 5% CO₂. For transfection of recombinant proteins, 1 – 4 x10⁵ cells were seeded the day before onto individual 35 mm culture dishes with 175 µm glass bottoms. Cells were transfected with 0.4 µg DNA using the Effectene transfection reagent (Qiagen), following the manufacturer's recommended protocol and grown for 18 – 22 hours before imaging.

2.4 Microscopy and protein detection

2.4.1 *E. coli* cell imaging

For visualisation of *de novo* proteins in *E. coli*, from the overnight culture, 50 ml of LB was inoculated with 500 µl and grown (37 °C, 200 rpm) to OD₆₀₀ = 0.4 – 0.6. For fixed cell imaging, after protein expression was induced with IPTG, the cultures were grown at 18 °C (200 rpm) for 5 hours. Aliquots of expression cultures were taken (typically 1 ml), and the cells collected by centrifugation (3000 xg, 5 minutes). The cell pellet was then washed 3 times in 1 ml PBS, followed by fixing with 1 ml 2% formaldehyde in PBS and incubating for 15 minutes. The cells were then pelleted (3000 xg, 5 minutes) and the formaldehyde discarded. The cells were washed 3 more times with 1 ml PBS before resuspending in 100 µl PBS. Two drops of ProLong Gold Antifade Mountant (Invitrogen) were added to a clean glass slide, and 10 µl of the fixed cells added to the mountant before covering with a coverslip and leaving to set overnight.

To image cells live, variable growth and imaging temperatures were used after induction of expression (18 °C – 37 °C). Cell pellets were collected by centrifugation (3000 xg, 5 minutes) and instead directly resuspended in 50 µl PBS and 15 µl applied to a slide pre-equilibrated to match the growth and imaging temperature. The coverslip was sealed with clear nail polish to prevent evaporation and carried to the microscope on a temperature controlled heat block. Microscope temperature was controlled using a variable temperature controlled stage (Linkam), pre-heated to the chosen imaging temperature at least 30 minutes before imaging. Microscopy images were analysed using ImageJ and presented as maximum intensity projections, assembled using Adobe Illustrator.^{205,206}

2.4.2 Mammalian cell imaging

All mammalian cell imaging was performed on live HeLa cells grown at 37 °C, and the temperature was maintained before and during imaging as described above. 18 – 22 hours after transfection, cells were washed 3 times in preheated PBS and imaged live in the glass bottomed culture dishes.

2.4.3 Automated image analysis

Cell samples for automated image analysis of condensate formation were collected as described above for *E. coli* imaging of fixed cells. Cells were collected beginning at the induction of protein expression in the mid-exponential phase of growth ($OD_{600} = 0.4 - 0.6$). This was established as timepoint 0, and cells collected and fixed for confocal microscopy every hour for the following 6 hours, when the cultures reached $OD_{600} = 2 - 2.5$. Both brightfield and fluorescence microscopy images were collected for automated cell detection and quantification.

Automated cell detection and condensate quantification used the custom ModularImageAnalysis (MIA) plugin for ImageJ (version 0.21.11), created and adapted by Dr Stephen Cross for this purpose.^{207,208} In brief, individual cells were detected from Z-stack images of *E. coli*. For this, the brightfield Z-stacks were background subtracted using a sliding paraboloid with a radius of 10 pixels. Then, single Z-slices giving the best focal plane were selected using the Stack Focuser ImageJ plugin.²⁰⁹ The focused images were then normalised for their brightfield intensity and background corrected further.²¹⁰ The corrected brightfield images were then passed to the convolutional neural network-based plugin StarDist for detection of *E. coli*, using a model trained on the DeepBacs *E. coli* dataset.²¹¹ Overlapping cells were discounted to avoid conflation of foci from multiple cells. Foci within the identified individual cells were detected using maximum intensity projections from the fluorescence microscopy Z-stacks. Background intensity was removed, and TrackMate’s LoG spot detector was used to convolve the image to enhance foci-like features, and detect foci that are brighter than a set threshold.²¹² The number of foci, their area, and their total fluorescence intensity were measured for each cell and recorded for statistical analysis.

2.4.4 TC-ReAsH II labelling

For specific labelling of proteins encoding tetracysteine sequences (CCPGCC), the bi-arsenical dye TC-ReAsH II (Invitrogen) was used.²¹³ For *in vitro* labelling, tetracysteine encoding proteins were each incubated at 100 μ M with 1 μ M TC-ReAsH II and 1 mM

TCEP in the dark, to prevent photobleaching. For measurement of enrichment within HERD-2.2-GFP droplets, the proteins were then mixed with 2 mM HERD-2.2-GFP, before inducing phase separation by addition of buffer containing 250 mM NaCl, 8% PEG 3350, and 20 mM Tris pH 7.5. The droplets formed were measured by excitation at 561 nm (TC-ReAsH II). For subtraction of background, HERD-2.2-GFP droplets containing no tetracysteine labelled proteins were incubated and imaged in parallel alongside each sample, and the fluorescence intensity within non-labelled droplets subtracted as background.

For in-cell labelling, cells were grown as described previously (18 °C, 200 rpm) for 5 hours after induction of protein expression. Here, plasmids encoding HERD-2.2-GFP and the tetracysteine labelled protein were co-transformed and co-expressed, along with cells expressing only HERD-2.2-GFP as a control for non-specific labelling. After 5 hours, 1 μ M TC-ReAsH II and 1 mM 1,2-ethanedithiol (EDT) were added for specific labelling of tetracysteine encoding proteins. Cells were then imaged live at room temperature.

2.4.5 Western blotting

For western blotting, transformed *E. coli* were grown overnight and inoculated 1:100 into 50 ml of fresh LB plus AMP or CMP. Cells were grown as described at 18 °C, 200 rpm, for 5 hours after the induction of protein expression. To normalise protein aliquots to cell density, aliquot volume was normalised to OD_{600} of the culture according to the following equation:

$$Aliquot\ volume\ (\mu l) = \frac{Volume\ of\ lysis\ buffer\ (\mu l) \times 400}{OD_{600}} \quad (2.1)$$

The 400 multiplier was chosen empirically as it gave good protein concentrations for visualisation by SDS-PAGE. Typically, 50 μ l was the volume of lysis buffer used for resuspension. The aliquots were collected by centrifugation (3000 xg, 15 minutes) and resuspended in BugBuster lysis buffer (Sigma Aldrich), with 1 μ l benzonase, before incubating at 37 °C for 30 minutes. The solution was then freeze thawed using liquid nitrogen 3 times to ensure complete lysis. Where cell lysates were separated into their soluble and insoluble fractions, the solution was then centrifuged (18000 xg, 20 minutes). The supernatant was removed and collected as the soluble protein fraction. The pellet was washed once in PBS, before resuspending again in 50 μ l of BugBuster as the insoluble protein fraction.

Samples were run on SDS-PAGE gels as described previously, but without staining with Coomassie total protein stain. Proteins were transferred to a 0.2 μ m PVDF western blot-

ting membrane with a Power Blotter XL (Invitrogen) transfer system using the recommended transfer buffer (Invitrogen) for 15 minutes at 1.3 A. The membrane was blocked using 5% milk powder in PBS-T for 1 hour, before incubating with the primary antibody (H1029, Sigma Aldrich) in 4% milk in PBS-T (1:5000) for 1 hour. The membrane was then washed 3 times for 5 minutes each in PBS-T to remove the primary antibody before incubating with the HRP coupled secondary anti-mouse antibody (J1430, Invitrogen) in 4% milk in PBS-T (1:10000) for 1 hour. The membrane was then washed 3 more times with PBS-T. Finally, 2 ml of SuperSignal West Pico Plus chemiluminescent substrate was added, and incubated for 1 minute before imaging.

2.4.6 Transmission electron microscopy

Purified protein samples for transmission electron microscopy (TEM) were prepared and deposited onto a glow-discharged 300-mesh carbon grid coated with pioloform. 10 μ l of sample was applied to a grid and incubated for 1 minute, then samples stained using a 2% uranyl acetate (UA) solution. Grids were swept through a drop of UA, then dropped sample-side down into a second drop and incubated for 3 minutes. The grid was then blotted with filter paper to remove excess stain, and swept through a third UA drop, and blotted again. The grid was then swept through two drops of ultrapure water, before blotting and leaving to dry. Stained grids were then imaged on a 120 kV Tecnai 12 electron microscope. Staining, sectioning and imaging of cell sections by TEM was performed by Dr Lorna Hodgson. 1 ml of cell suspension was pelleted by centrifugation, and 1 μ l of pellet was vitrified using a Leica EM PACT2 high pressure freezer with a rapid transfer system. The vitrified cells were freeze substituted with 0.2% UA, 5% H₂O in acetone for 5 hours at -90 °C using a Leica AFS2 automated freeze substitution system. The samples were then warmed to -45 °C, and kept for 2 hours at that temperature before washing in acetone for 30 minutes. Resin (Lowicryl HM20) was then infiltrated into the freeze substituted samples at increasing dilutions (25%, 50%, 75%) for 3 hours each, before embedding in 100% resin for 16 hours, followed by 3 changes of resin, left for 2 hours each. The infiltrated resin was then polymerised using UV for 48 hours. The resin blocks were sectioned using an EM UC6 microtome with a diamond knife at 45 °. For TEM, slices were placed onto TEM grids for imaging. For correlative light electron microscopy (CLEM), slices were first imaged using a Leica SP8 AOBS confocal microscope with a 63x oil immersion objective, before negative staining with 2% UA as described for purified samples, and imaging by TEM.

2.5 Soft-matter biophysics

2.5.1 Cloud-point measurements

Phase separation of HERD-2.2-GFP was measured in 4% PEG 3350, 125 mM NaCl, and 20 mM Tris pH 7.5, with the concentration of HERD-2.2-GFP varied from 2.7 – 37 mg/ml. The protein sample was filtered using a 0.2 μm filter, and all the materials, including the cuvette holder, cuvette, and protein sample were pre-heated in an incubator to 40 °C for at least 20 minutes to equilibrate. The solution temperature during measurement was measured using a temperature probe placed in the reference cuvette, with an identical buffer solution. Phase separation was measured by detecting the change in percent transmission (%T) at 600 nm, while the temperature was initially decreased from 40 °C to 5 °C, and then restored from 5 °C to 40 °C. The 50% transmission value as the temperature was reduced was identified as T_{cloud} , while the 50% transmission value as the temperature was increased again was identified as T_{clear} . The binodal phase boundary at each protein concentration point is the mean of T_{cloud} and T_{clear} .

2.5.2 Fluorescence recovery after photobleaching

For measurement of fluorescence recovery, imaging was recorded at 400 hz bidirectionally, giving an interval of 648 ms between frames. HERD-2.2-GFP droplets *in vitro* were formed by 33 mg/ml HERD-2.2-GFP, with 125 mM NaCl, 20 mM Tris-HCl pH 7.5, and 4% PEG 3350. *E. coli* expressing HERD-GFP proteins for in-cell measurement were grown and prepared as described for live cell confocal microscopy imaging. Prior to bleaching 3 measurements were recorded of the bleach point, giving the pre-bleach fluorescence intensity as their mean. Bleaching was performed using a single 40% laser burst for 100 ms (for *in vitro* bleaching) or 1 ms (for in-cell bleaching), and fluorescence recovery recorded over a period of 20 – 30 s. The background fluorescence intensity was subtracted from all measurements. To allow comparison between different bleaching experiments, recovery was normalised to the pre-bleach fluorescence intensity and to the fluorescence intensity measured immediately after bleaching. To correct for spontaneous bleaching during measurement, a second independent reference area was measured during the experiment. *In vitro* this was a distal area of the same large droplet, and in-cell this was an independent condensate in a different *E. coli* cell. This reference area was background subtracted and normalised in the same manner as the bleach area, and the bleach area divided by the reference area to account for spontaneous bleaching. $t_{1/2}$ values were calculated by fitting FRAP data in OriginPro. An exponential of $f(t) = A \cdot (1 - e^{-\tau t})$ was used. Here, A is the plateau intensity, τ is the fitted parameter, and t is the time

after bleaching. Using this exponential, half-lives were determined using the equation:
 $t_{1/2} = \frac{\ln(0.5)}{\tau}$.

2.6 Enzymology measurements

2.6.1 Measurement of indigo production

Indigo was produced in *E. coli* by the recombinantly expressed tryptophanase (TnaA) and flavin containing monooxygenase (FMO). A TnaA knockout strain of *E. coli* was used (Δ tnaa BL21 DE3), kindly given for this purpose by Dr Chong Zhang, so that all product formation was due to the recombinantly expressed enzymes.²¹⁴ In addition, for cell density measurements OD₇₀₀ was used instead of OD₆₀₀, to avoid conflation with the absorbance spectrum of indigo (measured by absorbance at 610 nm).²¹⁵ For co-expression of HERD proteins and the two enzymes, Δ tnaa *E. coli* were transformed with two vectors, one with the HERD protein under the T7 promoter (AMP selection), and a second polycistronic vector with the two complimentary enzymes TnaA and FMO under the arabinose promoter (CMP selection). A single colony from the transformation plate was used to inoculate an overnight culture, supplemented with both AMP and CMP antibiotics. From the overnight culture, 50 ml of fresh LB supplemented with AMP and CMP was inoculated 1:100 and grown (37 °C, 200 rpm) to OD₇₀₀ = 0.4 – 0.6. Protein expression was then induced with 400 μ M IPTG to induce expression of the HERD protein, and varying concentrations of D-arabinose (0% – 0.2 % w/v) to induce titratable expression of the TnaA and FMO fusions. Cultures were then grown at 18 °C or 33 °C, 200 rpm for 22 hours. Cell density was then recorded by measurement at OD₇₀₀. For indigo measurement, a 2 ml aliquot of each culture was collected by centrifugation (3000 xg, 5 minutes). The cell pellet was resuspended in NMP and sonicated (3x 5 seconds, 60% amplitude) to lyse the cells and resuspend the indigo. The solutions were centrifuged (13000 xg, 20 minutes) to remove cell debris before measurement of indigo concentration by absorbance at 610 nm. Finally, the expression levels of the two enzymes TnaA and FMO were measured by western blotting, for normalisation both to cell density and enzyme expression. For this, samples for western blotting were collected and performed as described previously, and blotted against the *N*-terminal His tags of the two enzymes. Despite their very similar molecular weights, FMO runs consistently lower on SDS-PAGE than TnaA due to its significantly different net charge, permitting analysis of both enzymes from a single blot. Enzyme expression was quantified from western blots using triplicate cultures in Image Studio Lite. Background subtraction used a 3-point top-and-bottom subtraction around the band of interest to subtract non-specific background particularly visible in the HERD samples. The concentration of FMO was used for the final normalisation as its catalysed

oxidisation of the indole to the indoxyl has been characterised as the rate-limiting step in this reaction pathway, with a $k_{\text{cat}}/K_{\text{M}}$ of $7.8 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ compared to the $2.7 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ of TnaA.^{216,217} The final relative indigo production was reported relative to the amount of indigo produced by the free enzymes, which was set as 1.

2.7 Structural biophysics

2.7.1 Solid phase peptide synthesis

Solid phase peptide synthesis (SPPS) was performed using a Liberty Blue peptide synthesiser from CEM. Peptides were synthesised as *C*-terminal amides on a rink amide resin using Fmoc protected amino acids, coupled with DIC/Oxyma. Amino acids were deprotected using solutions of 20% morpholine in DMF. Synthesised peptides were *N*-terminally acetylated by addition of pyridine (5%) and acetic anhydride (3%) in DMF and incubated for 1 hour at room temperature. Peptides were then cleaved from the resin by adding 95% trifluoroacetic acid (TFA), 2.5% H₂O, and 2.5% triisopropylsilane and incubated on a rotating platform for 2 hours at room temperature. The cleaved peptide was collected and excess TFA evaporated using a stream of N₂. Residual TFA was removed by precipitation of the peptide with ice cold diethyl ether, and centrifugation (4000 xg, 10 minutes). The peptide pellet was then resuspended in deionised water.

2.7.2 Peptide purification

Peptides made by SPPS were purified by high pressure liquid chromatography (HPLC) on a reverse phase C18 column with a 5 μm particle size and 100 Å pore size. In general, the crude peptide was dissolved in 20% acetonitrile with 0.1% TFA, and separated using a 20 – 100% gradient of acetonitrile over 30 minutes, with elution monitored at 280 nm. Peptide mass was confirmed by mass spectrometry and analytical HPLC.

2.7.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were recorded on a JASCO J-810 spectropolarimeter with a Peltier temperature controller, in a 10 mm path length reduced volume cuvette. Single measurements of ellipticity were performed at 5 °C. Measurements of ellipticity with respect to temperature (melting and cooling spectra), were recorded by collecting initial spectra at 5 °C, with ellipticity measured at 222 nm every 1 °C and full spectra measured every 5 °C as the temperature was increased to 90 °C, and then decreased again to 5 °C. Full spectra measured ellipticity between 190 nm and 260 nm in 1 nm intervals,

with a 100 nm/min scanning rate, 1 nm bandwidth and 1 s response time. A reference spectrum using the same cuvette, parameters and buffer at 5 °C was subtracted from the measured ellipticity. Ellipticity (deg) values were converted to mean residue ellipticity (MRE) ($\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}\cdot\text{res}^{-1}$) by normalisation to the number of peptide bonds in the protein, and the path length using the following equation:

$$MRE (\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}\cdot\text{res}^{-1}) = \frac{\theta \times 100}{c \times l \times b} \quad (2.2)$$

Where θ is the difference in absorbed circularly polarized light in millidegrees, c is the protein concentration in mM, l is the path length in cm, and b is the number of amide bonds in the protein. Fraction helicity was calculated using the MRE at 222 nm (MRE_{222}) using the following equation:

$$Fraction\ helix\ (\%) = 100 \times \frac{MRE_{222} - MRE_{coil}}{-42500 \times (1 - \frac{3}{n}) - MRE_{coil}} \quad (2.3)$$

Where MRE_{coil} is 640 - 45T, T is the temperature in degrees Celcius, and n is the number of amide bonds in the sample.

2.7.4 X-ray fibre diffraction

X-ray fibre diffraction was performed and analysed with the help of Prof Louise Serpell at the University of Sussex. Fibres were prepared in solution without NaCl or PEG to prevent diffraction peaks due to salt rings. Solutions containing purified fibres (16 mg/ml protein, 20 mM Tris pH 7.5) were hung between two glass capillaries with the ends sealed with paraffin, placed approximately 1 cm apart. These droplets were left to dry for 18 hours to produce dry protein fibres suspended between the capillaries. Fibres were aligned in the detector at 0 and 90° orientations, before diffraction using a Rigaku copper rotating anode X-ray source with a Saturn CCD detector at a distance of 50 mm for 30 – 60 s. Diffraction patterns were analysed using CLEARER to identify distances.

2.8 Computational tools

Protein parameters, including molecular weight, pI, net charge, and extinction coefficient (ϵ) at 280 nm were calculated from primary amino acid sequences using ExPASy ProtParam (<https://web.expasy.org/protparam>).²¹⁸ Models of homomeric coiled coils were generated using CCBuilder 2.0 using ISAMBARD (<http://coiledcoils.chm.bris.ac.uk/ccbuilder2/builder>).^{219,220} Predictions of protein in-

trinsic disorder were performed using IUPred3 (<https://iupred3.elte.hu>).^{221,222} Protein structure prediction and modelling was performed using AlphaFold2 ColabFold notebooks (<https://colab.research.google.com/github/sokrypton/ColabFold/>).²²³ All data analyses except those mentioned explicitly were performed in Python using the Pandas and Numpy libraries, and visualised using the Matplotlib and Seaborn libraries.

Chapter 3

Rational design of proteins for LLPS

The following work in this chapter comprises part of the publication: Assembling membraneless organelles from de novo designed proteins, Alexander T. Hilditch, Andrey V. Romanyuk, Stephen J. Cross, Richard Obexer, Jennifer J. McManus & Derek N. Woolfson, Nature Chemistry (2023).²²⁴ All of the work described in this chapter is that of the author, with the exception of those mentioned explicitly (construction and training of the MIA model written by Dr Stephen Cross, and used by the author for Figs. 3.12 and 3.16, and CD spectroscopy performed by Dr Andrey Romanyuk, presented in Fig. 3.15).

3.1 Chapter introduction

Phase separation of proteins is no longer the mysterious physical phenomenon that it first appeared when identified in the early 2000s. Recent physical and chemical studies on the properties of phase-separating proteins have begun to dissect the molecular architecture of their interactions and understand how they function *in vivo*.⁸³ Overall, phase-separating proteins can be considered to form a network of intermolecular interactions that drive de-mixing.⁸⁶ Therefore, phase separation can be considered to be an interplay between interaction valency, and affinity.⁸⁵ As molecular valency increases, the propensity of a protein to phase separate increases.^{88,225–227} As a result, many natural and engineered proteins for LLPS use repetitive PPI motifs to increase the overall molecular valency.^{90,95} Complementary to valency is interaction affinity. Overall, the higher the interaction affinity, the lower the protein concentration required to induce phase separation, known as the saturation concentration (C_{sat}).^{228–230} The interaction affinity of phase-separating proteins is challenging to elucidate, as the majority of systems use highly dynamic IDRs as scaffolds. However, engineered protein condensates using structurally defined PPIs have demonstrated these principles, that protein condensation is an interplay between valency

and affinity of interactions.^{188,199,231} In addition, more nuanced concepts surrounding the interactions that drive phase separation have begun to emerge. The interplay between intramolecular and intermolecular interactions is being recognised, as a high propensity for intramolecular interactions reduces the overall extent of network formation, and therefore attenuates phase separation.²²⁶ Taken together, these studies begin to create an image of the molecular architecture of proteins that drive phase separation. Nonetheless, the construction of minimal, *de novo* designed systems for LLPS from the bottom-up remains a formidable challenge.

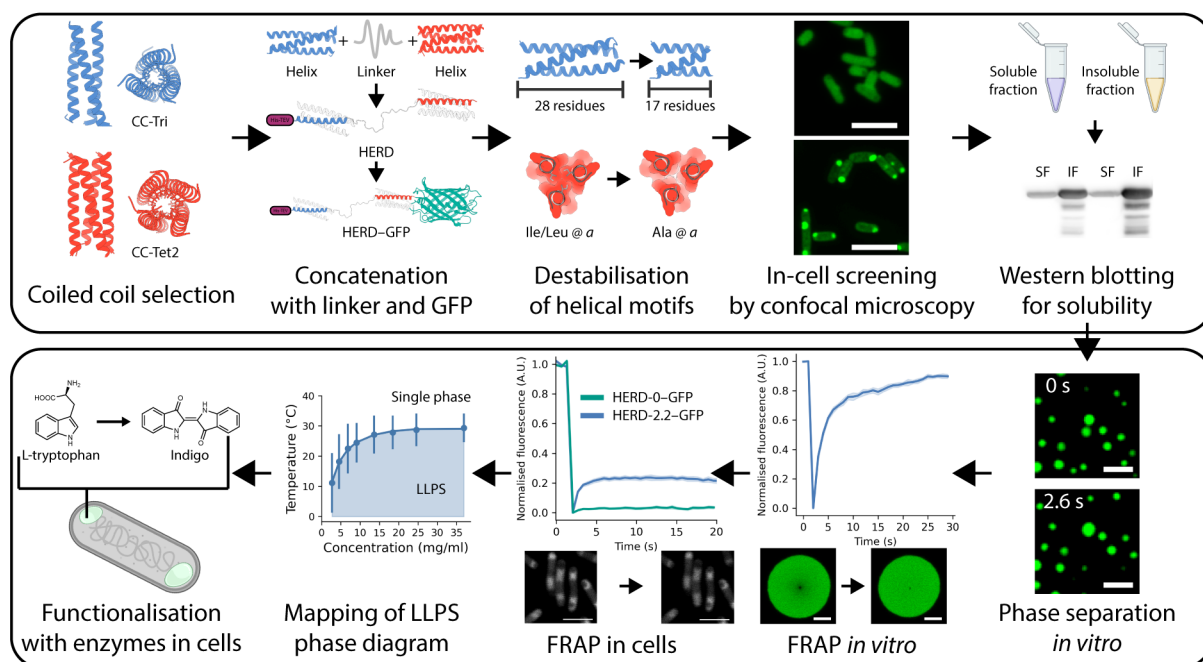


Figure 3.1: Workflow for the design and characterisation of proteins for LLPS. Rational design of concatenated coiled coils was followed by in cell and in vitro characterisation, for the ultimate creation of organelles in cells that can be functionalised with enzymes.

This chapter describes the design of a minimal polypeptide for LLPS in *E. coli*. PPI motifs are designed using extant *de novo* coiled-coil motifs, and their interactions rationally tuned to produce dynamic liquid condensates rather than arrested aggregates. Protein phase separation is characterised using soft-matter biophysical techniques both *in vitro* and in cells to confirm it forms compartments that behave like de-mixed liquids. This design process unifies many of the concepts about the molecular architecture of biomolecular condensates into a single design pipeline, demonstrating that the bottom-up design of LLPS is tractable and can be created from *de novo* protein motifs (Fig. 3.1).

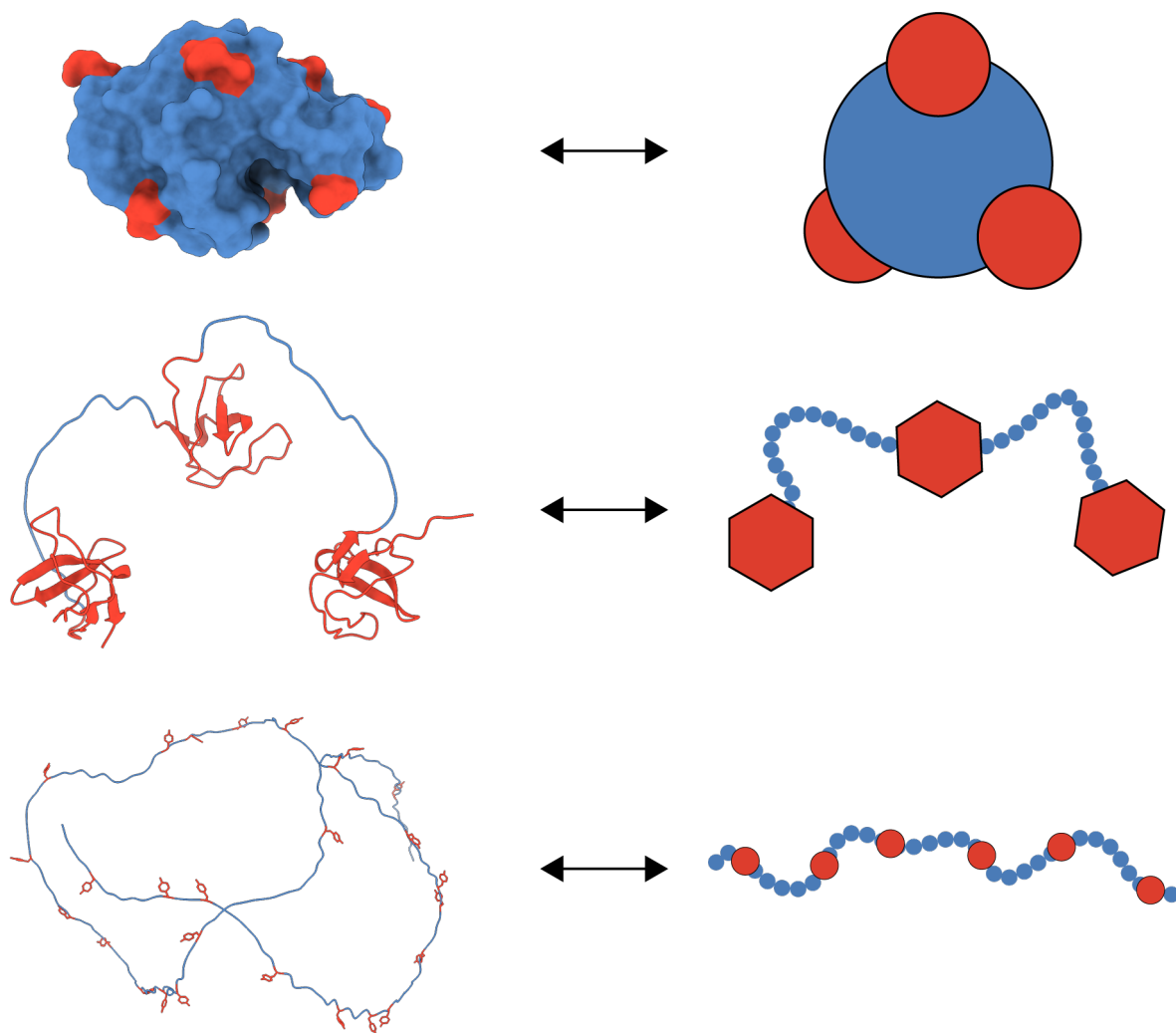


Figure 3.2: The stickers and spacers model for phase-separating proteins. Stickers and spacers can be used to represent a range of architectures of phase-separating proteins. Three different models are shown, with stickers highlighted in red, and spacers in blue. Top: the globular protein lysozyme undergoes LLPS with surface interactions acting as stickers. Middle: Poly SH3 domains have been used as stickers in engineered phase-separating proteins. Bottom: the *N*-terminal IDR from FUS is rich in tyrosine residues acting as single residue stickers. Figure adapted from Choi, Holehouse, and Pappu (2020) with permission from Annual Reviews.⁸⁹

3.2 Bottom-up design of *de novo* stickers and spacers

To create the initial framework for the design of a phase-separating protein, a stickers and spacers arrangement was used (Fig. 3.2).⁸⁹ This places PPI motifs (stickers) along a linear polypeptide, separated by linkers (spacers). This arrangement of motifs has been described previously as a way of generating the high valency of interactions required for LLPS.^{91,232} Further, it is a broad framework, which permits modularity of motifs and domains. For instance, sticker motifs can be folded, globular PPI domains, as in the case of poly-SH3 proteins.^{89,233} Alternatively, stickers can be individual residues in an IDR, as with the repeated tyrosine residues in FUS.³¹ Sticker and spacer frameworks have

even been used to describe the phase separation of monomeric globular proteins, such as lysozyme, which phase separates at high NaCl concentrations.²³⁴ However, in this case the stickers are attractive patches on the globular protein surface, physically separated by the protein's surface.

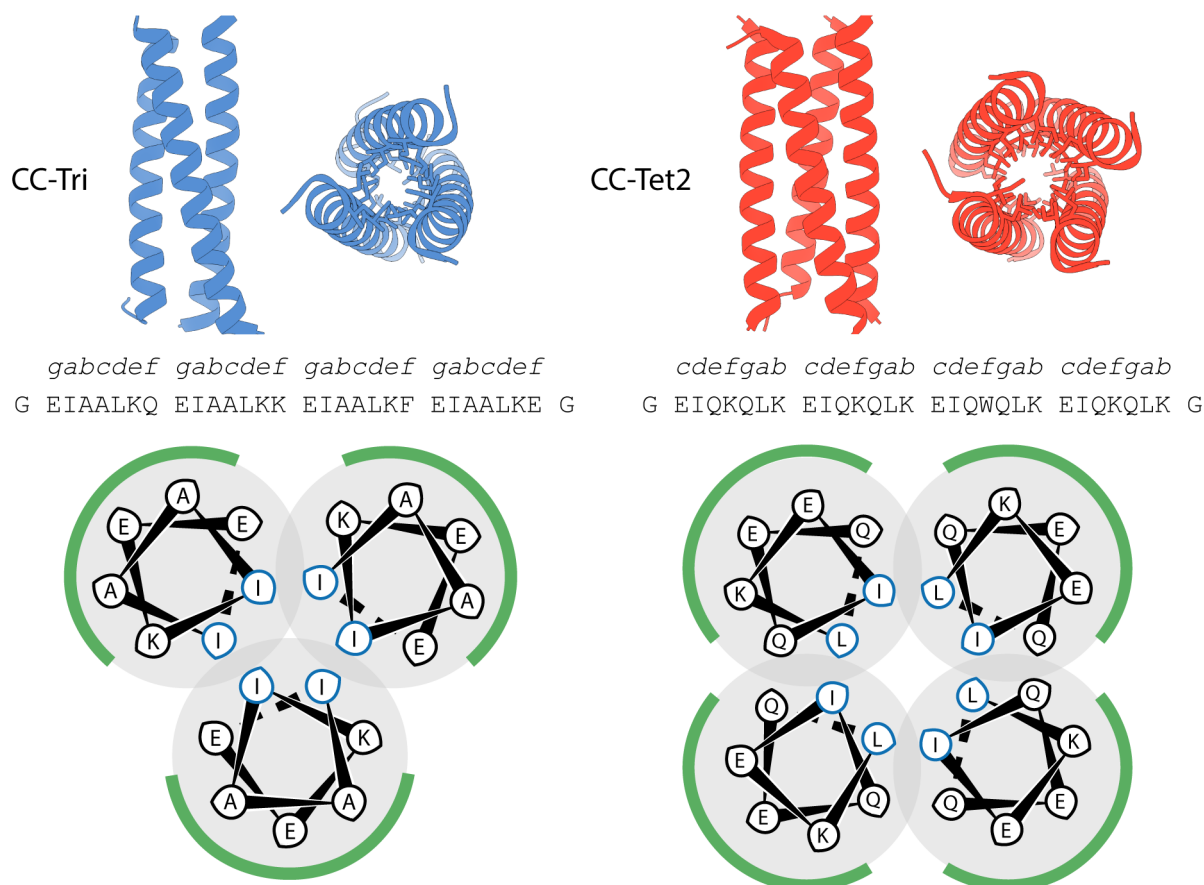


Figure 3.3: Orthogonal coiled coils as peptide stickers. Crystal structures of the coiled coils CC-Tri (blue, PDB: 4DZN) and CC-Tet2 (red, PDB: 6XY1), shown from the side, and top down, with the interacting *a* and *d* residues shown as sticks. Sequences for the peptides are given below, with the heptad repeat in italics, along with a helical wheel representation of the coiled coils.

3.2.1 Coiled-coil stickers

To create a *de novo* protein for LLPS, rather than relying on natural PPIs or phase-separating motifs, coiled-coil motifs were used as PPIs for the stickers. This not only ensures that phase separation has been designed truly from the bottom-up, but also enables as detailed as possible rationalisation of the contributing PPIs. Coiled-coil motifs are a tractable choice for stickers in such a framework for several reasons. Firstly, as valency has been demonstrated to be key for protein condensation, the stickers should aim to maximise the valency of the overall molecule.²²⁷ Coiled coils have been created to form

dimers to octamers in solution, giving an extremely high potential valency in a minimal sequence motif.^{111,142,235} Secondly, within coiled-coil sequence and structure space, there are a wide range of designed elements available. Both homomeric and heteromeric assemblies have been designed, as well as coiled coils with interaction affinities ranging from high μM to low nM.^{145,149,150,236} Thirdly, coiled coil design is accessible rationally, as well as computationally.^{139,219} This is important for the design of proteins for LLPS, as there are no existing computational frameworks for the design of such unstructured proteins, and computational design cannot yet produce the weak PPIs that would be required for LLPS. Instead, the rational mutation of coiled-coil sequences provides a tractable route to the design of proteins for LLPS. Indeed, coiled coils have recently begun to be implicated as oligomerisation motifs for phase separation.^{111,237–239}

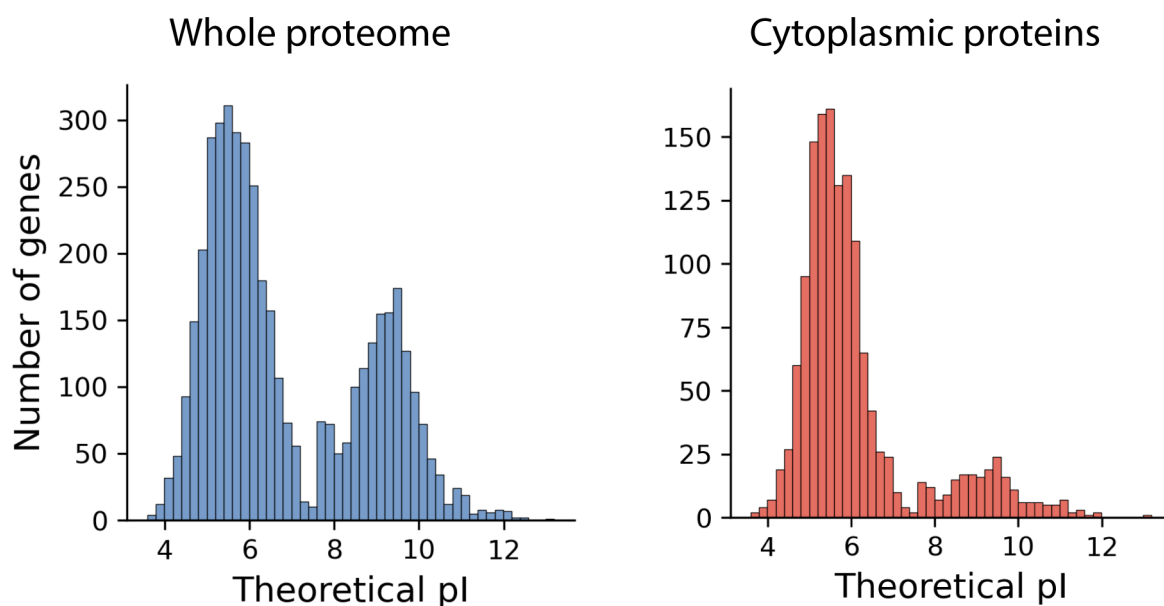


Figure 3.4: Theoretical pI distribution of the *E. coli* proteome. Histograms of the theoretical pI of the whole *E. coli* proteome (UniProt proteome:UP000000625; 4403 unique proteins), and only *E. coli* proteins flagged as being localised to the cytoplasm (GO:0005737) or cytosol (GO:000582; 1448 unique proteins). Theoretical pI was calculated using ExPASy Compute pI tool.²⁴⁰

To create as minimal a system as possible, only two coiled-coil sequences were used in the initial stickers and spacers arrangement. Longer sequences with more stickers would offer greater valency, but as the *de novo* coiled-coil peptides designed previously typically have dissociation constants (K_D) in the low nM range, the affinities of these stickers would likely be high enough to cause self-assembly with only two motifs.¹⁴² In addition, rather than using two identical coiled-coil motifs for the stickers, two different coiled coils were chosen in an effort to improve orthogonality between the two stickers. This would have the effect of reducing intramolecular interactions between the two stickers, and instead maximise intermolecular interactions.⁸³ Maximising intermolecular interactions permits

the greatest potential for intermolecular network formation, which has been suggested to be important for driving phase separation.⁸³ Furthermore, coiled coils with pentameric or greater oligomeric states were excluded from the design, as they have accessible hydrophobic cavities within the helical barrel that could cause hydrophobic collapse within the crowded phase-separated environment.^{151,235}

Name	Sequence	pI
	<i>gabcdef gabcdef gabcdef gabcdef</i>	
CC-Tri	G EIAAIKQ EIAAIKK EIAAIKW EIAAIKQ G	8.38
CC-Tri-EEWE	G EIAAIKE EIAAIKE EIAAIKW EIAAIKE G	4.60
	<i>cdefgab cdefgab cdefgab cdefgab</i>	
CC-Tet2	G EIQKQLK EIQKQLK EIQWQLK EIQKQLK G	9.60
CC-Tet2-EEWE	G EIQEQLK EIQEQLK EIQWQLK EIQEQLK G	4.60

Table 3.1: Theoretical pI of *de novo* coiled coils. Theoretical pIs of *de novo* coiled coils calculated from their primary amino acid sequence using ExPASy ProtParam.²⁴⁰

With pentamers and above excluded, the highest oligomeric states available are trimeric and tetrameric. Established and ongoing work has explored the creation of robust, hyperstable *de novo* coiled coils with dimeric, trimeric, and tetrameric oligomeric states (CC-Di, CC-Tri, and CC-Tet, respectively), that appear ideal for the creation of coiled-coil stickers (Fig. 3.3).¹⁴² However, following work by Dr Freddie Martin indicates that while these designed sequences are highly specific for their respective oligomeric states, the helices within some of these bundles can exchange with each other when mixed, indicating that they are not orthogonal to each other in all combinations.²⁴¹ This could potentially lead to increased intramolecular interactions, and a reduced intermolecular valency. Therefore, to maximise orthogonality between the two stickers, a more recently designed tetrameric coiled coil was used (CC-Tet2). Compared with CC-Tri and CC-Tet, CC-Tet2 uses a different arrangement of salt bridging residues as well as an alternative helical register, creating a highly robust homotetrameric coiled coil.¹⁴⁹ In addition, this redesigned CC-Tet2 was found to exchange far less with CC-Tri, indicating that the redesign had made the two coiled coils interfaces more orthogonal.²⁴¹ Therefore, the coiled coils CC-Tri and CC-Tet2 were selected for the stickers and spacers design.

Finally, the overall net charge and isoelectric point (pI) of the coiled coils were considered before their use in *E. coli* cells. Historically, to aid purification by high pressure liquid chromatography (HPLC), the peptides CC-Tri and CC-Tet2 had been designed with an overall positive charge, and pI of 8.5 – 9.5.^{142,149} While convenient for *in vitro* work, this is in contrast to the vast majority of proteins within the *E. coli* proteome.²⁴² Here, cytoplasmically distributed proteins are predominantly negatively charged, with an overall pI

of approximately 5.5 – 6.0, to prevent association with the highly negative nucleoid (Fig. 3.4).²⁴³ Only a small population of positively charged proteins exists, and is largely made up of DNA and RNA binding proteins.^{244,245} In order to prevent undesirable complex coacervation with nucleic acids, the net charge of the coiled coils was inverted by replacement of lysines at the solvent exposed f positions with glutamates. These acidic coiled coils, denoted CC-Tri-EEWE and CC-Tet2-EEWE, now had pIs of 4.6, much closer to that of endogenous *E. coli* proteins (Table 3.1).

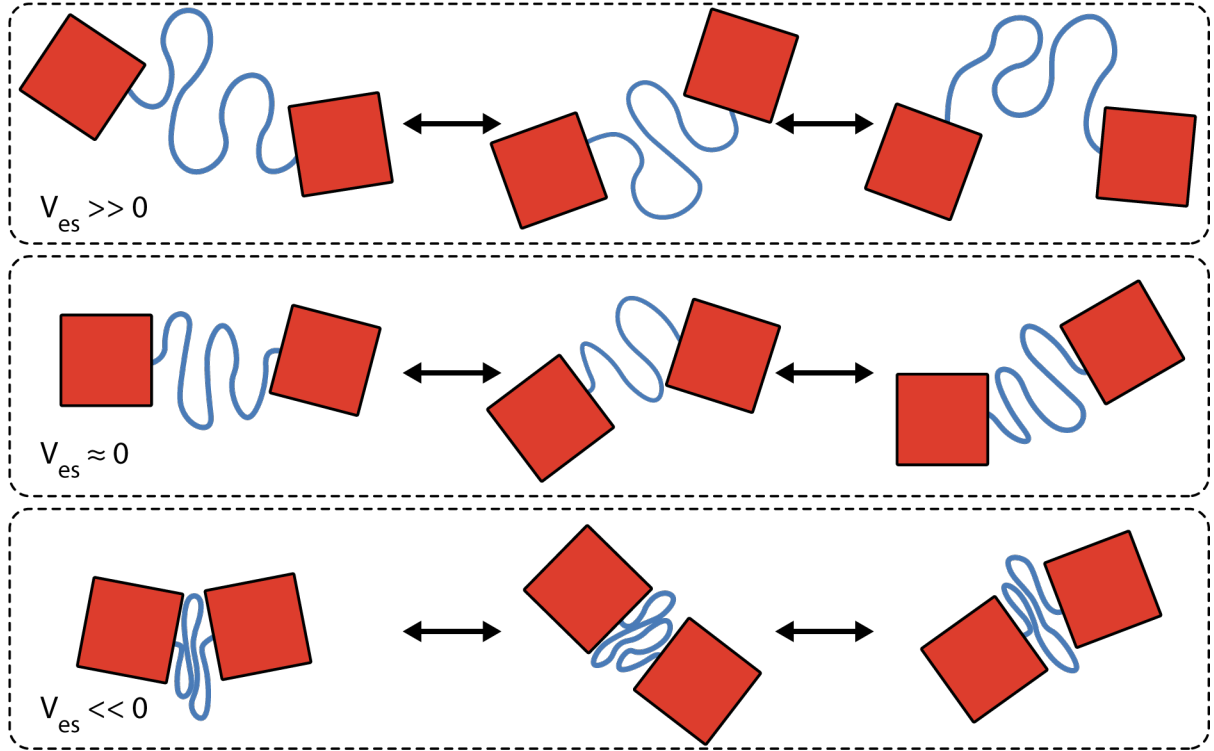


Figure 3.5: The effect of linker effective solvation on sticker cooperativity. Stickers are represented by red squares connected by blue spacers. Large positive v_{es} favours linker elongation and poor overall cooperativity between stickers. A large negative v_{es} instead favours linker compaction. v_{es} close to 0 gives a dynamic tether with both global and local cooperativity. Figure adapted from Harmon *et al.* (2017) with permission from eLife.²⁴⁶

3.2.2 Unstructured spacers

Next, a spacer was created to connect the two helical motifs. To create flexibility between the two stickers, the spacer was designed to be disordered. However, in order to rationalise as many of the interactions as possible using the *de novo* coiled coils, PPI motifs or attractive patches were excluded from the spacer design. Further, natural IDRs were also excluded from the design process, so that both the stickers and spacers were rationally designed. Several features of linkers between PPI motifs for phase separation have already been established. There are strong indications that the physical properties,

and in particular, the effective solvation volume (v_{es}), of spacers in a stickers and spacers model is critical to phase separation (Fig. 3.5).^{246,247} If the linker is efficiently solvated ($v_{es} > 0$), it will tend to elongate and stretch out, to maximise spacer-solvent interactions. By contrast, if the linker is poorly solvated ($v_{es} < 0$), it will contract to bury its surface and minimise spacer-solvent interactions.²⁴⁶ Highly solvated linkers are poor drivers of phase separation, as they reduce the overall cooperativity between the sticker motifs. Meanwhile, very poorly solvated linkers can have the effect of acting like additional PPI motifs due to their self-interactions. Instead, it has been suggested that the ideal linker scaffold should have a neutral effective solvation ($v_{es} \approx 0$), and act much like a random tether. v_{es} is largely determined by the fraction of charged residues in the sequence, with approximately 30% charged residues giving a v_{es} of 0.

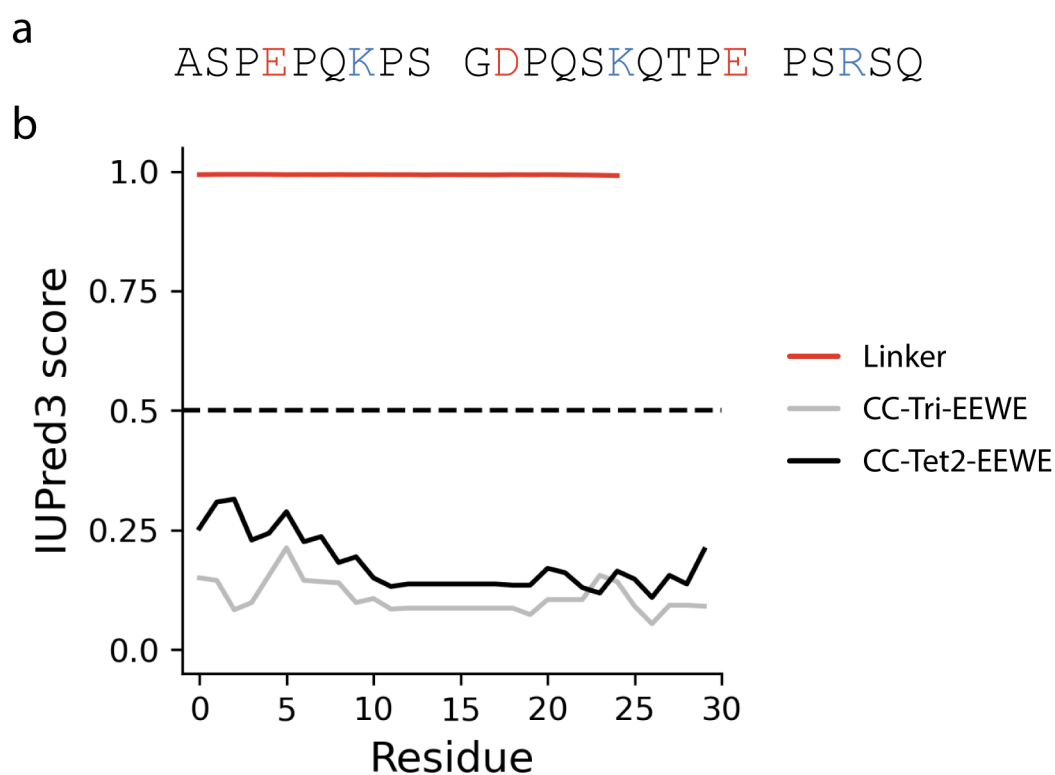


Figure 3.6: The designed linker, and disorder prediction of the coiled coils and linker by IUPred3. **a**, Primary amino acid sequence of the designed linker, with negatively charged residues in red, and positively charged residues in blue. **b**, Disorder prediction of the coiled coils CC-Tri-EEWE and CC-Tet2-EEWE, and the designed linker, by IUPred3. Probability of the sequences being disordered is given between 0 and 1. The threshold value for disorder (a score of 0.5 or greater) is shown by a hashed line. IUPred3 predictions used long disorder parameters without smoothing.

As well as effective solvation, the level of cooperativity between the stickers is also determined by the length of the spacer.²⁴⁶ Short spacers increase local cooperativity between the stickers, however, there is reduced global cooperativity, as stickers are less likely to make intermolecular interactions, and the network is less likely to grow and drive phase

separation. Increasing spacer length increases the probability that the molecular network will grow, as there is a greater global cooperativity. However, the stickers within the molecule are now locally less cooperative, and at infinitely long lengths are essentially independent. Therefore, linker length must balance global and local cooperativity.²⁴⁶

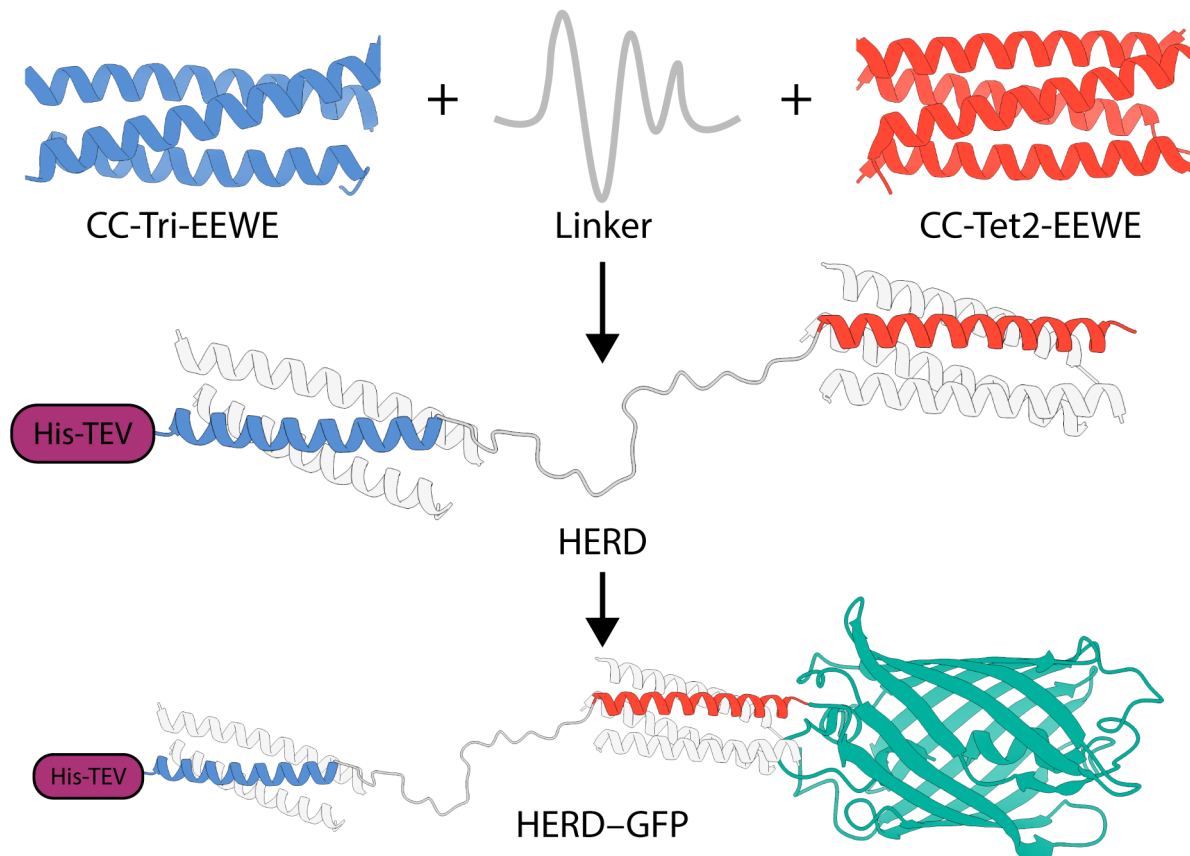


Figure 3.7: HERD protein architecture. Model of HERD-0 showing concatenation of the coiled coils CC-Tri-EEWE and CC-Tet2-EEWE with the designed linker, to create a linear polypeptide. The implied oligomeric states of the coiled coils are represented by white helices in the HERD protein.

These established concepts were considered when designing the *de novo* linker. Initially, residue composition was chosen to create a sequence with no secondary structure. Regular repeats of hydrophobic amino acids were excluded, and proline and glycine residues used to break secondary structure. Further, the sequence composition was chosen to be serine and proline rich to mimic natural IDRs, with an overall charged residue weight of 30% to give a v_{es} of ≈ 0 .^{248,249} The overall net-charge of the linker was designed to be 0, and charged residues spaced evenly across the sequence to prevent patchy ionic interactions. Hydrophobic and aromatic residues were largely excluded to avoid interactions similar to those in natural IDRs.⁹⁴ In addition, the linker was made to be 25 residues in length, to balance both global and local cooperativity. A single linker sequence was parameterised and tested using these constraints, to focus the design on the coiled-coil motifs used as stickers (Fig. 3.6a). Structural predictions of both the two coiled-coil motifs, and the

designed linker, were generated using IUPred3.²²¹ The designed linker was predicted to be entirely unstructured, with a 100% disorder prediction, while the two coiled coils are both predicted to form secondary structure (Fig. 3.6b).

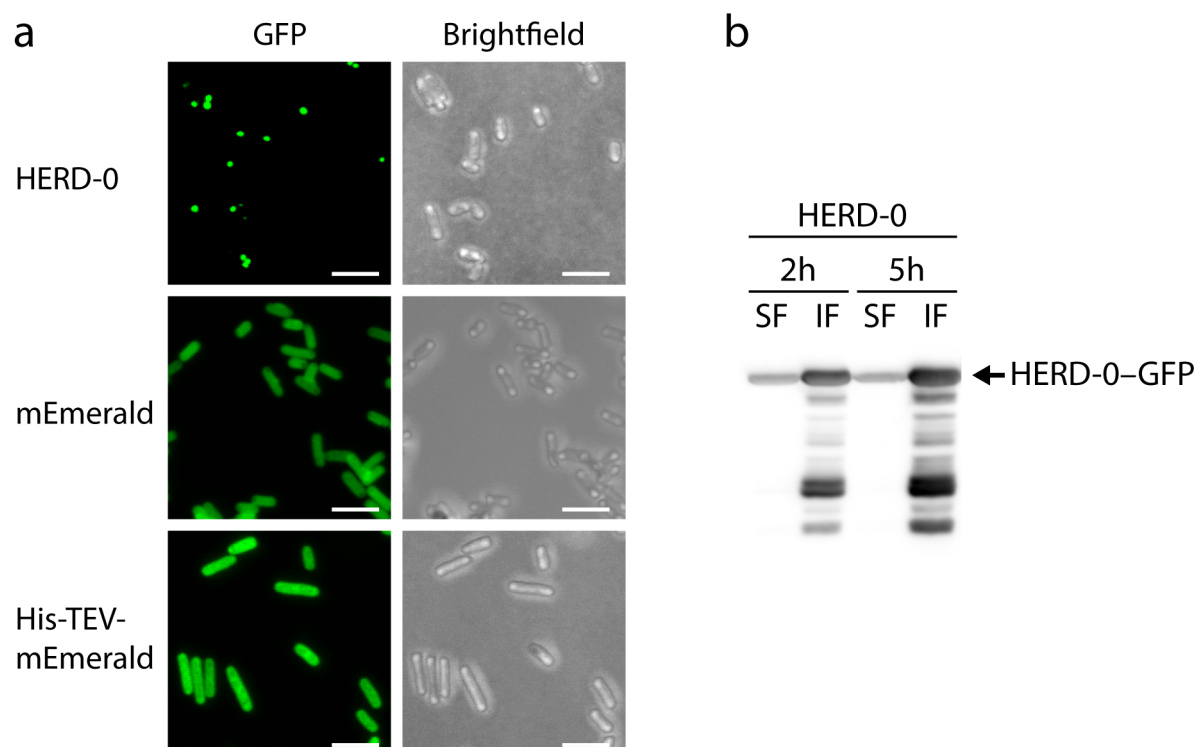


Figure 3.8: HERD-0-GFP forms fluorescent subcellular condensates. **a**, Fixed cell confocal microscopy images of *E. coli* grown at 18 °C expressing HERD-0-GFP, and the soluble proteins mEmerald and His-TEV-mEmerald. Scale bars are 5 μ m. **b**, Western blot of *E. coli* expressing HERD-0-GFP. Cell samples were collected 2 hours (2h) and 5 hours (5h) after induction of protein expression, and separated into the soluble (SF) and insoluble (IF) cellular fractions. Laddering visible below the band corresponding to HERD-0-GFP is due to protein degradation in the insoluble fraction.

3.3 HERD proteins form subcellular protein condensates

The chosen coiled coils, CC-Tri and CC-Tet2, and the designed linker were concatenated into a single polypeptide to create the final design framework, which was called a helical repeat domain (HERD) (Fig. 3.7). The monomeric green fluorescent protein (GFP) mEmerald was attached to the *C* terminus to facilitate imaging of subcellular localisation, and an *N*-terminal His-TEV site was added to enable protein purification, creating HERD-0-GFP (Tables 8.1 and 8.5).²⁵⁰ Expression of this initially designed construct, and examination of intracellular localisation by confocal fluorescence microscopy indicated the formation of fluorescent intracellular foci, while expression of mEmerald alone gave evenly distributed fluorescence, demonstrating that protein condensation was specific to HERD-0-GFP (Fig. 3.8a).

3.3.1 Destabilisation of HERD proteins creates soluble condensates

To assess whether the condensates were liquid-like or formed arrested aggregates, the cells were lysed and separated into their soluble and insoluble fractions, and localisation of HERD-0-GFP determined by western blotting. HERD-0-GFP localised almost exclusively to the insoluble fraction, indicating that it was unlikely to be forming dynamic protein condensates (Fig. 3.8b). The formation of insoluble aggregates was attributed to the extremely high affinities (low nM) of the *de novo* coiled-coil motifs.¹⁴²

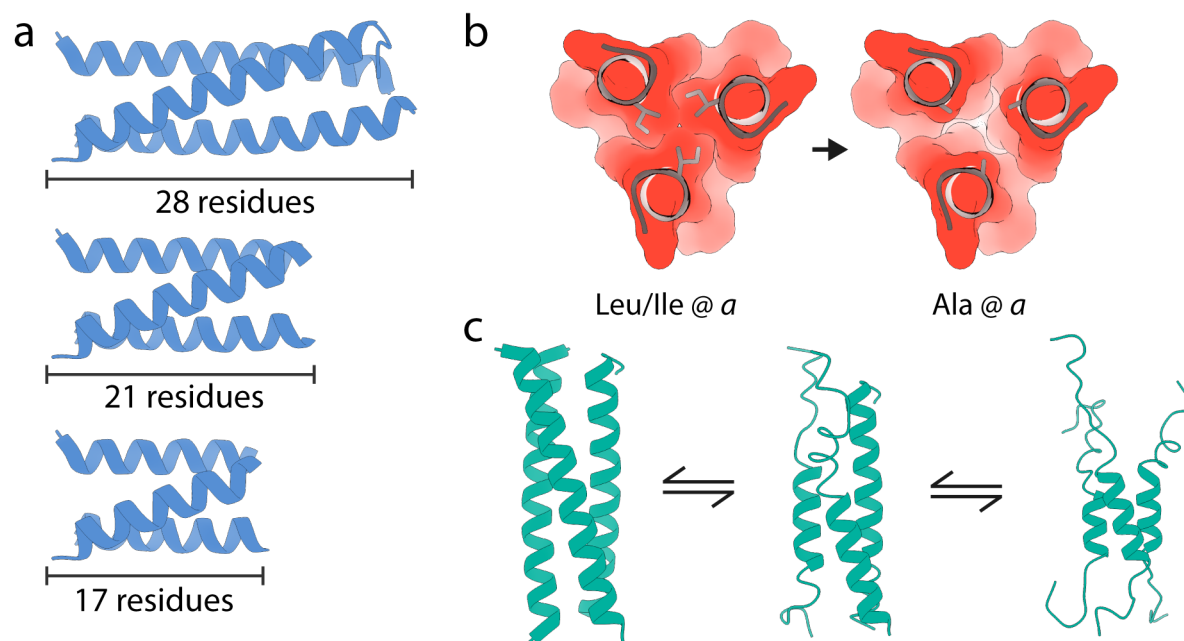


Figure 3.9: Destabilisation of *de novo* coiled coil stickers. Mutations were made to destabilise the helical regions by: **a**, reduction of helical length, **b**, mutation of core leucine and isoleucine residues to alanine, **c**, mutation of coiled coil surface residues to residues with lower helical propensities.

Once formed, these oligomers would be unlikely to dissociate again, and rather than forming dynamic condensates instead become irreversibly aggregated. To create HERD designs that exchanged dynamically, the affinities of the helical motifs were reduced by destabilisation of the coiled-coil interface. Initially, the helical regions were weakened by reduction of the helical length from the initial 4 heptads (28 residues) to 3 heptads (21 residues; Fig. 3.9a) to give HERD-1.1-GFP (Fig. 3.10, Table 8.1). However, when assessed in cells, this design behaved similarly to HERD-0-GFP, also producing insoluble aggregates in *E. coli*. Therefore, the further reduction of affinities between the helical regions was considered. To weaken the hydrophobic interface further, 3 different strategies were employed in parallel. Firstly, the helical length was shortened further from the initial 4 heptad (28 residue) length of HERD-0, down to a minimum of 2.5 heptads (17 residues). Secondly, the interface was disrupted by replacing some of the hydrophobic residues in

the core of the coiled coil with alanine, to reduce the intimate van der Waals packing within the interface (Fig. 3.9b).

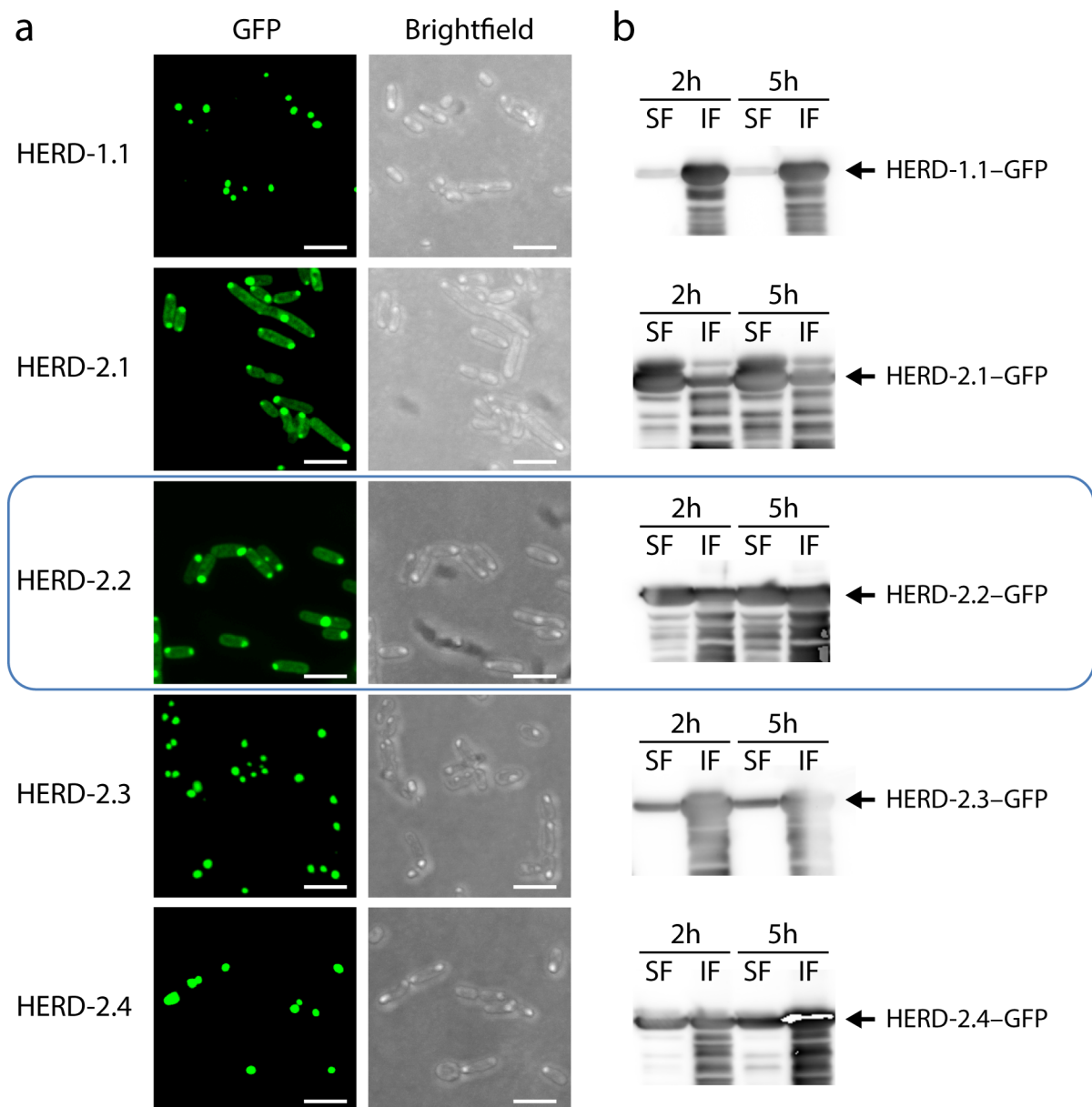


Figure 3.10: Destabilised HERD proteins create soluble condensates. **a**, Fixed cell confocal microscopy images of *E. coli* grown at 18 °C expressing HERD-1.1-GFP through HERD-2.4-GFP (Table 8.1). Scale bars are 5 μ m. **b**, Western blots of *E. coli* expressing HERD-1.1-GFP through HERD-2.4-GFP. Cell samples were collected 2 hours (2h) and 5 hours (5h) after induction of protein expression, and separated into the soluble (SF) and insoluble (IF) cellular fractions.

Here, rather than by mutating both the core *a* and *d* positions, only the *a* positions were mutated to retain some potential for helix-helix knobs-into-holes packing. Thirdly, alanine to serine mutations were made to the solvent-exposed surface of the helical regions to reduce their overall helical propensity, and improve their solvation (Fig. 3.9c, Table 8.1).^{251,252} These designs, HERD-2.1-GFP through HERD-2.4-GFP, were then screened

in cells for protein condensation and solubility. They retained the ability to form protein condensates, but had improved protein solubility (Fig. 3.10). Further, the complete destabilisation of the helical regions beneath the limit of coiled-coil formation by reducing the helical lengths to two heptads (14 residues) or less, resulted in the loss of condensate formation and largely evenly distributed fluorescence (HERD-2.5–GFP through HERD-2.8–GFP; Fig. 3.11).²⁵³

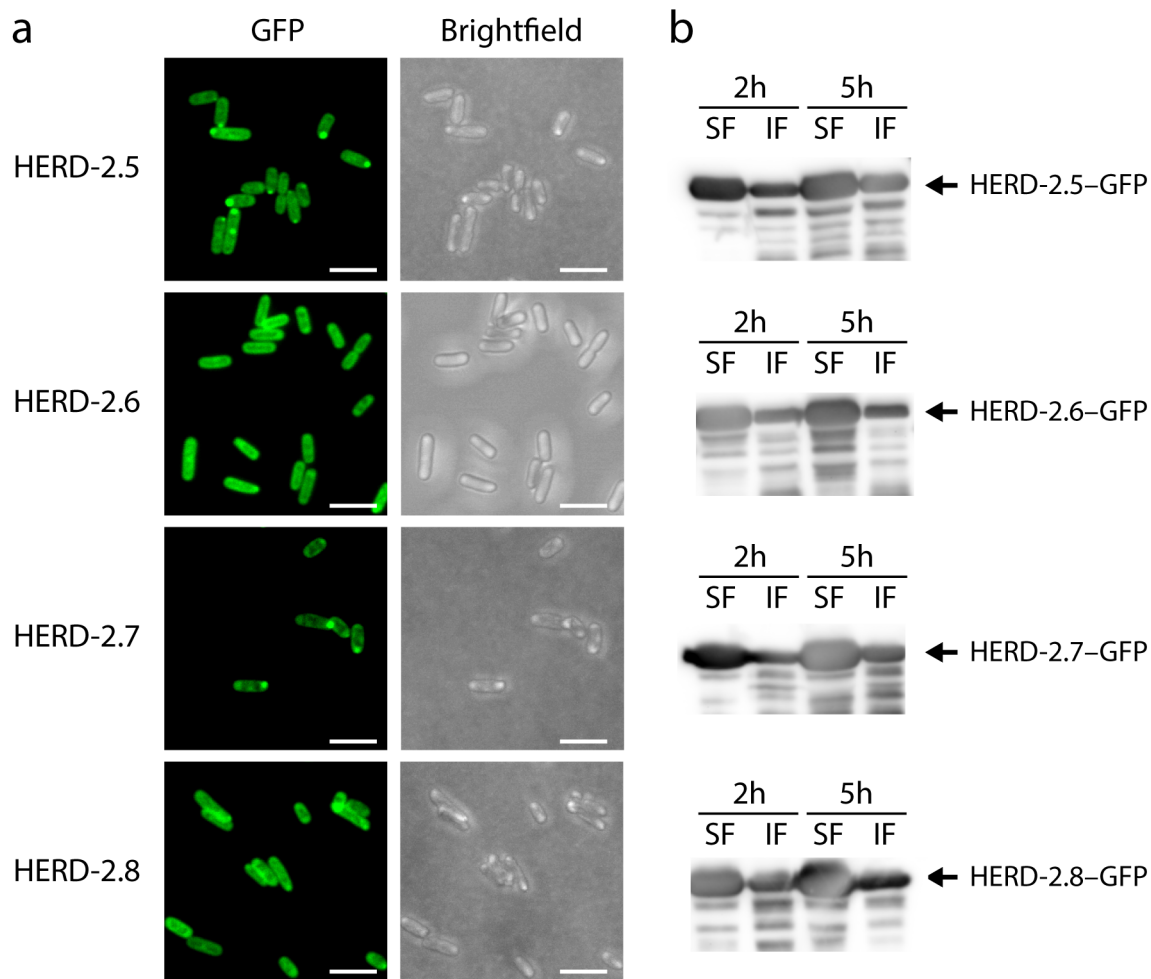


Figure 3.11: HERD proteins with severely reduced helical interactions produce fewer condensates. **a**, Fixed cell confocal microscopy images of *E. coli* grown at 18 °C expressing HERD-2.5–GFP through HERD-2.8–GFP. Scale bars are 5 μ m. **b**, Western blots of *E. coli* expressing HERD-2.5–GFP through HERD-2.8–GFP. Cell samples were collected 2 hours (2h) and 5 hours (5h) after induction of protein expression, and separated into the soluble (SF) and insoluble (IF) cellular fractions.

3.3.2 HERD-2.2–GFP undergoes a temperature-dependent phase transition

Automated image analysis was used to quantitatively determine changes in cellular localisation of the HERD proteins. The custom MIA plugin for ImageJ detects protein condensates automatically within *E. coli*, developed for this process by Dr Stephen Cross.^{207,208}

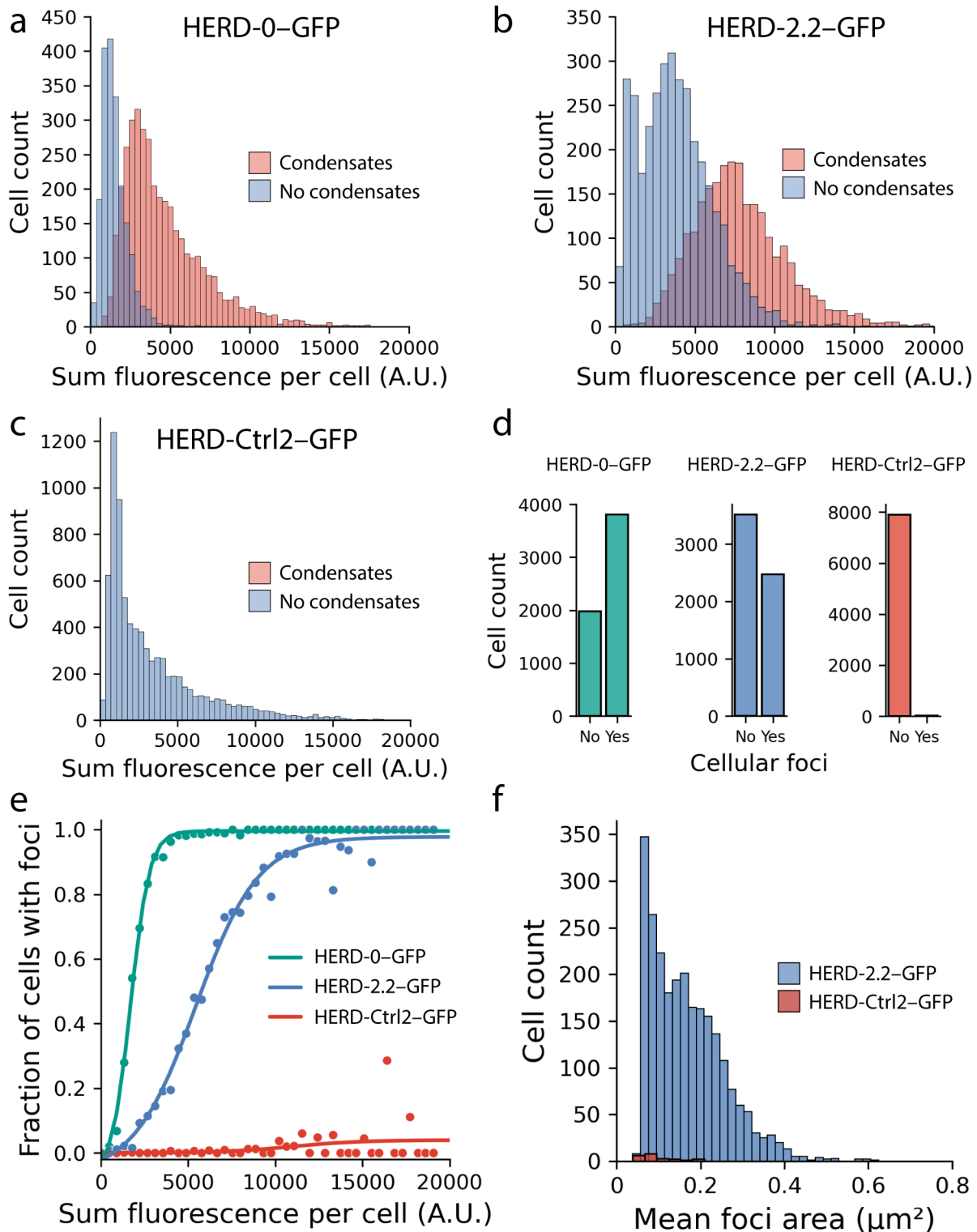


Figure 3.12: Automated image analysis of condensate formation in *E. coli*. **a – c**, Histograms of cells expressing HERD-0-GFP (**a**, $n = 5782$), HERD-2.2-GFP (**b**, $n = 5993$), and HERD-Ctrl2-GFP (**c**, $n = 7923$), with cells identified as containing foci by automated image analysis in red, and cells not containing condensates in blue, against total cellular fluorescence intensity. **d**, Total counts of cells with and without condensates, expressing HERD-0-GFP, HERD-2.2-GFP, and HERD-Ctrl2-GFP. **e**, Fraction of cells identified as displaying intracellular foci against total cellular fluorescence. **f**, Mean foci area (μm^2) of identified foci in cells expressing HERD-2.2-GFP and HERD-Ctrl2-GFP.

This identifies individual *E. coli* cells, and determines the overall fluorescence intensity and number of foci within the cell, as well as their size. To evaluate how protein localisation changes during expression, images of cells expressing HERD proteins were collected beginning at the induction of protein expression, and then every hour for the following 6 hours. Measurement of HERD-0-GFP indicated that foci formation occurred almost immediately, at low protein concentrations (Fig. 3.12a). In contrast, the destabilised HERD-2.2-GFP, with helical regions 17 residues in length, and alanine at the *a* positions and isoleucine at the *d* positions, formed condensates only after a threshold concentration had been reached, suggesting a phase transition (Fig. 3.12b). In addition, live-cell imaging of HERD-2.2-GFP indicated that condensation was temperature dependent. *E. coli* expressing HERD-2.2-GFP grown and imaged at 37 °C showed no condensates, while reducing the temperature to 33 °C resulted in the formation of enriched protein condensates, which became more dense as the temperature was reduced to 18 °C (Fig. 3.13).

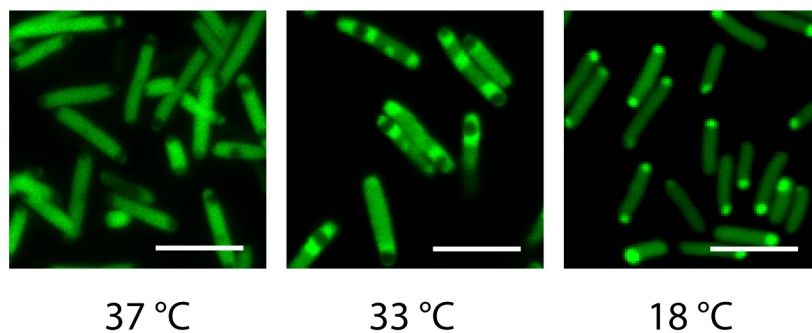


Figure 3.13: HERD-2.2-GFP condensation in cells is temperature dependent. Live cell confocal microscopy images of *E. coli* expressing HERD-2.2-GFP at varied growth and imaging temperatures (37 °C – 18 °C). Scale bars are 5 μ m.

In addition, to confirm that condensation was due to the helical regions, two HERD constructs were created where all of the core *a* and *d* residues had been mutated to alanine in one (HERD-Ctrl1-GFP) or both (HERD-Ctrl2-GFP) of the helical regions, to resemble the *de novo* designed monomeric α helix CC-Mono (Table 8.1).²⁵⁴ Neither of these proteins formed large intracellular condensates, indicating that the hydrophobic positions at *a* and *d* are essential for protein condensation. Moreover, this demonstrates that these positions are required in both of the helical regions, confirming the original design hypothesis (Fig. 3.14). Measurement of HERD-Ctrl2-GFP by automated image analysis confirmed that protein condensation had been abolished in this construct, and those foci that were identified by automated image analysis were far smaller than the mean foci area of HERD-2.2-GFP condensates (Fig. 3.12c-f).

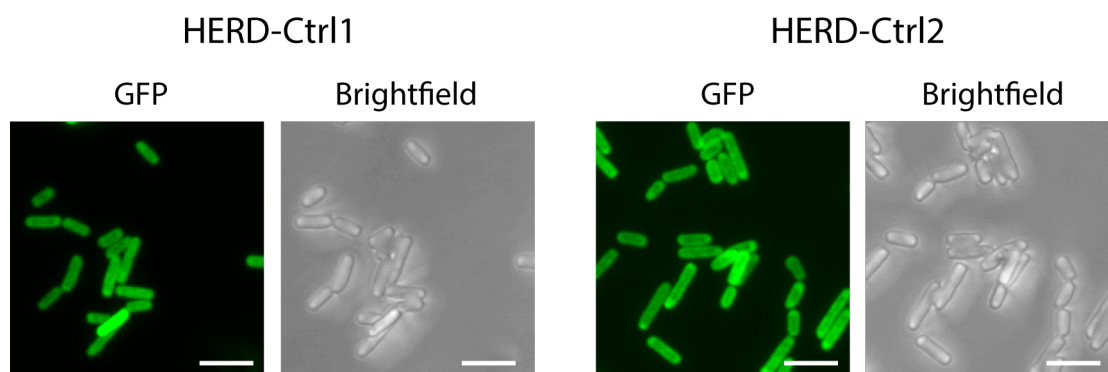


Figure 3.14: Fully mutated HERD constructs do not form condensates. Fixed cell confocal microscopy of *E. coli* grown at 18 °C expressing HERD-Ctrl1–GFP and HERD-Ctrl2–GFP. Scale bars are 5 μm .

3.3.3 The helical regions of HERD-2.2 are unstructured in solution

The helical regions used in the HERD-2.2 design had been significantly altered and destabilised compared to the parent CC-Tri and CC-Tet2 peptides. While these changes produced condensates that appear much more soluble, it was unclear whether they would still be forming coiled coils or even helical interactions. Dr Andrey Romanyuk investigated the solution-phase behaviour of the component peptides of HERD-2.2. For this, the HERD assembly was broken down into its constituent peptides (helical region 1: CC-Tri, linker, and helical region 2: CC-Tet2), and chemically synthesised *in vitro* by SPPS, followed by circular dichroism (CD) spectroscopy to examine their secondary structure in solution.²²⁴ The components of the parent HERD-0 showed the expected secondary structural characteristics by CD spectroscopy: the helical regions CC-Tri-EEWE and CC-Tet2-EEWE displayed strong α -helical profiles, with minima at 208 and 222 nm (Fig. 3.15a). Moreover, the designed linker was confirmed to be unstructured in solution, with overall low ellipticity between 190 and 260 nm.

However, measurement of the destabilised HERD-2.2 helical regions showed that the secondary structure had been nearly entirely lost from these sequences (Fig. 3.15b). Moreover, the synthesised single polypeptide comprising the two helical regions and linker also showed no secondary structure. These data were puzzling, and appeared contrary to the design hypothesis that helical interactions were driving LLPS of HERD-2.2–GFP. However, it was proposed that rather than highly ordered and well-folded coiled coils, the helical regions in HERD-2.2 could be interacting by weak, nascent helical interactions that are poorly folded in solution. Weak α -helical interactions have been implicated in phase separation of natural proteins such as elastin, where there is some evidence for protein folding assisted phase separation, however this remains an area of active research.^{255,256} To elucidate the potential for weak helical interactions, Dr Romanyuk investigated the secondary structure of the HERD-2.2 peptides in trifluoroethanol (TFE).

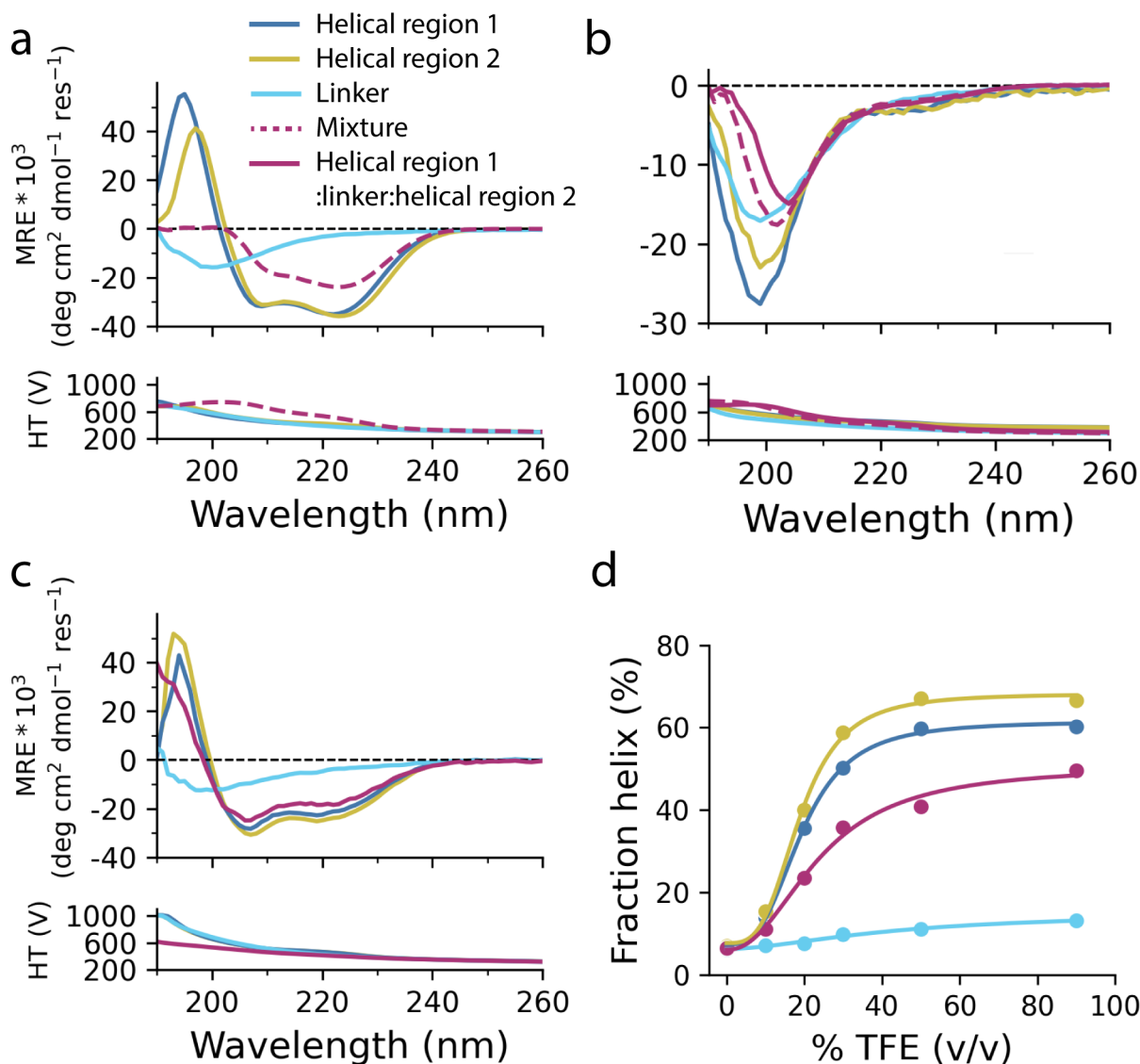


Figure 3.15: The helical regions of HERD-2.2 are unstructured in solution. a-b, CD spectra for the component peptides of HERD-0 (a) and HERD-2.2 (b) at 500 μM in 125 mM NaCl, 20 mM Tris pH 7.5. c, CD spectra for the component peptides of HERD-2.2 in 50% TFE, with 100 μM peptide. d, Fraction helicity of HERD-2.2 component peptides at increasing TFE concentration (0 – 90%), at 100 μM peptide.

TFE stabilises hydrogen bonding of peptides in solution, and therefore their secondary structure.²⁵⁷ Measurement of the HERD-2.2 components in TFE now showed a clear α -helical profile for the two helical regions, while the linker remained remarkably unstructured even up to 90% TFE (Fig. 3.15c–d). These data indicate that the helical regions of HERD-2.2 are unstructured in solution, but have a propensity to form α -helical secondary structures, which could potentially drive the weak interactions necessary for phase separation.

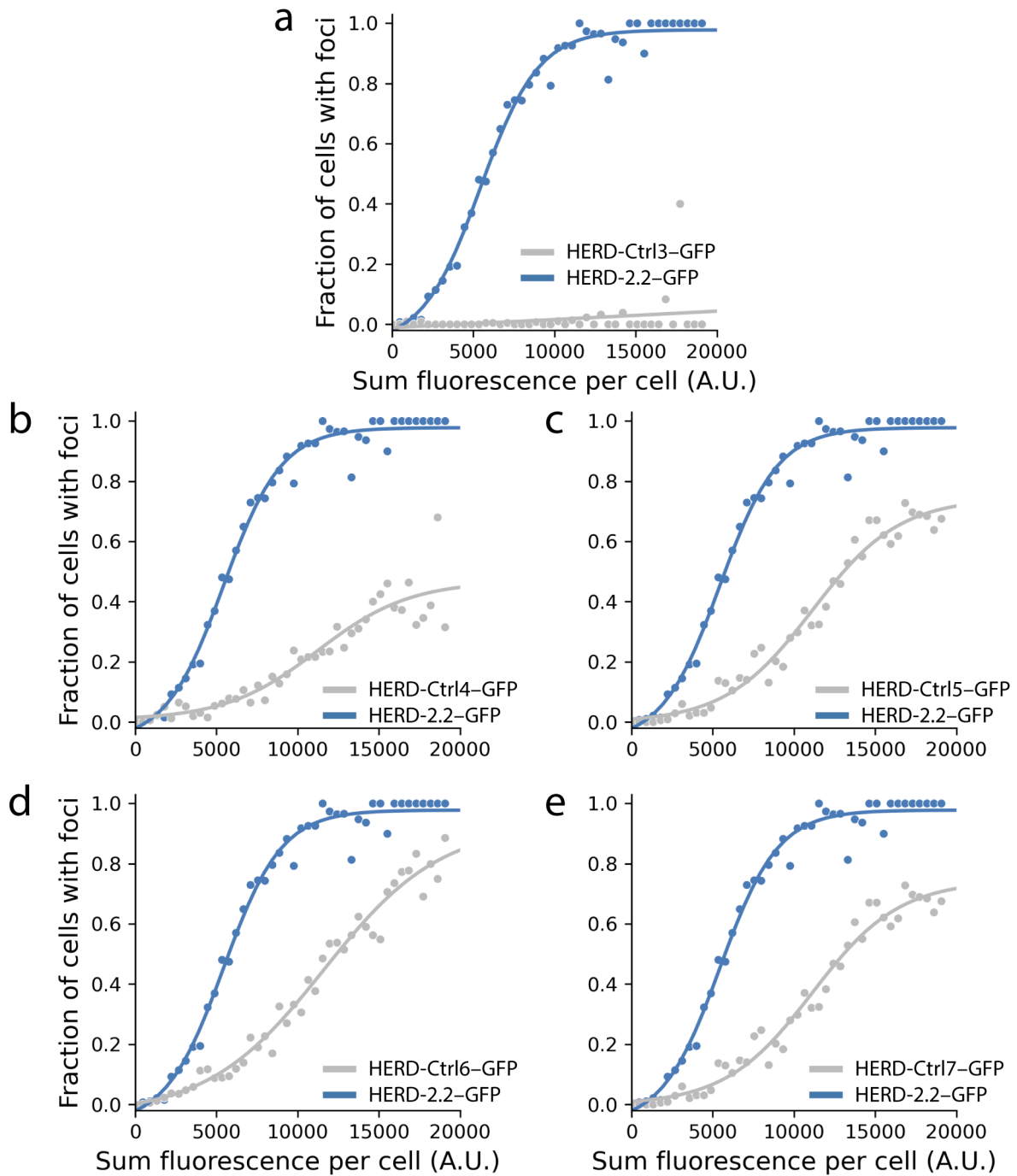


Figure 3.16: Fully mutated HERD constructs do not form condensates. Automated image detection of condensates within *E. coli* expressing HERD-Ctrl3-GFP (a, n = 4141), HERD-Ctrl4-GFP (b, n = 5530), HERD-Ctrl5-GFP (c, n = 5554), HERD-Ctrl6-GFP (d, n = 4339), and HERD-Ctrl7-GFP (e, n = 4113) against cells expressing HERD-2.2-GFP (blue, all panels, n = 5993).

3.3.4 Helical interactions of HERD-2.2 are important for condensation in cells

In vitro analysis on the component peptides implicated extremely weak, transient α -helical structure in the helical regions of HERD-2.2, but this was yet to be determined to be a factor on condensation in cells. To determine whether α -helical secondary structure could be implicated in protein condensation, several additional variations on HERD-2.2 were constructed where the helical regions were disrupted such that they would no longer be expected to form helices (HERD-Ctrl4-GFP through HERD-Ctrl7-GFP, Table 8.1). These constructs included helix breaking mutations (proline and glycine residues) at various positions in both helical regions, as well as an overall scramble of the hydrophobic residues in the heptad repeat, so that the helical regions would no longer be able to form an amphipathic helix.

Each of these designs was assessed for condensation by automated image analysis and compared to the HERD-2.2-GFP design (Fig. 3.16). All of the disrupted helical designs showed significantly reduced or nearly completely abolished protein condensation in *E. coli*, corroborating the idea that helical interactions were important for protein condensation. However, it should be noted that this does not conclusively prove α -helical interactions are present, and studies similar to those performed on elastin, such as following phase separation by nuclear magnetic resonance (NMR) would be required to investigate this further.²⁵⁸

3.3.5 AlphaFold2 predicts α -helical secondary structure for HERD-2.2

Since helical interactions could be implicated in HERD-2.2 interactions, the sequence was modelled using AlphaFold2 to see whether it would predict the now destabilised helical regions as α -helical.^{118,223} Modelling of the entire HERD-2.2-GFP sequence predicted the designed helical regions with α -helical secondary structure, and correctly predicted the well-folded β -barrel GFP (Fig. 3.17). Further, the linker was predicted to be unstructured, as was the *N*-terminal His-TEV tag. However, despite predicting the helical regions as α -helical, the confidence of the prediction made to the HERD sequence was low, demonstrated by a low predicted local distance difference test (pLDDT) score. Therefore, structural predictions of these helical regions should not be considered accurate. The low confidence may be due to the low sequence coverage for this *de novo* protein, and helical representations of these sequences could be due to bias from the published *de novo* α -helical coiled coils in the PDB, on which AlphaFold2 was trained. Overall though, modelling of HERD-2.2 using AlphaFold2 was not informative for predicting its secondary structure.

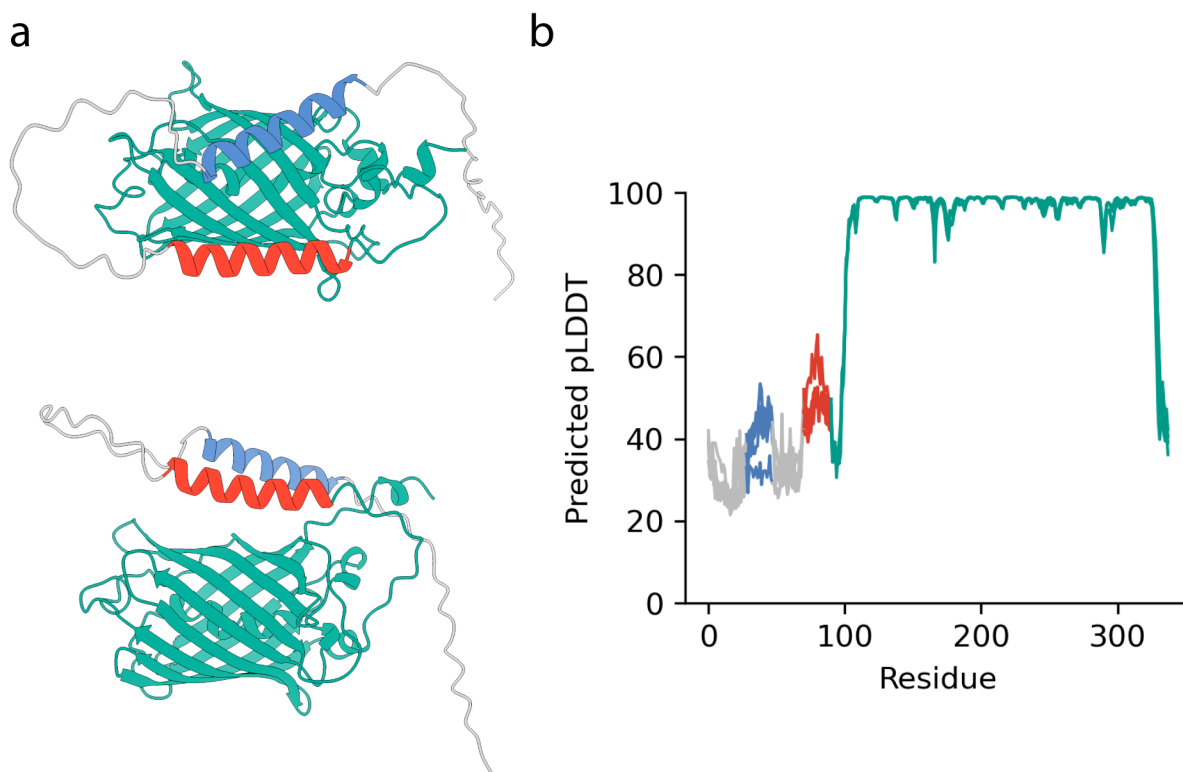


Figure 3.17: AlphaFold2 predicts α -helical secondary structure for HERD-2.2–GFP. **a**, Representative AlphaFold2 models of HERD-2.2–GFP, with the helical regions coloured in red and blue, and mEmerald in green. **b**, pLDDT plots for the top 5 AlphaFold2 models of HERD-2.2–GFP. Sections are coloured to correspond to the models shown in **a**.

3.3.6 Changes to the linker do not significantly alter condensation

Having interrogated several changes to the helical motifs that appear critical for protein condensation, several linker variations were also tested (Table 8.1). These experiments aimed to determine how important the length and sequence of the linker were to condensation of the *de novo* protein. Theoretical models of stickers and spacers suggest that protein condensation is highly dependent on changes to both the linker length and its effective solvation.²⁴⁷ Therefore, linkers were constructed to vary the length, while keeping the overall primary sequence composition the same. Or alternatively, to keep the length the same but vary v_{es} by inclusion or removal of additional charged residues. For this, linkers 8 residues and 50 residues long were designed to match the overall sequence composition of the original linker while varying the length, and two additional 25-residue linkers were designed with either a greater (50%) proportion of charged residues, or no charged residues to vary v_{es} .

These changes were made to the linker using the helical regions from HERD-2.2, which appeared to form condensates readily that were sensitive to their environment, making HERD-3.1–GFP through HERD-3.4–GFP. When screened in *E. coli*, all of these designs formed condensates that matched those made by HERD-2.2–GFP, indicating that con-

densation in cells had not been significantly perturbed by the changes to the linker (Fig. 3.18). However, it is highly likely that these changes could have impacts on the phase behaviour of the *de novo* protein that may not be revealed by in cell screening. Instead, a full assessment could be elucidated with more sensitive assays or *in vitro* measurements. However, as the design process focused on changes to the helical regions of the *de novo* protein, further investigation into the impact of the linker was not pursued.

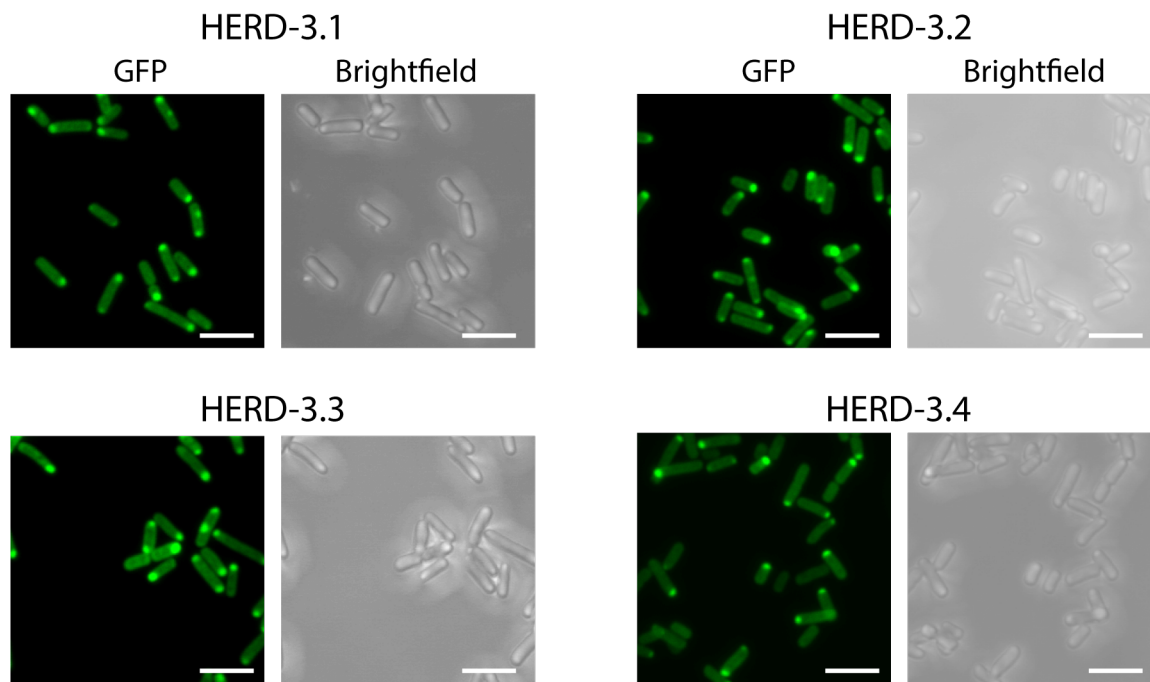


Figure 3.18: Changes to the linker do not significantly alter condensation. Fixed cell confocal microscopy of *E. coli* grown at 18 °C expressing HERD-3.1–GFP through HERD-3.4–GFP. Scale bars are 5 μm .

3.4 HERD-2.2–GFP undergoes LLPS *in vitro*

Following the observation that HERD-2.2–GFP underwent a phase transition in cells, the *de novo* protein was purified in order to evaluate its phase behaviour conclusively *in vitro*. Initially, phase separation of HERD-2.2–GFP was screened using buffers of varying ionic strength and molecular crowding. Varying concentrations of NaCl were used to modulate ionic strength, while PEG was used to simulate the crowded environment of the bacterial cytoplasm.^{259,260} The pH was maintained at 7.5 to emulate physiological pH. HERD-2.2–GFP did not phase separate at low NaCl and PEG concentrations, but formed a single soluble phase (Fig. 3.19). However, at increasing molecular crowding and ionic strength the formation of two de-mixed phases was evident. Promisingly, in several cases, this took the form of spherical macroscopic droplets, characteristic of liquid-like de-mixing. Further increases in molecular crowding resulted in the formation of amorphous

aggregates, indicative of a second phase transition. Optimal conditions within the regime where liquid-like droplets were observed were 250 mM NaCl, 4% PEG 3350 in Tris (pH 7.5).

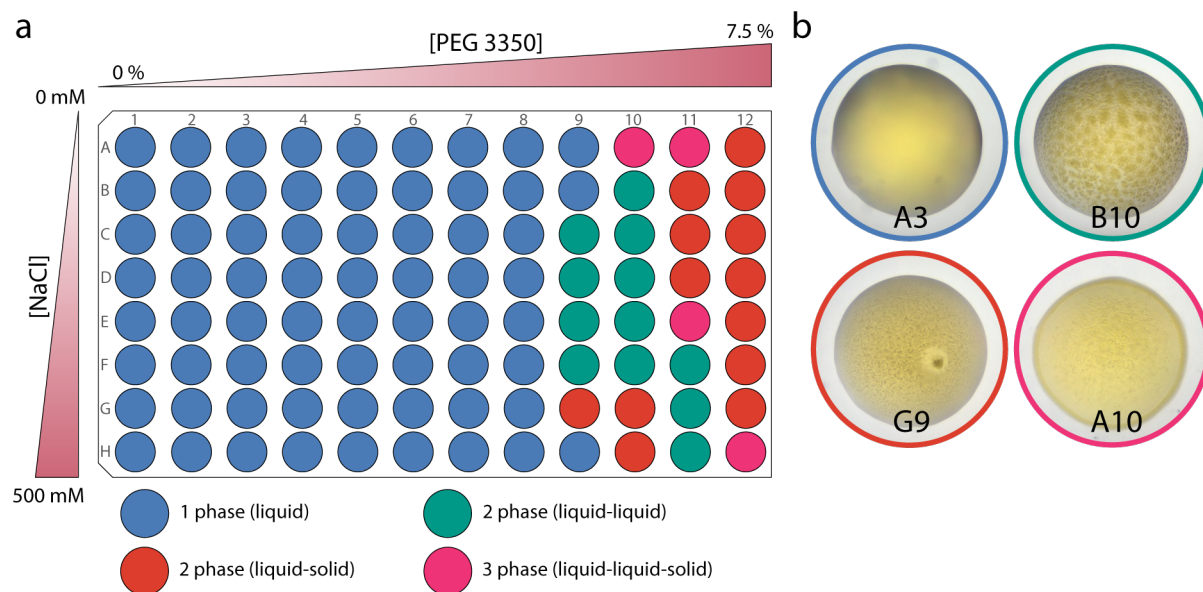


Figure 3.19: HERD-2.2-GFP forms de-mixed droplets *in vitro*. **a**, A screen for HERD-2.2-GFP phase separation was set up using a 2 dimensional gradient of NaCl (0 – 500 mM) and PEG 3350 (0 – 7.5%) in a 96 well MRC crystallography plate with HERD-2.2-GFP (400 μ M) in a total volume of 0.8 μ l, and 20 mM Tris pH 7.5. The sitting droplets were automatically imaged in a Formulatrix RI-1000 crystallisation hotel with a 5 MP camera and categorised based on appearance. **b**, Representative images of the phase behaviours described in **a**. Droplets were classified into 4 groups based on the several criteria: A single (1) phase had no observable precipitation or phase separation. 2 phases (liquid-liquid) formed largely spherical de-mixed particles with smooth borders between phases. 2 phases (liquid-solid) formed non-spherical particles with coarse (non-smooth) boundaries between phases. 3 phases (liquid-liquid-solid) had both morphologies within the same droplet.

3.4.1 HERD-2.2-GFP droplets are reversible and dynamic

Observation of HERD-2.2-GFP under these conditions (125 mM NaCl, 4% PEG 3350, 20 mM Tris pH 7.5) by confocal microscopy confirmed the presence of macroscopic de-mixed droplets, clearly enriched in the *de novo* fluorescently tagged protein (Fig. 3.20a). The droplets were highly spherical, and rapidly coalesced in under 2 seconds to form larger droplets that reformed a spherical shape (Fig. 3.20b). Moreover, variable-temperature microscopy measurements captured droplet nucleation as the temperature was reduced from 40 $^{\circ}$ C to 5 $^{\circ}$ C (Fig. 3.20c). This phase transition recapitulates the observed temperature dependence of condensation observed in cells, with a similar temperature range of nucleation, indicating that the chosen conditions mimicked those in cells. In addition, reheating of the de-mixed droplets *in vitro* fully reversed de-mixing, reforming a sin-

gle phase, demonstrating that phase separation of HERD-2.2–GFP was highly reversible (Fig. 3.20d). In turn, each of these measurements indicate that the designed HERD-2.2–GFP undergoes LLPS *in vitro* under these conditions, forming dynamic and reversible de-mixed droplets.

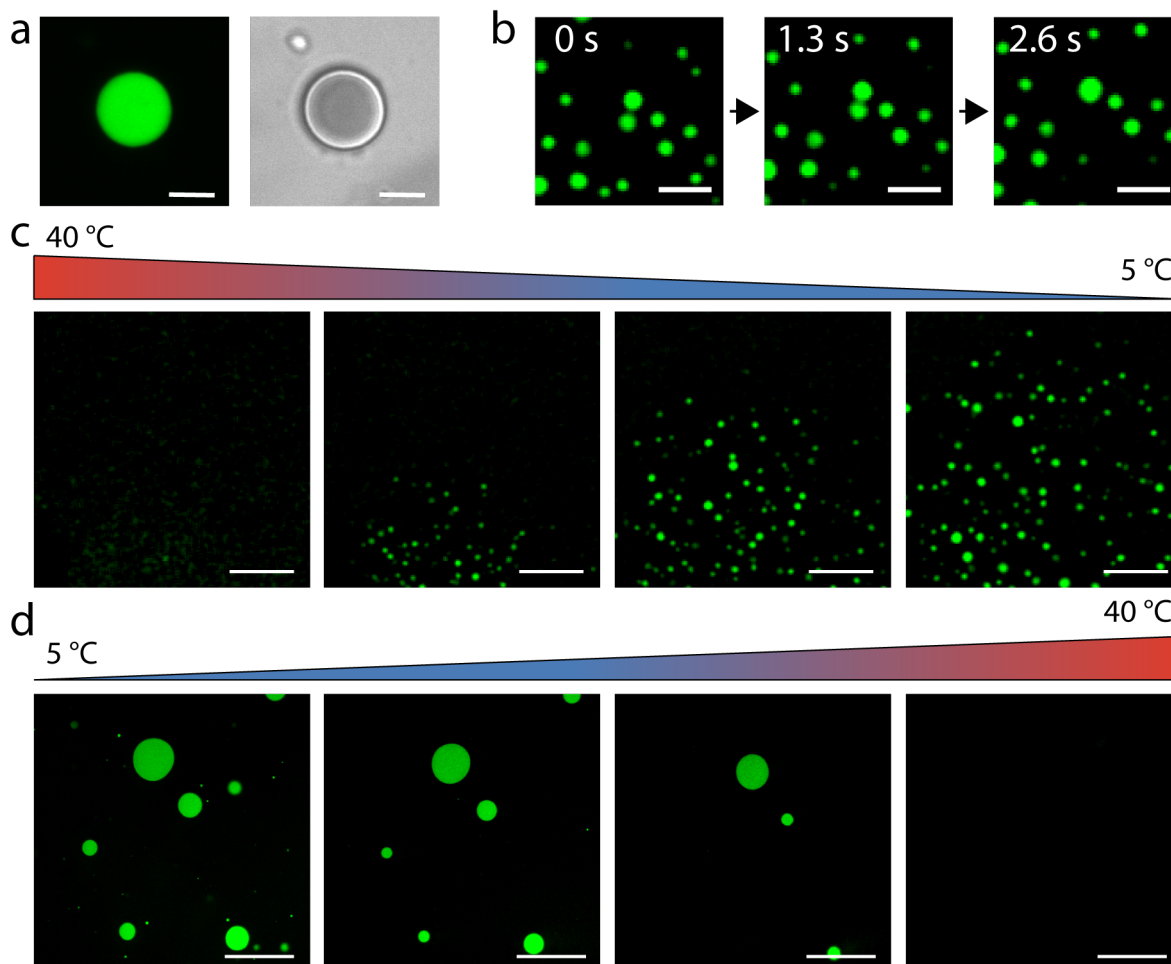


Figure 3.20: HERD-2.2–GFP droplets are reversible and dynamic. **a**, Confocal microscopy of de-mixed HERD-2.2–GFP droplets. Scale bar is 5 μm . **b**, Coalescence of HERD-2.2–GFP droplets. Coalescence was complete after 2 – 3 seconds, imaging every 648 ms. Scale bar is 5 μm . **c**, Nucleation of HERD-2.2–GFP droplets as the temperature is reduced from 40 $^{\circ}\text{C}$ to 5 $^{\circ}\text{C}$. Frames are 6.5 s apart. Scale bar is 20 μm . **d**, Dissolution of HERD-2.2–GFP droplets as the temperature is increased from 40 $^{\circ}\text{C}$ to 5 $^{\circ}\text{C}$. Frames are 64 s apart. Common conditions: 1 mM HERD-2.2–GFP, 125 mM NaCl, 4% PEG 3350, 20 mM Tris pH 7.5.

3.4.2 HERD-2.2–GFP still undergoes LLPS following TEV cleavage

To ensure that all of the interactions making up HERD-2.2–GFP were robustly interrogated, the contribution of the *N*-terminal His-TEV tag on phase separation was assessed. Initially, the tag had been retained for the *in vitro* measurements to more closely match the protein measured in cells. To assess its contribution, the tag was cleaved using the

TEV protease site. Following cleavage, the tag-free HERD-2.2–GFP still underwent phase separation, forming enriched droplets similar to the protein with the *N*-terminal tag, but required greater molecular crowding reagent, with 10% PEG 3350 used instead of 4% (Fig. 3.21). This increase in molecular crowding is noteworthy, and indicates that the tag does contribute to the attractive interactions of the protein, though it is not the main or a required driving force for LLPS. The HERD proteins described from this point retain this tag, as this combination of interactions gave the optimum conditions for LLPS in cells.

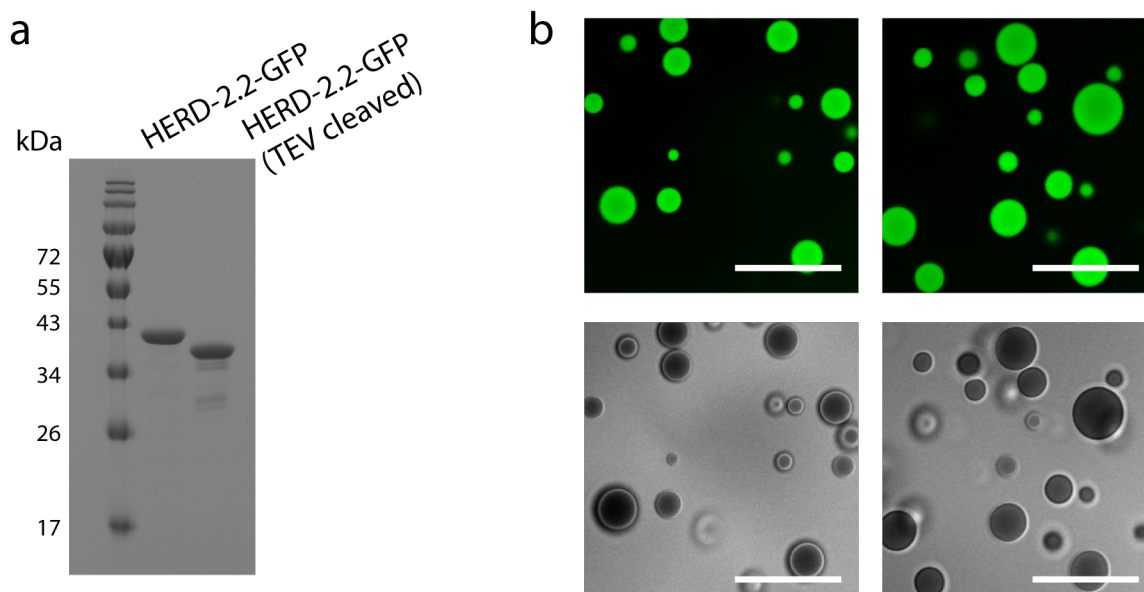


Figure 3.21: HERD-2.2–GFP still undergoes LLPS following TEV cleavage. **a**, SDS-PAGE of HERD-2.2–GFP and HERD-2.2–GFP following TEV cleavage and purification, confirming removal of the tag. **b**, Confocal microscopy of HERD-2.2–GFP droplets following TEV cleavage. Scale bar is 15 μ m. Conditions: 2 mM HERD-2.2–GFP (cleaved), 125 mM NaCl, 10% PEG 3350, 20 mM Tris pH 7.5.

3.4.3 The LLPS boundary of HERD-2.2–GFP can be mapped

To characterise the phase transition of HERD-2.2–GFP robustly, the binodal LLPS phase boundary was mapped by cloud-point measurements. These measurements use the light scattering effect produced by protein de-mixing to measure nucleation.^{261,262} As temperature had been identified to trigger LLPS of HERD-2.2–GFP, it was used to induce nucleation. The change in cloud-point temperature was measured as a function of HERD-2.2–GFP concentration. Beginning with a single phase solution heated to 40 $^{\circ}$ C in the identified 4% PEG 3350 crowding conditions, the temperature was lowered and nucleation measured by the change in percent transmission (%T; Fig. 3.22a). As the temperature was decreased there was a sharp decrease in %T, characteristic of nucleation-driven LLPS, with the temperature at 50% transmission giving the cloud-point temperature (T_{cloud}).

Further, on reheating the sample back to 40 °C the %T nearly completely recovered, returning to a largely single phase system. Here, the temperature at 50% transmission gives the solution clearing temperature (T_{clear}). This reversibility is again characteristic of LLPS, which is an entirely reversible phase transition, unlike irreversible aggregation or spinodal decomposition.⁴⁴

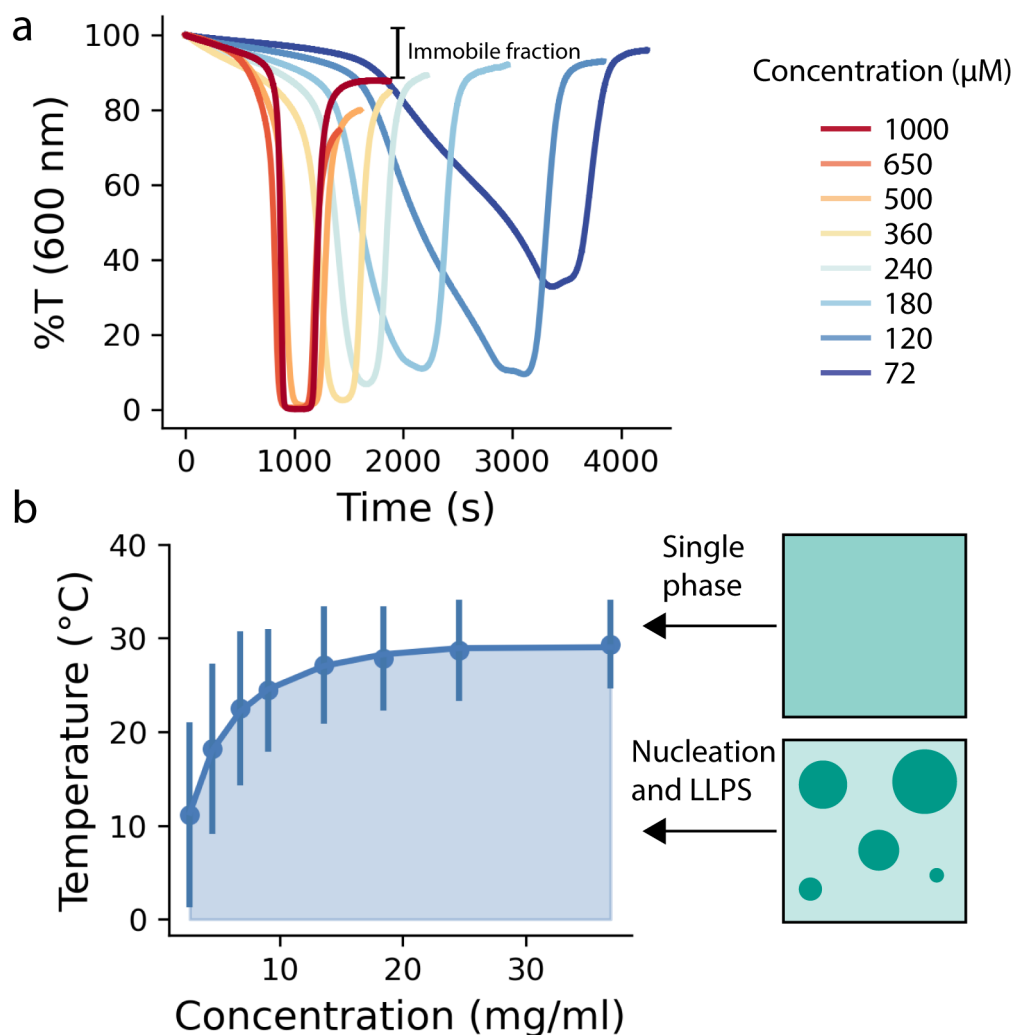


Figure 3.22: Measurement of the binodal phase boundary for HERD-2.2-GFP. **a**, Measurements of the change in transmission (600 nm) as the temperature was reduced from 40 °C to 5 °C, and then returned to 40 °C. **b**, The measured portion of the binodal phase boundary of HERD-2.2-GFP. Bars represent the measured difference between T_{cloud} and T_{clear} , with the midpoint being the phase boundary.

Several other aspects of these data confirm that the phase transition observed was nucleation driven LLPS. Firstly, at high protein concentrations the changes in %T were rapid, giving extremely sharp phase transitions. However, as the protein concentration was decreased the transitions became protracted, as droplet nucleation and coalescence was slowed by the reduced protein concentration in solution. Further, the measured transitions T_{cloud} and T_{clear} displayed the characteristic hysteresis associated with nucleation

(Fig. 3.22b).^{263,264} As there is an energy barrier associated with nucleation and LLPS, the temperature T_{cloud} is always lower than T_{clear} , except at the critical point.^{30,33} For HERD-2.2–GFP, as expected for the binodal phase boundary, as the protein concentration was increased the difference between T_{cloud} and T_{clear} was reduced, indicating that it was moving towards the critical point. However, as there was still a large difference between T_{cloud} and T_{clear} at the highest protein concentration measured (12.5 °C at 37 mg/ml), the critical point likely lies at a much higher protein concentration.

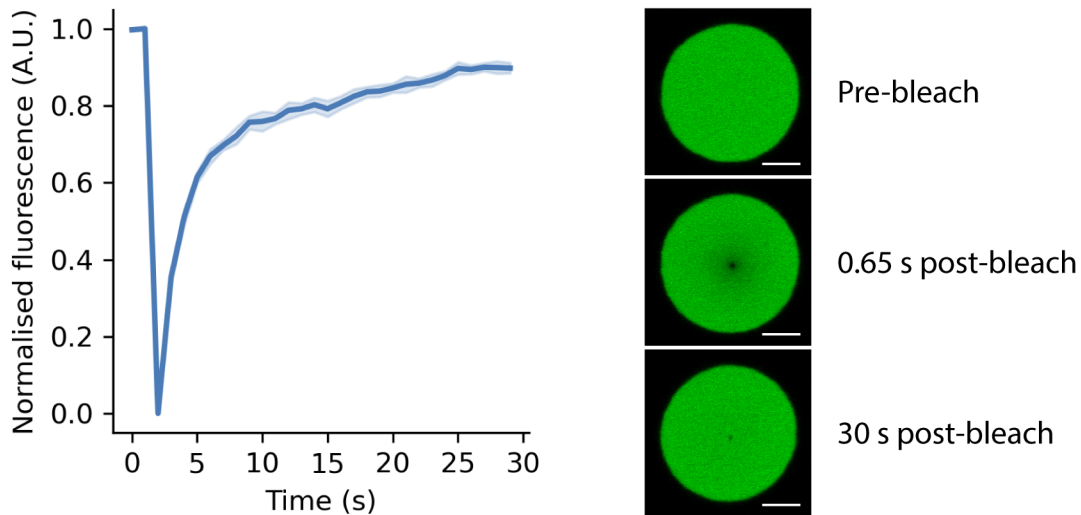


Figure 3.23: HERD-2.2–GFP droplets recover rapidly after photobleaching. Measured fluorescence recovery and representative images of the droplets before bleaching, immediately after bleaching, and 30 s after bleaching. $n = 13$. Scale bar is 5 μm . Conditions: 1 mM HERD-2.2–GFP, 125 mM NaCl, 4% PEG 3350, 20 mM Tris pH 7.5.

It is noteworthy that at the highest protein concentrations the reversibility of phase separation was incomplete, with %T only returning to approximately 90% of the starting value. This incomplete reversal of phase separation is likely due to a second, irreversible phase transition occurring after the initial nucleation-limited LLPS.²⁶⁵ LLPS of proteins is by definition a metastable process. This is because LLPS occurs due to weak, relatively non-directional interactions. However, proteins are generally anisotropic molecules, with patchy or directional interactions. While these interactions can help to drive LLPS, within the dense environment of a phase-separated droplets molecules can come into close proximity. Now, the stronger structural interactions can predominate and induce a second irreversible phase transition, either through spinodal decomposition to form a gel, or aggregation, or even crystallisation. Therefore, it is likely that the incomplete recovery of %T is due to a small amount of protein aggregation or gelation occurring as a result of the initial LLPS.

3.5 HERD-2.2–GFP droplets recover rapidly after bleaching

Having confirmed that HERD-2.2–GFP undergoes a liquid-liquid phase transition *in vitro*, the molecular motility within the de-mixed droplets was probed by fluorescence recovery after photobleaching (FRAP).^{266,267} In this experiment, a small region of fluorescence within the de-mixed droplet is bleached by a high-intensity laser burst, and the recovery of fluorescence within the bleached area measured with respect to time.²⁶⁷ This provides information on whether the macromolecules within the droplets are immobile or whether they can move freely.²⁶⁸

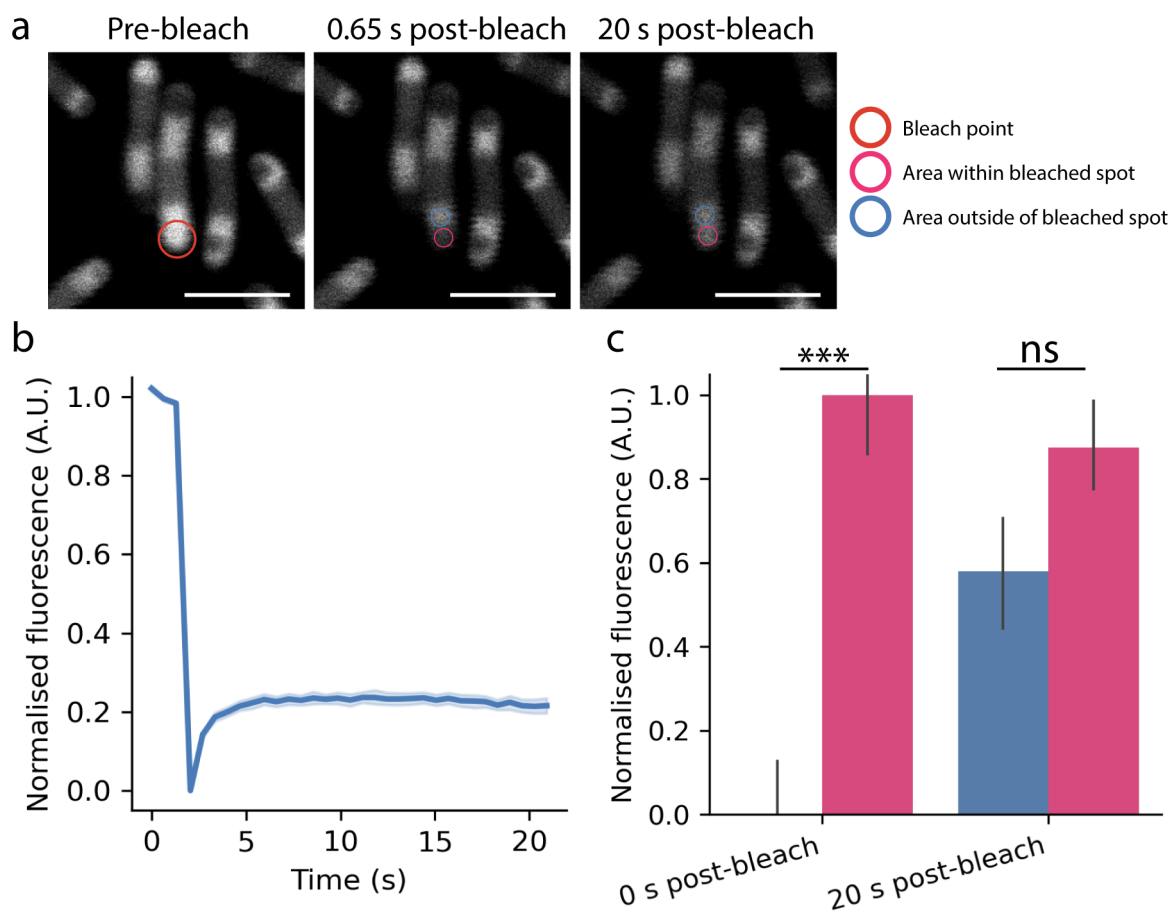


Figure 3.24: HERD-2.2–GFP forms liquid-like droplets in cells. **a**, Representative images of *E. coli* expressing HERD-2.2–GFP before bleaching, immediately after bleaching, and 20 s after bleaching. The bleach point is circled in red, and the two measured areas within the same condensate (within the bleached area, and outside of the bleached area) are circled in blue and magenta respectively. Scale bar is 5 μm . **b**, Measured fluorescence recovery within the bleached area of HERD-2.2–GFP condensates in *E. coli* grown at 37 $^{\circ}\text{C}$. $n = 13$. **c**, Relative normalised fluorescence intensity between the measured areas directly within the bleached spot (blue) and outside of the bleached spot within the same condensate (magenta). $n = 13$, error bars represent the standard error. Statistical testing used a two-tailed T-test with P-values of < 0.001 for 0 s post-bleach, and 0.0958 for 20 s post-bleach.

3.5.1 FRAP *in vitro* gives rapid and complete recovery of fluorescence

The rate of recovery within HERD-2.2–GFP droplets *in vitro* was assessed. To facilitate bleaching, relatively high protein concentrations were used (1 mM) so that the droplets were large enough that recovery was not limited by the amount of fluorescent protein within each droplet. Bleaching of a region of protein within each droplet caused a drop in fluorescence that rapidly recovered, with a 50% recovery time ($t_{1/2}$) of approximately 1.54 s (Fig. 3.23). However, it should be considered that the rate of imaging was relatively slow (every 648 ms) to prevent spontaneous bleaching, so this value likely has additional error. Nonetheless, recovery of fluorescence was nearly complete; only 10% of the initial fluorescence signal did not recover 30 s after bleaching. These data demonstrate that LLPS of HERD-2.2–GFP produces motile and dynamic droplets under these conditions.

3.5.2 HERD-2.2–GFP produces liquid-like condensates in cells

Further, and to demonstrate that HERD-2.2–GFP condensates could behave like demixed liquids in cells, FRAP was performed on *E. coli* expressing HERD-2.2–GFP. Initially, and because of the temperature dependence of LLPS for HERD-2.2–GFP, FRAP was performed on cells grown at 37 °C that were then chilled at 5 °C for 5 minutes to induce phase separation. These cells initially do not form HERD-2.2–GFP condensates, however after the brief cooling, enriched, fluorescent condensates were readily visible. By inducing phase separation immediately before imaging it was proposed that these condensates would be most likely to still be liquid-like, rather than having already matured into arrested, immobile materials. It was noted that condensates formed by this method were larger than those formed by HERD-2.2–GFP when grown at 18 °C. This was an additional advantage, as smaller condensates are more challenging to photobleach accurately in live cells.^{44,269}

Bleaching of HERD-2.2–GFP condensates in *E. coli* again showed a rapid recovery of fluorescence within the bleached area, with a $t_{1/2}$ of <1 s (Fig. 3.24). Precise interpretation of the exact value was not performed as the rate of imaging precludes accurate analysis. However, in contrast to the *in vitro* experiments, the magnitude of fluorescence recovery was modest, with recovery only reaching approximately 25% of the initial fluorescence after 20 s. The cause of this weak magnitude of recovery was initially puzzling, as visually the condensates appeared to have relatively uniformly distributed fluorescence after bleaching, suggesting that there was not in fact a large immobile fraction. Indeed, condensates that had been bleached had highly unequally distributed fluorescence immediately after bleaching. The fluorescence intensity directly within the bleached area was significantly lower than that within the same condensate, but outside of the bleach point.

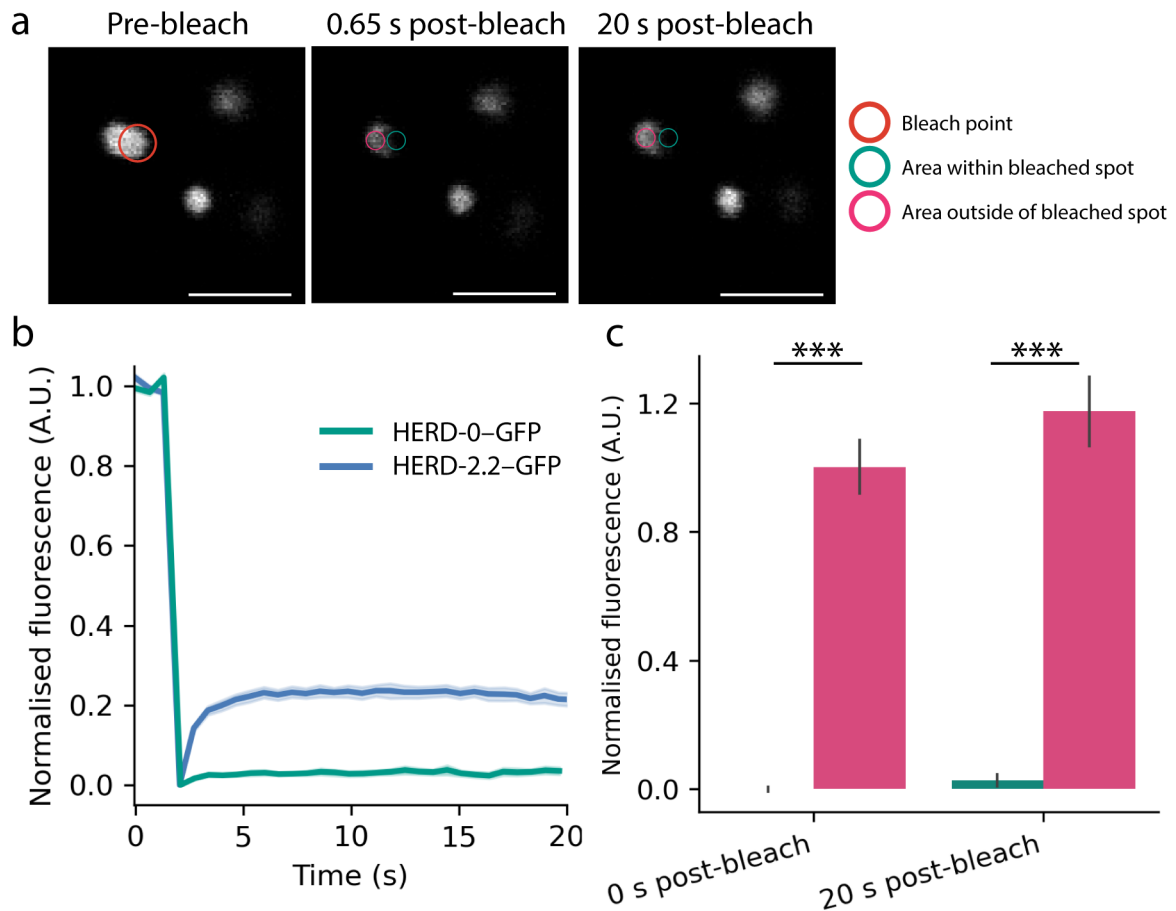


Figure 3.25: HERD-0-GFP forms static aggregates in *E. coli*. **a**, Representative images of *E. coli* expressing HERD-0-GFP before bleaching, immediately after bleaching, and 20 s after bleaching. The bleach point is circled in red, and the two measured areas within the same condensate (within the bleached area, and outside of the bleached area) are circled in teal and magenta respectively. Scale bar is 2.5 μm . **b**, Measured fluorescence recovery within the bleached area of HERD-0-GFP condensates in *E. coli* grown at 37 $^{\circ}\text{C}$. $n = 19$. **c**, Relative normalised fluorescence intensity between the measured areas directly within the bleached spot (teal) and outside of the bleached spot within the same condensate (magenta). $n = 13$, error bars represent the standard error. Statistical testing used a two-tailed T-test with P-values of < 0.001 for both comparisons.

However, 20 s after bleaching the distribution of fluorescence had nearly completely re-equilibrated so that these two areas were now almost equal again. This suggests that in fact there had been near-complete recovery of fluorescence, but that there had been a significant amount of fluorescent protein lost in the bleach process, causing the magnitude of fluorescence to appear much lower. This is likely a result of the droplets in *E. coli* being far smaller than those used *in vitro*, where there was a large available pool of protein available for recovery. Here, bleaching the cells reduced the available pool of fluorescent protein significantly, and was only slightly corrected by normalisation for background or spontaneous bleaching, where fluorescence of a nearby independent cell was used to correct for spontaneous bleaching during measurement. Comparison of these data with

literature experiments of FRAP on phase separated droplets in *E. coli*, corroborated the notion that fluorescence does not recover completely, giving the false impression of a large immobile fraction.²⁷⁰

3.5.3 HERD-0–GFP produces immobile aggregates in *E. coli*

From these data, HERD-2.2–GFP appears to form condensates in *E. coli* that can behave with liquid-like properties, with rapid recovery of fluorescence after photobleaching that matches that measured *in vitro*. To assess the design strategy and screening method used for selection of HERD-2.2–GFP, condensates made by HERD-0–GFP were also measured by FRAP (Fig. 3.25). *E. coli* expressing HERD-0–GFP were also grown at 37 °C and chilled at 5 °C, however in contrast to HERD-2.2–GFP, HERD-0–GFP formed condensates prior to cooling. In addition, when tested by FRAP, condensates made by HERD-0–GFP displayed nearly no fluorescence recovery following bleaching, indicating that protein within these condensates was almost entirely arrested. Further, following the same comparison as made for HERD-2.2–GFP, HERD-0–GFP condensates did not return to an equal distribution of fluorescence following bleaching, demonstrating that there was conclusively no fluorescence recovery. These results show that the design strategy had been successful in transitioning from immobile aggregates to dynamic protein condensates in cells.

3.6 Chapter summary

The design and engineering of proteins for phase separation has made significant progress in recent years.¹⁸⁷ Now, protein designers have begun to explore the potential for the bottom-up construction of artificial MLOs.⁸³ Concepts like the stickers-and-spacers framework have proved extremely useful in this regard, and have been implemented in the engineering of natural phase-separating proteins and the design of synthetic multivalent assemblies.⁸⁹ This chapter has explored the potential for *de novo* protein design to create a stickers and spacers array from the bottom-up. This incorporates elements of rational coiled-coil design, using sequence-to-structure relationships for these α -helical assemblies to modify their interfaces and affinities to an extent currently inaccessible to computational design methods (Figs. 3.25 & 3.26) Further, these helical motifs are concatenated by a designed unstructured linker.

Changes to the primary sequences of these polypeptides translate directly to the material properties of the resulting condensates, as demonstrated by the FRAP measurements on HERD-2.2–GFP and the parent HERD-0–GFP. The optimised HERD-2.2–GFP forms

droplets that are highly dynamic, and behave like de-mixed liquids, indicative of LLPS (Fig. 3.24). It is likely that the resulting subcellular compartments, which are among the first MLOs to be created in prokaryotes, could be permissive to functionalisation.²⁷¹ This could allow the sequestration of endogenous cellular proteins, or the compartmentalisation of functional client proteins within an organelle-like structure in *E. coli*. These ideas form the basis for the following chapter of this thesis.

It is interesting to highlight how this bottom-up designed construct resembles, and differs from, natural phase-separating proteins. By following a stickers-and-spacers framework the design incorporates significant elements of disorder, and in many ways emulates natural intrinsically disordered proteins.²⁷² Its success however, is that the unstructured linker does not appear to be a significant factor for phase separation, and so more of the interactions can be rationalised through helical sequence rules. However, it is also clear that some of the design concepts used in the selection of the coiled coils CC-Tri and CC-Tet2 are no longer valid in the destabilised HERD-2.2. It is well recognised that as helical length and knobs-into-holes packing are attenuated, so is the specificity for oligomeric state and interaction interfaces.²⁷³ This means that while the original CC-Tri and CC-Tet2 may be largely orthogonal to one another, the destabilised motifs used in HERD-2.2 are highly unlikely to be. In fact, Dr Andrey Romanyuk examined the secondary structure of the components of HERD-2.2, and found that due to the extensive destabilisation of the coiled coils they were largely unstructured in solution, requiring TFE to stabilise their α -helical secondary structure (Fig. 3.15).²²⁴

It is also noteworthy that in contrast to natural phase-separating proteins, HERD-2.2-GFP requires molecular crowding to induce LLPS, whereas some natural proteins form de-mixed droplets readily in dilute solutions.²⁶⁰ This may be because HERD-2.2-GFP was designed to condense in the prokaryotic cytoplasm. Prokaryotes not only have much higher levels of molecular crowding, but are also capable of greater recombinant expression levels than eukaryotes.^{274,275} It follows that a protein designed for LLPS under these conditions would require higher concentrations and additional molecular crowding for phase separation *in vitro* as well. However, despite this, the rational HERD design strategy is appealing as a basis for creating proteins that will phase separate under different conditions, for instance for the development of *de novo* MLOs in eukaryotic cells. This will be the basis for chapter 6 of this thesis.

Further investigation into the components of the HERD proteins could present interesting evidence into the precise roles of the components in this stickers-and-spacers framework, including the unstructured linker. These experiments may be of interest to corroborate the theoretical frameworks developed on phase-separating domains. In addition, functional motifs or additional PPIs could be inserted into the linker, for instance to create

multiple layers of control over recruitment. These ideas are outside of the scope of this thesis, but it will be interesting to see how protein design begins to interface with the conceptual and theoretical frameworks for phase separation in the future, as it has already increasingly done so.









HERD #	Structural model	Description
0		4 heptads (28 residues) Original coiled coil designs
1.1		3 heptads (21 residues)
2.1		3 heptads (21 residues) Ala at <i>a</i> position
2.2		2.5 heptads (18 residues) Ala at <i>a</i> position
2.3		3 heptads (21 residues) Low helical propensity residues in HR1
2.4		3 heptads (21 residues) Low helical propensity residues in HR1 1 Ala at <i>a</i> position
2.5		2 heptads (14 residues) Ala at <i>a</i> position
2.6		4 heptads (28 residues) Low helical propensity residues in HR1 Ala at <i>a</i> position

Figure 3.26: Graphical representation of HERD designs 0 through 2.6. Constructs are shown as models with alanine mutations represented by teal spheres in the place of the hydrophobic core residues. Low helical propensity is represented by splaying of the helical bundles.

HERD #	Structural model	Description
2.7		3 heptads (21 residues) Low helical propensity residues in HR1 2 Ala at <i>a</i> position
2.8		3 heptads (21 residues) Low helical propensity residues in HR1 3 Ala at <i>a</i> position
3.1		Long (50 residue) linker
3.2		Short (8 residue) linker
3.3		Increased ionic residue density linker
3.4		Uncharged linker
Ctrl1		3 heptads (21 residues) Ala at <i>a</i> and <i>d</i> positions in HR1 Ala at <i>a</i> position in HR2
Ctrl2		3 heptads (21 residues) Ala at <i>a</i> and <i>d</i> positions in HR1 and 2

Figure 3.27: Graphical representation of HERD designs 2.7 through 3.4, and Ctrl1 and 2. Constructs are shown as models with alanine mutations represented by teal spheres in the place of the hydrophobic core residues. Low helical propensity is represented by splaying of the helical bundles. HERD-Ctrl1 and 2 are shown in their presumed monomeric oligomeric states.

Chapter 4

Functionalisation of *de novo* MLOs

The following work in this chapter comprises part of the publication: Assembling membraneless organelles from de novo designed proteins, Alexander T. Hilditch, Andrey V. Romanyuk, Stephen J. Cross, Richard Obexer, Jennifer J. McManus & Derek N. Woolfson, Nature Chemistry (2023).²²⁴ All of the work described in this chapter is that of the author.

4.1 Chapter introduction

The discovery of protein phase separation and LLPS in biology unlocks a variety of potential applications. Compartmentalisation of proteins within synthetic organelle-like compartments offers the potential to enrich functional enzymes and reaction intermediates, or sequester them away from endogenous proteins and other cellular processes.^{190,193} These membraneless compartments also present a route to functional protocells due to their versatility and facility of assembly.²⁷⁶ As described in chapter 3, both engineered and entirely synthetic scaffolds for protein phase separation have been developed that offer a route to artificial condensates in living cells.^{188,198,199} The challenge now is to functionalise these condensates, creating artificial cellular compartments that behave like organelles.

The uptake of client proteins into condensates has now been shown relatively routinely.^{201,277} Both recombinantly expressed proteins and endogenous macromolecules can be enriched within synthetic membraneless organelles. Further, and as described in chapter 1, several systems for controlled uptake and release of client proteins have been described using small molecule or photo stimulation to trigger switchable PPIs.^{182,190} It is now clear that protein localisation can be robustly controlled by recruitment to protein condensates. Moreover, these condensates can be leveraged to produce far-reaching effects on

cell physiology and metabolism. However, it is still unclear how the compartmentalisation of macromolecules within these condensates changes their physical and chemical properties.²⁷⁸ *In vitro*, LLPS has been used to enhance the rate of enzymatic reactions, or channel reaction intermediates through branched metabolic pathways.^{193,193,279,280} This is largely attributed to the clustering of enzymes within a pathway leading to increased turnover and pathway efficiency.^{281,282} However, LLPS does not always confer a benefit on catalysis.^{278,283} Protein condensates are dense environments, with a viscosity similar to that of glycerol.^{34,284,285} This can slow mobility of both enzymes and substrate to the extent that phase separation is deleterious to reaction kinetics.²⁷⁴ Further, crowding can slow or halt enzyme dynamics, or even block active site accessibility, again reducing rates of reactions.²⁸⁶ Overall, it is challenging to predict how specific enzymatic reactions will react to phase separation. This is particularly true for reactions performed in their native environment – in living cells.

This chapter describes the functionalisation of the designed *de novo* HERD proteins with catalytic enzymes. The two enzyme pathway for producing indigo dye from L-tryptophan is used as a model reaction to measure the effect of protein condensation on product formation in *E. coli*. The enzymes TryptophanaseA (TnaA), and Flavin containing monooxygenase (FMO), are demonstrated to be readily enriched in HERD-2.2–GFP droplets *in vitro* and in cells. Co-condensation of this enzymatic pathway in *E. coli* produces up to 6-fold more indigo than the free enzymes under equivalent conditions. Further, the material properties of the condensates are demonstrated to be critical to catalysis: while dynamic condensates confer significant benefits on pathway efficiency, static, arrested condensates have the opposite effect, greatly reducing product formation.

4.2 HERD-2.2 can colocalise multiple proteins in *E. coli*

Initially, to demonstrate that the designed HERD-2.2 polypeptide could colocalise multiple client proteins, an additional HERD-2.2 protein was designed, HERD-2.2–mCherry, where mEmerald was replaced with the red fluorescent protein mCherry (Tables 8.2 and 8.5). When expressed in *E. coli*, HERD-2.2–mCherry formed condensates similarly to HERD-2.2–GFP (Fig. 4.1a–b). Further, co-expression of HERD-2.2–GFP and HERD-2.2–mCherry produced condensates that were enriched in both fluorescent proteins, indicating that fusion to HERD-2.2 was sufficient not only to produce synthetic condensates, but also to colocalise multiple proteins in cells (Fig. 4.1c).

However, it is true that although both monomeric and evolutionarily somewhat divergent, mCherry and mEmerald are structurally highly similar, both being β -barrel fluorescent proteins. Therefore, it could be argued that this colocalisation could be biased due to

the use of two highly similar client proteins. To demonstrate co-condensation of a greater diversity of proteins, several additional fusion proteins were designed. To visualise cellular localisation, the HERD-2.2–mCherry scaffold was retained, and additional proteins fused to the *C*-terminus of mCherry. The model client enzymes alcohol dehydrogenase, glucose-6-phosphatase, TnaA, and FMO were fused to HERD-2.2–mCherry. Each of these enzymes have different sequences, structures, and oligomeric states. When co-expressed with HERD-2.2–GFP, all of these fusion proteins produced condensates that were enriched in both fluorescent proteins, demonstrating effective colocalisation by HERD-2.2 (Fig. 4.2, Tables 8.2 and 8.5).

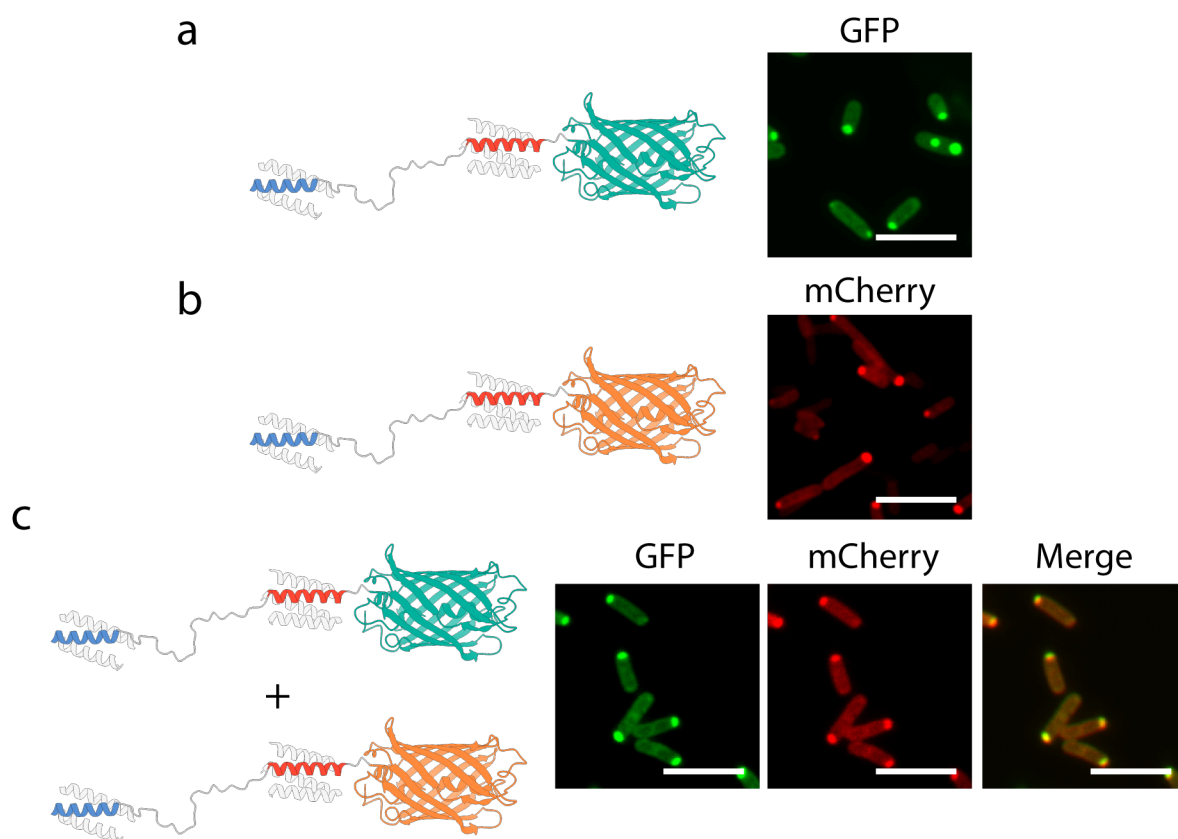


Figure 4.1: HERD-2.2–GFP colocalisation of GFP and mCherry. **a**, Fixed cell confocal microscopy of HERD-2.2–GFP. **b**, Fixed cell confocal microscopy of HERD-2.2–mCherry. **c**, Fixed cell confocal microscopy of co-expressed HERD-2.2–GFP and HERD-2.2–mCherry. Scale bar is 5 μm for all panels.

4.3 Enzymatic production of indigo as an example pathway

The successful colocalisation of multiple proteins by HERD-2.2 confirmed that fusion of HERD-2.2 to client enzymes could be used as a route to creating functional MLOs. The next step was to apply this concept to a functional catalytic pathway. Here, several options were considered. Co-condensation of a large number of enzymes for a complex

biosynthetic pathway was tempting. MLOs appear to be the ideal scaffold for such an application, as their overall lack of structure makes them amenable to a wide range and number of client proteins.²⁷⁸ However, ensuring equivalent expression levels of each of the enzymes would likely prove challenging to robustly demonstrate an improvement on catalysis.

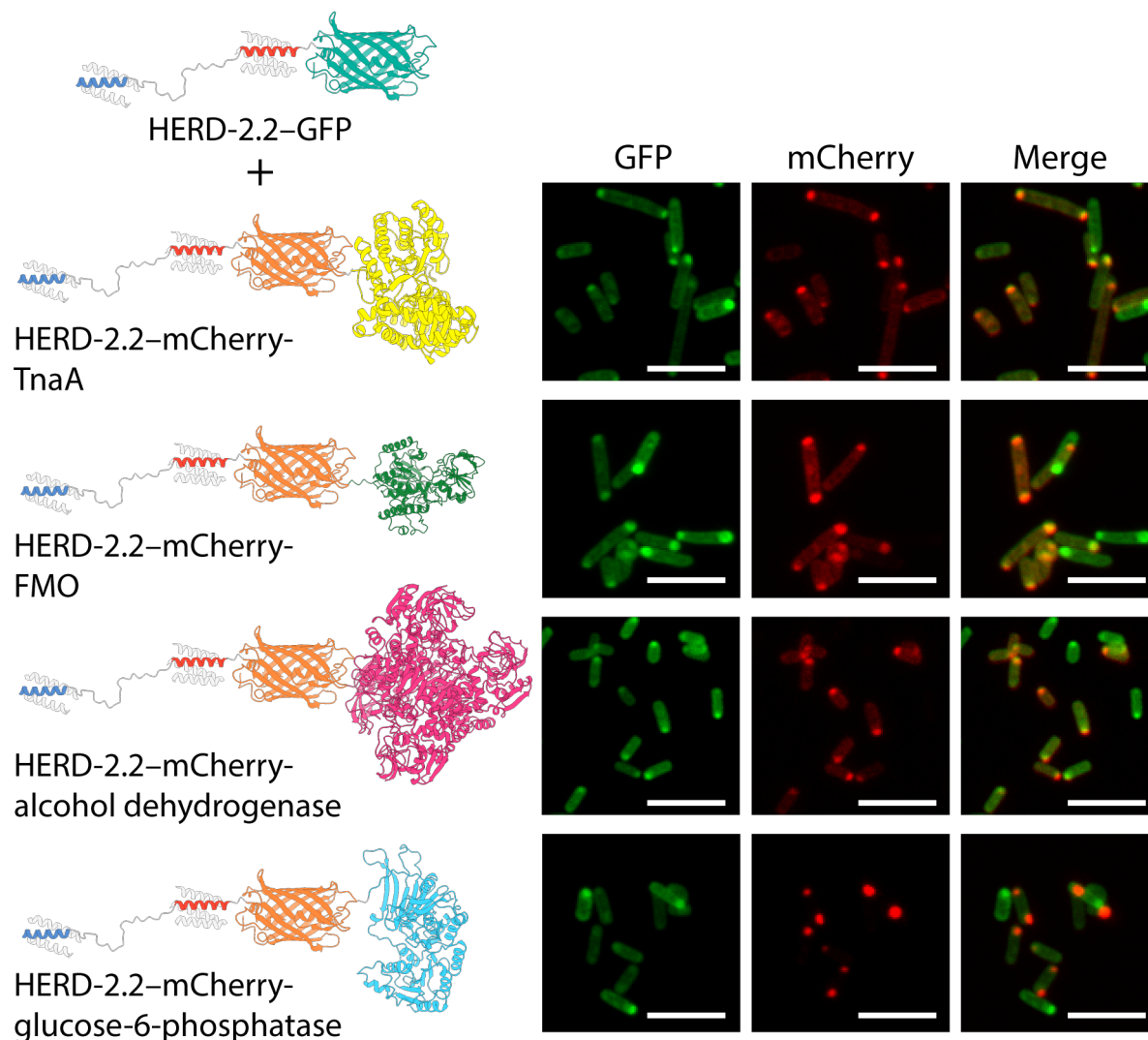


Figure 4.2: HERD-2.2-GFP colocalisation of functional enzymes. Fixed cell confocal microscopy of co-expressed HERD-2.2-GFP and HERD-2.2-mCherry fusions. Each of the enzymes is represented as a monomer for clarity, but all are oligomeric (TnaA, alcohol dehydrogenase – tetrameric, FMO, glucose-6-phosphatase – dimeric). Scale bar is 5 μm .

Instead, as a proof-of-concept enzymatic pathway, the three step reaction for the production of indigo dye from L-tryptophan was chosen (Fig. 4.3).²⁸⁷ Indigo is a fluorescent small molecule used within the textile industry. Indigo is made industrially by petrochemical synthetic pathways, and as a result carries environmental and health concerns due to the inorganic catalysts and reducing agents used. Instead, there have been recent efforts in green chemistry to replace petrochemical processes with biochemical reactions.

The biochemical pathway to produce indigo from L-tryptophan uses two enzymes, TnaA and FMO, followed by a final spontaneous chemical dimerisation step to produce indigo dye.²⁸⁵ TnaA initially cleaves the indole ring from L-tryptophan, followed by oxidation by FMO to produce the indoxyl. This reaction pathway was chosen for several reasons. Firstly, as the reaction product is fluorescent with a relatively high extinction coefficient (ϵ , 22140 M⁻¹ cm⁻¹ at 610 nm in dimethylformamide (DMF)), quantification of reaction progression is possible spectroscopically.²¹⁵ Secondly, both TnaA and FMO are amenable to modification by recombinant molecular biology techniques, such as C- or N-terminal fusion to other polypeptides.²⁸⁷ Thirdly, previous work has demonstrated that colocalisation of TnaA and FMO in lipid-containing scaffolds can confer a mild enhancement on pathway efficiency, suggesting that the pathway could be improved by colocalisation.²⁸⁸

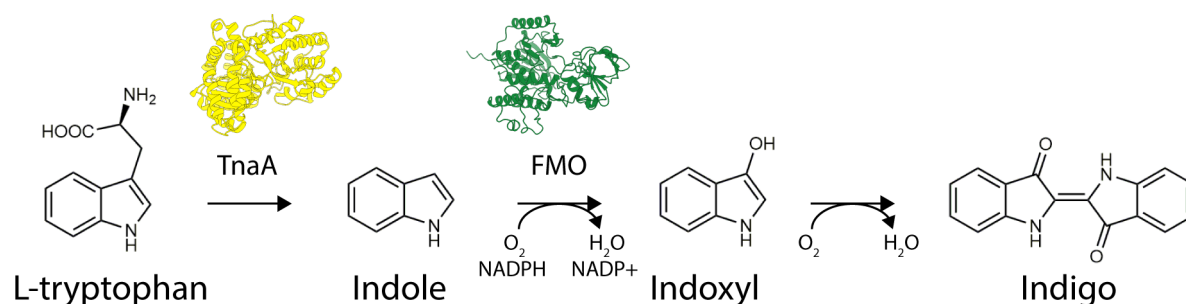


Figure 4.3: Enzymatic production of indigo dye. Indigo is produced from L-tryptophan by the enzymes TnaA and FMO by a three step biochemical pathway.

4.3.1 HERD-2.2–TnaA but not FMO undergoes LLPS *in vitro*

To colocalise TnaA and FMO within *de novo* condensates, both proteins were fused to the C-terminus of HERD-2.2, creating HERD-2.2–TnaA and HERD-2.2–FMO. Both fusions expressed well in *E. coli* and were purified using the N-terminal His tag. First, it was investigated whether HERD-2.2–TnaA and HERD-2.2–FMO could undergo LLPS *in vitro* similarly to HERD-2.2–GFP. The phase behaviour of both fusion proteins was tested under conditions with varying ionic strength, molecular crowding, and pH. While HERD-2.2–TnaA formed de-mixed droplets under similar conditions to HERD-2.2–GFP, HERD-2.2–FMO did not form droplets under any of the tested conditions (Fig. 4.4a). Examination of the surface properties of these two enzymes suggested an explanation for this difference (Fig. 4.4b). TnaA has a very similar net-charge to mEmerald (both -6 at pH 7.5), while FMO is much more negatively charged (-21 at pH 7.5). Due to the weak interactions driving LLPS it is likely that this increase in net-charge could be sufficient to disrupt LLPS by adding additional repulsive interactions.

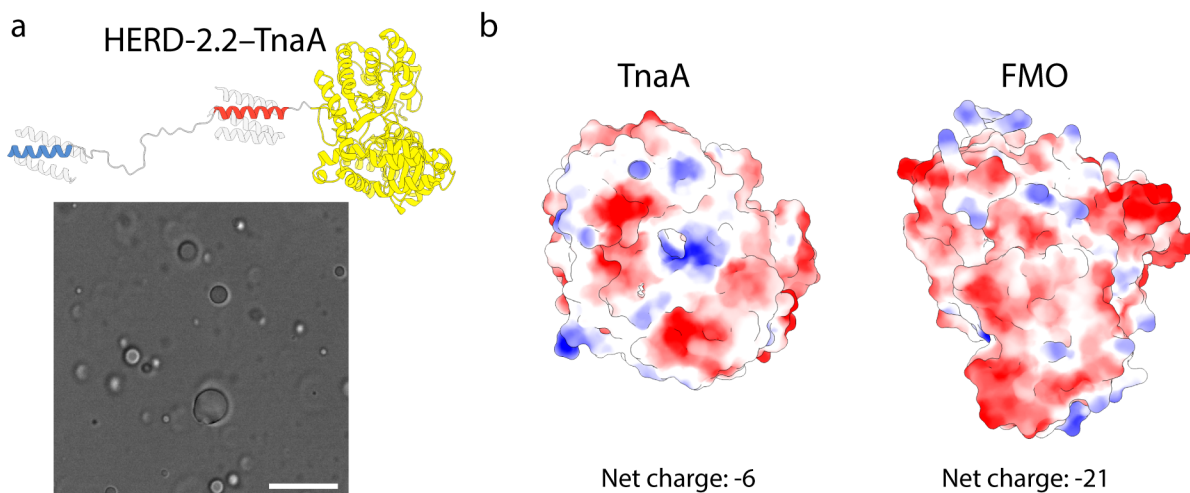


Figure 4.4: HERD-2.2-TnaA de-mixed droplets *in vitro*. **a**, Confocal microscopy brightfield image of de-mixed droplets formed by HERD-2.2-TnaA. Scale bar is 15 μm . **b**, Models of TnaA and FMO coloured by electrostatics, with negatively charged side chains coloured in red, and positively coloured side chains coloured in blue.

4.3.2 HERD-2.2-TnaA and FMO are enriched in HERD-2.2-GFP droplets

Having established that HERD-2.2-TnaA but not FMO would undergo LLPS *in vitro*, phase separation of HERD-2.2-GFP was instead used to create the MLO scaffold. In this arrangement, HERD-2.2-TnaA and HERD-2.2-FMO are recruited as client enzymes to extant HERD-2.2-GFP droplets (Fig. 4.5). This has several benefits over using the catalytic enzymes themselves to induce LLPS. Firstly, if over-expressed to the concentrations required to phase separate, TnaA and FMO would be present in near stoichiometric concentrations to the substrate, and changes to pathway efficiency would be challenging to determine in such an enzyme-rich environment. Secondly, and as demonstrated by the altered conditions required for phase separation of HERD-2.2-TnaA, droplets made by these enzymes would likely have altered material properties or phase transitions to the previously characterised HERD-2.2-GFP. By using HERD-2.2-GFP as the scaffold and injecting small concentrations of HERD-2.2-TnaA and HERD-2.2-FMO, the properties of the condensates are likely to closely resemble the liquid-like droplets already characterised by soft matter techniques. Finally, the inclusion of a fluorescent scaffold permits the monitoring of protein condensation with respect to loading of client enzymes.

First, it was tested if HERD-2.2-TnaA and HERD-2.2-FMO would indeed be enriched within HERD-2.2-GFP droplets. To do this with minimal changes to the polypeptide, a tetra-cysteine (TC) tag (CCPGCC) was included into the unstructured linker of HERD-2.2 to create TC-HERD-2.2-TnaA and TC-HERD-2.2-FMO (Table 8.2).²⁸⁹ To test enrichment within HERD-2.2-GFP droplets *in vitro*, both these proteins were first expressed and purified. Addition of a bi-arsenical dye (TC-ReAsH II) was then

used to site-specifically label the tetra-cysteine tagged enzymes within HERD-2.2-GFP droplets (Fig. 4.6a).²⁹⁰ HERD-2.2-GFP was included in excess to create the de-mixed droplets (500 μ M) while TC-HERD-2.2-TnaA and TC-HERD-2.2-FMO were added in catalytic concentrations (25 μ M). De-mixing was then induced by addition of PEG to induce molecular crowding. Measurement of TC-ReAsH II fluorescence within droplets containing TC-HERD-2.2-TnaA and TC-HERD-2.2-FMO indicated that both enzymes were enriched within the droplets, though to differing amounts (Fig. 4.6b-c). TC-HERD-2.2-TnaA produced over 20 times greater fluorescence intensity than droplets containing TC-ReAsH II, but no tetra-cysteine tagged enzymes. TC-HERD-2.2-FMO, while still enriched, only had around a 6-fold greater fluorescence intensity than the background. Again, suggesting that FMO is less readily enriched within the *de novo* condensates using the HERD-2.2 based system.

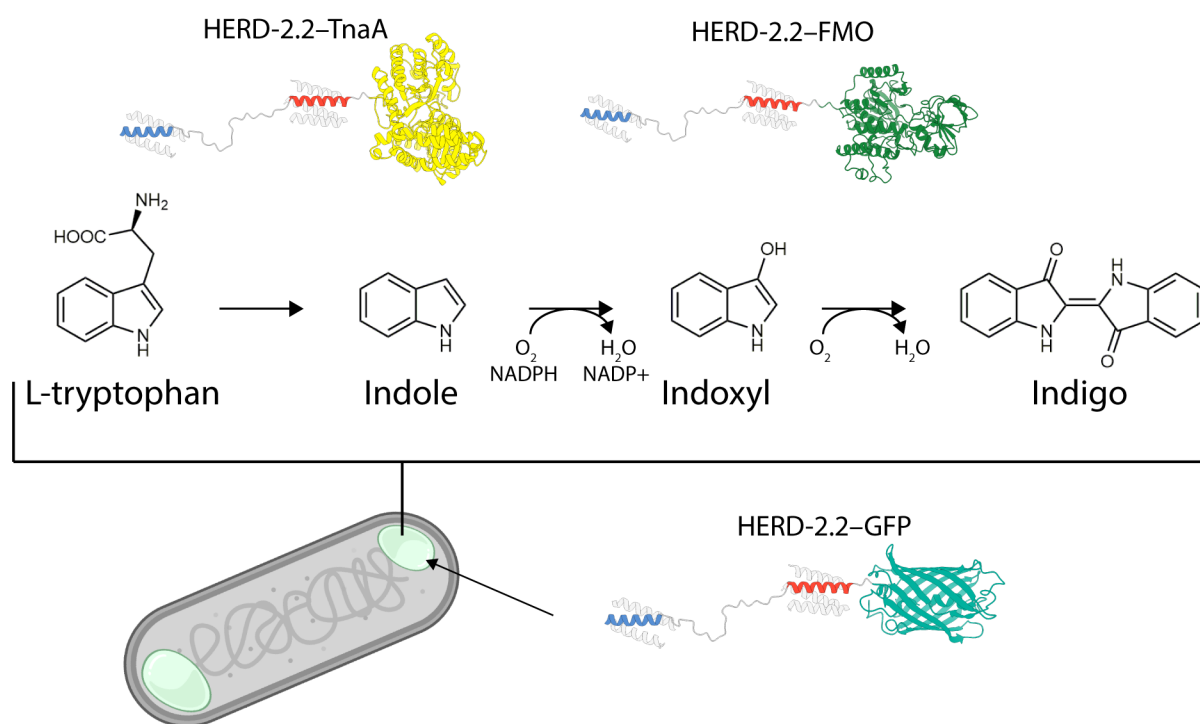


Figure 4.5: Enzymatic production of indigo in *de novo* condensates. Schematic of the *de novo* condensates assembled by HERD-2.2 for the production of indigo dye in *E. coli*, co-condensing both HERD-2.2-TnaA and HERD-2.2-FMO within HERD-2.2-GFP MLOs.

Further, to demonstrate co-condensation of TC-HERD-2.2-TnaA and TC-HERD-2.2-FMO in cells, the tetra-cysteine tagged enzymes were co-expressed with HERD-2.2-GFP in *E. coli*, and the cell-permeable TC-ReAsH II used to specifically label the enzymes.²⁸⁹ Here again, both cells expressing TC-HERD-2.2-TnaA and TC-HERD-2.2-FMO showed TC-ReAsH II fluorescence within HERD-2.2-GFP condensates (Fig. 4.6d-e). However, there was a strong background signal in cells only expressing HERD-2.2-GFP, without a tetra-cysteine label.

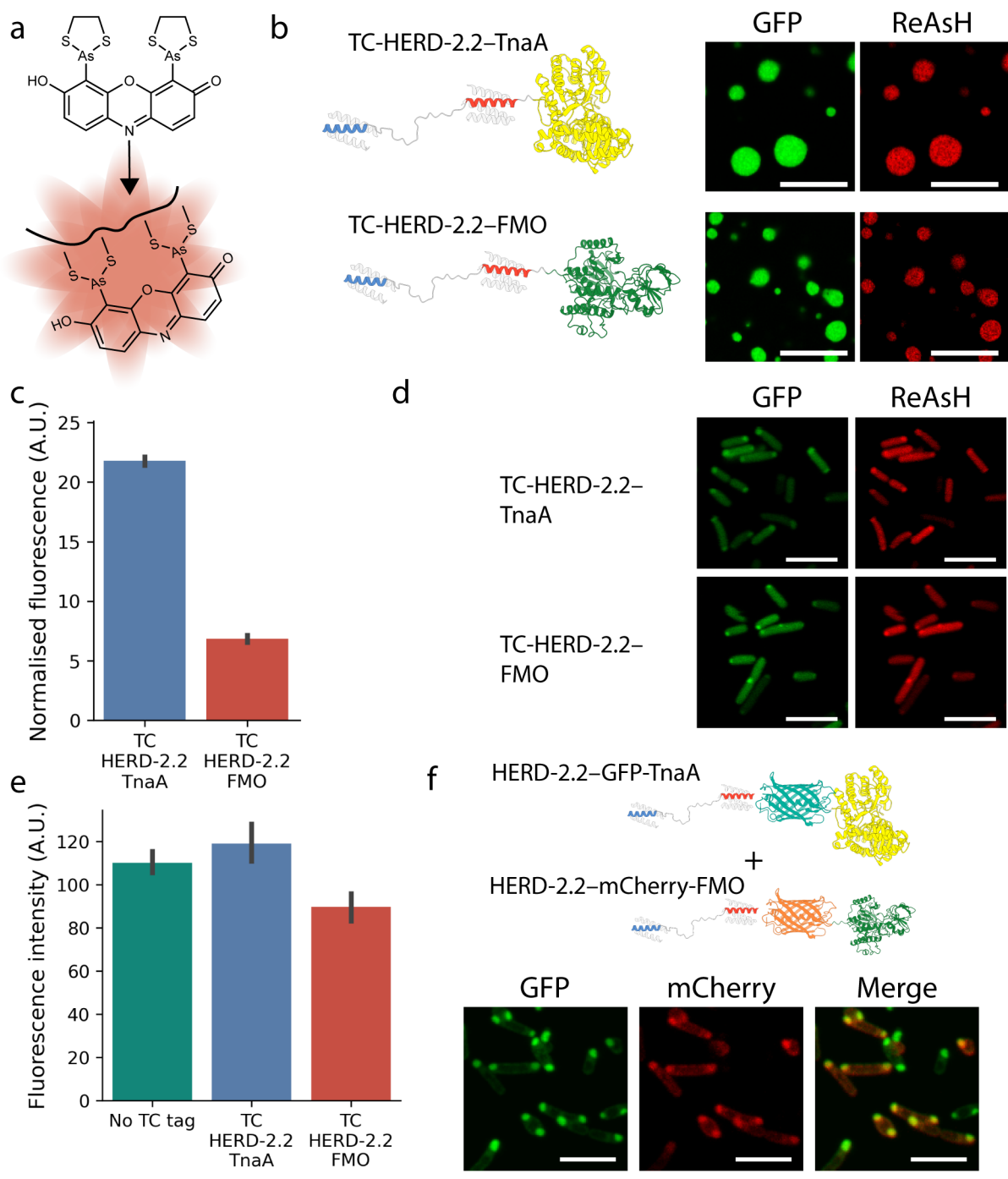


Figure 4.6: HERD-2.2–TnaA and HERD-2.2–FMO can be co-condensed *in vitro* and in cells. **a**, Chemical structure of ReAsH dye, showing fluorescence induced by binding to a tetracysteine motif. **b**, Confocal microscopy images of HERD-2.2–GFP droplets enriched with the tetracysteine-labelled enzymes TC-HERD-2.2–TnaA and TC-HERD-2.2–FMO, and labelled with ReAsH. **c**, Normalised fluorescence intensity of HERD-2.2–GFP droplets containing TC-HERD-2.2–TnaA and TC-HERD-2.2–FMO after excitation of ReAsH dye (561 nm). **d**, Confocal microscopy images of *E. coli* co-expressing HERD-2.2–GFP and TC-HERD-2.2–TnaA and TC-HERD-2.2–FMO, following labelling with ReAsH dye. **e**, Fluorescence intensity of HERD-2.2–GFP condensates in *E. coli* with no co-expressed tetracysteine labelled enzymes, or co-expressing TC-HERD-2.2–TnaA and TC-HERD-2.2–FMO after excitation of ReAsH dye (561 nm). **f**, Fixed cell confocal microscopy of *E. coli* co-expressing HERD-2.2–GFP–TnaA and HERD-2.2–mCherry–FMO. Scale bars are 5 μ m for all panels.

In theory, TC-ReAsH II is only fluorescent when covalently conjugated to a tetra-cysteine tag, and so this background was surprising. However, it may be that dye fluorescence is affected by the environment within the HERD-2.2–GFP condensates. Due to the level of background, and being unable to precisely normalise expression levels in cells, TC-ReAsH II labelling was not used to measure enrichment of HERD-2.2–TnaA and HERD-2.2–FMO in cells. Instead, and to confirm that both enzymes would co-condense when fused to HERD-2.2, TnaA was fused *C*-terminally to HERD-2.2–GFP, creating HERD-2.2–GFP–TnaA, and co-expressed with HERD-2.2–mCherry–FMO. *E. coli* co-expressing these proteins produced condensates enriched in both fluorescent proteins, demonstrating that both enzymes could be co-condensed using the HERD-2.2 system (Fig. 4.6f).

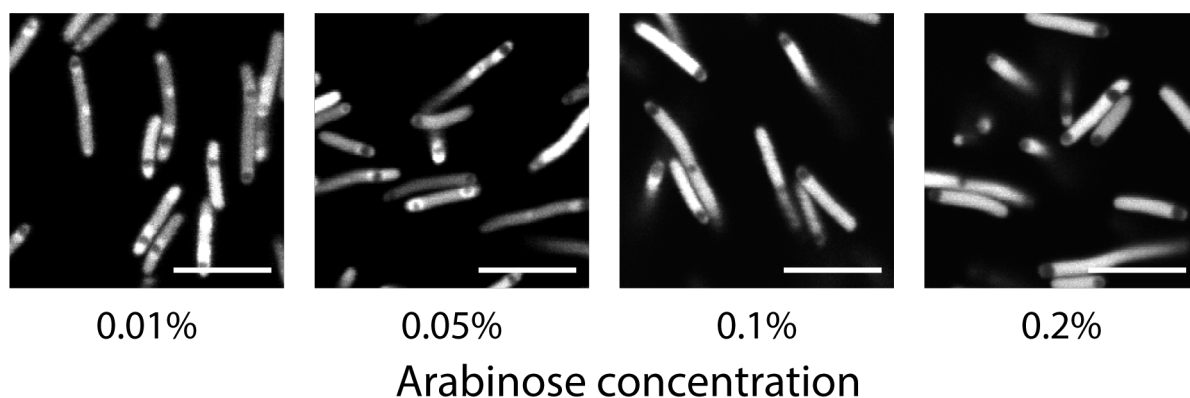


Figure 4.7: HERD-2.2–GFP condensates are sensitive to enzyme loading. Live cell confocal microscopy images of HERD-2.2–GFP condensates in *E. coli* co-expressed with the enzymes HERD-2.2–TnaA and HERD-2.2–FMO. Scale bars are 5 μ m.

4.4 Indigo production in HERD-2.2 condensates is sensitive to enzyme loading

With a *de novo* scaffold for LLPS in hand, and having demonstrated that it was possible to enrich functional enzymes within the condensates in cells, the potential for these MLOs to accelerate enzymatic pathway efficiency was investigated. For indigo production in *E. coli*, HERD-2.2–GFP was expressed at high levels under the viral T7 promoter, while the two enzymes HERD-2.2–TnaA and HERD-2.2–FMO were expressed at lower concentrations under the titratable arabinose promoter on a single polycistronic vector. In addition, as TnaA is an endogenous protein in *E. coli*, expression was performed in Δ tnaA BL21 (DE3) *E. coli* kindly donated by Dr Chong Zhang, Tsinghua University, so that all the available TnaA present was under the control of the exogenous plasmid.²¹⁴ The relative amounts of indigo produced were normalised both to cell density (using OD₇₀₀ to avoid conflation of cell density and absorbance due to indigo), and enzyme expression levels.²¹⁵

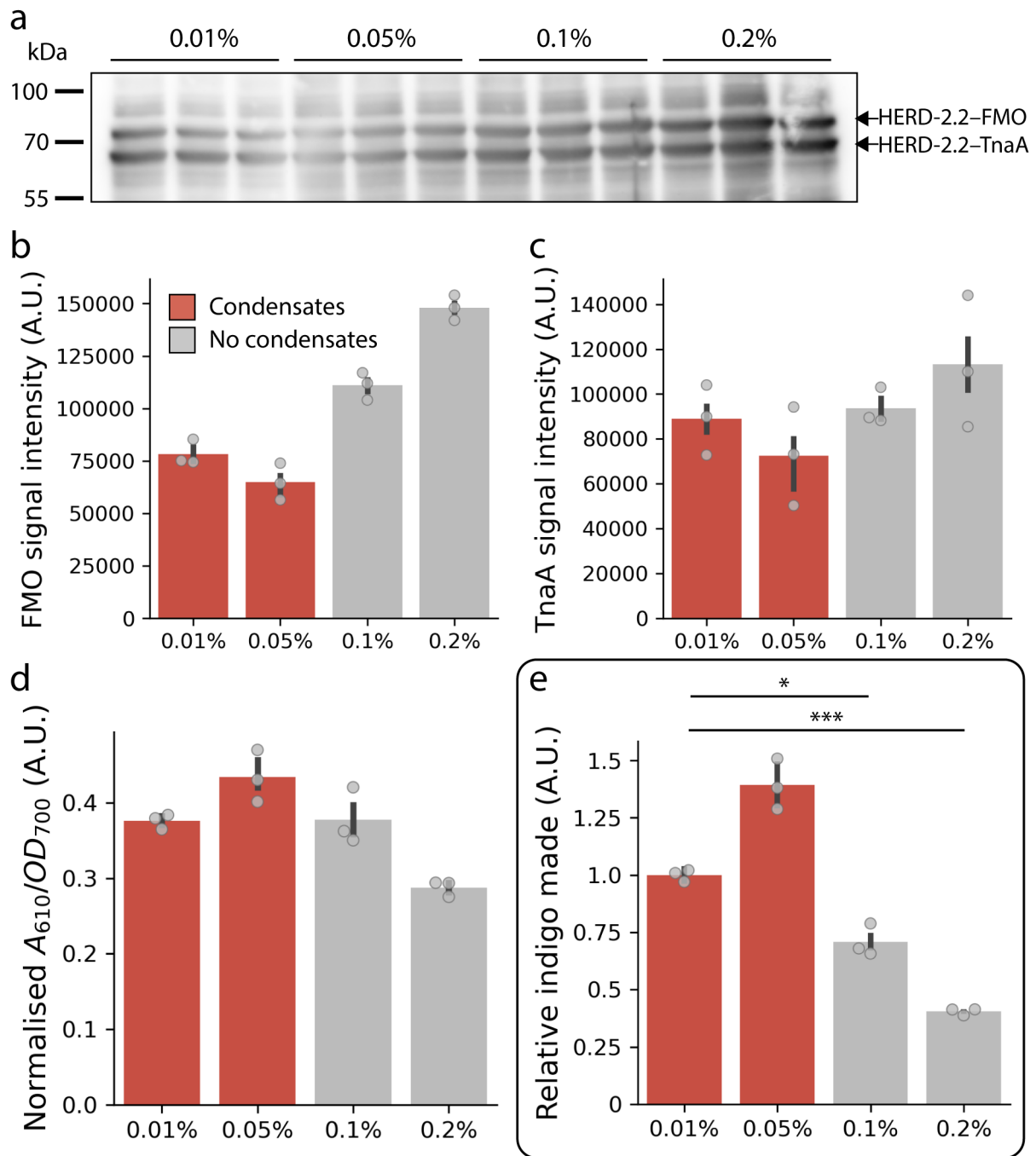


Figure 4.8: Indigo production is accelerated by condensation. **a**, Western blot for quantification of HERD-2.2-FMO and HERD-2.2-TnaA, blotted against the *N*-terminal His tag. Details of western blot quantification are given in the materials and methods. **b-c**, Quantification of HERD-2.2-FMO (**b**) and HERD-2.2-TnaA (**c**) from western blotting. **d-e**, Quantification of indigo production (A_{610}) normalised only to cell density (OD_{700} ; **d**) and normalised both to cell density and to FMO expression levels as measured by western blotting (A.U.; **e**). Data for all panels are represented as the mean \pm the standard error from $n = 3$ biologically independent experiments for each sample. Statistical testing was performed using one-way ANOVA and TukeyHSD post-hoc test with **e**: 0.01% vs 0.1% $P = 0.0318$ (*); 0.01% vs 0.2% $P = 0.001$ (***). For all panels samples identified as forming condensates are coloured in red, and samples forming exclusively a single phase are coloured in grey.

Enzyme expression levels were quantified using western blotting against the TnaA and FMO constructs using the *N*-terminal His tags. For normalisation, the concentration of FMO was used as this has been characterised as the rate-limiting enzyme in this pathway.^{216,217} Initially, the effect of client enzyme concentration on condensation was assessed. Having determined that in particular FMO was poorly tolerated within HERD-2.2 condensates, it was possible that its presence at high concentrations could be disruptive to protein condensation. *E. coli* expressing HERD-2.2-GFP and HERD-2.2-TnaA and HERD-2.2-FMO were induced with varying amounts of D-arabinose (0% – 0.2%) to modulate the enzyme concentrations, while keeping the concentration of IPTG constant (400 μ M). Cells were grown and imaged live at 33 °C, as these conditions were believed to be the most likely to produce dynamic HERD-2.2-GFP condensates with liquid-like properties. As demonstrated previously, HERD-2.2-GFP forms enriched fluorescent condensates under these conditions (Fig. 4.7). Further, on induction of enzyme expression with low concentrations of D-arabinose, these fluorescent condensates are maintained. However, and as postulated, high expression levels (above 0.1% D-arabinose) of HERD-2.2-TnaA and HERD-2.2-FMO resulted in the disruption of HERD-2.2-GFP condensates, returning to a single phase system.

Interestingly however, cells that formed HERD-2.2-GFP condensates appeared to produce more indigo than those with disrupted condensates. Cells induced with lower D-arabinose concentrations (0.01% and 0.05%) that retained their condensates produced materially more indigo than those induced with higher concentrations (0.1% and 0.2%) where the condensates were disrupted (Fig. 4.8). Normalisation for enzyme expression levels, which were confirmed to be higher in cells induced with 0.1% and 0.2% D-arabinose by western blotting, accentuated this difference in indigo production (Fig. 4.8e). These data confirm other reports that protein condensation can be disrupted by excess loading of client enzymes that are poorly tolerated. However they also suggest that HERD-2.2-GFP condensates confer an improvement on indigo production for this two-enzyme pathway.

4.5 HERD-2.2-GFP condensates accelerate pathway efficiency

To confirm the acceleration of pathway efficiency by HERD-2.2 condensates, several additional experiments were performed. Indigo production in cells with HERD-2.2 condensates was first compared to those expressing free enzymes labelled only with a His tag for quantification by western blotting. For the HERD-2.2 labelled enzymes, 0.01% arabinose was used to induce expression, so that condensation was not disrupted by high expression levels. In addition, the expression levels of the free enzymes was matched as closely as possible to the HERD-2.2 fusions to ensure as robust a comparison as possible.

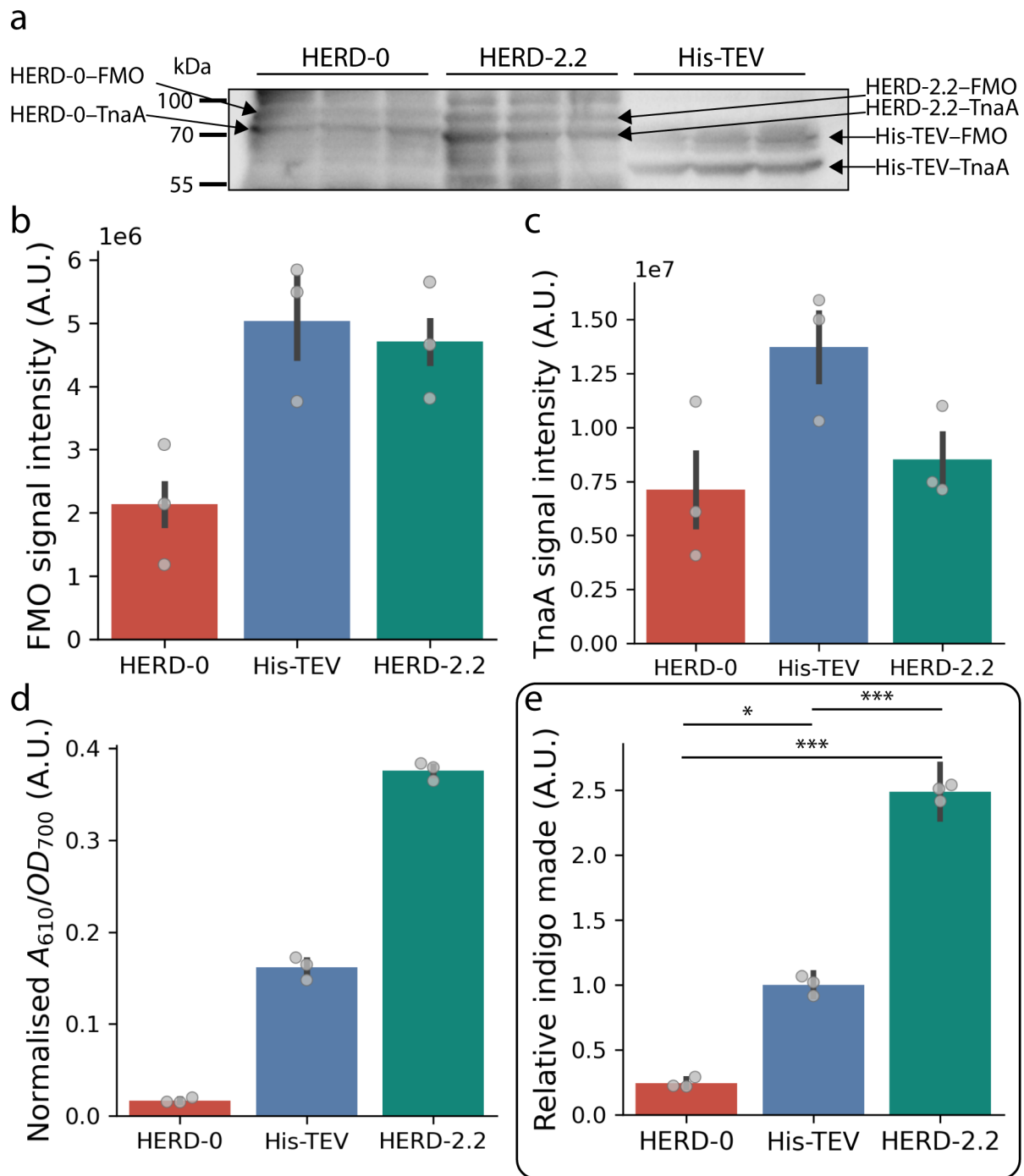


Figure 4.9: HERD-2.2 labelled enzymes produce more than 2-fold more indigo. **a**, Western blot for quantification of the HERD-2.2, HERD-0, and His-TEV FMO and TnaA fusions, blotted against the *N*-terminal His tag. Details of western blot quantification are given in the materials and methods. **b-c**, Quantification of the FMO (**b**) and TnaA (**c**) fusions from western blotting. **d-e**, Quantification of indigo production (A_{610}) normalised only to cell density (OD_{700} ; **d**) and normalised both to cell density and to FMO expression levels as measured by western blotting (A.U.; **e**). Data for all panels are represented as the mean \pm the standard error from $n = 3$ biologically independent experiments for each sample. Statistical testing was performed using one-way ANOVA and TukeyHSD post-hoc test with **e**: P = 0.0285 (*); His-TEV vs HERD-2.2 P = 0.001 (***) ; HERD-0 vs HERD-2.2 P = 0.001 (***) .

Here, the free enzymes expressed less readily than the HERD-2.2 fusions, and so were induced with a higher arabinose concentration (0.2%). Comparison of indigo production in cells grown at 33 °C demonstrated a clear improvement in indigo production due to condensation. The HERD-2.2 fusion enzymes produced approximately 2.3-fold more indigo over the same time period as the free enzymes, after normalisation for cell density and enzyme expression levels (Fig. 4.9). This fold improvement is modest, but approximately matches that shown by previous in-cell colocalisation of the same enzymes in lipid-containing scaffolds.²⁸⁸ Further, this fold change is consistent with the decrease in indigo production shown earlier by over-loading with client proteins disrupting condensate formation.

4.5.1 HERD-0–GFP condensates restrict indigo production

As it was postulated that the material properties of the designer condensates could impact enzymatic pathway efficiency, the effect of colocalisation within HERD-0–GFP condensates was assessed. These condensates had been measured by FRAP to be static and likely arrested aggregates, and so would be expected to behave poorly as reaction crucibles (Fig. 3.25). To colocalise TnaA and FMO in HERD-0 condensates, two additional proteins were designed, where TnaA and FMO were now fused to the *C*-terminus of HERD-0. These proteins expressed well, and were co-expressed with HERD-0–GFP and compared to the free enzymes. In this system however, there was a significant decrease in indigo production compared to the free enzymes: following normalisation, there was an approximately 4-fold overall decrease in indigo production compared to the free-enzyme control (Fig. 4.9). This decrease confirms the hypothesis that enzymes within HERD-0 condensates have diminished activity, possibly due to attenuated substrate accessibility or even protein misfolding within the arrested condensates.

4.6 Pathway efficiency correlates with material properties

The potential dependence of pathway efficiency on condensate material properties was striking, and warranted further investigation. While the HERD-0 system suggested a correlation, these irreversibly aggregated condensates made conclusive determination challenging. Instead, the strong temperature dependence of HERD-2.2–GFP condensation was used to corroborate the influence of material properties on enzymatic pathway efficiency.

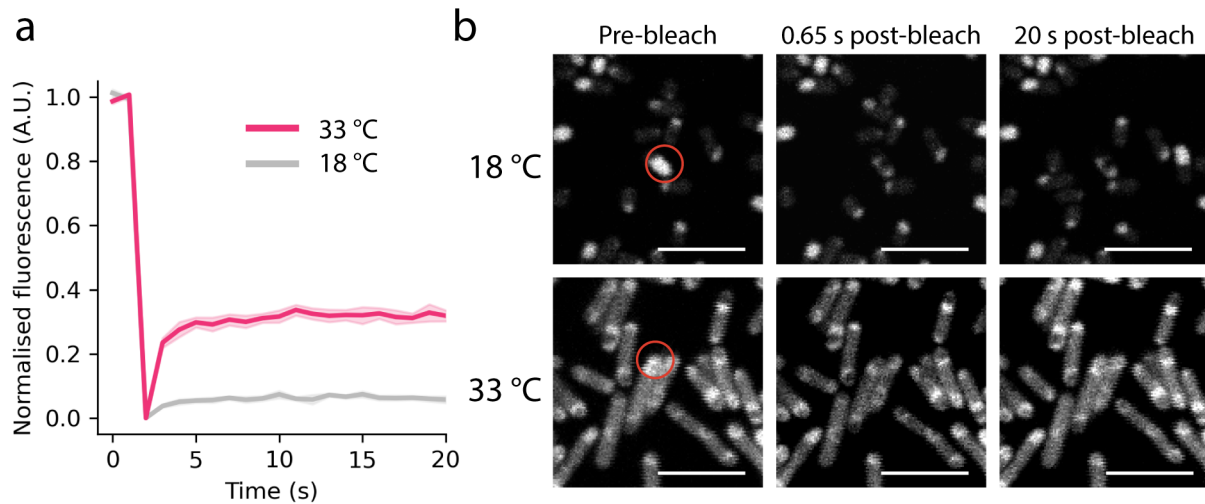


Figure 4.10: HERD-2.2-GFP condensates grown at 18 °C are not dynamic. **a**, Measured fluorescence recovery within the bleached area of HERD-2.2-GFP condensates in *E. coli* grown at 33 °C (pink, $n = 16$) and 18 °C (grey, $n = 15$). **b**, Representative images of *E. coli* expressing HERD-2.2-GFP condensates in *E. coli* grown at the indicated temperature, imaged before bleaching, immediately after bleaching, and 20 s after bleaching. The red circle indicates the bleach area. Scale bars are 5 μm .

4.6.1 HERD-2.2 condensates grown at 18 °C do not recover after bleaching

HERD-2.2-GFP appeared to form different condensate morphologies at different growth temperatures. At 37 °C, the protein was entirely cytoplasmically distributed, while at 33 °C large, relatively sparse, condensates were formed, and at 18 °C the condensates formed were smaller and appear more dense (Fig. 3.13). It was posited that these changes in morphology could relate to changes in material properties. To test this, FRAP was performed on HERD-2.2-GFP in *E. coli* grown at 33 °C and at 18 °C (Fig. 4.10). As anticipated, the condensates formed at 33 °C gave rapid recovery after photobleaching, with a rate and magnitude of recovery similar to that measured for HERD-2.2-GFP in *E. coli* grown at 37 °C. These data confirm that at 33 °C HERD-2.2-GFP displays liquid-like properties. In contrast, in cells grown at 18 °C, HERD-2.2-GFP gave extremely limited fluorescence recovery after photobleaching. From the change in growth temperature alone, the condensates had switched from behaving like de-mixed liquids to static condensates.

4.6.2 HERD-2.2 condensates grown at 18 °C inhibit enzymatic activity

Having determined that HERD-2.2-GFP condensates switch from liquid-like behaviour when grown at 33 °C to arrested materials at 18 °C, these principles were applied to enzymatic pathway efficiency. The amount of indigo made by HERD-2.2 co-condensed TnaA and FMO, and the free enzymes, was compared in cells grown at 18 °C.

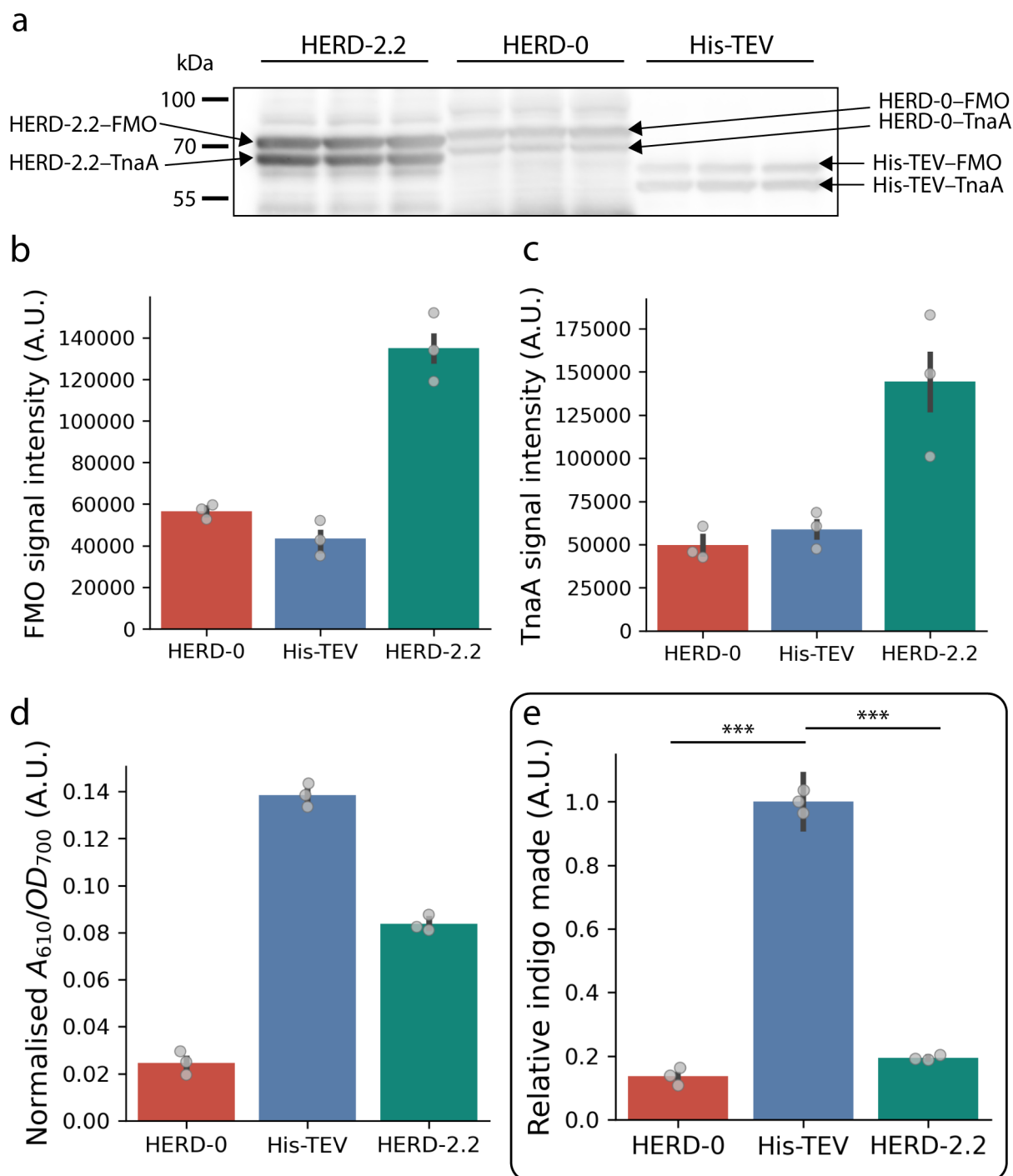


Figure 4.11: HERD-2.2 condensates perform worse than free enzymes at 18 °C. **a**, Western blot for quantification of the HERD-2.2, HERD-0, and His-TEV FMO and TnaA fusions, blotted against the *N*-terminal His tag. Details of western blot quantification are given in the materials and methods. **b-c**, Quantification of the FMO (**b**) and TnaA (**c**) fusions from western blotting. **d-e**, Quantification of indigo production (A_{610}) normalised only to cell density (OD_{700} ; **d**) and normalised both to cell density and to FMO expression levels as measured by western blotting (A.U.; **e**). Data for all panels are represented as the mean \pm the standard error from $n = 3$ biologically independent experiments for each sample. Statistical testing was performed using one-way ANOVA and TukeyHSD post-hoc test with **e**: HERD-0 vs His-TEV $P = 0.001$ (***) ; His-TEV vs HERD-2.2 $P = 0.001$ (***) .

Here, instead of conferring a benefit on pathway efficiency, the HERD-2.2 localised enzymes produced significantly less indigo than the free enzymes: after normalisation, HERD-2.2 condensed enzymes were determined to have produced only around 20% of the indigo made by the free enzymes (Fig. 4.11). This dramatic shift from an overall 230% increase in indigo production, to an 80% decrease, is correlated with the measured material properties of the HERD-2.2 condensates. The lack of fluorescence recovery suggests that the condensates have now become immobile and are likely restricting enzyme activity or substrate accessibility similarly to the HERD-0 condensates. Indeed, measurement of HERD-0 localised enzymes in cells grown at 18 °C again showed very poor indigo production, now only slightly less than that made by HERD-2.2 condensates under the same conditions.

4.7 HERD-2.2 re-design boosts FMO co-condensation and indigo production

HERD-2.2 has been demonstrated to be a useful scaffold for condensate formation. Overall, it confers a benefit on the chosen enzymatic pathway that is comparable to other examples of colocalisation of two enzyme pathways. However, the improvement in indigo production was relatively modest. It was postulated that the effect due to condensation within an MLO could be improved by more effective enrichment of the client enzymes.

In particular, the poor enrichment of FMO within HERD-2.2 droplets was identified as a possible weakness of the designed colocalisation strategy. HERD-2.2–FMO was measured to be enriched within HERD-2.2–GFP droplets much less than HERD-2.2–TnaA (Fig 4.6c). As the rate limiting enzyme of the two-enzyme pathway, this relatively weak co-condensation could be limiting the overall benefits of LLPS on indigo production. As noted previously, FMO is highly negatively charged, which is potentially accentuated by it being dimeric.²⁹¹ To improve co-condensation of FMO within HERD-2.2–GFP droplets, a final HERD protein was designed to adjust the overall net charge of the protein. Rather than by changing the surface charge of FMO, which would necessitate characterising the enzyme to ensure its activity and kinetics had not been altered, all of the changes to the protein were made to the designed HERD sequence. To adjust the net charge to make it overall less negative, all 10 negatively charged residues in HERD-2.2 (9 glutamate residues and 1 aspartate residue) were replaced with glutamine (Fig. 4.12a). These residues were all positioned either in the solvent exposed positions of the helical regions or in the unstructured linker, and so would not be expected to have an effect on the attractive interactions of the HERD sequence. This new HERD, denoted HERD-4.1, had a net charge when fused to FMO of -13, closer to the net charge of HERD-2.2–GFP (-8).

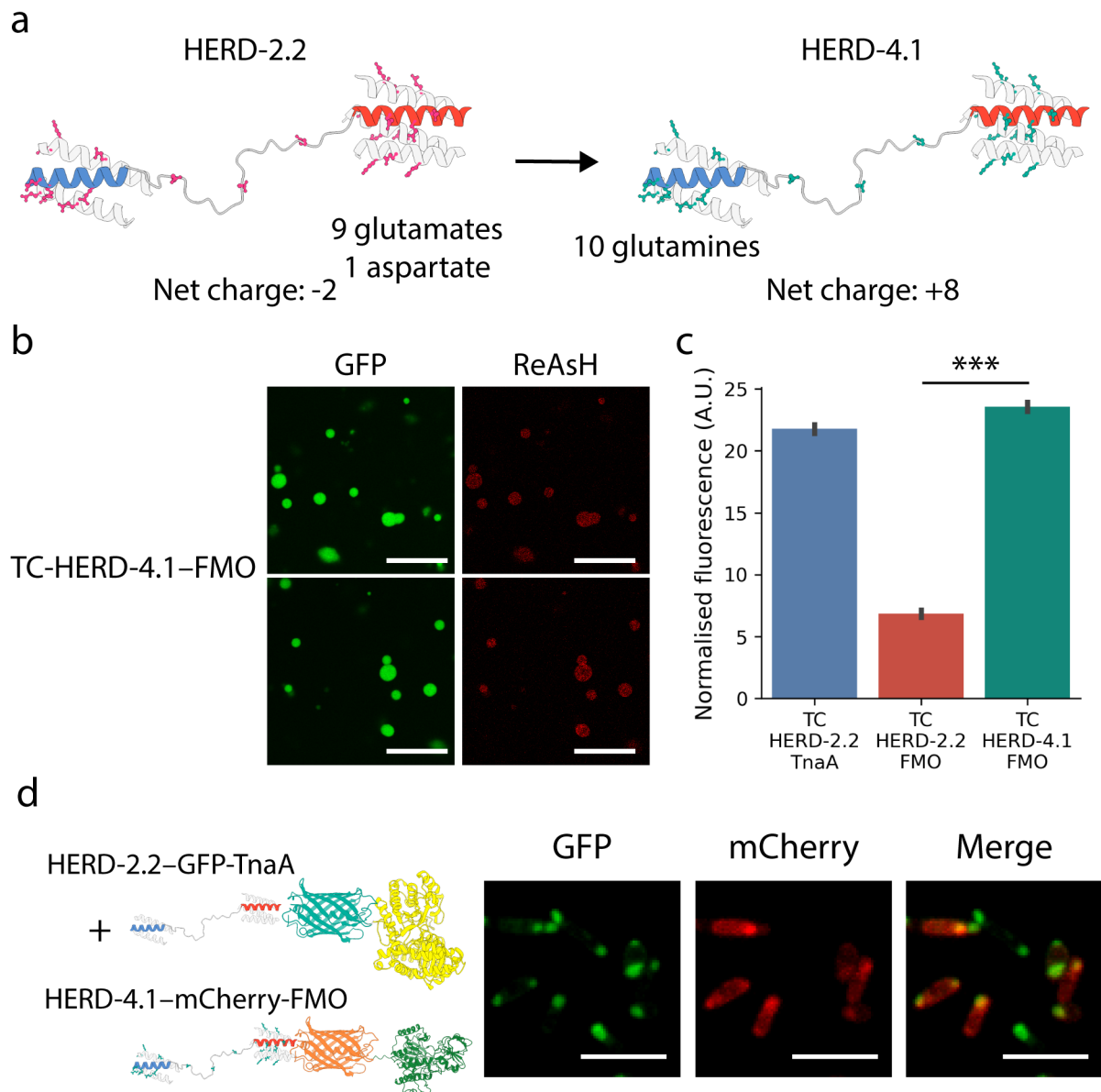


Figure 4.12: HERD-2.2 charge re-design boosts FMO co-condensation. **a**, Models of HERD-2.2 and HERD-4.1, highlighting the glutamate and aspartate residues in HERD-2.2 (magenta sticks) and mutation to glutamine residues in HERD-4.1 (teal sticks). **b**, Confocal microscopy images of HERD-2.2-GFP droplets enriched with the tetracysteine labelled enzyme TC-HERD-4.1-FMO, and labelled with ReAsH. Scale bars are 20 μm . **c**, Normalised fluorescence intensity of HERD-2.2-GFP droplets containing TC-HERD-2.2-TnaA, TC-HERD-2.2-FMO and TC-HERD-4.1-FMO after excitation of ReAsH dye (561 nm). Data are represented as the mean \pm the standard error from $n = 12$ (TC-HERD-2.2-TnaA, TC-HERD-2.2-FMO) or $n = 18$ (TC-HERD-4.1-FMO) measurements. $P = 0.001$ (***) by one way ANOVA and TukeyHSD post-hoc test. **d**, Fixed cell confocal microscopy images of co-expressed HERD-2.2-GFP-TnaA and HERD-4.1-mCherry-FMO. Scale bars are 5 μm .

To evaluate whether this re-design had improved co-condensation of FMO into HERD-2.2-GFP droplets, a tetra-cysteine tagged HERD-4.1-FMO protein was designed and purified as described previously for analysis by TC-ReAsH II labelling (Table 8.2).

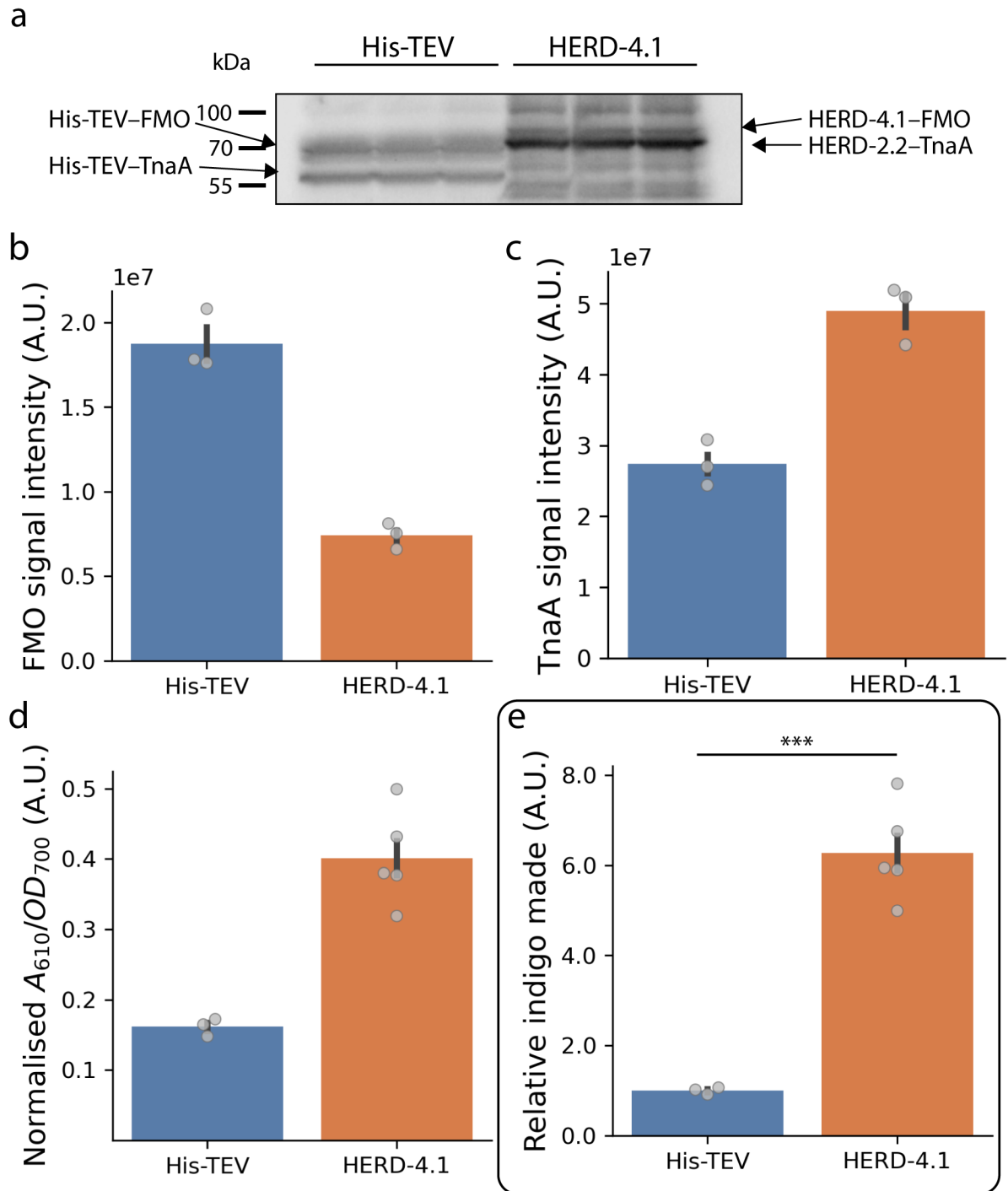


Figure 4.13: HERD-4.1 based condensates improve indigo production 6-fold. **a**, Western blot for quantification of the HERD-2.2/4.1, HERD-0, and His-TEV FMO and TnaA fusions, blotted against the *N*-terminal His tag. Details of western blot quantification are given in the materials and methods. **b-c**, Quantification of the FMO (**b**) and TnaA (**c**) fusions from western blotting. **d-e**, Quantification of indigo production (A_{610}) normalised only to cell density (OD_{700} ; **d**) and normalised both to cell density and to FMO expression levels as measured by western blotting (A.U.; **e**). Data for panels **b-c** are represented as the mean \pm the standard error from $n = 3$ biologically independent experiments for each sample, while data for panels **d-e** are from $n = 3$ (His-TEV) or $n = 5$ (HERD-4.1) biologically independent experiments. Statistical testing was performed using one-way ANOVA and TukeyHSD post-hoc test with **e**: His-TEV vs HERD-4.1 $P = 0.001$ (***)

Compared to TC-HERD-2.2-FMO, the re-designed TC-HERD-4.1-FMO was now enriched approximately 3.5 times more in the HERD-2.2-GFP droplets, showing a clear and significant increase in co-condensation of this enzyme (Fig. 4.12b-c). In addition, HERD-4.1-FMO was confirmed to colocalise with HERD-2.2-TnaA in *E. coli*. HERD-4.1-mCherry-FMO was constructed and co-expressed with HERD-2.2-GFP-TnaA. Again, and as for HERD-2.2-mCherry-FMO, this produced condensates enriched in both fluorescent proteins, demonstrating that both enzymes were still co-localised in the condensates consistent with the re-design (Fig. 4.12d). Measurement of improved enrichment within these condensates was prohibited however, due to changes in overall expression levels between the two designs.

Finally, with the re-designed HERD-4.1 in hand, the effect of co-condensation of this enzymatic cascade within HERD-2.2 condensates was measured. As described previously, HERD-2.2-TnaA and HERD-4.1-FMO were placed together under the control of the arabinose promoter on a single polycistronic vector, and co-expressed with HERD-2.2-GFP under the control of the T7 promoter. When grown at 33 °C, these cells produced the most indigo of all the combinations measured: more than 2.5-times more than the free enzymes, which, after normalisation, translated to an overall more than 6-fold improvement over the free enzymes, partially owing to slightly weaker expression of HERD-4.1-FMO (Fig. 4.13). These data confirm that rational sequence re-design using simple concepts such as protein net charge, can have powerful effects on pathway efficiency.

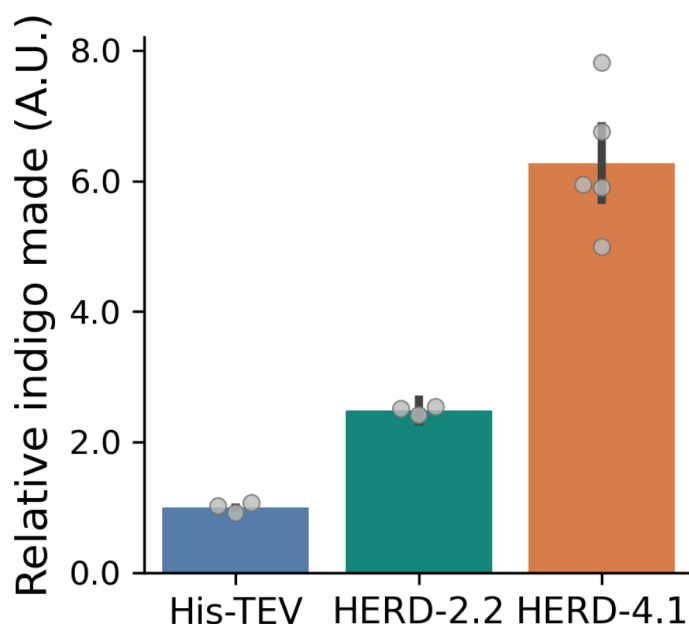


Figure 4.14: *De novo* condensates can accelerate enzymology. Relative amount of indigo made by cells using free enzymes (His-TEV), enzymes condensed using HERD-2.2 based *de novo* condensates, or enzymes condensed using HERD-4.1 based condensates.

4.8 Chapter summary

As the design and engineering of protein condensates has become accessible to synthetic biologists, there has been increased focus on exploring their potential for function within living cells.¹⁷⁹ The compartmentalisation or sequestration of proteins has been demonstrated robustly by several different studies, and now the focus is on how biomolecular condensates can alter the physical and chemical properties of the encapsulated macromolecules.^{97,200,292} This chapter has described the functionalisation of the designed HERD-2.2–GFP biomolecular condensates. The inert condensates have been charged both *in vitro* and in cells with functional enzymes, creating active MLOs. Further, the condensates in *E. coli* were co-enriched with the two enzymes TnaA and FMO in the multi-step pathway for the production of indigo dye from L-tryptophan. These *de novo* condensates ultimately offer significant improvements on indigo production over the free enzymes (Fig. 4.14).

Through a series of experiments, the properties of the HERD-2.2–GFP condensates have been directly linked to pathway efficiency. These are concepts that have been considered elsewhere, but are challenging to directly assess in living cells. Here, due to the rational design approach used, and the temperature sensitivity of the HERD-2.2–GFP design, the material properties of the condensates can be manipulated directly. MLOs formed under the optimal conditions for liquid-like behaviour, giving fast FRAP and dynamic droplets, offer improvements on pathway efficiency and indigo formation (Figs. 4.9 & 4.10). However, condensates formed under less-favourable conditions such as at lower temperatures, giving slow or limited FRAP, instead likely render enzymes inaccessible to substrate and significantly reduce pathway efficiency (Fig. 4.11). This chapter has also confirmed the design principle of creating increasingly weak PPIs. The initial HERD-0–GFP scaffold, although able to form condensates at lower concentrations and higher temperatures than HERD-2.2–GFP, is flawed as an MLO for enzymology, almost entirely inhibiting catalysis of the enzymes localised to it (Fig. 4.9).

These results confirm that condensates can vary widely in their material properties even through small changes to their environment. In the functionalisation of MLOs, it is critical that the condensates chosen have the correct material properties to condense client and scaffold proteins without restricting their functionality.^{293,294} This process can be challenging, as changes to the scaffold to introduce additional PPI motifs or change its subcellular localisation can also change the overall interactions of the protein and, therefore, its propensity to phase separate as a de-mixed liquid. It is likely that assessment of condensate dynamics by FRAP, or more sensitive techniques such as fluorescence correlation spectroscopy (FCS) or single molecule measurements, will remain an essential part of the MLO design-build-test toolkit.^{268,269}

Designer phase-separating scaffolds may be the solution to these challenges. By designing the scaffold from the bottom-up, its interactions and architecture can be modelled to suit the function from the outset.¹⁹⁹ Here, the potential for the HERD scaffold to be re-designed to tailor it towards a specific application has been assessed in a proof-of-concept study. Re-balancing the overall net charge of the designer polypeptide was demonstrated to have a significant impact on co-condensation of the rate limiting enzyme, and as a result, improved the fold change in indigo production from a 2.3-fold increase to more than a 6-fold increase over the free enzymes (Figs. 4.12 & 4.13). This not only demonstrates that simple principles can still guide protein design of MLOs, but also that the HERD scaffold is amenable to changes and tuning to the desired function.

Overall, these data corroborate the concept that LLPS can enhance the rate of product formation within a biochemical pathway. This process is generally understood to be largely due to enrichment of the enzymes within the same cellular location, giving improved substrate and intermediate transfer between enzymes. However, there are additional contributions that could be at play, for instance the unique behaviour of proteins at the de-mixed droplet interface. These physical effects are not directly considered here, however it is noteworthy that the overall 6-fold enhancement in indigo formation matches the theoretical value for colocalisation of a two-enzyme pathway.²⁹⁵ This hints that much greater enhancements on product formation could be made by co-condensation of a three-enzyme pathway, with a much greater 110-fold theoretical maximum improvement.²⁹⁵ In addition, future work could examine the effect of condensation on enzyme kinetics. Here, product formation was considered as an end-point value, due to the relatively slow rate and challenges associated with assessing product formation in cells. However, accurate measurement of substrate turnover, for instance by monitoring NADPH depletion *in vitro*, could enable accurate measurement of Michaelis-Menten enzyme kinetics in single-phase and two-phase systems.

Chapter 5

Investigation into fibre formation by HERD proteins

5.1 Chapter introduction

While LLPS is being explored by synthetic biologists for its potential to augment biology, the same processes are being interrogated clinically for their potential roles in disease and degenerative processes.^{45,46,51,87} LLPS is characterised by weak and dynamic interactions, forming a de-mixed droplet where the macromolecules are enriched, but still exchange and diffuse quickly. However, the same interactions that stimulate self-assembly as inert droplets can also drive the formation of other ordered assemblies that act pathologically.

This process is due to the inherent metastability of LLPS. Protein phase separation is not a stable state, instead lower-energy states are accessible under certain conditions (Fig. 5.1).³⁶ Further, unlike LLPS, these lower energy states are irreversible processes, so that once formed they are permanently arrested and cannot return to form the dynamic liquid-like droplets.^{296,297} This phenomenon of undergoing a second phase transition from LLPS has been termed the maturation or molecular aging of protein condensates.⁴⁴

In some cases, droplet maturation is an essential part of the biological role of a protein. For elastin, droplets are the nucleation point for the formation of extracellular filaments that are an essential part of the extracellular matrix.^{255,256} However, droplet maturation has also been implicated in the formation of several pathological assemblies. The ALS-associated protein FUS forms liquid droplets under a number of conditions, but over time these mature into amyloid fibres characteristic of neurodegeneration.⁶⁰ Further, mutations linked to ALS accelerate droplet maturation, suggesting a link between aberrant phase separation, fibre formation, and neurodegeneration.^{65,73,74} This formation of pathological assemblies by droplet maturation is a recurring theme, with many of the classic

examples of phase-separating proteins (FUS, TDP-42, Tau, hnRNPA1) undergoing similar secondary transitions to amyloid-like states.^{298–300} Despite its clinical significance, the mechanisms by which droplet maturation occurs are still relatively poorly understood, and developing this understanding of how phase separated droplets can mature could enable routes to reduce the accumulation of amyloidogenic structures in neurons.

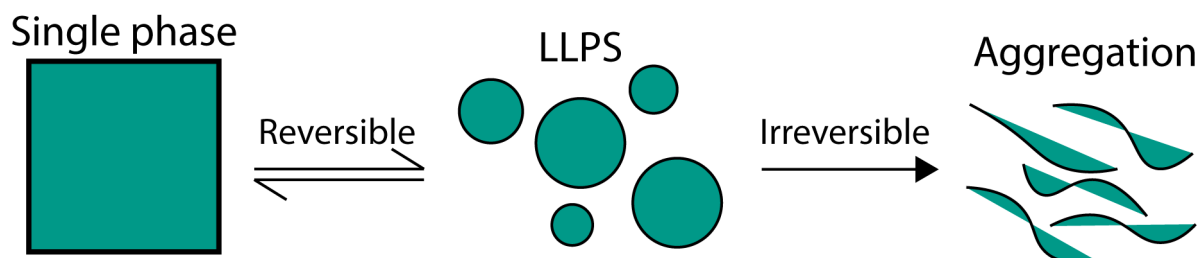


Figure 5.1: Maturation of LLPS is due to its metastability. LLPS is a reversible process, but in proteins it is metastable to other irreversible phases, such as aggregation or fibre formation.

This chapter describes the formation of supramolecular fibres by the designed HERD-2.2 protein. Fibre formation is characterised both *in vitro* and in *E. coli*, and the potential for their use as supramolecular scaffolds for protein recruitment is discussed. Finally, the interplay between fibre formation and LLPS is discussed in this *de novo* system, and the potential to switch between the two states.

5.2 HERD-2.2 can form fibres as well as drive LLPS

While the HERD-2.2 polypeptide had been designed to drive LLPS in *E. coli*, it was discovered that it is also capable of assembling into ordered fibres. As described in chapters 3 and 4, the HERD series of proteins were designed as minimal polypeptides for easy fusion to client proteins for LLPS. This enables the rapid portability of the designed sequence to different use cases. For this purpose, the design and characterisation of the HERD proteins used a monomeric fluorescent protein (mEmerald) as a model client protein fusion, and to enable detection of cellular localisation of the recombinant protein. However, to fully assess the contribution of all the elements of the design, a final HERD-2.2 polypeptide was expressed and purified without the *C*-terminal mEmerald for characterisation *in vitro*. Without the attached client protein, *in vitro* HERD-2.2 did not form de-mixed liquid droplets. Instead, the formation of protein fibres several microns in length were evident (Fig. 5.2).

This switch in phase behaviour was puzzling, as the fluorescent client protein was not implicated in participating in PPIs, and mEmerald has been characterised as behaving monomerically.³⁰¹ Instead, it is possible that steric bulk from the client protein hinders

full saturation of the helical regions in the HERD protein, and therefore prevents fibre formation, which would likely require tight, regular packing of many polypeptide chains. GFP has been used before as a solubility tag for attractive protein domains, and it may be that a similar mechanism is the cause of the change in phase behaviour between HERD-2.2 and HERD-2.2-GFP.³⁰²

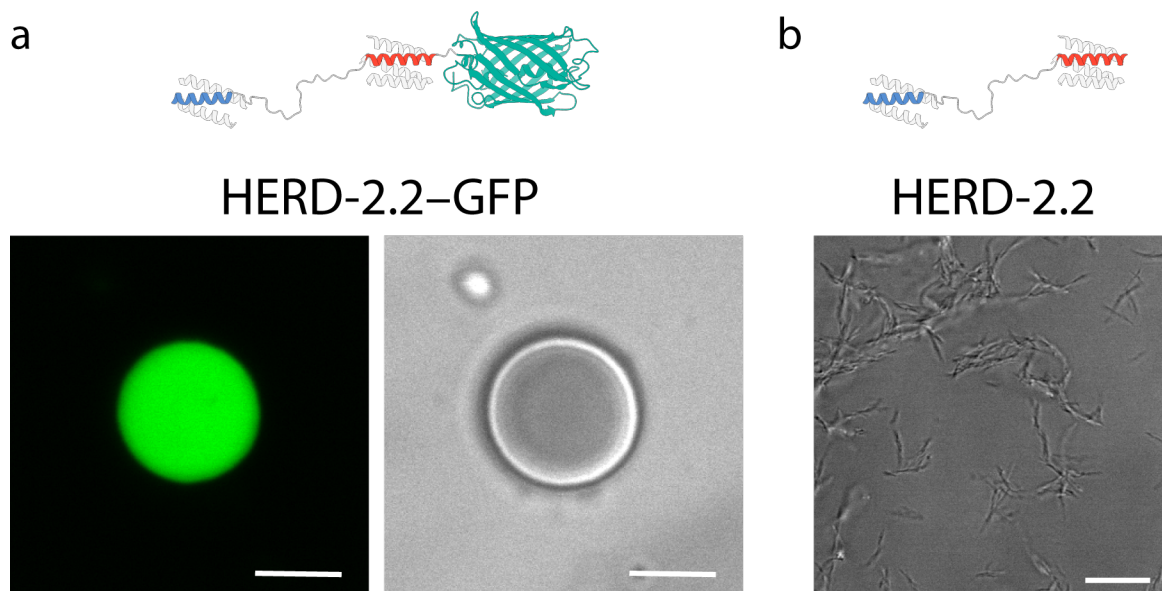


Figure 5.2: HERD-2.2 forms micron length fibres *in vitro*. a, Fluorescence and brightfield confocal microscopy images of HERD-2.2-GFP *in vitro*. Conditions: 1 mM HERD-2.2-GFP, 125 mM NaCl, 4% PEG 3350, 20 mM Tris pH 7.5. Scale bar 5 μ m. b, Brightfield confocal microscopy image of HERD-2.2 fibres *in vitro*. Conditions: 500 μ M HERD-2.2, 125 mM NaCl, 10% PEG 3350, 20 mM Tris pH 7.5. Scale bar 10 μ m

5.2.1 HERD-2.2 forms protein fibres 13 nm in diameter

To explore the molecular structure of the HERD-2.2 fibres, the purified protein solution was examined by transmission electron microscopy (TEM) with Ms Judith Mantell. This revealed the formation of individual protein fibres in solution, with diameters of approximately 13 nm (Fig. 5.3a–b). In buffers without additional molecular crowding reagents or added salt, these fibres were arranged randomly on the TEM grid, forming a dense mat of fibres. However, when PEG and NaCl were added at pH 6 (10% PEG 3350, 125 mM NaCl, 50 mM Bis-Tris pH 6), the formation of even larger assemblies was visible, which were detectable by confocal microscopy (Fig. 5.3c). Measurement by TEM showed that with additional molecular crowding the individual fibrils assembled laterally to form thicker fibres, that extended over several microns. Examination of these laterally assembled fibres confirmed that they were made up of individual fibrils. However, there was insufficient resolution to detect spacing within the individual fibrils that might relate to the molecular structure of the fibres themselves.

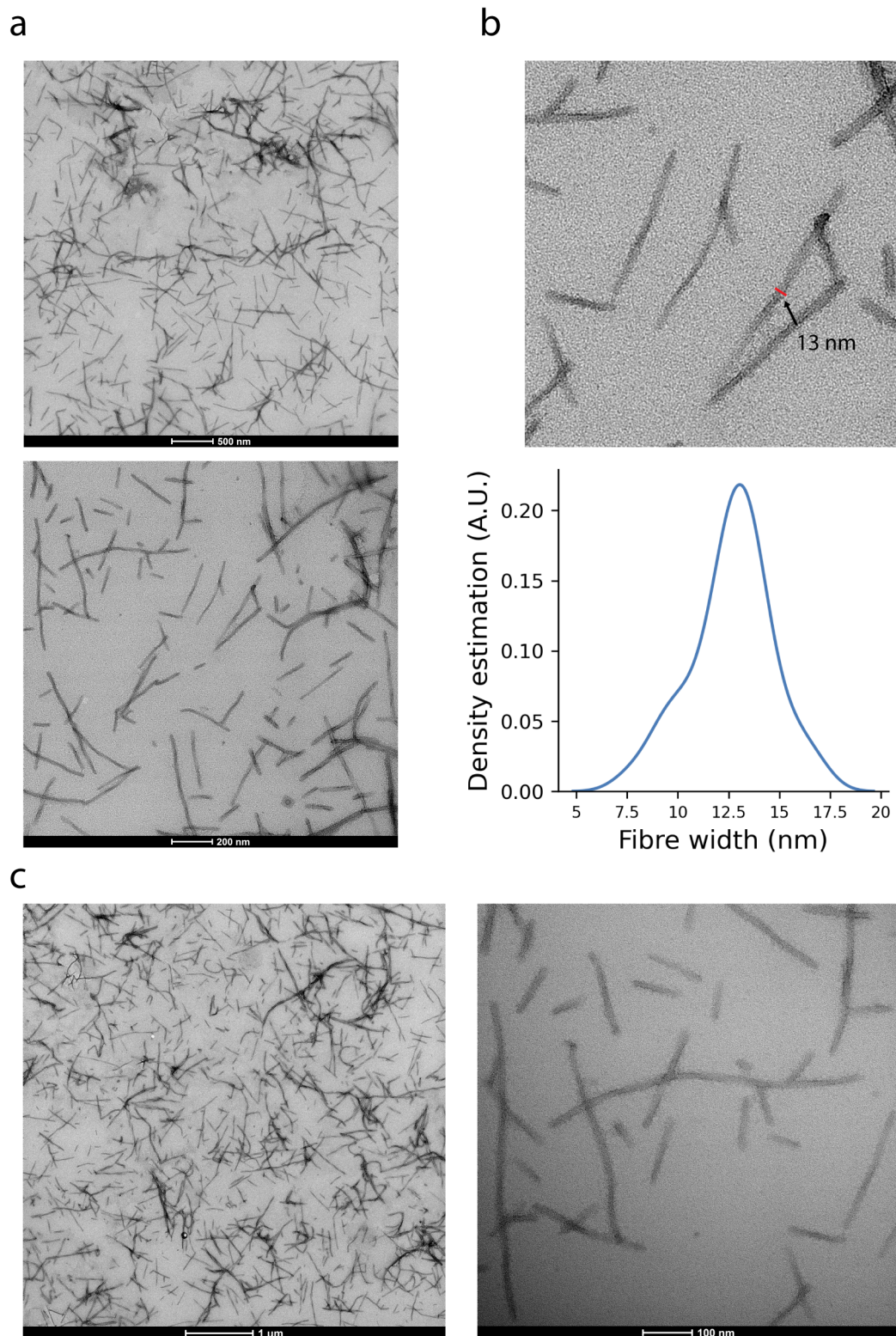


Figure 5.3: HERD-2.2 forms fibres 13 nm in diameter. a, TEM images of HERD-2.2 fibres *in vitro*. Conditions: 25 μM HERD-2.2, 20 mM Tris pH 7.5. b, Measurement of HERD-2.2 fibre diameter *in vitro*. The red line shows the measurement area. Fibre diameter is plotted using a kernel density estimation from measurements of $n=43$ fibres. c, TEM images of HERD-2.2 fibres with the addition of molecular crowding reagents. Conditions: 25 μM HERD-2.2, 10% PEG 3350, 125 mM NaCl, 50 mM Bis-Tris pH 6.

The slightly altered pH that was required for self-assembly of HERD-2.2 (pH 6) compared to HERD-2.2-GFP (pH 7.5) was curious, as charged interactions were not expected to play a role in assembly. However, it was posited that the *C*-terminal GFP could be altering the overall pI of the protein slightly, as mEmerald has a pI of 5.85. To adjust the self-assembly conditions of HERD-2.2, a single glutamate residue was added to the *C*-terminus, to increase the overall net-charge of HERD-2.2 and bring it closer in line with that of HERD-2.2-GFP (Table 8.3). This construct, named HERD-2.2-E, also formed fibres similarly to HERD-2.2, but now at the same pH as HERD-2.2-GFP (Fig. 5.4). Both of these proteins were considered for further investigation.

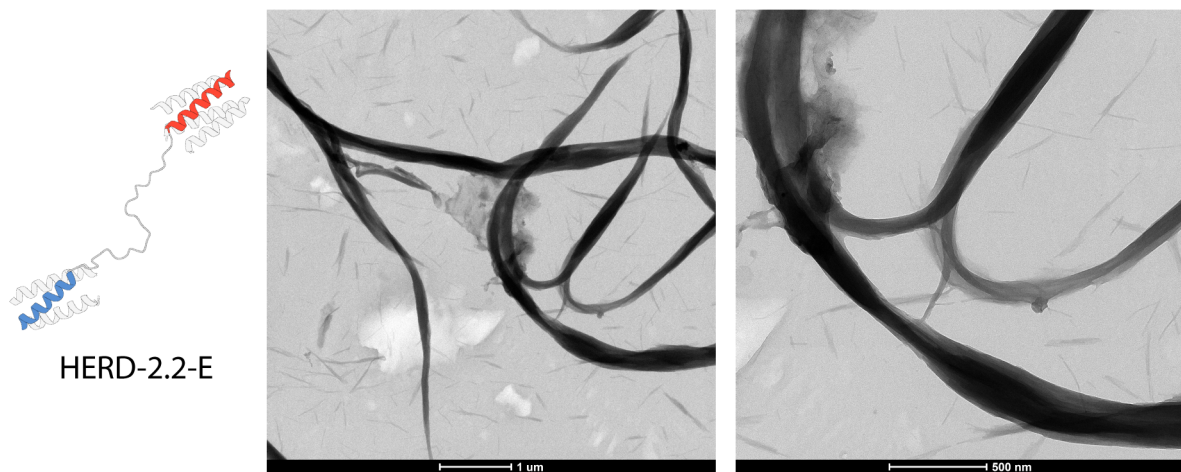


Figure 5.4: HERD-2.2-E forms fibres at a shifted pH. TEM images of the HERD-2.2-E fibres *in vitro*. Conditions: 50 μ M HERD-2.2-E, 10% PEG 3350, 125 mM NaCl, 20 mM Tris pH 7.5.

5.2.2 HERD-2.2-E forms fibres within the *E. coli* cytoplasm

Having determined that HERD-2.2 and HERD-2.2-E could form fibres *in vitro*, it was tested whether similar behaviour could occur in cells. As HERD-2.2 was not fluorescently labelled, its subcellular localisation could not be assessed by fluorescence microscopy. Instead, the formation of fibres was examined by TEM with Dr Lorna Hodgson. *E. coli* expressing HERD-2.2-E were grown and vitrified by high-pressure freezing, and embedded in resin for sectioning by ultramicrotomy so that thin layers of the block could be imaged by TEM. TEM on sections of *E. coli* expressing HERD-2.2-E showed the formation of fibres within the cytoplasm (Fig. 5.5). These fibres appeared to be assembled laterally, similarly to the fibres measured *in vitro* with simulated molecular crowding reagents. The formation of such long fibres was remarkable, and it was surprising that they were not toxic to the *E. coli* by physically puncturing the plasma membrane. Instead, the fibres were able to bend, potentially with elastic properties, as cells grown for an extended period of time (21 hours after induction of protein expression), developed fibres that

grew from one cell pole, curved around the cell periphery to the other cell pole, and then curved back towards the original cell pole, showing a remarkable capacity for flexibility. Cells expressing HERD-2.2-GFP did not form fibres, and instead dense regions of protein were found at the cell poles where the protein condensates localised. Cells expressing no recombinant proteins showed no areas of enrichment or intracellular fibres.

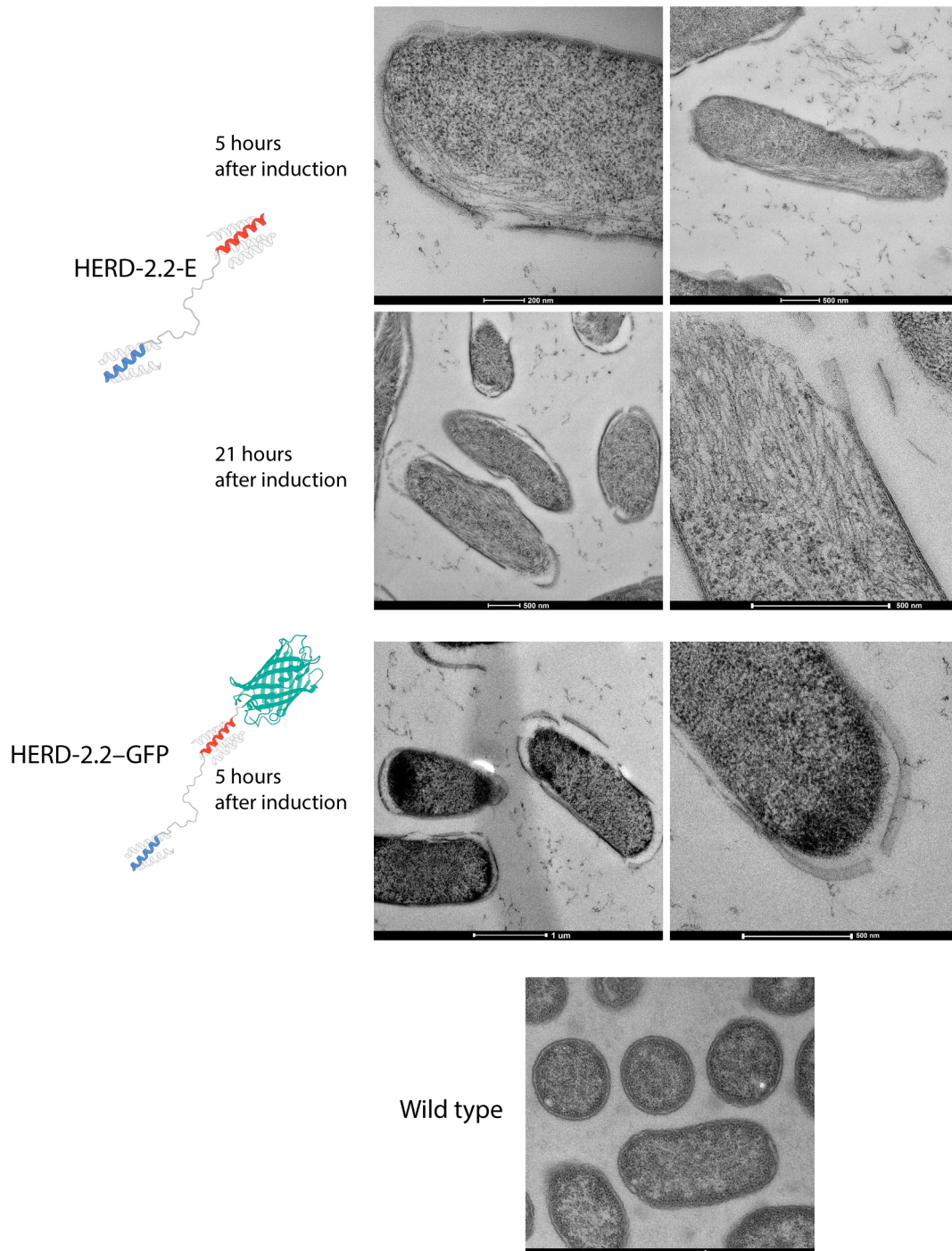


Figure 5.5: HERD-2.2-E forms fibres *in vitro* and in cells. TEM images of the expressed HERD-2.2-E and HERD-2.2-GFP in *E. coli*. Cells expressing HERD-2.2-E were grown for 5 and 21 hours after induction, and before high pressure freezing and staining, and cells expressing HERD-2.2-GFP were grown for 5 hours. *E. coli* were imaged by Dr Lorna Hodgson.

5.3 HERD-2.2–GFP can be recruited to HERD fibres

The formation of such large fibres within the cramped *E. coli* cytoplasm was remarkable, and reminiscent of engineered systems for subcellular recruitment using fibrous structures.¹⁵² Therefore, the capacity for HERD-2.2-E fibres to act as recruitment scaffolds was assessed. To test recruitment of proteins to HERD-2.2 fibres in cells, HERD-2.2-E and HERD-2.2–GFP were co-expressed together within the same cells. This was also a test of which phase behaviour would predominate when the two proteins were expressed simultaneously. It was hypothesised that at high enough HERD-2.2–GFP concentrations, fibre formation may be disrupted and spherical protein condensates might predominate. However, at lower HERD-2.2–GFP concentrations fibre formation may be possible with some incorporation of the fluorescent protein – indeed, co-expression of the two proteins indicated that the latter was the case in this experiment. Confocal fluorescence microscopy of *E. coli* co-expressing HERD-2.2-E and HERD-2.2–GFP showed the formation of long fluorescently labelled fibres that ran the entire length of the cells (Fig. 5.6a).

Remarkably, and as they were now visible by fluorescence microscopy, the full scale of these fibres could be visualised. TEM of sections from these cells demonstrated that fibres were indeed formed by co-expression of both proteins (Fig. 5.6b). These fibres appeared more dense than those formed by expression of HERD-2.2-E alone, possibly by the inclusion of the bulky fluorescent protein. However, they were still arranged laterally throughout the cell. To determine conclusively that the fibres were labelled with the fluorescent HERD-2.2–GFP, correlative light-electron microscopy (CLEM) was performed with Dr Lorna Hodgson.³⁰³ Here, cells were embedded in resin under milder conditions to retain the fluorescence from GFP, so that fluorescence confocal microscopy could be used to identify regions of fluorescence, followed by TEM to characterise the molecular structure at nanometre resolution.¹⁵² Indeed, CLEM of *E. coli* co-expressing HERD-2.2 and HERD-2.2–GFP demonstrated that the green fluorescence overlaid clearly with the subcellular fibres visible by TEM, showing specific colocalisation of the two proteins (Fig. 5.6c).

5.3.1 Cells expressing HERD-2.2 fibres become elongated

One of the most striking aspects of the *E. coli* co-expressing HERD-2.2 and HERD-2.2–GFP was their length. It was evident that expression of these fibres had caused the cells to grow much longer than the natural length expected for *E. coli* (typically 1 – 2 μm). Indeed, measurement of *E. coli* expressing HERD-2.2 alone indicated that they grew to a mean length of $7.4 \pm 1.8 \mu\text{m}$ (Fig. 5.7).

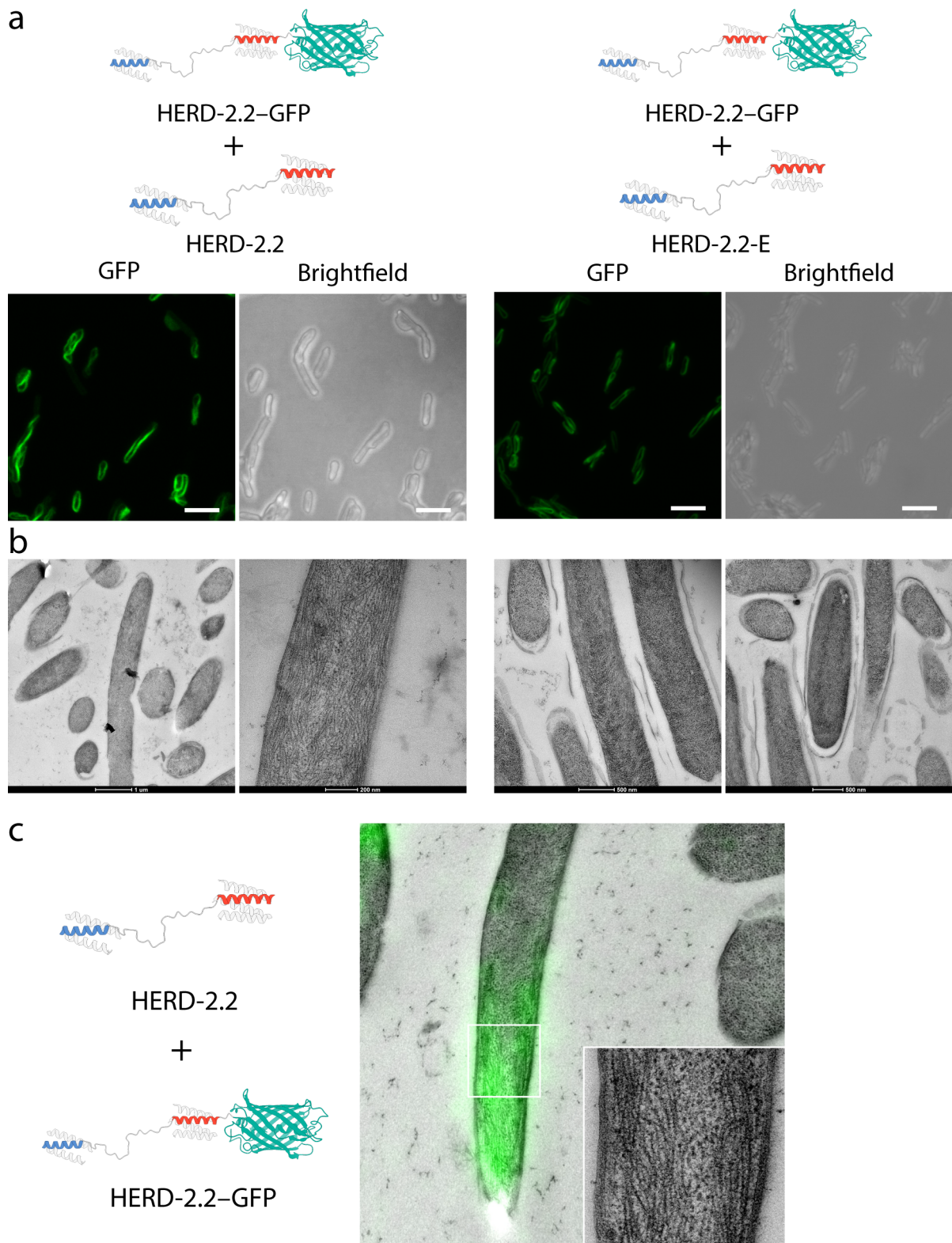


Figure 5.6: HERD-2.2-GFP can be recruited to HERD fibres in cells. a-b, Confocal fluorescence microscopy (a) and TEM (b) images of the co-expressed HERD-2.2 and HERD-2.2-GFP, and HERD-2.2-E and HERD-2.2-GFP in *E. coli*. Scale bars for confocal microscopy images are 5 μm . c, CLEM overlay of fluorescence confocal microscopy and TEM of *E. coli* co-expressing HERD-2.2 and HERD-2.2-GFP. Overlay was created by Dr. Lorna Hodgson.

Moreover, when HERD-2.2 and HERD-2.2-GFP were co-expressed together, the elongated cell length became even more dramatic, with the mean length increasing to $12.8 \pm 4.5 \mu\text{m}$. This increase in cell length was determined to be unique to the expression of HERD-2.2, with cells expressing HERD-2.2-GFP alone growing to a mean length of only $2.0 \pm 0.6 \mu\text{m}$, only slightly shorter than cells transformed with an empty vector (wild type – $3.1 \pm 0.7 \mu\text{m}$). The reason for this super elongated phenotype is unclear, however it is likely due to a significant defect in cell division due to the formation of HERD-2.2 fibres. Indeed, it is possible that fibre formation prevents cell septation and division, as has been characterised in strains deficient in aspects of the cell division machinery.³⁰⁴

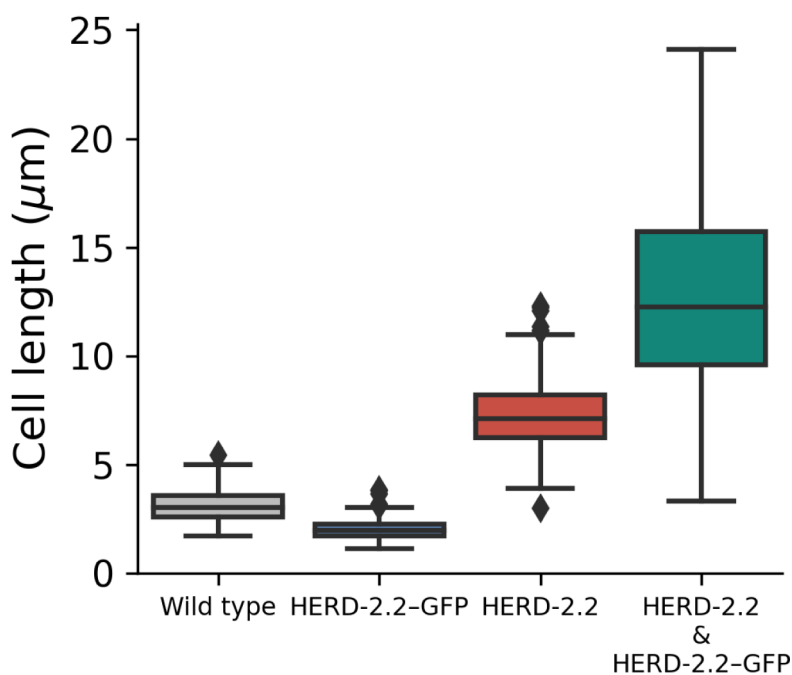


Figure 5.7: Cells expressing HERD-2.2 fibres are elongated. Quantification of cell length from wild type cells (n=106), and cells expressing HERD-2.2-GFP (n=100), HERD-2.2 (n=107), and HERD-2.2-GFP and HERD-2.2 co-expressed (n=100).

5.4 CD spectroscopy of HERD-2.2 fibres

Having observed the phenotype associated with HERD-2.2 fibres in *E. coli*, the structural basis of their assembly was examined. It was considered that these could be α -helical fibres, because of the helical regions that make up the stickers in the designed HERD proteins. Although less prevalent than β fibrils, α -helical fibres have been designed and robustly characterised previously. These can self-assemble by homotypic or heterotypic interactions, creating long individual fibrils or branched hydrogels.^{169–171,305} To attempt to determine their secondary structure, the HERD-2.2 fibres were measured by CD spectroscopy. To reduce the effect of scattering on the measurements due to the fibres in

solution, only 25 μM protein was used, and no NaCl or molecular-crowding reagents were introduced, so that the fibres were evenly distributed in solution without assembling laterally (Fig. 5.8a). HERD-2.2 fibres displayed some, relatively weak, secondary structure by CD spectroscopy, characterised by the weak minimum at around 222 nm (Fig. 5.8b).

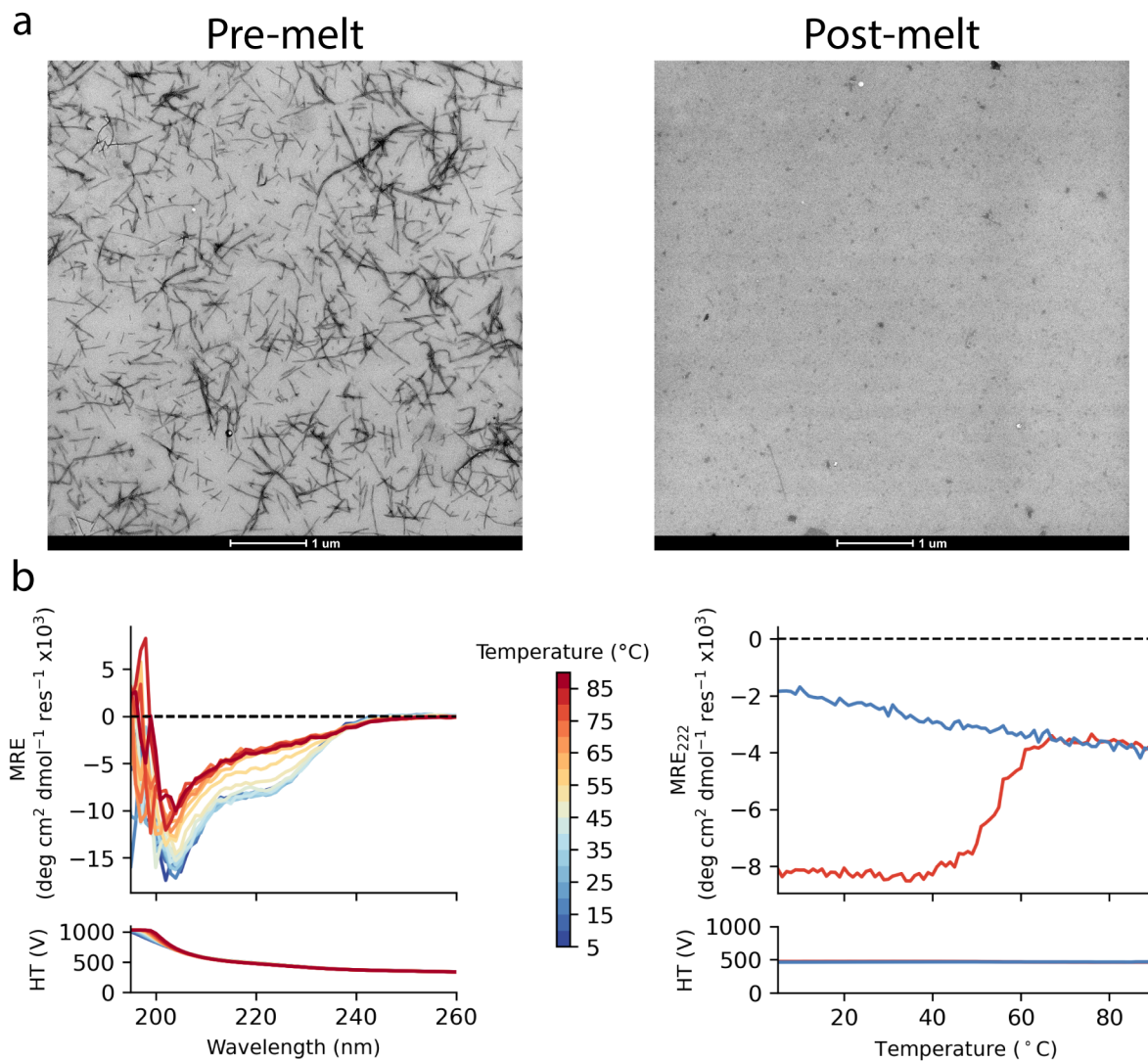


Figure 5.8: CD spectroscopy of HERD-2.2 fibres indicates secondary structure. **a**, TEM images of HERD-2.2 fibres before melting by heating to 90 °C and after melting and returning to 5 °C. **b**, Left: CD spectra recorded every 5 °C as the temperature was increased from 5 °C to 90 °C. Right: MRE at 222 nm during melting (5 °C – 90 °C) and cooling (90 °C – 5 °C). Conditions: 25 μM HERD-2.2, 50 mM Tris pH 7.5.

The weak secondary structure may be due to measurement of insoluble fibres. Overall, it was not possible to conclude what secondary structure was present in the fibres. Heating of the solution indicated that secondary structure was lost upon heating, measured by an increase in MRE at 222 nm. TEM following melting confirmed that the loss of secondary structure was accompanied by a loss of fibres in solution. However, cooling of the denatured solution did not induce re-folding, and fibres were not re-formed.

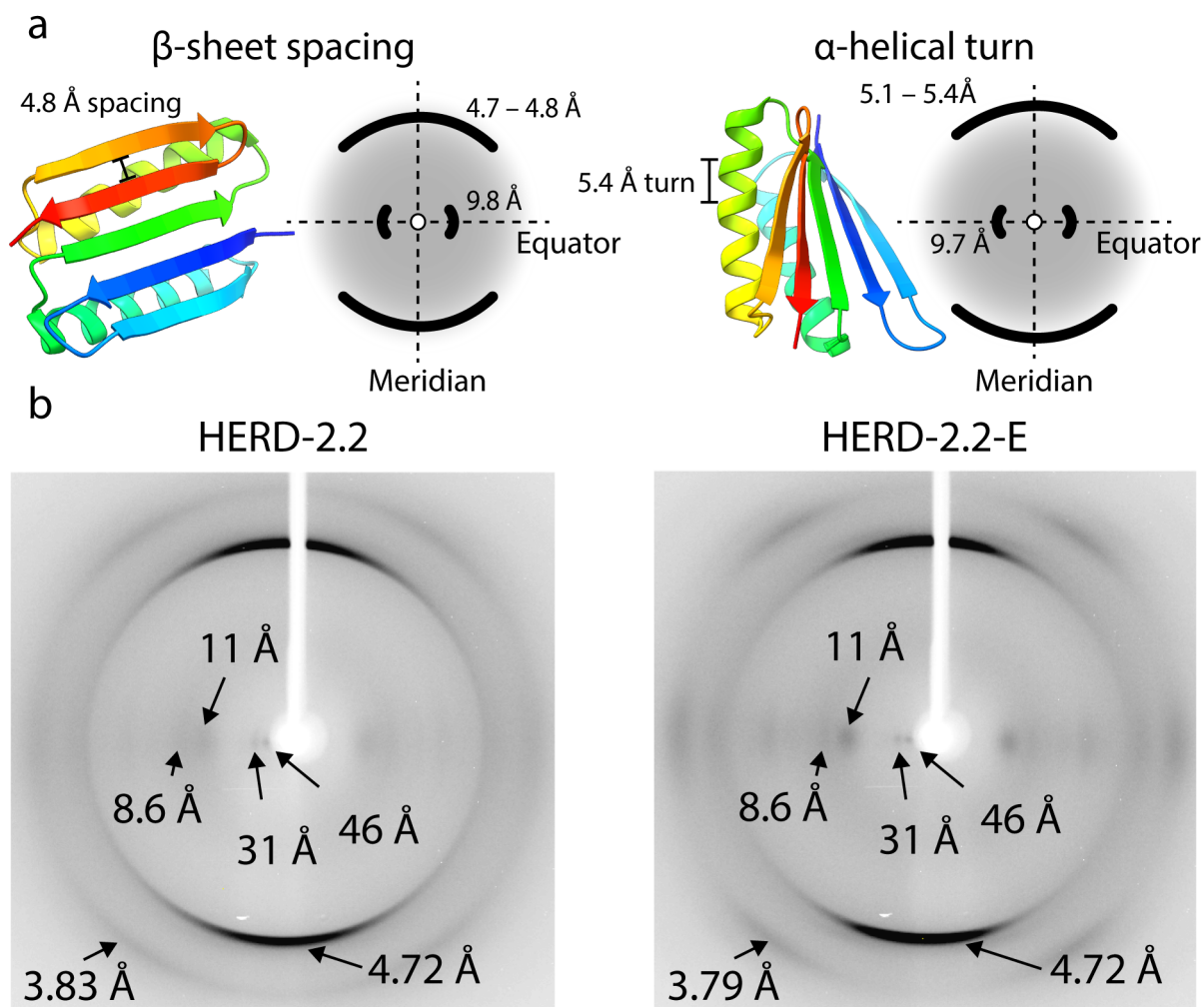


Figure 5.9: X-ray fibre diffraction of HERD-2.2 fibres. **a**, Cartoons of simple diffraction patterns from β and α fibres, with the characteristic spacing of both. Additional diffraction peaks at 9.8 and 9.7 Å correspond to the inter-sheet and inter-helix distances respectively. **b**, X-ray fibre diffraction patterns of HERD-2.2 and HERD-2.2-E, with annotated diffraction peaks. Diffraction data were collected and analysed with the help of Prof. Louise Serpell.

5.4.1 X-ray fibre diffraction indicates β secondary structure

To build a greater idea of the secondary structure within the HERD fibres, they were analysed by X-ray fibre diffraction. Solutions of HERD-2.2 and HERD-2.2-E were prepared (16 mg/ml protein, 20 mM Tris pH 7.5), and suspended between glass capillaries before drying overnight, to produce dry, elongated fibres of the purified proteins. Diffraction of both HERD-2.2 and HERD-2.2-E fibres produced similar diffraction patterns, as expected from their near-identical sequences. The critical diffraction peaks are along the meridian of the diffraction pattern. The spacing of these reflections is indicative of either a cross- β architecture, with a characteristic spacing of 4.7 Å from the distance between strands in a β sheet, or α -helical secondary structure, with a slightly larger spacing of 5.1 Å from the α -helical turns (Fig. 5.9a).^{306–308} Measurement of the diffraction patterns

of the HERD-2.2 and HERD-2.2-E fibres indicated primary meridional peaks at 4.72 Å, indicating β secondary structure (Fig. 5.9b). Additional diffraction peaks at longer distances that would correspond to the inter-sheet or inter-helical distance were faintly discernible but did not distinguish additional structural information.

This structure is at odds with the design hypothesis that these sequences were constructed from α -helical elements. Indeed, the peptides have been demonstrated to fold as α -helical in solution with added TFE (Fig. 3.15). Therefore, it is unclear what is driving the assembly of β secondary structure in these proteins, however it could be interesting to investigate in the context of LLPS and amyloidogenesis.

5.5 HERD-2.2-E fibres can pattern LLPS droplets

Ultimately, the HERD-2.2 fibres were interesting because of the potential interplay between LLPS and fibre formation, which has been demonstrated to be critical in a number of natural systems. Therefore, the association between de-mixed liquid droplets and structured fibres was considered. While in cells the co-expression of the two proteins led to the patterning of HERD-2.2 fibres with GFP, *in vitro* their association could be more nuanced, due to the less restricted volume and lower molecular crowding. For *in vitro* investigation the altered HERD-2.2-E protein was used, as it formed fibres under more similar conditions to LLPS of HERD-2.2–GFP.

Initially, the two purified proteins HERD-2.2 and HERD-2.2–GFP were mixed *in vitro*, and phase separation was induced by addition of 10% PEG and 125 mM NaCl. Independently, this would result in the formation of fibres and de-mixed droplets. When now mixed and phase separated together, green labelled fibres were formed, similarly to those observed in cells (Fig. 5.10a). However, now, in addition to labelling the length of the fibre, some de-mixed droplets remained, visible by fluorescent puncta that patterned the surface of the fibres. Examination of these fibres by TEM confirmed the formation of both fibres and de-mixed droplets within the solution (Fig. 5.10b). These interactions appear to indicate some interplay between the phase separated droplets and fibres, however the precise manner of these interactions was not yet clear.

5.6 Fibres are not nucleated from HERD-2.2–GFP droplets

Because of the interplay between LLPS and fibre formation observed by the HERD-2.2 design, it was considered whether this could act as a model for droplet maturation.³⁰⁹ For this, the capacity for HERD-2.2–GFP droplets to nucleate the formation of HERD-

2.2 fibres was assessed. As the difference between these constructs was the addition or removal of the *C*-terminal fluorescent protein, it was reasoned that cleavage of the fluorescent protein from HERD-2.2–GFP could lead to the formation of fibres *in vitro*. Further, cleavage from within de-mixed liquid droplets could allow monitoring of fibre nucleation from within a large, easily observable phase separated environment, creating an accessible model for fibre nucleation by LLPS. To induce protease digestion of HERD-2.2–GFP, the TEV protease digestion site was moved from the *N*-terminus (where it formed part of the His-TEV tag), to the linker between HERD-2.2 and mEmerald. This would allow addition of TEV protease for cleavage of the *C*-terminal fluorescent protein.

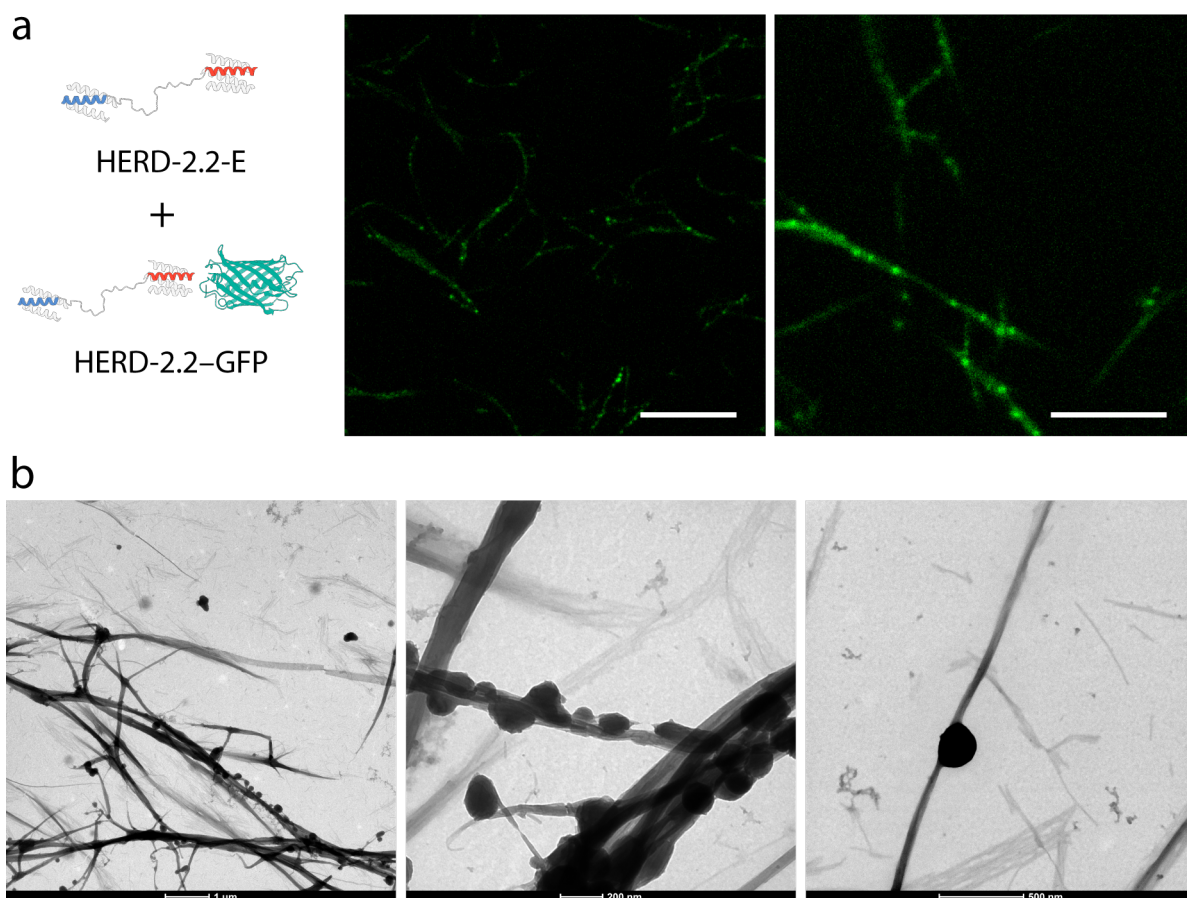


Figure 5.10: HERD-2.2-E fibres can pattern LLPS droplets. a, Confocal microscopy images of the mixed HERD-2.2-E and HERD-2.2–GFP proteins *in vitro*. Scale bars are 10 μm (left) and 5 μm (right). b, TEM images of the mixed HERD-2.2-E and HERD-2.2–GFP proteins. Conditions for both panels: 50 μM HERD-2.2-E, 150 μM HERD-2.2–GFP, 10% PEG 3350, 125 mM NaCl, 20 mM Tris pH 7.5

For this, both the original HERD-2.2 design and the additional HERD-2.2-E design were tested. Expression and purification of these constructs, named HERD-2.2-TEV–GFP and HERD-2.2-E-TEV–GFP respectively, indicated that they both still formed de-mixed droplets under the same conditions as HERD-2.2–GFP, demonstrating that phase separation had not been perturbed (Fig. 5.11a). However, despite efficient digestion of the

protein by TEV protease, forming two independent polypeptides, there was no evidence of fibre formation by either of the individual HERD proteins (Fig. 5.11b). This may be because the protein concentrations used were insufficient for fibre formation, or that the conditions did not allow for self-assembly. Indeed the assembly process of HERD-2.2 is still unclear, with the irreversible melting by CD spectroscopy indicating that there may be a concentration dependence. Future work could investigate how fibres are formed in this system, and this may lead to real-time observation of fibre nucleation by LLPS.

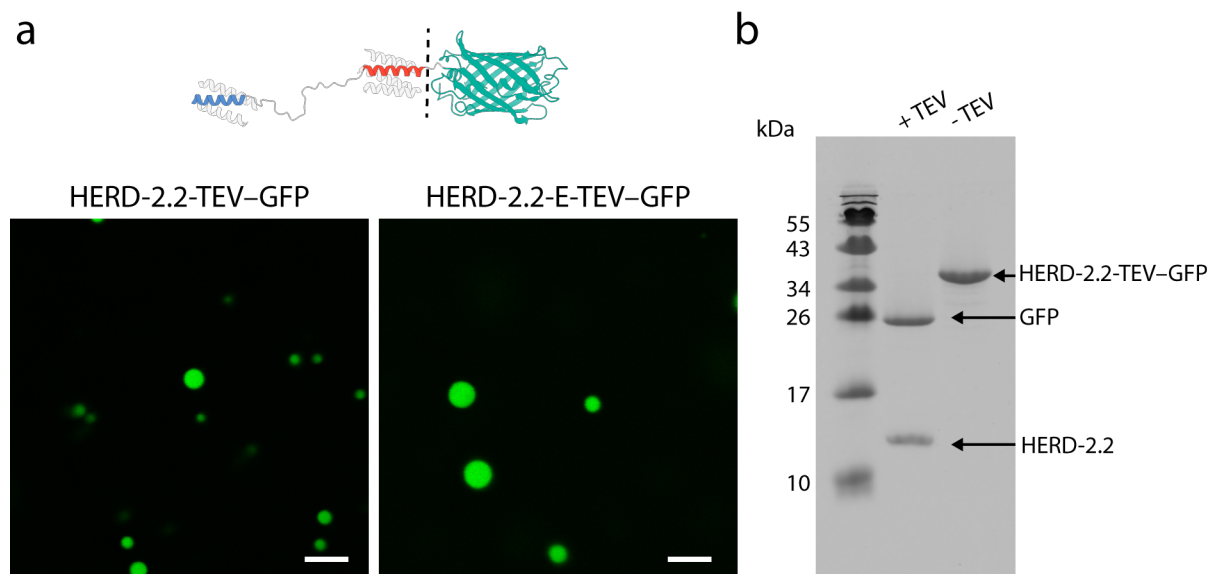


Figure 5.11: TEV cleavage of GFP does not induce fibre formation. a, Confocal microscopy images of HERD-2.2-TEV-GFP and HERD-2.2-E-TEV-GFP *in vitro*. Scale bars are 5 μm . Conditions: 1 mM HERD protein, 4 % PEG 3350, 125 mM NaCl, 20 mM Tris pH 7.5. b, SDS-PAGE stained with Coomassie total protein stain on HERD-2.2-TEV-GFP solution with added TEV protease (+TEV) and solution without added TEV protease (-TEV).

5.7 Chapter summary

Many proteins that undergo LLPS have also been identified to form other proteinaceous materials. These assemblies vary greatly in their structures and functions, but have been particularly implicated in disease because of the potential for pathogenic amyloidogenesis or aggregation.^{35,46,74} This chapter has described the capacity for the designed HERD-2.2 polypeptide to form macromolecular fibres. These fibres have been characterised to assemble both *in vitro* and in *E. coli*. The resulting structures have significant effects on cell physiology, creating elongated rod-shaped bacteria far longer than wild type *E. coli* (Fig. 5.7). Moreover, these fibres can be patterned with fluorescently labelled HERD-2.2, acting as a rigid scaffold for client recruitment.

Further, the molecular architecture of the fibres was investigated by CD spectroscopy and

X-ray fibre diffraction. The structure of these fibres was determined to be different to that of the designed HERD-2.2-GFP. As indicated by X-ray fibre diffraction, the fibres formed by HERD-2.2 have β secondary structure components (Fig. 5.9). There are indications that the peptides used in HERD-2.2-GFP, and HERD-2.2, have α -helical properties in solution, as they formed α -helical assemblies after stabilisation with TFE (Fig. 3.15c-d). Therefore, there are suggestions of an interesting shift in secondary structure from α to β .³¹⁰

Overall, the mechanism of fibre formation by HERD-2.2 is still unclear. This polypeptide was designed as an *N*-terminal fusion to client proteins, such as GFP, for LLPS. However, without a client protein not only is phase separation lost, but instead the protein forms structured fibres. This is certainly an unintended feature of the HERD design, and its role and prevalence could be the subject of future research. The distinguishing feature of the HERD-2.2 fibres is the interplay between LLPS and fibre formation. This is a phenomenon that has been observed in a number of natural proteins, and has been particularly identified as a pathway to amyloid formation.⁶⁷ Understanding how aberrant phase separation can lead to droplet maturation and the formation of pathogenic assemblies could lead to more effective prevention or treatment of neurodegenerative disorders.

Fibres made by the designed HERD-2.2-E protein were demonstrated to interact with demixed liquid droplets formed by HERD-2.2-GFP *in vitro* (Fig. 5.10). However, attempts to re-create droplet maturation by cleavage of the *C*-terminal fluorescent protein were unsuccessful (Fig. 5.11). It is possible that this is due to the differing structures of these proteins. In solution, the peptides constituting HERD-2.2 have a propensity to form α helices, while the HERD-2.2 fibres are β assemblies. There are cases where proteins undergo a change in secondary structure from α to β , particularly with respect to fibre formation, and it is possible that this could be the case with this designed protein.^{306,310,311} However, more detailed and time-resolved measurements of their secondary structure would be required to investigate these ideas. Measurements that can capture the nucleation process of HERD-2.2 could be exciting as an alternative way to monitor amyloidogenesis. It is possible that HERD-2.2 still undergoes LLPS as part of its assembly process, but that this quickly matures into the fibrous structures visible by TEM and confocal microscopy. Imaging of fibre nucleation could be informative to understanding how these structures self-assemble in cells, though this would be challenging to investigate *in situ*.

Chapter 6

De novo proteins for LLPS in eukaryotes

6.1 Chapter introduction

Chapter 3 described the design of a protein for LLPS in *E. coli* from the bottom-up. This system, named HERD-2.2–GFP, has been characterised to have all of the properties associated with de-mixed liquid condensates. Chapter 4 described the functionalisation of these condensates with two enzymes, to create a functional MLO in bacteria. Together, these present a tractable route to the bottom-up construction of cellular compartments in bacteria by LLPS. However, the scale and simplicity of prokaryotic cells means that the potential applications are limited. Instead, the design of proteins for LLPS in eukaryotes would enable a wider range of functions. Understanding how recombinant protein expression and the cellular environment differs between prokaryotes and eukaryotes, and how this impacts PPIs, is central to this task.

It is significant that it took far longer for natural phase-separating proteins to be identified in prokaryotes than eukaryotes.²⁷¹ Indeed, only a handful of phase-separating proteins have been characterised in prokaryotes.^{312,313} This makes the designed HERD-2.2–GFP system somewhat unusual, having been designed to phase separate in *E. coli*.^{270,314,315} The reason for the much higher prevalence of LLPS in eukaryotes is not yet clear, however it may be due to the similar greater abundance of disordered proteins. Predictions indicate that as many as 30% of the residues in eukaryotic proteins are likely to be disordered, in contrast to only 10% of those in prokaryotes.³¹⁶ IDRs are strongly implicated as drivers of LLPS in nature, and this abundance of disorder may be linked to an increased capacity for phase separation.⁸⁶

In addition, prokaryotic cells have significantly different internal macromolecule concen-

trations, estimated to be 3 times more crowded than eukaryotic cells: while in prokaryotes around 30% of their internal volume is occupied by macromolecules, in eukaryotes it is 10%.³¹⁷ This much higher crowding, in addition to the much smaller overall volume in prokaryotes, can significantly affect the relative strength of PPIs required for intermolecular interactions.^{263,318}

This chapter describes the design and characterisation of HERD proteins for LLPS in eukaryotes. Beginning from the successful HERD-2.2-GFP design used in *E. coli*, rational coiled-coil design was again used to make predictions about PPI strength and condensation, this time for LLPS in mammalian HeLa cells. This process highlights the differences in internal chemistries between prokaryotes and eukaryotes, and the effect this has on PPIs in cells. Further, the experiments described here begin to offer additional insights into why LLPS is so prevalent in eukaryotes, but not in prokaryotes.

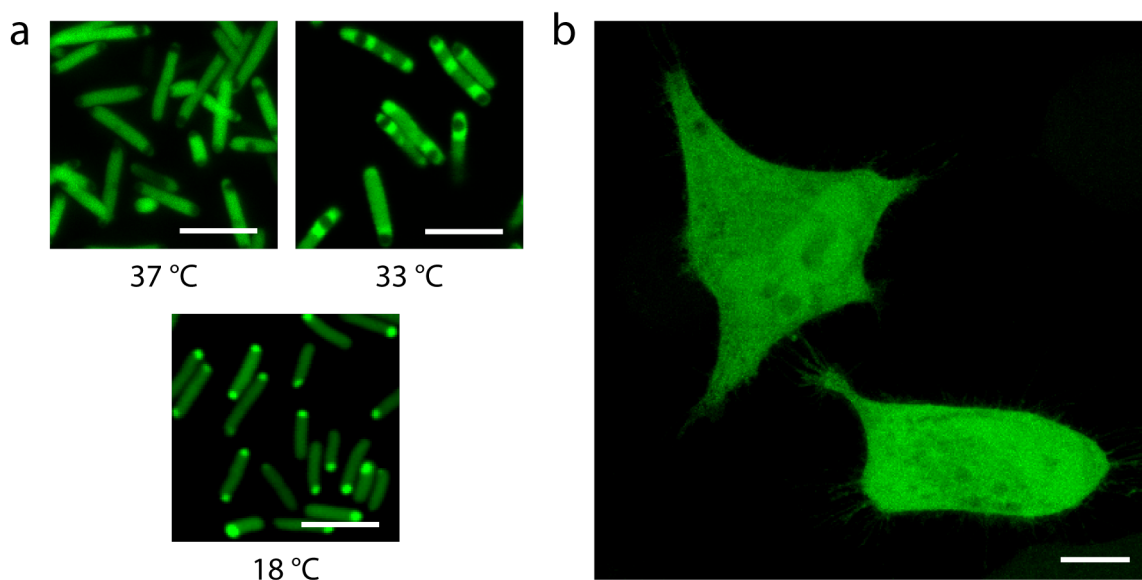


Figure 6.1: HERD-2.2-GFP does not phase separate in HeLa cells. a, HERD-2.2-GFP shows temperature sensitivity of condensation in *E. coli*. Scale bars are 5 μm . b, HERD-2.2-GFP does not form condensates in HeLa cells. Scale bar 10 μm .

6.2 HERD-2.2-GFP does not phase separate in HeLa cells

Initially, and to test whether the designed HERD-2.2-GFP protein would phase separate directly in eukaryotic cells, this designed protein was cloned into a mammalian cell vector for transient transfection in HeLa cells. The sequence for HERD-2.2-GFP was transferred into the viral cytomegalovirus vector (pCMV) by PCR using oligonucleotides described in Table 2.1, followed by restriction digest and ligation using the 5' and 3' NotI and BamHI restriction sites.

Transient transfection of HERD-2.2–GFP in HeLa indicated that the designed protein expressed well in mammalian cells, producing a range of green fluorescence indicative of successful transfection. Cells were imaged live at 37 °C. However, unlike expression in *E. coli*, no cellular condensates were visible in the HeLa cells (Fig. 6.1). From the characterisation of HERD-2.2–GFP described in chapter 3, this was not an entirely unexpected result. Even in *E. coli*, HERD-2.2–GFP was determined to be highly temperature sensitive, and does not form condensates in cells at 37 °C. As mammalian cells are less tolerant of temperature changes, cells were grown at 37 °C, and so it was unlikely that condensates would be formed under these conditions. Further, as introduced above, eukaryotic cells have much lower levels of molecular crowding than prokaryotes, and so this effective drop in crowding would reduce the overall propensity of HERD-2.2–GFP to phase separate still further.^{317,318} Not only that, but transient transfection in HeLa cells is unlikely to reach the protein concentrations created by T7 expression of recombinant proteins in *E. coli*, where the available volume is also much lower.³¹⁸ The combination of these factors is the likely reason why the HERD-2.2–GFP system phase separates in prokaryotes but not in eukaryotes.

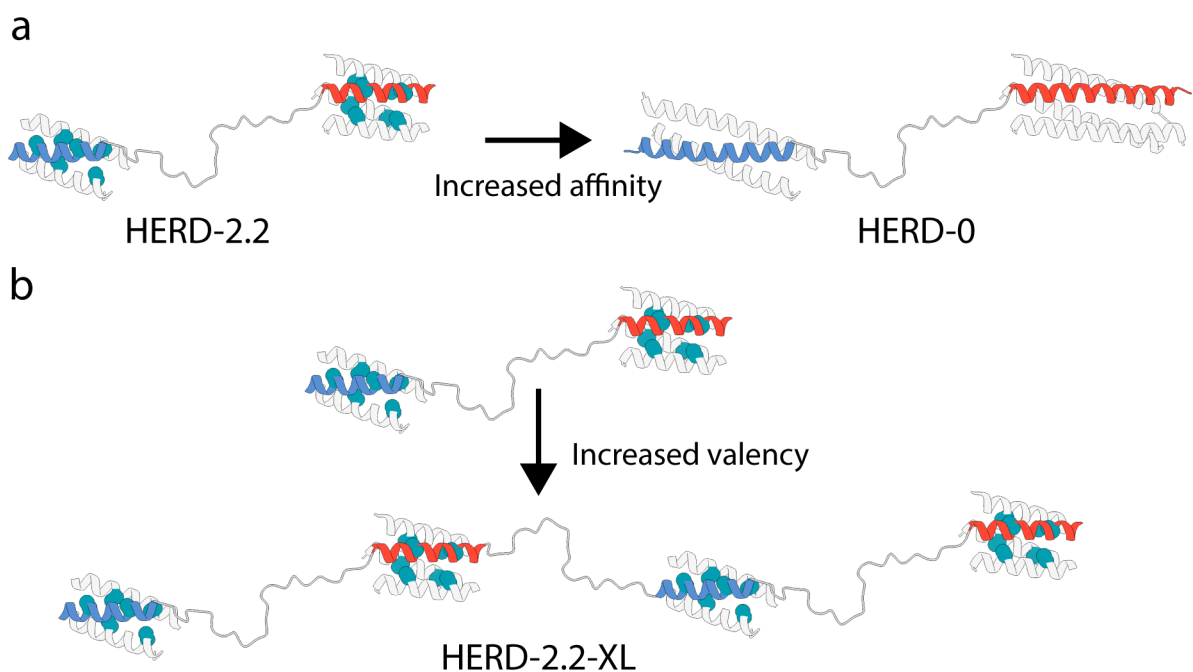


Figure 6.2: HERD re-design to introduce additional valency or affinity. Phase separation can be introduced either through (a), an increase in interaction affinity, or (b), an increase in the overall valency of interactions.

6.3 HERD re-design is required for LLPS in mammalian cells

The rational re-design of the HERD system for phase separation in mammalian cells was considered. To increase the propensity of HERDs for LLPS in eukaryotes, two approaches were attempted: firstly, the assembly of HERD constructs with additional stickers and spacers, to increase the overall valency of the design; secondly, design of HERD constructs with higher affinity helical regions, to increase the overall affinity of the design (Fig. 6.2).

6.3.1 Helical regions with tighter affinities

Initially, to test designs with higher affinities, the originally designed HERD-0 construct was tested in HeLa cells. This has 4 heptad (28 residue) helical regions with fully intact hydrophobic cores. In prokaryotes, this produces insoluble aggregates that do not recover after photobleaching (Fig. 3.25). However, when HERD-0-GFP was expressed in HeLa it still did not produce condensates, giving evenly distributed fluorescence indicative of soluble cytoplasmically distributed protein (Fig. 6.3). This result was surprising, given the highly insoluble condensates that HERD-0-GFP formed in *E. coli*.

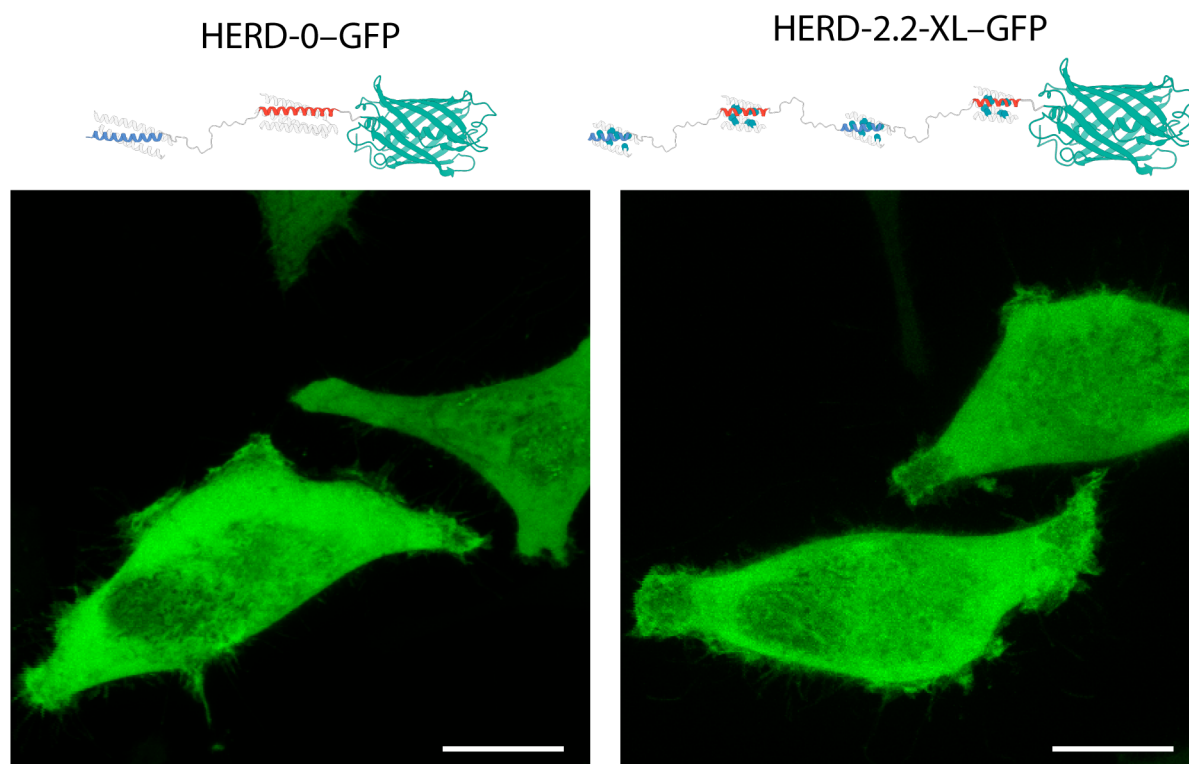


Figure 6.3: Neither HERD-0-GFP nor HERD-2.2-XL-GFP phase separate in HeLa cells. Live cell confocal microscopy images of HERD-0-GFP, and HERD-2.2-XL-GFP expressed in HeLa cells. Scale bars are 15 μm .

6.3.2 Additional stickers and spacers

As HERD-0 was insufficient for condensation in HeLa cells, it was tested whether increasing valency could instead drive phase separation. Here, more of the existing helical regions from HERD-2.2 were added to introduce a greater overall valency of interactions, in an extended stickers-and-spacers arrangement.

For this design, as the interactions given by the helical regions of HERD-2.2 had proven ideal for phase separation in prokaryotes, these stickers were used again but in a longer HERD design. This design, named HERD-2.2-XL-GFP, used 4 helical regions duplicated from the two used in HERD-2.2-GFP, separated by 3 identical linkers. The introduction of additional stickers should increase the overall attractive interactions of the protein, and potentially lower the saturation concentration. However, expression of HERD-2.2-XL-GFP in HeLa did not result in condensate formation, indicating that these additional stickers were insufficient to drive de-mixing (Fig. 6.3).

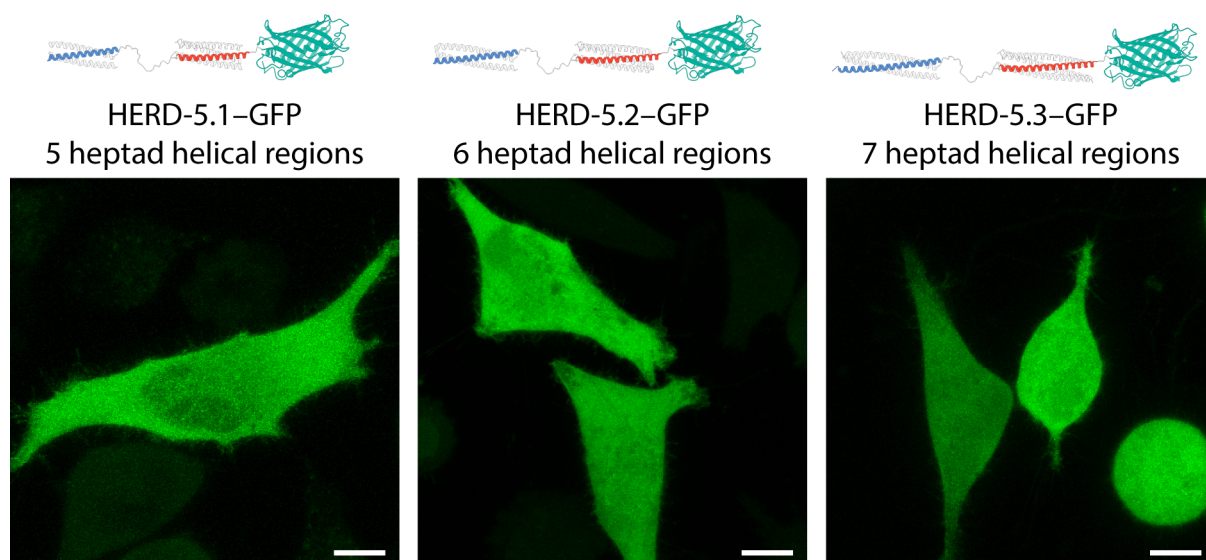


Figure 6.4: Longer coiled coils do not drive condensation. Live cell confocal microscopy images of HERD-5.1-GFP through HERD-5.3-GFP, expressed in HeLa cells. Scale bars are 10 μm .

6.3.3 Longer coiled coils for LLPS

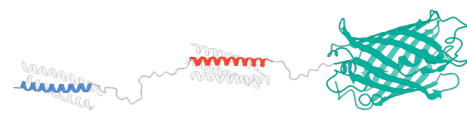
In chapter 3, three methods of destabilisation were used to attenuate the interaction potentials between the helical regions in the HERD proteins: shortening the helical regions, introducing mutations into the hydrophobic core, and introducing helix-destabilising mutations onto the predicted surface of the bundles. For the new designs, it appeared that rather than removing net interactions from the HERD design for LLPS, additional interactions would have to be introduced. Therefore, the opposite approaches were considered.

However, even HERD-0 was insufficient to drive condensation, and this design has helical regions with fully intact hydrophobic cores, with leucine and isoleucine at the *a* and *d* positions, and no destabilising mutations made to the surface of the bundles. Therefore, the only remaining parameter to adjust was length. Therefore, the helical lengths in the HERD-0 design were increased from 4, to 5, 6, and 7 heptads in length (35, 42, and 49 residues), creating HERD-5.1–GFP through HERD-5.3–GFP (Table 8.4). In addition, at this point for in-cell screening the *N*-terminal His tag was removed to ensure all the interactions were coming from the designed HERD elements. These coiled-coil motifs would be predicted to be highly strong and specific PPIs, owing to their extended hydrophobic interfaces. However, when expressed in HeLa cells, none of these HERDs produced condensates, but gave diffuse fluorescence (Fig. 6.4).

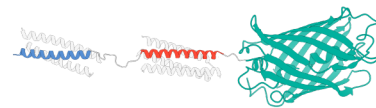
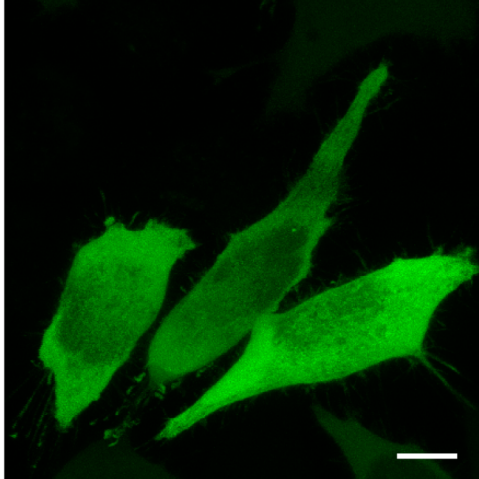
The absence of any protein condensates with even these strong PPI motifs was puzzling, and suggested that the design hypothesis for phase separation using HERDs was not appropriate for eukaryotic cells. Such long coiled coils should be almost guaranteed to interact, even at very low concentrations, and therefore it suggested that some aspect of the design was not functioning as intended. To exclude the role of unforeseen elements, which had not been robustly interrogated in the original design, several additional HERD designs were tested. To eliminate the potential for steric hindrance by the client GFP obstructing coiled-coil formation, designs with longer linkers between the HERD and GFP were screened. In addition, designs with shorter (8 residue) linkers between the coiled-coil motifs were also tested, in line with increasing their cooperativity. These permutations were tested with both 4 heptad and 7 heptad coiled-coil motifs, to ensure that affinity between the coiled coils was not limiting to self-assembly. These designs were called HERD-5.4–GFP through HERD-5.7–GFP (Table 8.4). Transfection of these designs in HeLa cells confirmed that none of these proteins formed condensates, demonstrating that these designs did not differ from HERD-0–GFP or HERD-5.3–GFP (Fig. 6.5).

6.4 LLPS and membrane localisation

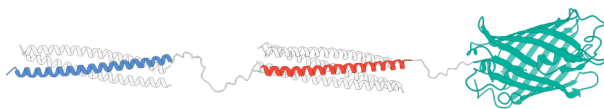
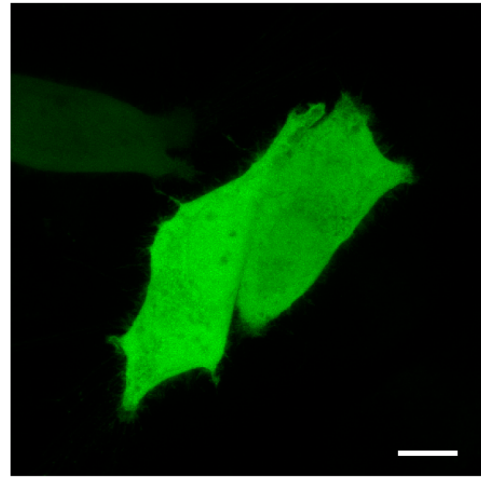
As neither HERD-2.2–GFP nor HERD-0–GFP were sufficient to phase separate in HeLa on its own, the addition of orthogonal localisation mechanisms was considered. The most compelling of these was membrane localisation (Fig. 6.6a). Membrane localisation can be induced by the addition of small *N*- or *C*-terminal peptides. The effect of membrane localisation is not only the trafficking of protein to the cell periphery in the case of the plasma membrane, but also a dramatic increase in the effective concentration of the protein (Fig. 6.6b). This is because the protein is now effectively constrained in a 2D rather than 3D space.



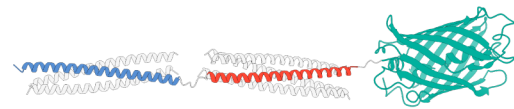
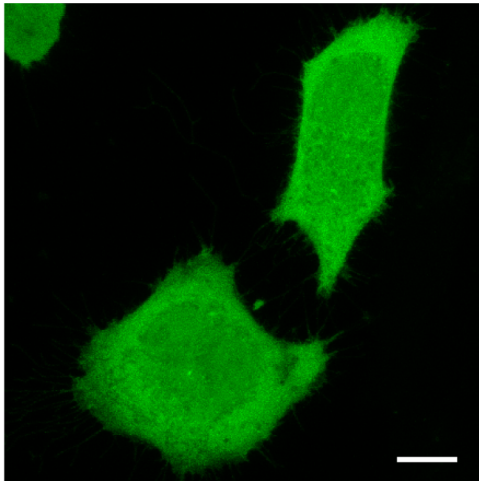
HERD-5.4-GFP
4 heptad helical regions
Longer linker



HERD-5.5-GFP
4 heptad helical regions
Short linker



HERD-5.6-GFP
7 heptad helical regions
Longer linker



HERD-5.7-GFP
7 heptad helical regions
Short linker

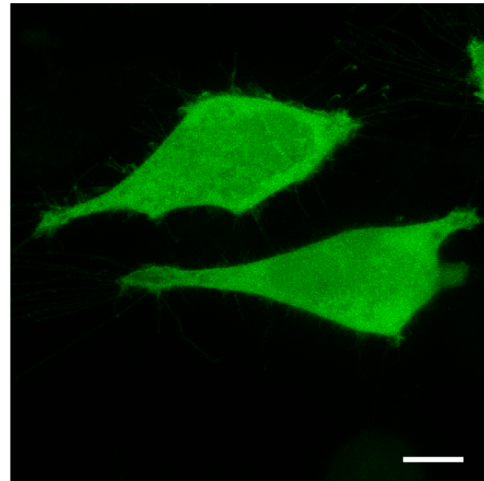


Figure 6.5: Expression of HERD-5.4 through 5.7-GFP in HeLa. Live cell confocal microscopy images of HERD-5.4-GFP through HERD-5.7-GFP, expressed in HeLa cells. Scale bars are 10 μm .

Indeed, theoretical analysis on PPIs of soluble proteins that are then constrained on a membrane indicates that fixing the proteins on a membrane increases their effective affinities by up to 1000-fold.³¹⁹ This increase can have the effect of making a weak PPI,

which would normally have too low of an affinity to find its binding partner in the 3D cytoplasm, more than sufficient for association on the membrane. This mechanism was used to great effect for the construction of orthogonally translating organelles (Fig. 6.6c).¹⁹⁶ In their construction, both phase-separating and membrane-localising proteins were used as localisation mechanisms.

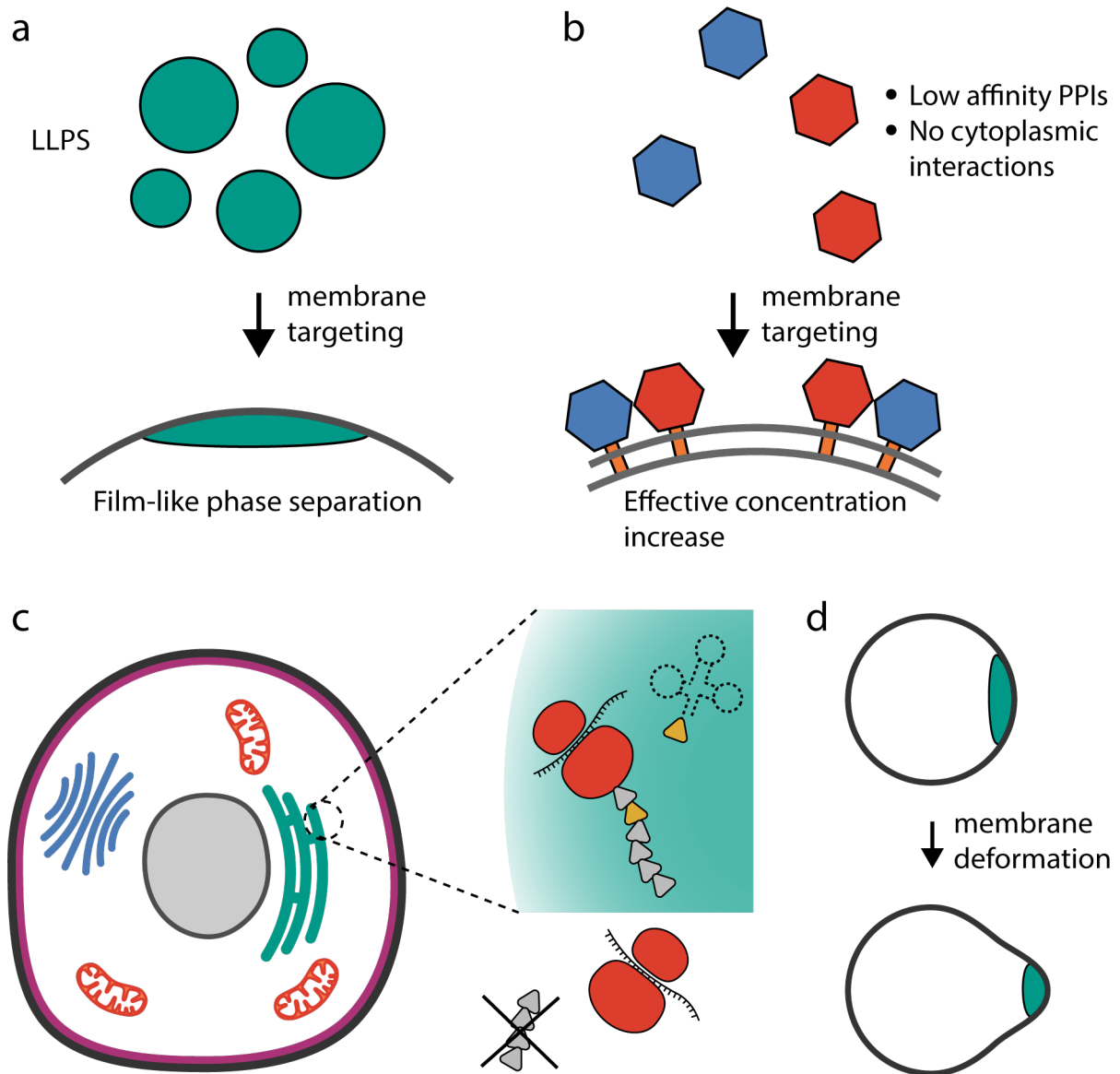


Figure 6.6: LLPS and membrane interactions. **a**, De-mixed liquid droplets can be targeted to cellular membranes to create film-like organelles. **b**, Constraining proteins on a 2-dimensional membrane can increase their effective concentration by up to 1000-fold, and induce PPIs that would otherwise be too weak to form in the 3-dimensional cytoplasm. **c**, Phase separation and membrane localisation has been used to target LLPS to several different subcellular locations, including the inner cell membrane (magenta), the endoplasmic reticulum (teal), the golgi apparatus (blue), and the outer mitochondrial membrane (orange). This has been used for selective and orthogonal amber stop codon replacement. Adapted from Reinkemeier *et al.* (2021) with permission from Elsevier.¹⁹⁶ **d**, LLPS can deform otherwise spherical phospholipid membranes.

This combination created an organelle that would selectively replace amber stop codons with an exogenous tRNA.¹⁹⁵ Individually, these systems perform poorly, with large amounts of background and poor selectivity.¹⁹⁵ However, the combination of both of these systems to create a phase separated film on a phospholipid membrane gives improved orthogonality and selectivity. Therefore, it seems tractable that the combination of two complementary localisation methods can be used to create an artificial organelle in mammalian cells.

In addition, there are functional reasons why localisation of phase separated droplets to the plasma membrane might be of interest. LLPS has been implicated in several processes associated with membrane remodelling. In particular, Wiskott–Aldrich syndrome protein (WASP) is a family of actin nucleation factors that promote actin polymerisation and membrane deformation.³²⁰ Neural-WASP (N-WASP) has been implicated in LLPS on the membrane, using phase separation to drive actin nucleation and create filopodia-like structures.^{321,322} Moreover, other natural proteins, such as Cavin1, have extensive IDRs that are proposed to drive phase separation, and are crucial to the formation of spherical membrane nanodomains called caveolae.³²³ Even cytoplasmically localised phase-separating proteins have been described to be able to induce membrane deformation (Fig. 6.6d).^{324,325} *In vitro* studies on synthetic and cell-derived vesicles have demonstrated that recruitment of phase-separating proteins, such as FUS and LAF-1, to membranes is sufficient to bend the membrane inwards and form long tubules extending into the vesicle.^{326,327} Together, these studies demonstrate that phase separation on the membrane is a key element in membrane deformation and remodelling.

6.4.1 HERD designs can be localised to the plasma membrane

To localise HERD-2.2–GFP to the plasma membrane, two membrane-targeting sequences were considered: Fyn, and CAAX (Table 8.5). The Fyn membrane tag comes from the *N*-terminus of the Fyn tyrosine kinase (GCVQCKDK).³²⁸ This sequence is endogenously acetylated by fatty acids on the glycine and two cysteine residues to create a sequence that is targeted specifically to the inner leaflet of the plasma membrane. The CAAX tag meanwhile is a *C*-terminal sequence used for membrane targeting of monomeric GTPases.³²⁹ Similarly, the final CAAX residues signal for prenylation of the cysteine that drives membrane insertion.

Each of these targeting sequences were first fused to HERD-2.2–GFP as *N*-terminal (Fyn) or *C*-terminal (CAAX) fusions, and the proteins transiently transfected in HeLa. These proteins, denoted Fyn-HERD-2.2–GFP and HERD-2.2–GFP-CAAX respectively, displayed plasma membrane localisation, with enhanced fluorescence at the cell periphery (Fig. 6.7). The CAAX tag however appeared to be detrimental to cell viability, and the

majority of fluorescent cells identified were rounded. In contrast, the Fyn tag appeared to be well tolerated and induced membrane localisation, and so was chosen for further analysis.

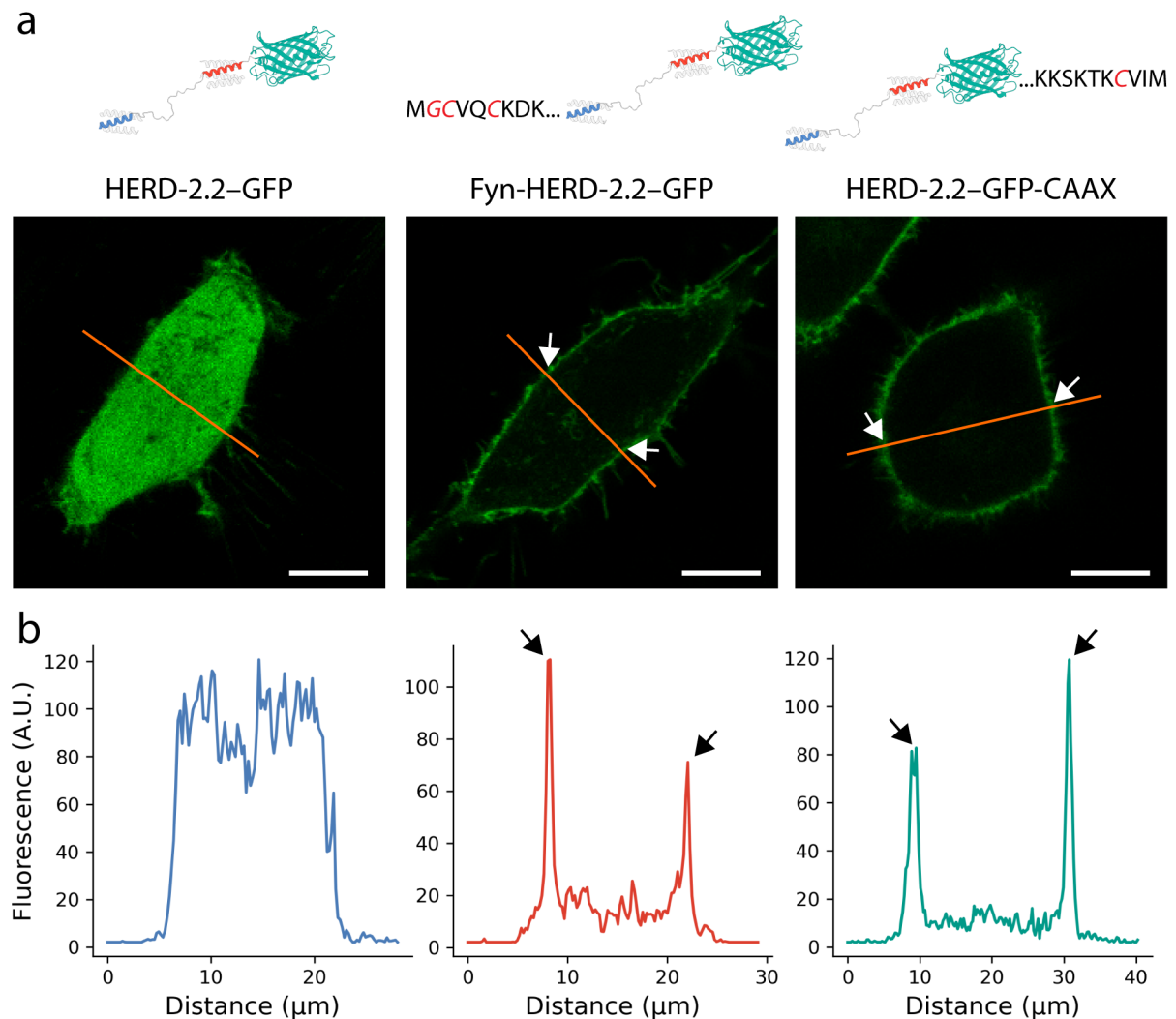


Figure 6.7: HERD-2.2-GFP can be localised to the plasma membrane. **a**, Live cell confocal microscopy images of HERD-2.2-GFP, and the membrane localised Fyn-HERD-2.2-GFP and HERD-2.2-GFP-CAAX expressed in HeLa cells. Scale bars are 10 μm . **b**, Fluorescence intensity profile of the orange line in **a**, showing the distribution of fluorescence shifting from cytoplasmic to membrane localised with the inclusion of either the Fyn or CAAX tag. Black arrows correspond to the position of white arrows in **a**.

Fyn-HERD-2.2-GFP gave strong membrane localisation, but it was challenging to identify whether phase separation was happening in addition to membrane targeting. Areas of enriched fluorescence were visible around the cell periphery, but distinguishing phase separation from areas of increased membrane density, due to membrane invaginations or protrusions, was not possible. However, these areas of enrichment may be challenging to visualise by confocal microscopy, and alternative techniques, such as total internal reflection (TIRF) microscopy could extract information about enrichment on the membrane.

6.5 Intramolecular interactions may inhibit phase separation

To inform the design of future proteins for phase separation in mammalian cells, models of HERD-0-GFP and HERD-5.3-GFP were investigated using AlphaFold2. Modelling of these designs indicated that in both cases, the coiled-coil motifs were predicted to be robustly helical. This is corroborated by the solution phase data for the helical regions of HERD-0 (Fig 3.14a). However, in the AlphaFold2 models these helical regions are implicated to interact intramolecularly, forming coiled coils between the two helices (Fig. 6.8). This contrasts to data by Dr Freddie Martin, which indicated that these trimeric (CC-Tri-EEWE) and tetrameric (CC-Tet2-EEWE) coiled coils behave orthogonally.²⁴¹ These interactions, although models without experimental evidence, are predicted by AlphaFold2 with a high level of confidence, as indicated by the low predicted aligned error (PAE) score given by their cross-interaction. Indeed, it is possible that while these coiled coils are orthogonal in solution as independent peptides, when combined into a single polypeptide, the resulting increase in local concentration means that their propensity to interact intramolecularly is very high.

If the potential to form intramolecular interactions is too high, there will be too few intermolecular interactions to drive phase separation. Further, it suggests why increasing the length of the coiled coils does not improve their propensity to condense – stronger coiled coils would simply form stronger intramolecular interactions, as indicated by the model of HERD-5.3-GFP, which forms an extended intramolecular coiled coil. It is interesting to note that these models are in contrast to those of HERD-2.2-GFP, where the interactions are so weak that the coiled-coil motifs are not predicted to interact intramolecularly, and this may be partially the cause of this design’s success in *E. coli* (Fig. 3.17).

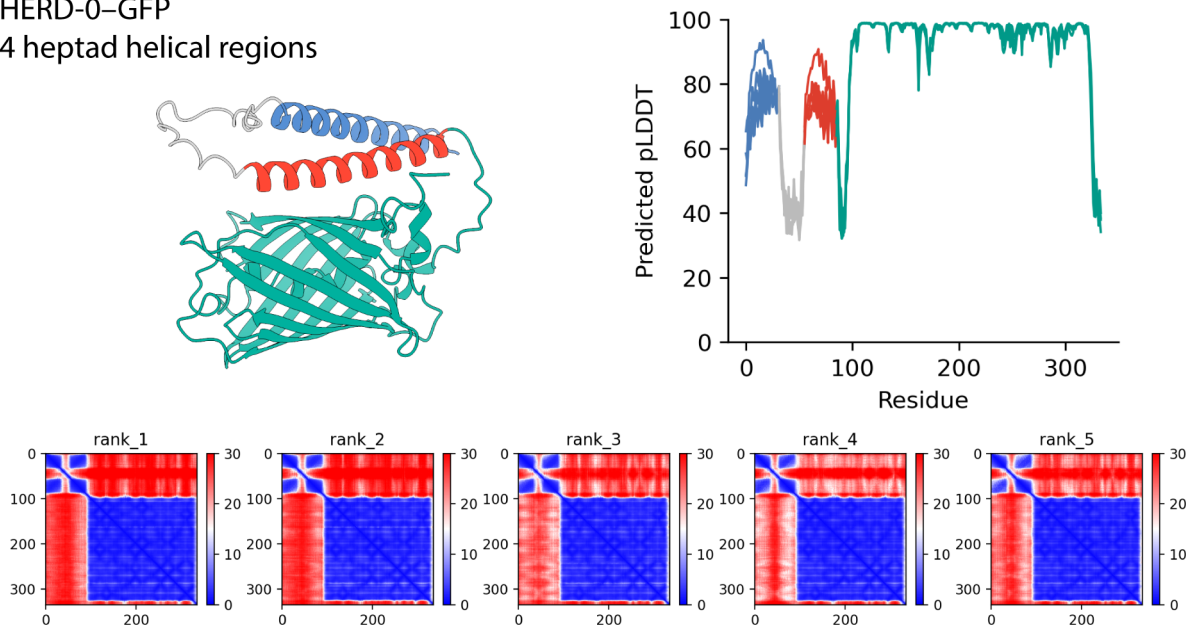
6.6 Chapter summary

Following the successful design and functionalisation of *de novo* proteins for LLPS in *E. coli* their application in eukaryotic cells was considered. Potentially, this offers a greater range of applications because of their more complex cell biology, as well as much greater cell volume. The majority of natural proteins identified to undergo LLPS have been found in eukaryotes, so this appears a natural environment for these disordered and flexible proteins.^{35,271} This chapter has described the expression and initial visualisation of HERD proteins in HeLa cells. The weak, unstructured, HERD-2.2-GFP that was successful in producing de-mixed liquid droplets in *E. coli* was not found to phase separate in HeLa. Instead it formed a single soluble phase (Fig. 6.1). Using the principles of valency and

affinity to drive LLPS, HERD designs with more coiled-coil motifs, or stronger PPIs, were tested for phase separation. These designs used coiled-coil motifs from 4 up to 7 heptads in length (28 – 49 residues). Surprisingly, none of these designs formed condensates either (Fig. 6.5).

HERD-0-GFP

4 heptad helical regions



HERD-5.3-GFP

7 heptad helical regions

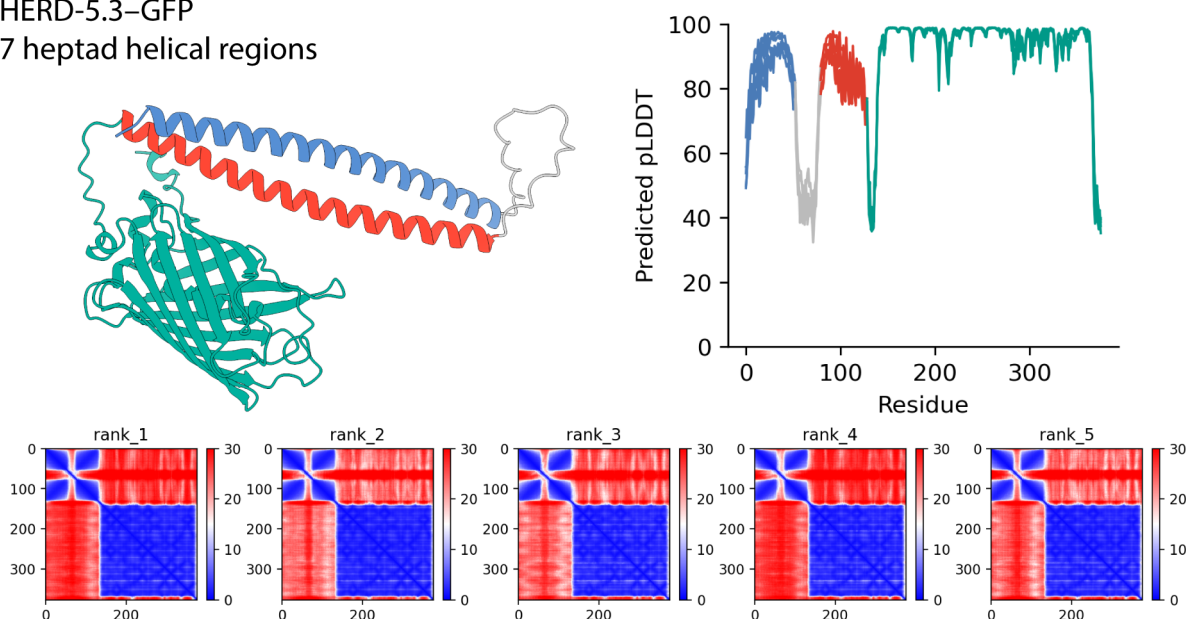


Figure 6.8: Intramolecular interactions may be inhibiting phase separation. Representative AlphaFold2 models of HERD-0-GFP and HERD-5.4-GFP with pLDDT and PAE plots showing confidence scores for each model.

These results reiterate the differences between proteins designed for phase separation in bacterial and mammalian cells. As noted previously, eukaryotic cells are significantly larger, less crowded, and express recombinant proteins to lower concentrations than *E.*

coli.^{317,318} These compounding effects have the result that proteins with weak PPIs that self-assemble within *E. coli*, can be entirely soluble within HeLa cells. Introducing additional PPI motifs or increasing the affinity between the existing motifs seemed like a tractable approach to drive phase separation. These have been used previously in synthetic phase-separating proteins, where increasing the affinity of the PPIs reduced the saturation concentration required for phase separation.¹⁹⁹ Further, increasing the number of PPI motifs has also been used to similar effects by an overall increase in valency.^{180,190} However, these approaches did not induce phase separation in the HERD framework. Proteins with 4 coiled-coil motifs, or motifs with up to 7 heptads, were still found to only form a single phase. Indeed, it is remarkable that these proteins did not form insoluble aggregates when expressed. If the intended network of interactions was assembling as designed, these PPIs would be hyper-thermostable, and likely cause severe aggregation. Instead, the maintenance of a single, soluble phase suggests that the HERD design is not assembling as intended in mammalian cells.

Some suggestions for this observation come from modelling of HERD-0 and HERD-5.3 structures (Fig. 6.8). These indicate the formation of strong intramolecular interactions that could be inhibiting the network of intermolecular interactions designed to drive phase separation. This intramolecular folding is an interesting idea, and future work could attempt to tie these models to experimental evidence. For instance, measurement of oligomeric state and hydrodynamic radius by dynamic light scattering could be used to test if HERD designs are oligomerising into large assemblies, or become saturated at low oligomeric states.

As the HERD designs did not phase separate as cytoplasmically localised proteins, the combination of orthogonal subcellular targeting motifs was tested. Membrane targeting of phase-separating proteins has been exploited previously to improve selectivity.¹⁹⁶ Further, constraining proteins on a 2D membrane can have significant effects on PPIs by compounding concentration effects.³¹⁹ Here, the HERD proteins were targeted to the plasma membrane of HeLa cells, forming enriched areas around the cell periphery (Fig. 6.7). However, using confocal fluorescence microscopy it was not possible to conclude whether phase separation was occurring in addition to membrane localisation. Further studies using more localised microscopy techniques, such as TIRF microscopy, or single molecule measurements, could be the solution to measuring phase separation on the cell membrane.

Chapter 7

Discussion and future work

7.1 Principal conclusions of this thesis

In a remarkably short time, phase separation has become recognised as a ubiquitous feature of cell biology. Phase separation has been implicated in almost every cellular process, from chromatin organisation, cell signalling, and genetic regulation, to microtubule organisation, cell division, and cellular stress responses.^{35,43,91,312,330,331} However, despite these widespread associations, there are still significant gaps in our understanding of how phase-separating proteins are organised at a molecular level. This thesis describes the bottom-up design of proteins for LLPS in cells, using rational protein design principles. Here, a summary of the work is presented, along with a discussion of the various strengths and limitations of the study, and suggestions for future experiments to build on these concepts.

7.1.1 *De novo* design delivers proteins for LLPS

Previous engineered proteins for LLPS have largely, with select notable exceptions, focused on the repurposing of existing phase-separating proteins.¹⁷⁹ These engineered assemblies have achieved remarkable success, proving to be robust scaffolds for phase separation for a range of applications. More recently, phase-separating systems of increasing complexity have been developed, introducing several variations of controlled assembly and disassembly mechanisms, phase separation on cellular membranes, and a host of different applications for the control of molecular biology.^{180,188,196} However, the bottom-up design of scaffold proteins for phase separation has remained relatively unexplored. Recent studies have highlighted the molecular features of proteins undergoing phase separation, and these concepts present the opportunity to create proteins that drive LLPS from first principles.^{198,199}

In chapter 3, the design of a *de novo* protein for LLPS in *E. coli* is described. For this assembly, rationally designed α -helical coiled-coil motifs are concatenated together with a designed linker to create a linear polypeptide, and the helical affinities tuned to drive dynamic LLPS (Fig. 7.1). This design process highlights the molecular principles of multivalency and weak intermolecular interactions as critical for the construction of a scaffold protein from the bottom-up.²⁶ Characterisation of the designed HERD-2.2-GFP protein confirms that the design trajectory has been successful in producing *de novo* de-mixed droplets that display all of the macroscopic features of LLPS: droplet coalescence, reversible de-mixing, and rapid fluorescence recovery after photobleaching (Figs. 3.19, 3.21, & 3.22). Crucially, these HERD-2.2-GFP proteins produce liquid-like condensates in cells, with rapid fluorescence recovery after photobleaching (Fig. 3.24). The formation of MLOs in bacteria is particularly interesting, as unlike eukaryotes they do not have canonical membrane-bound organelles, making this system one of a very small number of mechanisms to form subcellular organelle-like compartments in prokaryotes.

This design approach could present a framework for the general construction of *de novo* condensates, potentially offering a route to even more tightly controlled phase separation. The further tuning of scaffold affinity and valency could be used to create condensates with altered phase boundaries, giving phase separation at elevated or lowered temperatures. In addition, by using rationally designed coiled coils there is the potential to introduce orthogonality into the design. Despite their short sequences and structural similarity, several studies have now demonstrated that coiled coils with different heptad repeats can be orthogonal to each other.^{145,241} This could be a potential route into creating multiple orthogonal condensates using the same design architecture.²⁴¹ However, it should be considered that in the HERD-2.2 design used in *E. coli*, this is likely not possible, as the extensive destabilisation of the helical regions will likely have significantly reduced the specificity of these helical regions. Indeed, it is unclear how much helicity is present in the HERD-2.2 PPI motifs. Helical repeats appear to be crucial to condensation in cells, however, it is challenging to assess the secondary structure of very weakly folded protein domains. Similarly to elastins, an in-depth investigation of structural changes during phase separation could shed light on how helicity is implicated in LLPS in this system.³³²

In addition to expanding the repertoire of helical regions, there is still significant scope for investigating the unstructured linker. A selection of altered linker compositions is also described in chapter 3 (Fig. 3.18). However, this is not comprehensive, and designs focus instead on the helical regions. However, this *de novo* system could be used to provide experimental data for the numerous *in silico* studies performed on sticker-and-spacer systems.^{90,225} Providing experimental data on how changes to the linker in a stickers and spacers framework could inform modelling and prediction of such systems in the future.

7.1.2 *De novo* droplets are highly efficient reaction crucibles

Several highly impressive studies demonstrating the potential applications of protein condensates have already been described.^{180,182,188,195} However, again these have focused on repurposing natural phase-separating scaffolds, and typically employ the capacity for condensates to passively compartmentalise client macromolecules. This sequestration approach is still powerful, and can have significant impacts on cell biology by depletion of endogenous pools of a target protein. Nonetheless, *in vitro* experiments have demonstrated that simple and complex coacervates also have the potential to accelerate enzymatic reactions, and this presents the opportunity to compartmentalise and accelerate an enzymatic pathway using phase separation in living cells.³³³

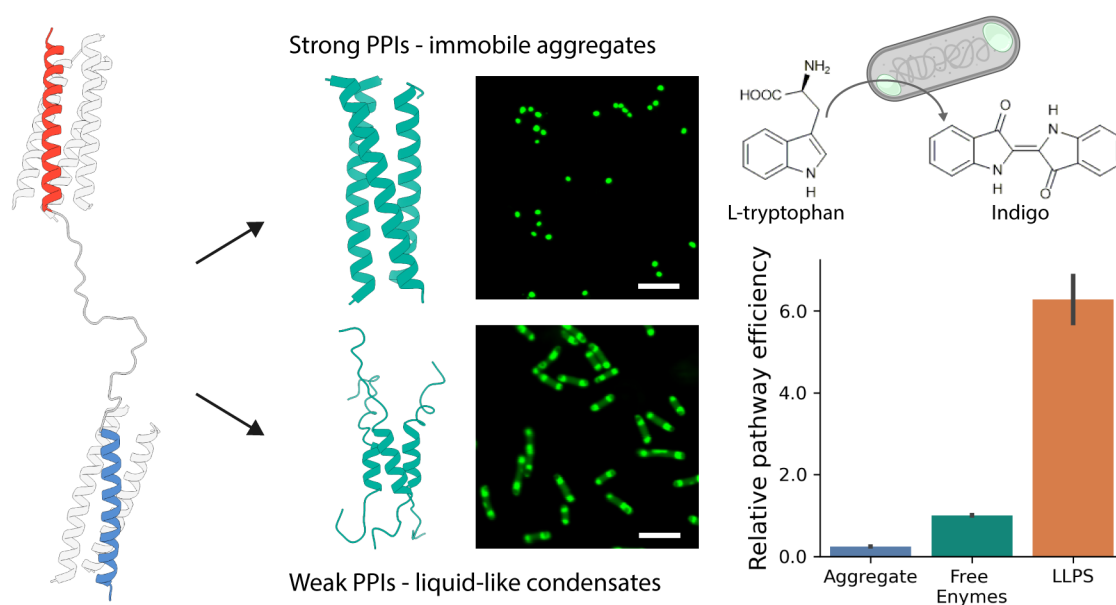


Figure 7.1: *De novo* design delivers solutions for LLPS. The HERD proteins designed for LLPS in bacteria are created by concatenation of *de novo* coiled-coil motifs, into a multivalent linear polypeptide. Through a process of destabilisation of the coiled-coil motifs the phase behaviour of these proteins is altered, to transition from immobile aggregates to ultimately create liquid-like condensates that undergo LLPS in *E. coli*. Finally, these condensates are used to co-compartmentalise the two enzyme pathway for the production of indigo dye from tryptophan, and produce a 6-fold improvement in product formation over the comparable free enzymes.

Chapter 4 describes the application of the designed HERD-2.2 framework to the compartmentalisation of an enzymatic pathway. The designer protein is used to co-condense the enzymes TnaA and FMO for the enzymatic production of indigo dye. The HERD-2.2–TnaA and HERD-2.2–FMO fusions are readily enriched within HERD-2.2–GFP droplets *in vitro*, and fusion to the orthogonal fluorescent protein mCherry further demonstrates their simultaneous enrichment in condensates in *E. coli* (Figs. 4.2 & 4.6). Further, co-condensation of the two enzymes in HERD-2.2–GFP droplets results in an overall 6-

fold enhancement on indigo formation (Fig. 7.1). This demonstrates a clear benefit to enzymatic pathway efficiency due to protein condensation in the designed MLOs.

Using phase separation for enrichment of client enzymes is particularly prudent for this pathway. Both of the enzymes TnaA and FMO are oligomeric (TnaA – tetrameric, FMO – dimeric).^{217,291} Classical methods of colocalisation of these enzymes, such as by fusion to dimerisation domains or assembly on a rigid scaffold, would likely be unsuccessful or function poorly because of the potential for indefinite oligomerisation.¹⁶⁶ The combination of two symmetric oligomeric enzymes is even reminiscent of strategies to create protein nanoparticles, however, without careful alignment of geometry it would likely result in amorphous aggregation.^{161,334} Instead, colocalisation within dynamic droplets permits enzyme assembly without forming an indefinite network of PPIs that could restrict substrate accessibility or enzyme dynamics. However, it is also true that using dynamic protein condensates can present additional considerations on enzyme colocalisation. For instance, the HERD-2.2 condensates have a maximum enzyme loading capacity, beyond which the condensates are dispelled (Fig. 4.7). This could restrict the concentration of enzymes that can be enriched, and therefore their utility as enzymatic reaction crucibles. Future work could assess how condensation and enzyme loading are coupled, to understand how best to leverage the potential of MLOs in enzymatic reactions. Some early indications of this approach are given by the re-designed HERD-4.1, which used a simple net charge adjustment to significantly improve co-condensation of the rate-limiting FMO, and boosted the improvement in indigo production from a modest 2.3-fold to over 6-fold enhancement (Figs. 4.12 & 4.13). It is likely that future works on a more systematic analysis of protein enrichment in MLOs could provide further improvements on co-condensation.

The use of *de novo* condensates for enzymology also permits the interrogation of how condensate material properties affect enzymatic reactions. The impact of condensate rheology and dynamics on enzymatic function has been considered for some time, but is challenging to investigate in cells due to the lack of control over condensate material properties.¹⁹¹ Here, in the designed HERD system, the design trajectory provides additional condensates that do not behave like de-mixed liquids, and are likely arrested aggregates. In addition, HERD-2.2–GFP is highly temperature sensitive, and can switch between behaving like a dynamic liquid and a static aggregate depending on the growth temperature (Fig. 4.10). Using these handles on the material properties of the condensates, product formation was determined to be highly dependent on enrichment within dynamic condensates, rather than arrested materials. Indeed, condensation of catalytic enzymes within arrested condensates significantly reduces the amount of indigo produced by the reaction, likely to be due to reduced substrate accessibility or disrupted protein dynamics (Fig. 4.11). These data provide a direct link between the material properties

of MLOs *in vivo* with pathway efficiency and product formation.

Chapter 4 provides proof-of-concept that MLOs designed from the bottom-up can be used to enrich functional enzymes, and enhance pathway efficiency. The next step would be to apply this to even more complex enzymatic pathways. While indigo dye is a relevant small molecule used in the textile industry, it has an established high-yield chemical synthesis route. Instead, applying MLOs to the production of relevant pharmaceuticals that can only be synthesised enzymatically could give meaningful improvements to the production of small molecule therapeutics, with no additional metabolic burden. Further, and as suggested by theoretical modelling of enzyme colocalisation, enriching three or more enzymes could enhance product formation by as much as two orders of magnitude, much more than is possible using the two enzyme pathway described here.²⁹⁵

7.1.3 The interplay between LLPS and fibre formation remains elusive

Droplet maturation and fibre formation from LLPS are recurring themes in pathological phase separation. A number of proteins that undergo LLPS have also been identified to form fibres, gels, or aggregates.³⁵ In particular, amyloid formation has been closely linked to aberrant phase separation.³⁰⁹ In chapter 5, fibres formed by the designed HERD-2.2 protein are described. This protein was originally designed to undergo LLPS with an accompanying client protein. However, without clients, these polypeptides form fibres (Fig. 5.2). The fibres assemble both *in vitro* and in *E. coli*, forming assemblies that stretch the entire length of the cell (Fig. 5.5). Indeed, they even cause the host cell to elongate, potentially by an inability to divide, to far beyond its natural length (Fig. 5.7).

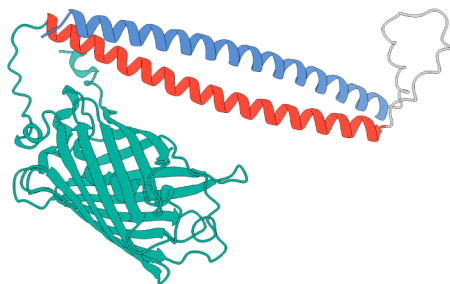
The scale of these fibres in living cells is remarkable. Not only are they readily expressed and form bundles several microns in length, but they appear to be tolerated by the host cell without causing obvious defects in viability. This may be because of the more rigid bacterial cell wall, preventing damage to the cell membrane by the fibres. These fibres have also been demonstrated to act as recruitment scaffolds for globular proteins. When co-expressed, the HERD-2.2 fibres recruit and pattern the GFP tagged HERD-2.2–GFP (Fig. 5.6). This suggests that the fibres could recruit other, functional proteins to scaffold enzymatic reactions, or control cell physiology, similar to the studies performed in chapter 4 with liquid droplets. However, it is unclear whether these assemblies would confer a benefit on catalysis, and could instead trap the enzymes in a non-functional conformation or restrict substrate accessibility. Further, several studies have already used fibrous proteins as scaffolds for enzymatic reactions in *E. coli*, so the potential scope for advancement may be limited.^{152,288} Instead, future work could investigate whether these fibres are tolerated in mammalian cells. Instead of expression in the small, and crowded, bacterial cytoplasm, expression in mammalian cells could create an even more dramatic

phenotype with the synthetic fibres. These could potentially even be manipulated to create a *de novo* cytoskeleton, with the introduction of binding sites or functional proteins into the scaffold.

These HERD-based fibres are interesting because of their scale and phenotype, but they were not designed. Instead, and reminiscent of many natural proteins, the designed HERD-2.2 protein undergoes LLPS, but also appears to have a propensity to form fibres. The structural mechanism for this rearrangement remains unclear. It is possible that some steric block has been removed when HERD-2.2 is expressed without a client protein. However, cleavage of the *C*-terminal GFP does not result in fibre nucleation from demixed liquid droplets, though this may be due to experimental conditions (Fig. 5.11). Melting of HERD-2.2 fibres also indicates that they do not re-fold at low concentrations, and it is possible that significant molecular crowding or high concentrations are required for assembly of these fibres (Fig. 5.8). Further investigation of the assembly dynamics of these fibres could present a route into reversible folding of the fibres. With this, the interplay between LLPS and fibre formation could potentially be investigated.

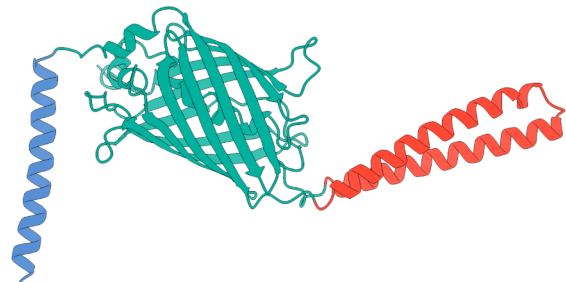
HERD-5.3-GFP

7 heptad helical regions

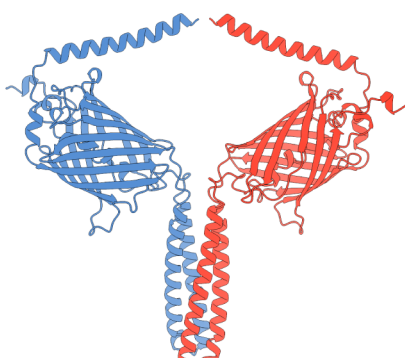


Sterically occluded coiled coils

4 heptad helical regions



Sterically occluded coiled coils
4 heptad helical regions - dimer



Sterically occluded coiled coils
4 heptad helical regions - trimer

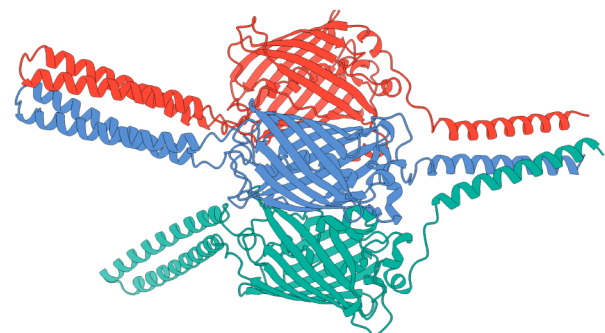


Figure 7.2: Redesign of proteins for LLPS in eukaryotes. AlphaFold2 models of the HERD-5.3-GFP design, and alternative sterically occluded coiled-coil designs, where one coiled-coil motif has been inserted into the loop of GFP distal from the termini.

7.1.4 Alternative designs are required for LLPS in mammalian cells

Most designed and engineered phase-separating proteins have focused on mammalian cells as host organisms.^{185,200,201,335,336} This is likely particularly tractable because the vast majority of phase-separating proteins originate from mammalian cells.³¹³ However, there are other significant advantages, such as the much larger cellular lumen that not only facilitates imaging but also permits more significant effects due to sequestration. Further, due to the more complex cellular organisation of mammalian cells, there are a greater variety of potential applications within mammalian cells. Therefore, in chapter 6, the HERD system is re-designed to transport the designer phase-separating framework from a prokaryotic to a eukaryotic context. These redesigns use increasing valency and affinity as mechanisms to improve phase separation of the HERD design. However, none of the tested designs form condensates in mammalian cells (Figs. 6.3 & 6.4).²³¹

Some indications as to why these designs do not phase separate come from modelling of the designed polypeptides. Both HERD-0–GFP and HERD-5.3–GFP are suggested to form strong intramolecular interactions in models created by AlphaFold2 (Fig. 6.8). The coiled-coil motifs used to create the HERD framework have been carefully chosen to reduce the propensity for intramolecular interactions. Indeed, data on the peptides *in vitro* suggest that these motifs are orthogonal, and do not exchange helices between their respective bundles when mixed.²⁴¹ However, the effective concentration effect when these motifs are concatenated into a single polypeptide could mean that these interactions are far more prominent than those *in vitro*.

It is tempting to compare the valency and estimated affinities of the HERD designs with natural proteins, however, comparison of the HERD system with natural proteins is challenging. The extensive disorder of natural phase-separating proteins makes it difficult to quantify their affinities and valencies of interaction. Therefore, it is unclear whether the motifs used in the HERD design are comparable to natural proteins. Further, there are very few phase-separating systems that have been designed or engineered with a quantifiable valency of interaction. Some noteworthy systems include a two-component system with dimeric and tetrameric oligomerising motifs that undergoes phase separation in yeast.¹⁹⁹ These designs were tested with a range of interaction affinities for one of the interfaces. This suggests, as originally hypothesised, that trimeric and tetrameric PPI motifs should be more than sufficient for phase separation. Further, with the helical lengths tested their affinities should be sub-nanomolar, and unlikely to be too weak to find their interaction partner in cells.

Instead, it seems likely that intramolecular interactions between the coiled-coil motifs may be limiting phase separation. This would reduce the overall effective valency for intermolecular interactions. Indeed, the flexibility of the linker could be partially to

blame here, as it allows the coiled coils to fold over onto each other. Future designs for phase separation in mammalian cells will need to consider additional design elements to reduce intramolecular interactions. One such idea is to sterically block the two coiled coils from interacting, by placing them on opposite sides of a globular protein (Fig. 7.2). In this example design, one coiled coil has been placed on the *N*-terminus of GFP as used previously, but instead of placing a second coiled coil proximal to the first, a second PPI motif has instead been inserted into a loop of GFP distant from the termini. This design means that the two coiled-coil motifs are sterically occluded from interacting, and instead must make intermolecular interactions. Modelling of these assemblies using AlphaFold2 indicated that they could indeed oligomerise using the included coiled-coil motifs.

Overall, chapter 6 confirms that different design approaches are required for designing phase-separating systems for bacterial and mammalian cells. These concepts will become more important as the role of protein design in phase separation continues to grow, and more synthetic systems are developed.

7.2 Possible directions for future research beyond this thesis

After the explosion of research on protein phase separation in recent years, a complete picture is beginning to emerge of their biological roles, and molecular architecture. Protein design and engineering has played a significant part in that process, guiding the development of new tools for protein assembly and localisation. However, there is still scope for outstanding future discoveries to be made regarding the mechanism of protein phase separation, and its potential applications.

7.2.1 Globular proteins for LLPS

As described throughout this thesis, protein phase separation and intrinsic disorder appear to go hand in hand.^{93,337,337} Nearly all of the characterised natural proteins that undergo LLPS have some component of intrinsic disorder, and many of them are extensively unstructured. This has carried through to the synthetic systems developed, with many of them relying on IDRs or extended oligomerising motifs (Fig. 7.3a). Some designer systems have even sought to optimise repeats from natural IDRs of phase-separating proteins.¹⁹⁸ Indeed, the HERD system described in this thesis is no exception, with a designed intrinsically disordered linker and helical regions that appear are largely unstructured in solution. Therefore, it could appear that disorder is a requirement for protein phase separation. However, physically, there is no such pre-requisite. Instead, it appears that unstructured protein regions simply offer the best combination of weak

attractive interactions, with a high valency and flexibility to drive LLPS, and as a result are frequently implicated in phase separation.^{26,34} Nonetheless, if disorder is not a requirement for phase separation, it should be possible to construct globular proteins from the bottom-up that have the correct composition of attractive interactions to drive phase separation as well.

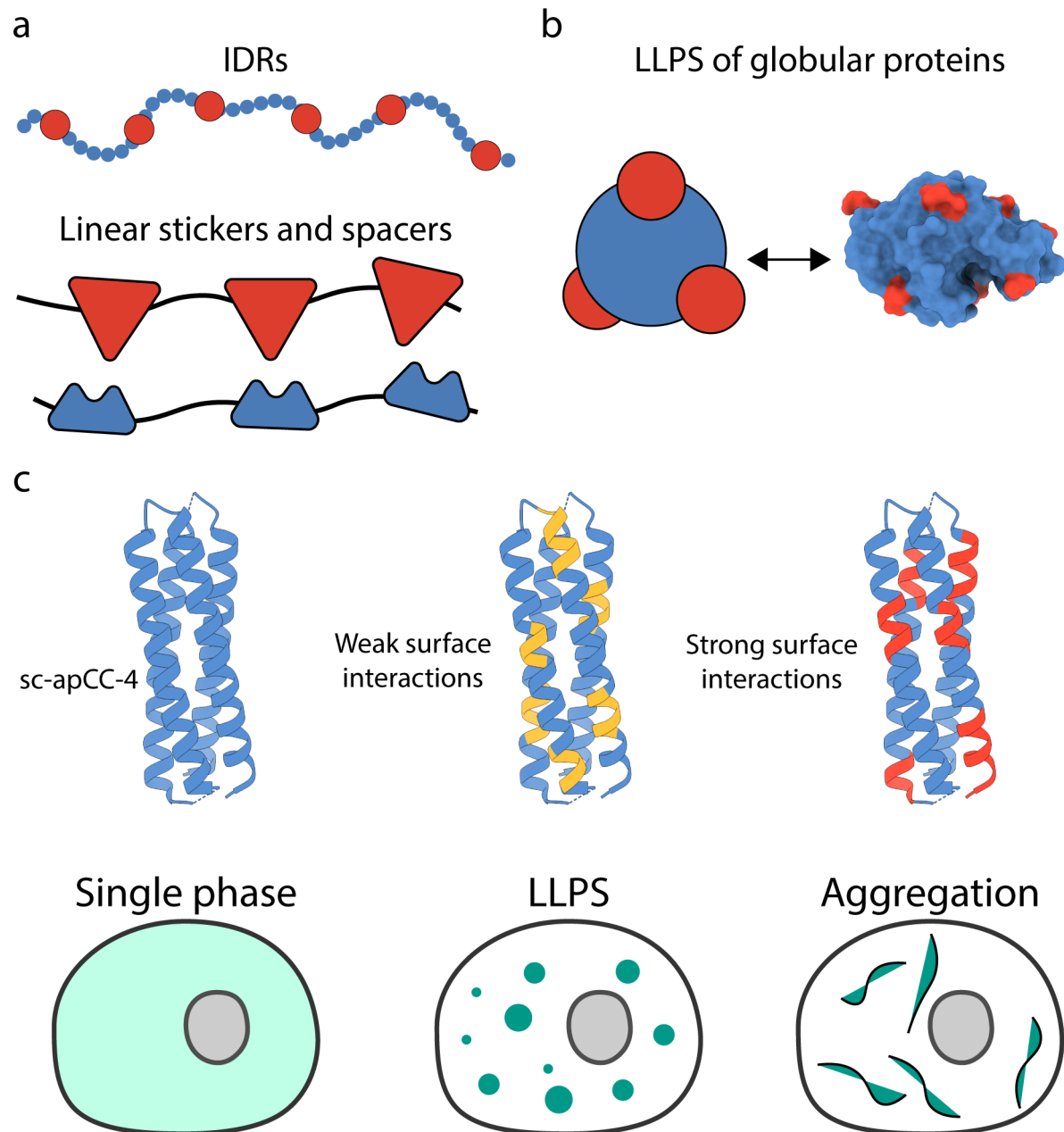


Figure 7.3: Design of globular proteins for LLPS. **a**, Existing strategies for LLPS currently rely almost exclusively on IDRs or other linear concatenations of PPI motifs. **b**, LLPS of globular proteins is possible, and has been observed *in vitro* and in some cases *in vivo* for a select number of proteins such as lysozyme, but is currently unexplored by design. **c**, *De novo* protein design could tackle this challenge by modification of protein surface properties, such as by using the designed single chain 4-helix bundle sc-apCC-4 (PDB: 8A3K). Rational design of surface interactions could be used to tune interaction affinity to drive LLPS.

LLPS of globular proteins is certainly possible and has been characterised *in vitro* by physicists for many years; the classic example being the monomeric globular protein lysozyme, which undergoes LLPS at high NaCl concentrations (Fig. 7.3b). Recent work has begun to explore this concept. Attachment of cationic tags to globular proteins, including GFP, has been demonstrated to induce condensation by adjusting their net charge.²⁷⁰ Future work could build on this concept of designing globular proteins for LLPS. However, rather than using a natural protein scaffold, a *de novo* globular scaffold would permit larger and more rationalisable changes to the protein sequence (Fig. 7.3c). This could be existing protein scaffolds, however, rationally designed proteins, such as the recently developed single chain 4-helix bundle, would enable a similar rational design approach as used in this thesis.³³⁸ Further, using a rationally designed scaffold would prevent inadvertent mutations to residues critical for scaffold assembly. Engineering of this scaffold for phase separation could be tractable by mutation of surface residues to introduce patchy ionic or hydrophobic interactions. In contrast to the design process used for the HERD assemblies, where mutations focused on the core *a* and *d* residues, this design would focus on the solvent exposed *b*, *c*, and *f* positions.

The successful design of an entirely *de novo*, globular protein for LLPS in cells would be the first of its kind, and a shift in focus away from relying on intrinsically disordered assemblies. Indeed, this could even help to rationalise how globular and disordered proteins differ in their propensities for phase separation.

7.2.2 Predictive computational frameworks for phase separation

Computational design of globular and oligomeric proteins has made significant advances in the last two years.³³⁹ However, there are still several areas of protein design where computational frameworks perform poorly, or don't yet exist at all. Most computationally designed oligomeric proteins are hyper-thermostable, with nM dissociation constants.³⁴⁰⁻³⁴³ Similarly, computationally designed PPI interfaces are selected for high-affinity interactions. This means that computational design of weak PPIs is still relatively unexplored, and challenging using existing frameworks that optimise ruthlessly for stability and affinity. Further, the design of PPIs within intrinsically disordered proteins is still in its infancy. Current techniques rely on high-resolution structural data, where IDRs and unstructured proteins are significantly under-represented, and as such designing proteins to interact specifically with IDRs remains extremely challenging, despite recent advances.^{118,129} Therefore, it is perhaps no surprise that computational frameworks for the design of phase-separating proteins do not yet exist.

However, significant efforts have been made to model phase separation *in silico*. Molecular dynamics simulations have been used to model changes in protein self-association

with changes in sticker and spacer affinity and valency.^{90,225} However, due to the timescale involved and the number of protein molecules that have to be modelled for self-assembly, these simulations are coarse grained, and do not have the resolution to capture detailed molecular interactions like knobs-into-holes packing of coiled coils. Notably, some recent studies have begun to explore atomistic simulations of phase separation to capture individual PPIs, which could elucidate more specific interactions, and has been demonstrated on characterised variations of natural sequences.²³⁸

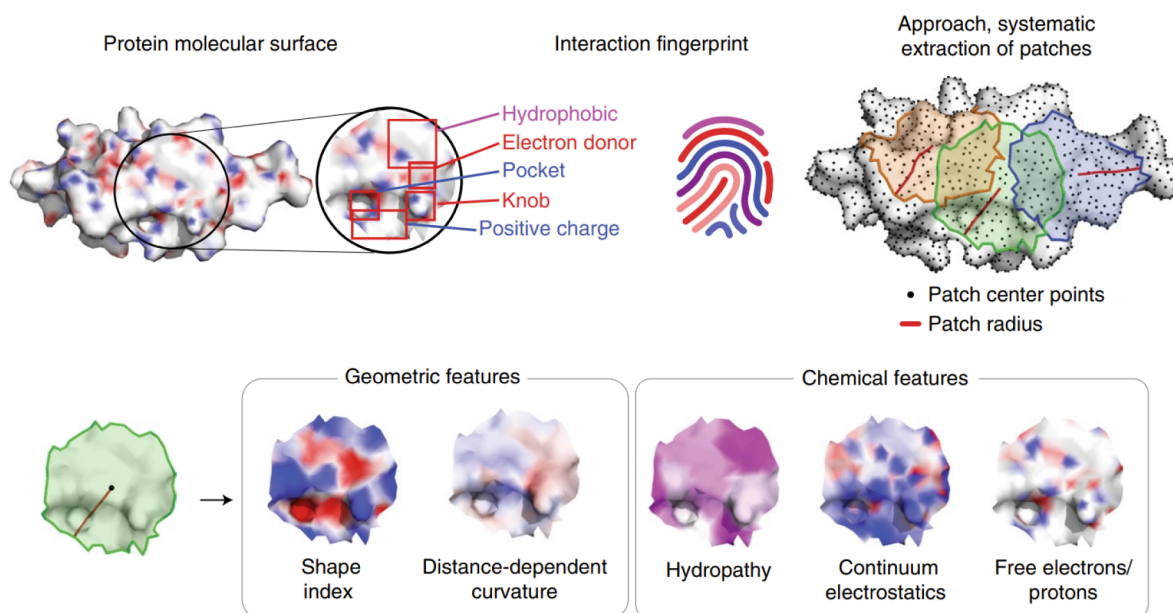


Figure 7.4: Creation of a predictive computational framework for LLPS. To create a predictive framework for phase separation, features like effective surface charge, hydrophobicity, and geometry could be combined using contemporary machine learning platforms like MaSIF. Adapted from Gainza *et al.* (2019) with permission from Springer Nature.³⁴⁴

Nonetheless, the ultimate goal is for simulations of phase separation to be predictive. Predictive modelling could guide mutation or treatment of natural proteins to alter their phase-separating behaviour. It could also provide a tractable route to the routine *de novo* design of proteins for phase separation. Such a model could be based on new machine learning based tools for predicting PPIs. A number of these tools have been released in recent years, and all use elements of a protein’s amino acid sequence and surface geometry to predict areas that are likely to be involved in PPIs (Fig. 7.4).^{344–347} While these frameworks are calibrated towards high-affinity PPIs, the re-purposing of these tools towards weak PPIs needed for phase separation could present a tractable route for the predictive modelling of proteins for phase separation. These models could be used to incorporate several elements of a protein’s properties that are implicated in LLPS, including its effective surface charge, the relative distribution of its surface charge (how patchy it is), its surface hydrophobicity, and overall surface area and geometry.

These could be tested on known phase-separating proteins, before attempting to apply it predicatively to a *de novo* scaffold.

The computational prediction of phase-separating propensity at an atomic resolution would be an entirely new tool for synthetic biologists. Using machine learning in this context would also be a complimentary marriage of computational and rational design, by applying data driven by machine learning to the rational design of PPI motifs. This would not only allow the predictive mutation of natural proteins, but also the evaluation of unknown proteins that have been implicated in LLPS, but are intractable to characterise. This could even be applied to the now available Alphafold2 models for proteins with no available structure.

Bibliography

1. Ho, B., Baryshnikova, A. & Brown, G. W. Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. *Cell Systems* **6**, 192–205 (2018).
2. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**, 4569–74 (2001).
3. Ellis, R. J. & Minton, A. P. Join the crowd. *Nature* **425**, 27–28 (2003).
4. Theillet, F.-X. *et al.* Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chemical Rev.* **114**, 6661–6714 (2014).
5. Ellis, R. J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597–604 (2001).
6. Mourão, M. A., Hakim, J. B. & Schnell, S. Connecting the dots: the effects of macromolecular crowding on cell physiology. *Biophys J.* **107**, 2761–2766 (2014).
7. Sharp, K. A. Unpacking the origins of in-cell crowding. *Proc. Natl. Acad. Sci. USA* **113**, 1684–5 (2016).
8. Watson, H. Biological membranes. *Essays Biochem.* **59**, 43–69 (2015).
9. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).
10. Clark, D. P., Pazdernik, N. J. & McGehee, M. R. Chapter 1 - cells and organisms. *Molecular Biology (Third Edition)*. 2–37 (2019).
11. Antonny, B. Composition, physical properties, and curvature. *Encyclopedia of Cell Biology*. 201–207 (2016).
12. Hohmann, T. & Dehghani, F. The cytoskeleton - a complex interacting meshwork. *Cells* **8** (2019).
13. Wickstead, B. & Gull, K. The evolution of the cytoskeleton. *J. Cell Biol.* **194**, 513–25 (2011).
14. Nath, D. The prokaryotic cytoskeleton. *Nat. Rev. Mol. Cell Biol.* **9**, s19–s19 (2010).
15. Mayer, F. Cytoskeletons in prokaryotes. *Cell Biol. Int.* **27**, 429–38 (2003).

16. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
17. Ochoa, J. M. & Yeates, T. O. Recent structural insights into bacterial microcompartment shells. *Curr. Opin. Microbiol.* **62**, 51–60 (2021).
18. Kirst, H. & Kerfeld, C. A. Bacterial microcompartments: catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering. *BMC Biol.* **17**, 79 (2019).
19. Stewart, A. M., Stewart, K. L., Yeates, T. O. & Bobik, T. A. Advances in the world of bacterial microcompartments. *Trends Biochem. Sci.* **46**, 406–416 (2021).
20. Kerfeld, C. A. & Melnicki, M. R. Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* **31**, 66–75 (2016).
21. Dank, A. *et al.* Bacterial microcompartment-dependent 1,2-propanediol utilization of *Propionibacterium freudenreichii*. *Front. Microbiol.* **12** (2021).
22. Pokhrel, A., Kang, S.-y. & Schmidt-Dannert, C. Ethanolamine bacterial microcompartments: from structure, function studies to bioengineering applications. *Curr. Opin. Microbiol.* **62**, 28–37 (2021).
23. Petit, E. *et al.* Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. *PLoS One* **8**, e54337 (2013).
24. Pederson, T. The nucleolus. *Cold Spring Harb. Perspec. Biol.* **3** (2011).
25. Boeynaems, S. *et al.* Protein phase separation: A new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).
26. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
27. Antifeeva, I. A. *et al.* Liquid–liquid phase separation as an organizing principle of intracellular space: overview of the evolution of the cell compartmentalization concept. *Cell. Mol. Life Sci.* **79**, 251 (2022).
28. Brangwynne, C. P. *et al.* Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729 (2009).
29. Hyman, A. A. & Simons, K. Beyond oil and water—phase transitions in cells. *Science* **337**, 1047 (2012).
30. Flory, P. J. *Principles of Polymer Chemistry* (Cornell University Press, 1953).
31. Brangwynne, C., Tompa, P. & Pappu, R. Polymer physics of intracellular phase transitions. *Nat. Physics* **11**, 899–904 (2015).
32. Huggins, M. L. Some properties of solutions of long-chain compounds. *J. Phys. Chem.* **46**, 151–158 (1942).
33. Flory, P. J. Thermodynamics of high polymer solutions. *J. Chem. Phys.* **10**, 51–61 (2004).

34. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
35. Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **22**, 196–213 (2021).
36. Hondele, M., Heinrich, S., De Los Rios, P. & Weis, K. Membraneless organelles: phasing out of equilibrium. *Emerg. Topics Life Sci.* **4**, 343–354 (2020).
37. Handwerger, K. E., Cordero, J. A. & Gall, J. G. Cajal bodies, nucleoli, and speckles in the xenopus oocyte nucleus have a low-density, sponge-like structure. *Mol. Biol. Cell* **16**, 202–211 (2004).
38. Strom, A. R. & Brangwynne, C. P. The liquid nucleome - phase transitions in the nucleus at a glance. *J. Cell Sci.* **132** (2019).
39. Sawyer, I. A., Bartek, J. & Dundr, M. Phase separated microenvironments inside the cell nucleus are linked to disease and regulate epigenetic state, transcription and RNA processing. *Semin. Cell Dev. Biol.* **90**, 94–103 (2019).
40. Nizami, Z., Deryusheva, S. & Gall, J. G. The Cajal body and histone locus body. *Cold Spring Harb. Perspect. Biol.* **2**, a000653 (2010).
41. Feng, Z. *et al.* Structural basis for mitotic centrosome assembly in flies. *Cell* **169**, 1078–1089.e13 (2017).
42. Raff, J. W. Phase separation and the centrosome: *A fait accompli?* *Trends Cell Biol.* **29**, 612–622 (2019).
43. Woodruff, J. B. *et al.* The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. *Cell* **169**, 1066–1077 (2017).
44. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).
45. Zhang, L. *et al.* Phase-separated subcellular compartmentation and related human diseases. *Int. J. Mol. Sci.* **23** (2022).
46. Alberti, S. & Dormann, D. Liquid–liquid phase separation in disease. *Annu. Rev. Genetics* **53**, 171–194 (2019).
47. Aguzzi, A. & Altmeyer, M. Phase separation: Linking cellular compartmentalization to disease. *Trends Cell Biol.* **26**, 547–558 (2016).
48. Uversky, V. N. Chapter 17 - MLOstasis: liquid–liquid phase separation and biomolecular condensates in cell competition, fitness, and aging. *Droplets of Life*. 485–504 (2023).
49. Xia, J. Liquid-liquid phase separation: A new perspective to understanding aging and pathogenesis. *Biosci. Trends* **16**, 359–362 (2022).

50. Zbinden, A., Pérez-Berlanga, M., De Rossi, P. & Polymenidou, M. Phase separation and neurodegenerative diseases: A disturbance in the force. *Dev. Cell* **55**, 45–68 (2020).
51. Nam, J. & Gwon, Y. Neuronal biomolecular condensates and their implications in neurodegenerative diseases. *Front. Aging Neurosci.* **15** (2023).
52. Lim, J. & Yue, Z. Neuronal aggregates: formation, clearance, and spreading. *Dev. Cell* **32**, 491–501 (2015).
53. Kempermann, G., van Praag, H. & Gage, F. H. Activity-dependent regulation of neuronal plasticity and self repair. *Prog. Brain Res.* **127**, 35–48 (2000).
54. Mofatteh, M. mRNA localization and local translation in neurons. *AIMS Neurosci.* **7**, 299–310 (2020).
55. Nagano, S. & Araki, T. Axonal transport and local translation of mRNA in neurodegenerative diseases. *Front. Mol. Neurosci.* **14** (2021).
56. Richter, J. D. & Zhao, X. The molecular biology of FMRP: new insights into fragile X syndrome. *Nat. Rev. Neurosci.* **22**, 209–222 (2021).
57. Kim, T. H. *et al.* Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* **365**, 825–829 (2019).
58. Tsang, B. *et al.* Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc. Natl. Acad. Sci.* **116**, 4218–4227 (2019).
59. Pakravan, D., Orlando, G., Bercier, V. & Van Den Bosch, L. Role and therapeutic potential of liquid-liquid phase separation in amyotrophic lateral sclerosis. *J. Mol. Cell Biol.* **13**, 15–28 (2021).
60. Milicevic, K., Rankovic, B., Andjus, P. R., Bataveljic, D. & Milovanovic, D. Emerging roles for phase separation of RNA-binding proteins in cellular pathology of ALS. *Front. Cell Dev. Biol.* **10** (2022).
61. Solomon, D. A., Smikle, R., Reid, M. J. & Mizielińska, S. Altered phase separation and cellular impact in C9orf72-linked ALS/FTD. *Front. Cell. Neurosci.* **15** (2021).
62. Babinchak, W. M. *et al.* The role of liquid–liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* **294**, 6306–6317 (2019).
63. Kwiatkowski, J., T. J. *et al.* Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–8 (2009).
64. Baron, D. M. *et al.* Amyotrophic lateral sclerosis-linked FUS/TLS alters stress granule assembly and dynamics. *Mol. Neurodegen.* **8**, 30 (2013).
65. Zhang, X. *et al.* *In vivo* stress granule misprocessing evidenced in a FUS knock-in ALS mouse model. *Brain* **143**, 1350–1367 (2020).

66. Ding, Q. *et al.* TDP-43 mutation affects stress granule dynamics in differentiated NSC-34 motoneuron-like cells. *Front. Cell Dev. Biol.* **9** (2021).
67. Gui, X. *et al.* Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* **10**, 2006 (2019).
68. Salapa, H. E., Johnson, C., Hutchinson, C., Popescu, B. F. & Levin, M. C. Dysfunctional RNA binding proteins and stress granules in multiple sclerosis. *J. Neuroimmun.* **324**, 149–156 (2018).
69. Bentmann, E. *et al.* Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* **287**, 23079–94 (2012).
70. Dormann, D. *et al.* ALS-associated fused in sarcoma (FUS) mutations disrupt transportin-mediated nuclear import. *EMBO J.* **29**, 2841–57 (2010).
71. Liu-Yesucevitz, L. *et al.* Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. *PLoS One* **5**, e13250 (2010).
72. Mackenzie, I. R. *et al.* TIA1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* **95**, 808–816.e9 (2017).
73. Murakami, T. *et al.* ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* **88**, 678–90 (2015).
74. Patel, A. *et al.* A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015).
75. Marcelo, A., Koppenol, R., de Almeida, L. P., Matos, C. A. & Nóbrega, C. Stress granules, RNA-binding proteins and polyglutamine diseases: too much aggregation? *Cell Death Disease* **12**, 592 (2021).
76. Kar, M., Posey, A. E., Dar, F., Hyman, A. A. & Pappu, R. V. Glycine-rich peptides from FUS have an intrinsic ability to self-assemble into fibers and networked fibrils. *Biochemistry* **60**, 3213–3222 (2021).
77. Dao, T. P. *et al.* ALS-linked mutations affect UBQLN2 oligomerization and phase separation in a position- and amino acid-dependent manner. *Structure* **27**, 937–951.e5 (2019).
78. Su, X. *et al.* Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595–9 (2016).
79. Spannll, S., Tereshchenko, M., Mastromarco, G. J., Ihn, S. J. & Lee, H. O. Biomolecular condensates in neurodegeneration and cancer. *Traffic* **20**, 890–911 (2019).
80. Bouchard, J. J. *et al.* Cancer mutations of the tumor suppressor SPOP disrupt the formation of active, phase-separated compartments. *Mol. Cell* **72**, 19–36.e8 (2018).

81. Heinrich, B. S., Maliga, Z., Stein, D. A., Hyman, A. A. & Whelan, S. P. J. Phase transitions drive the formation of vesicular stomatitis virus replication compartments. *mBio* **9** (2018).
82. Munder, M. C. *et al.* A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* **5** (2016).
83. Villegas, J. A., Heidenreich, M. & Levy, E. D. Molecular and environmental determinants of biomolecular condensate formation. *Nat. Chem. Bio.* **18**, 1319–1329 (2022).
84. Li, Q. *et al.* Protein databases related to liquid-liquid phase separation. *Int. J. Mol. Sci.* **21**, 6796 (2020).
85. Peran, I. & Mittag, T. Molecular structure in biomolecular condensates. *Curr. Opin. Struc. Biol.* **60**, 17–26 (2020).
86. Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *J. Biol. Chem.* **294**, 7115–7127 (2019).
87. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
88. Mittag, T. & Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* (2022).
89. Choi, J.-M., Holehouse, A. S. & Pappu, R. V. Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
90. Zheng, W. *et al.* Molecular details of protein condensates probed by microsecond long atomistic simulations. *J. Phys. Chem. B* **124**, 11671–11679 (2020).
91. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
92. Martin, E. W. & Mittag, T. Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry* **57**, 2478–2487 (2018).
93. Martin, E. W. & Holehouse, A. S. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerg. Top. Life Sci.* **4**, 307–329 (2020).
94. Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).
95. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
96. Courchaine, E. M., Lu, A. & Neugebauer, K. M. Droplet organelles? *EMBO J.* **35**, 1603–1612 (2016).

97. Schuster, B. S. *et al.* Biomolecular condensates: Sequence determinants of phase separation, microstructural organization, enzymatic activity, and material properties. *J. Phys. Chem. B* **125**, 3441–3451 (2021).
98. Schmit, J. D., Feric, M. & Dundr, M. How hierarchical interactions make membraneless organelles tick like clockwork. *Trends Biochem. Sci.* (2021).
99. Murthy, A. C. *et al.* Molecular interactions underlying liquid-liquid phase separation of the fus low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).
100. Sahli, L., Renard, D., Solé-Jamault, V., Giuliani, A. & Boire, A. Role of protein conformation and weak interactions on γ -gliadin liquid-liquid phase separation. *Sci. Rep.* **9**, 13391 (2019).
101. Fonin, A. V. *et al.* New evidence of the importance of weak interactions in the formation of PML-bodies. *Int. J. Mol. Sci.* **23** (2022).
102. Woolfson, D. N. A brief history of de novo protein design: Minimal, rational, and computational. *J. Mol. Bio.* **433**, 167160 (2021).
103. Regan, L. *et al.* Protein design: Past, present, and future. *Biopolymers* **104**, 334–350 (2015).
104. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F. & Lombardi, A. De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **68**, 779–819 (1999).
105. Woolfson, D. N. *et al.* De novo protein design: how do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* **33**, 16–26 (2015).
106. Korendovych, I. V. & DeGrado, W. F. De novo protein design, a retrospective. *Quarterly Rev. Biophys.* **53**, e3 (2020).
107. DeGrado, W. F. & Lear, J. D. Induction of peptide conformation at apolar water interfaces. 1. a study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **107**, 7684–7689 (1985).
108. Dawson, W. M., Rhys, G. G. & Woolfson, D. N. Towards functional de novo designed proteins. *Curr. Opin. Chem. Biol.* **52**, 102–111 (2019).
109. Korendovych, I. V. Rational and semirational protein design. *Methods Mol. Biol.* **1685**, 15–23 (2018).
110. Hellinga, H. W. Rational protein design: Combining theory and experiment. *Proc. Natl. Acad. Sci.* **94**, 10015–10017 (1997).
111. Woolfson, D. N. Understanding a protein fold: The physics, chemistry, and biology of α -helical coiled coils. *J. Biol. Chem.* **299**, 104579 (2023).
112. Nanda, V. *et al.* De novo design of a redox-active minimal rubredoxin mimic. *J. Am. Chem. Soc.* **127**, 5804–5 (2005).
113. Huang, P. S. *et al.* De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).

114. Vorobieva, A. A. *et al.* De novo design of transmembrane β barrels. *Science* **371** (2021).
115. Voet, A. R. D. *et al.* Computational design of a self-assembling symmetrical β -propeller protein. *Proc. Natl. Acad. Sci.* **111**, 15102–15107 (2014).
116. Thomson, A. R. *et al.* Computational design of water-soluble α -helical barrels. *Science* **346**, 485 (2014).
117. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697 (2019).
118. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
119. Bouatta, N., Sorger, P. & AlQuraishi, M. Protein structure prediction by AlphaFold2: are attention and symmetries all you need? *Acta Crystallogr. D. Struct. Biol.* **77**, 982–991 (2021).
120. Jumper, J. *et al.* Applying and improving AlphaFold at CASP14. *Proteins* **89**, 1711–1721 (2021).
121. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
122. Bertoline, L. M. F., Lima, A. N., Krieger, J. E. & Teixeira, S. K. Before and after AlphaFold2: An overview of protein structure prediction. *Front. Bioinformatics* **3** (2023).
123. Callaway, E. What’s next for AlphaFold and the AI protein-folding revolution. *Nature* **604**, 234–238 (2022).
124. Sorenson, J. M. & Head-Gordon, T. Matching simulation and experiment: a new simplified model for simulating protein folding. *J. Comput. Biol.* **7**, 469–81 (2000).
125. Leaver-Fay, A. *et al.* Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol.* **523**, 109–43 (2013).
126. Koehler Leman, J. *et al.* Better together: Elements of successful scientific software development in a distributed collaborative community. *PLoS Comput. Biol.* **16**, e1007507 (2020).
127. Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
128. Leaver-Fay, A. *et al.* Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–74 (2011).
129. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
130. Pan, X. & Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. *J. Biol. Chem.* **296**, 100558 (2021).

131. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* (2023).
132. John, I. *et al.* Illuminating protein space with a programmable generative model. *bioRxiv* 2022.12.01.518682 (2022).
133. Lupas, A. N. & Bassler, J. Coiled coils – a model system for the 21st century. *Trends Biochem. Sci.* **42**, 130–140 (2017).
134. Lupas, A. N. & Gruber, M. The structure of alpha-helical coiled coils. *Adv. Protein Chem.* **70**, 37–78 (2005).
135. Lupas, A. Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382 (1996).
136. Lupas, A. N., Bassler, J. & Dunin-Horkawicz, S. The structure and topology of α -helical coiled coils. *Subcell. Biochem.* **82**, 95–129 (2017).
137. Crick, F. The packing of α -helices: simple coiled-coils. *Acta Crys.* **6**, 689–697 (1953).
138. Crick, F. The fourier transform of a coiled-coil. *Acta Crys.* **6**, 685–689 (1953).
139. Woolfson, D. N. Coiled-coil design: Updated and upgraded. *Subcell. Biochem.* **82**, 35–61 (2017).
140. Rhys, G. G. *et al.* Navigating the structural landscape of de novo α -helical bundles. *J. Am. Chem. Soc.* **141**, 8787–8797 (2019).
141. Dawson, W. M. *et al.* Coiled coils 9-to-5: rational de novo design of α -helical barrels with tunable oligomeric states. *Chem. Sci.* **12**, 6923–6928 (2021).
142. Fletcher, J. M. *et al.* A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Syn. Bio.* **1**, 240–250 (2012).
143. Edgell, C. L., Savery, N. J. & Woolfson, D. N. Robust de novo-designed homotetrameric coiled coils. *Biochemistry* **59**, 1087–1092 (2020).
144. Wood, C. W. & Woolfson, D. N. Ccbuilder 2.0: Powerful and accessible coiled-coil modeling. *Protein Sci.* **27**, 103–111 (2018).
145. Lebar, T., Lainšček, D., Merljak, E., Aupič, J. & Jerala, R. A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nat. Chem. Bio.* (2020).
146. Scott, A. J. *et al.* Constructing ion channels from water-soluble α -helical barrels. *Nat. Chem.* **13**, 643–650 (2021).
147. Fletcher, J. M. *et al.* N@a and N@d: Oligomer and partner specification by asparagine in coiled-coil interfaces. *ACS Chem. Biol.* **12**, 528–538 (2017).
148. Bromley, E. H. C., Sessions, R. B., Thomson, A. R. & Woolfson, D. N. Designed α -helical tectons for constructing multicomponent synthetic biological systems. *J. Am. Chem. Soc.* **131**, 928–930 (2009).

149. Edgell, C. L., Savery, N. J. & Woolfson, D. N. Robust de novo-designed homotetrameric coiled coils. *Biochemistry* **59**, 1087–1092 (2020).
150. Edgell, C. L., Smith, A. J., Beesley, J. L., Savery, N. J. & Woolfson, D. N. De novo designed protein-interaction modules for in-cell applications. *ACS Syn. Bio.* (2020).
151. Dawson, W. M. *et al.* Structural resolution of switchable states of a de novo peptide assembly. *Nat. Commun.* **12**, 1530 (2021).
152. Lee, M. J. *et al.* Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat. Chem. Bio.* **14**, 142–147 (2018).
153. Matthaei, J. F. *et al.* Designing two-dimensional protein arrays through fusion of multimers and interface mutations. *Nano Letters* **15**, 5235–5239 (2015).
154. Mehrban, N. *et al.* Functionalized α -helical peptide hydrogels for neural tissue engineering. *ACS Biomater. Sci. Eng.* **1**, 431–439 (2015).
155. Signarvic, R. S. & DeGrado, W. F. De novo design of a molecular switch: Phosphorylation-dependent association of designed peptides. *J. Mol. Bio.* **334**, 1–12 (2003).
156. Langan, R. A. *et al.* De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019).
157. Xu, C. *et al.* Computational design of transmembrane pores. *Nature* **585**, 129–134 (2020).
158. Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013).
159. Courbet, A. *et al.* Computational design of mechanically coupled axle-rotor protein assemblies. *Science* **376**, 383–390 (2022).
160. Divine, R. *et al.* Designed proteins assemble antibodies into modular nanocages. *Science* **372**, eabd9994 (2021).
161. Padilla, J. E., Colovos, C. & Yeates, T. O. Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci.* **98**, 2217–2221 (2001).
162. Sasaki, E. *et al.* Structure and assembly of scalable porous protein cages. *Nat. Commun.* **8**, 14663 (2017).
163. Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365–1368 (2015).
164. Yeates, T. O., Liu, Y. & Laniado, J. The design of symmetric protein nanomaterials comes of age in theory and practice. *Curr. Opin. Struc. Biol.* **39**, 134–143 (2016).
165. Laniado, J. *et al.* Geometric lessons and design strategies for nanoscale protein cages. *ACS Nano* **15**, 4277–4286 (2021).

166. Yeates, T. O. & Padilla, J. E. Designing supramolecular protein assemblies. *Curr. Opin. Struc. Biol.* **12**, 464–470 (2002).
167. MacPhee, C. E. & Dobson, C. M. Formation of mixed fibrils demonstrates the generic nature and potential utility of amyloid nanostructures. *J. Am. Chem. Soc.* **122**, 12707–12713 (2000).
168. Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* **99**, 2754–9 (2002).
169. Ryadnov, M. G. & Woolfson, D. N. Engineering the morphology of a self-assembling protein fibre. *Nat. Materials* **2**, 329–332 (2003).
170. Pandya, M. J. *et al.* Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry* **39**, 8728–34 (2000).
171. Banwell, E. F. *et al.* Rational design and application of responsive α -helical peptide hydrogels. *Nat. Materials* **8**, 596–600 (2009).
172. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
173. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
174. Brouwer, P. J. M. *et al.* Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle. *Nat. Commun.* **10**, 4272 (2019).
175. Ueda, G. *et al.* Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *eLife* **9**, e57659 (2020).
176. Walls, A. C. *et al.* Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382.e17 (2020).
177. Edwardson, T. G. W. *et al.* Protein cages: From fundamentals to advanced applications. *Chemical Rev.* **122**, 9145–9197 (2022).
178. Jakobson, C. M., Slininger Lee, M. F. & Tullman-Ercek, D. De novo design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment. *Protein Sci.* **26**, 1086–1092 (2017).
179. Hastings, R. L. & Boeynaems, S. Designer condensates: a toolkit for the biomolecular architect. *J. Mol. Bio.* 166837 (2021).
180. Schuster, B. S. *et al.* Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9**, 2985 (2018).
181. Wei, S.-P. *et al.* Formation and functionalization of membraneless compartments in *Escherichia coli*. *Nat. Chem. Bio.* (2020).
182. Bracha, D. *et al.* Mapping local and global liquid phase behavior in living cells using photo-oligomerizable seeds. *Cell* **175**, 1467–1480.e13 (2018).

183. Reichheld, S. E., Muiznieks, L. D., Keeley, F. W. & Sharpe, S. Direct observation of structure and dynamics during phase separation of an elastomeric protein. *Proc. Natl. Acad. Sci. USA* **114**, E4408–e4415 (2017).
184. Dine, E., Gil, A. A., Uribe, G., Brangwynne, C. P. & Toettcher, J. E. Protein phase separation provides long-term memory of transient spatial stimuli. *Cell Syst.* **6**, 655–663.e5 (2018).
185. Reed, E. H., Schuster, B. S., Good, M. C. & Hammer, D. A. SPLIT: Stable protein coacervation using a light induced transition. *ACS Syn. Bio.* (2020).
186. Garabedian, M. V. *et al.* Protein condensate formation via controlled multimerization of intrinsically disordered sequences. *Biochemistry* (2022).
187. Qian, Z.-G., Huang, S.-C. & Xia, X.-X. Synthetic protein condensates for cellular and metabolic engineering. *Nat. Chem. Bio.* (2022).
188. Shin, Y. *et al.* Spatiotemporal control of intracellular phase transitions using light-activated optodroplets. *Cell* **168**, 159–171.e14 (2017).
189. Hernández-Candia, C. N., Pearce, S. & Tucker, C. L. A modular tool to query and inducibly disrupt biomolecular condensates. *Nat. Commun.* **12**, 1809 (2021).
190. Garabedian, M. V. *et al.* Designer membraneless organelles sequester native factors for control of cell behavior. *Nat. Chem. Bio.* **17**, 998–1007 (2021).
191. Lasker, K. *et al.* The material properties of a bacterial-derived biomolecular condensate tune biological function in natural and synthetic systems. *Nat. Commun.* **13**, 5643 (2022).
192. Lasker, K. *et al.* Selective sequestration of signalling proteins in a membraneless organelle reinforces the spatial regulation of asymmetry in *Caulobacter crescentus*. *Nat. Microbiol.* **5**, 418–429 (2020).
193. Zhao, E. M. *et al.* Light-based control of metabolic flux through assembly of synthetic organelles. *Nat. Chem. Bio.* **15**, 589–597 (2019).
194. Wang, Y. *et al.* Phase-separated multienzyme compartmentalization for terpene biosynthesis in a prokaryote. *Angew. Chem. Int. Ed. Engl.* **61**, e202203909 (2022).
195. Reinkemeier, C. D., Girona, G. E. & Lemke, E. A. Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science* **363**, eaaw2644 (2019).
196. Reinkemeier, C. D. & Lemke, E. A. Dual film-like organelles enable spatial separation of orthogonal eukaryotic translation. *Cell* **184**, 4886–4903.e21 (2021).
197. Reinkemeier, C. D. & Lemke, E. A. Condensed, microtubule-coating thin organelles for orthogonal translation in mammalian cells. *J. Mol. Bio.* **434**, 167454 (2022).
198. Dzuricky, M., Rogers, B. A., Shahid, A., Cremer, P. S. & Chilkoti, A. De novo engineering of intracellular condensates using artificial disordered proteins. *Nat. Chem.* (2020).

199. Heidenreich, M. *et al.* Designer protein assemblies with tunable phase diagrams in living cells. *Nat. Chem. Bio.* **16**, 939–945 (2020).
200. Yoshikawa, M., Yoshii, T., Ikuta, M. & Tsukiji, S. Synthetic protein condensates that inducibly recruit and release protein activity in living cells. *J. Am. Chem. Soc.* **143**, 6434–6446 (2021).
201. Yoshikawa, M. & Tsukiji, S. Modularly built synthetic membraneless organelles enabling targeted protein sequestration and release. *Biochemistry* **60**, 3273–3276 (2021).
202. Zhang, Q. *et al.* Visualizing dynamics of cell signaling in vivo with a phase separation-based kinase reporter. *Mol. Cell* **69**, 334–346.e4 (2018).
203. Liu, J. *et al.* Multifaceted cargo recruitment and release from artificial membraneless organelles. *Small* 2201721 (2022).
204. Hong, K., Song, D. & Jung, Y. Behavior control of membrane-less protein liquid condensates with metal ion-induced phase separation. *Nat. Commun.* **11**, 5554 (2020).
205. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
206. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
207. Cross, S. J. MIA DeepLearning (2021).
208. Cross, S. J. ModularImageAnalysis (MIA) (v1.0.3) (2022).
209. Umorin, M. Stack focuser (2022).
210. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
211. Spahn, C. *et al.* DeepBacs for multi-task bacterial image analysis using open-source deep learning approaches. *Commun. Biol.* **5**, 688 (2022).
212. Tinevez, J.-Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
213. Das Subash, C., Panda, D., Nayak, D. & Pattnaik Asit, K. Biarsenical labeling of vesicular stomatitis virus encoding tetracysteine-tagged M protein allows dynamic imaging of M protein and virus uncoating in infected cells. *J. Virology* **83**, 2611–2622 (2009).
214. Fang, M.-Y. *et al.* High crude violacein production from glucose by *Escherichia coli* engineered with interactive control of tryptophan pathway and violacein biosynthetic pathway. *Microb. Cell Factories* **14**, 8 (2015).
215. Seixas de Melo, J., Moura, A. P. & Melo, M. J. Photophysical and spectroscopic studies of indigo derivatives in their keto and leuco forms. *J. Phys. Chem. A* **108**, 6975–6981 (2004).

216. Alfieri, A., Malito, E., Orru, R., Fraaije, M. W. & Mattevi, A. Revealing the moonlighting role of NADP in the structure of a flavin-containing monooxygenase. *Proc. Natl. Acad. Sci.* **105**, 6572–6577 (2008).
217. Harris, A. P. & Phillips, R. S. Benzimidazole analogs of (L)-tryptophan are substrates and inhibitors of tryptophan indole lyase from *Escherichia coli*. *Febs J.* **280**, 1807–17 (2013).
218. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*. 571–607 (2005).
219. Wood, C. W. *et al.* ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design. *Bioinformatics* **33**, 3043–3050 (2017).
220. Wood, C. W. & Woolfson, D. N. CCBUILDER 2.0: Powerful and accessible coiled-coil modeling. *Protein Sci.* **27**, 103–111 (2018).
221. Erdős, G., Pajkos, M. & Dosztányi, Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297–W303 (2021).
222. Erdős, G. & Dosztányi, Z. Analyzing protein disorder with IUPred2A. *Curr. Prot. Bioinformatics* **70**, e99 (2020).
223. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
224. Hilditch, A. T. *et al.* Assembling membraneless organelles from de novo designed proteins. *Nature Chemistry* (2023).
225. Paloni, M., Bailly, R., Ciandrini, L. & Barducci, A. Unraveling molecular interactions in liquid–liquid phase separation of disordered proteins by atomistic simulations. *J. Phys. Chem. B* **124**, 9009–9016 (2020).
226. Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. & Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Comp. Biol.* **14**, e1005941 (2018).
227. Espinosa, J. R. *et al.* Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. *Proc. Natl. Acad. Sci.* **117**, 13238 (2020).
228. Sanders, D. W. *et al.* Competing protein-RNA interaction networks control multi-phase intracellular organization. *Cell* **181**, 306–324.e28 (2020).
229. Falkenberg, C. V., Blinov, M. L. & Loew, L. M. Pleomorphic ensembles: formation of large clusters composed of weakly interacting multivalent molecules. *Biophys. J.* **105**, 2451–60 (2013).
230. Dignon, G. L., Zheng, W., Best, R. B., Kim, Y. C. & Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci.* **115**, 9929–9934 (2018).

231. Nandi, S. K., Österle, D., Heidenreich, M., Levy, E. D. & Safran, S. A. Affinity and valence impact the extent and symmetry of phase separation of multivalent proteins. *Phys. Rev. Lett.* **129**, 128102 (2022).
232. Banjade, S. & Rosen, M. K. Phase transitions of multivalent proteins can promote clustering of membrane receptors. *eLife* **3**, e04123 (2014).
233. Banjade, S. *et al.* Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc. Natl. Acad. Sci. USA* **112**, E6426–35 (2015).
234. Schulze, J. *et al.* Phase behavior of lysozyme solutions in the liquid–liquid phase coexistence region at high hydrostatic pressures. *Phys. Chem. Chem. Phys.* **18**, 14252–14256 (2016).
235. Dawson, W. M. *et al.* Coiled coils 9-to-5: rational de novo design of α -helical barrels with tunable oligomeric states. *Chem. Sci.* **12**, 6923–6928 (2021).
236. Thomas, F., Boyle, A. L., Burton, A. J. & Woolfson, D. N. A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime. *J. Am. Chem. Soc.* **135**, 5161–5166 (2013).
237. Ford, L. K. & Fioriti, L. Coiled-coil motifs of RNA-binding proteins: Dynamicity in RNA regulation. *Front. Cell Dev. Biol.* **8** (2020).
238. Ramirez, D. A., Hough, L. E. & Shirts, M. R. Coiled-coil domains are sufficient to drive liquid-liquid phase separation of proteins in molecular models. *bioRxiv* (2023).
239. Newton, J. C. *et al.* Phase separation of the LINE-1 ORF1 protein is mediated by the N-terminus and coiled-coil domain. *Biophys. J.* **120**, 2181–2191 (2021).
240. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *2-D Proteome Analysis Protocols*. 531–552 (1999).
241. Martin, F. *Exploring the dynamic and conformational landscape of α -helical peptide assemblies*. Thesis, University of Bristol (2022).
242. Schavemaker, P. E., Śmigiel, W. M. & Poolman, B. Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome. *eLife* **6**, e30084 (2017).
243. Link, A. J., Robison, K. & Church, G. M. Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* **18**, 1259–313 (1997).
244. Cherstvy, A. G. Positively charged residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups. *J. Phys. Chem. B* **113**, 4242–7 (2009).
245. Nadassy, K., Wodak, S. J. & Janin, J. Structural features of protein-nucleic acid recognition sites. *Biochemistry* **38**, 1999–2017 (1999).

246. Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).
247. Harmon, T. S., Holehouse, A. S. & Pappu, R. V. Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. *New J. Phys.* **20**, 045002 (2018).
248. Theillet, F.-X. *et al.* The alphabet of intrinsic disorder. *Intrinsically Disord. Prot.* **1**, e24360 (2013).
249. Uversky, V. N. The alphabet of intrinsic disorder. *Intrinsically Disord. Prot.* **1**, e24684 (2013).
250. Balleza, E., Kim, J. M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods* **15**, 47–51 (2018).
251. Pace, C. N. & Scholtz, J. M. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75**, 422–427 (1998).
252. Myers Jeffrey, K., Pace, C. N. & Scholtz, J. M. A direct comparison of helix propensity in proteins and peptides. *Proc. Natl. Acad. Sci.* **94**, 2833–2837 (1997).
253. Dong, H. & Hartgerink, J. D. Short homodimeric and heterodimeric coiled coils. *Biomacromolecules* **7**, 691–695 (2006).
254. Leonard, D. J. *et al.* Scalable synthesis and coupling of quaternary α -arylated amino acids: α -aryl substituents are tolerated in α -helical peptides. *Chem. Sci.* **12**, 9386–9390 (2021).
255. Reichheld, S. E. *et al.* Conformational transitions of the cross-linking domains of elastin during self-assembly. *J. Biol. Chem.* **289**, 10057–10068 (2014).
256. Rauscher, S. & Pomès, R. The liquid structure of elastin. *eLife* **6**, e26526 (2017).
257. Buck, M. Trifluoroethanol and colleagues: cosolvents come of age. recent studies with peptides and proteins. *Quarterly Rev. Biophys.* **31**, 297–355 (1998).
258. Reichheld, S. E., Muiznieks, L. D., Keeley, F. W. & Sharpe, S. Direct observation of structure and dynamics during phase separation of an elastomeric protein. *Proc. Natl. Acad. Sci.* **114**, E4408–E4415 (2017).
259. van den Berg, B., Ellis, R. J. & Dobson, C. M. Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* **18**, 6927–6933 (1999).
260. André, A. A. M. & Spruijt, E. Liquid-liquid phase separation in crowded environments. *Int. J. Mol. Sci.* **21**, 5908 (2020).
261. Mohsen-Nia, M., Rasa, H. & Modarress, H. Cloud-point measurements for (water + poly(ethylene glycol) + salt) ternary mixtures by refractometry method. *J. Chem. Eng. Data* **51**, 1316–1320 (2006).

262. Elsheikh, M. A., Elnaggar, Y. S., Gohar, E. Y. & Abdallah, O. Y. Nanoemulsion liquid pre-concentrates for raloxifene hydrochloride: optimization and in vivo appraisal. *Int. J. Nanomedicine* **7**, 3787–802 (2012).
263. Cinar, H. & Winter, R. The effects of cosolutes and crowding on the kinetics of protein condensate formation based on liquid–liquid phase separation: a pressure-jump relaxation study. *Sci. Rep.* **10**, 17245 (2020).
264. Lu, J., Carpenter, K., Li, R. J., Wang, X. J. & Ching, C. B. Cloud-point temperature and liquid-liquid phase separation of supersaturated lysozyme solution. *Biophys. Chem.* **109**, 105–12 (2004).
265. McManus, J. J., Charbonneau, P., Zaccarelli, E. & Asherie, N. The physics of protein self-assembly. *Curr. Opin. Colloid Interface Sci.* **22**, 73–79 (2016).
266. Muzzopappa, F. *et al.* Detecting and quantifying liquid–liquid phase separation in living cells by model-free calibrated half-bleaching. *Nat. Commun.* **13**, 7787 (2022).
267. Rademacher, D. J., Cabe, M. & Bakowska, J. C. Fluorescence recovery after photobleaching of yellow fluorescent protein tagged p62 in aggresome-like induced structures. *J. Vis. Exp.* (2019).
268. Alshareedah, I., Kaur, T. & Banerjee, P. R. Chapter six - methods for characterizing the material properties of biomolecular condensates. *Methods Enzymology.* **646**, 143–183 (2021).
269. Titus, A. R. & Kooijman, E. E. Chapter two - current methods for studying intracellular liquid-liquid phase separation. *Current Topics in Membranes.* **88**, 55–73 (2021).
270. Yeong, V., Werth, E. G., Brown, L. M. & Obermeyer, A. C. Formation of biomolecular condensates in bacteria by tuning protein electrostatics. *ACS Cent. Sci.* **6**, 2301–2310 (2020).
271. Abbondanzieri, E. A. & Meyer, A. S. More than just a phase: the search for membraneless organelles in the bacterial cytoplasm. *Curr. Genetics* **65**, 691–694 (2019).
272. Liu, J., Zhorabek, F., Dai, X., Huang, J. & Chau, Y. Minimalist design of an intrinsically disordered protein-mimicking scaffold for an artificial membraneless organelle. *ACS Central Sci.* **8**, 493–500 (2022).
273. Lau, S. Y., Taneja, A. K. & Hodges, R. S. Synthesis of a model protein of defined secondary and quaternary structure. effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J. Biol. Chem.* **259**, 13253–61 (1984).
274. Tabaka, M., Kalwarczyk, T., Szymanski, J., Hou, S. & Holyst, R. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. *Front. Phys.* **2** (2014).
275. Mika, J. T. & Poolman, B. Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotech.* **22**, 117–126 (2011).

276. Altenburg, W. J. *et al.* Programmed spatial organization of biomacromolecules into discrete, coacervate-based protocells. *Nat. Commun.* **11**, 6282 (2020).
277. Kamagata, K. *et al.* Molecular principles of recruitment and dynamics of guest proteins in liquid droplets. *Sci. Rep.* **11**, 19323 (2021).
278. Nakashima, K. K., Vibhute, M. A. & Spruijt, E. Biomolecular chemistry in liquid phase separated compartments. *Front. Mol. Biosci.* **6**, 21 (2019).
279. Guan, M. *et al.* Incorporation and assembly of a light-emitting enzymatic reaction into model protein condensates. *Biochemistry* **60**, 3137–3151 (2021).
280. Prouteau, M. & Loewith, R. Regulation of cellular metabolism through phase separation of enzymes. *Biomolecules* **8** (2018).
281. Alberti, S. The wisdom of crowds: regulating cell function through condensed states of living matter. *J. Cell Sci.* **130**, 2789–2796 (2017).
282. Kojima, T. & Takayama, S. Membraneless compartmentalization facilitates enzymatic cascade reactions and reduces substrate inhibition. *ACS App. Mat. Interfaces* **10**, 32782–32791 (2018).
283. O’Flynn, B. G. & Mittag, T. The role of liquid–liquid phase separation in regulating enzyme activity. *Curr. Opin. Cell Biol.* **69**, 70–79 (2021).
284. Bouchard, J. J. *et al.* Cancer mutations of the tumor suppressor spop disrupt the formation of active, phase-separated compartments. *Mol. Cell* **72**, 19–36.e8 (2018).
285. Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci.* **112**, 7189–7194 (2015).
286. Popielec, A., Ostrowska, N., Wojciechowska, M., Feig, M. & Trylska, J. Crowded environment affects the activity and inhibition of the NS3/4A protease. *Biochimie* **176**, 169–180 (2020).
287. Fabara, A. N. & Fraaije, M. W. Production of indigo through the use of a dual-function substrate and a bifunctional fusion enzyme. *Enzyme Microb. Tech.* **142**, 109692 (2020).
288. Myhrvold, C., Polka, J. K. & Silver, P. A. Synthetic lipid-containing scaffolds enhance production by colocalizing enzymes. *ACS Syn. Bio.* **5**, 1396–1403 (2016).
289. Hoffmann, C. *et al.* Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nat. Protoc.* **5**, 1666–77 (2010).
290. Irtegun, S., Ramdzan, Y. M., Mulhern, T. D. & Hatters, D. M. ReAsH/FlAsH labeling and image analysis of tetracysteine sensor proteins in cells. *J. Vis. Exp.* (2011).
291. Cho, H. J. *et al.* Structural and functional analysis of bacterial flavin-containing monooxygenase reveals its ping-pong-type reaction mechanism. *J. Struc. Biol.* **175**, 39–48 (2011).

292. Ado, G. *et al.* Discovery of a phase-separating small molecule that selectively sequesters tubulin in cells. *Chem. Sci.* **13**, 5760–5766 (2022).
293. Wang, Z. *et al.* Material properties of phase-separated TFEB condensates regulate the autophagy-lysosome pathway. *J. Cell Biol.* **221** (2022).
294. Wang, Z., Lou, J. & Zhang, H. Essence determines phenomenon: Assaying the material properties of biological condensates. *J. Biol. Chem.* **298**, 101782 (2022).
295. Castellana, M. *et al.* Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotech.* **32**, 1011–1018 (2014).
296. Linsenmeier, M. *et al.* Dynamic arrest and aging of biomolecular condensates are modulated by low-complexity domains, RNA and biochemical activity. *Nat. Commun.* **13**, 3030 (2022).
297. Lin, Y., Protter, D. S., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–19 (2015).
298. Kanaan, N. M., Hamel, C., Grabinski, T. & Combs, B. Liquid-liquid phase separation induces pathogenic tau conformations in vitro. *Nat. Commun.* **11**, 2809 (2020).
299. Babinchak, W. M. *et al.* The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* **294**, 6306–6317 (2019).
300. Carey, J. L. & Guo, L. Liquid-liquid phase separation of TDP-43 and FUS in physiology and pathology of neurodegenerative diseases. *Front. Mol. Biosci.* **9** (2022).
301. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–6 (2002).
302. Su, W. W. Fluorescent proteins as tools to aid protein production. *Microb. Cell Factories* **4**, 12 (2005).
303. Hodgson, L., Verkade, P. & Yamauchi, Y. Correlative light and electron microscopy of influenza virus entry and budding. *Methods Mol. Biol.* **1836**, 237–260 (2018).
304. Mileykovskaya, E., Sun, Q., Margolin, W. & Dowhan, W. Localization and function of early cell division proteins in filamentous *Escherichia coli* cells lacking phosphatidylethanolamine. *J. Bacteriol.* **180**, 4252–7 (1998).
305. Petka, W. A., Harden, J. L., McGrath, K. P., Wirtz, D. & Tirrell, D. A. Reversible hydrogels from self-assembling artificial proteins. *Science* **281**, 389–392 (1998).
306. Kreplak, L., Doucet, J., Dumas, P. & Briki, F. New aspects of the alpha-helix to beta-sheet transition in stretched hard alpha-keratin fibers. *Biophys. J.* **87**, 640–7 (2004).

307. Ke, P. C. *et al.* Half a century of amyloids: past, present and future. *Chem. Soc. Rev.* **49**, 5473–5509 (2020).
308. Burgess, N. C. *et al.* Modular design of self-assembling peptide-based nanotubes. *J. Am. Chem. Soc.* **137**, 10554–10562 (2015).
309. Wu, H. & Fuxreiter, M. The structure and dynamics of higher-order assemblies: Amyloids, signalosomes, and granules. *Cell* **165**, 1055–1066 (2016).
310. Ciani, B., Hutchinson, E. G., Sessions, R. B. & Woolfson, D. N. A designed system for assessing how sequence affects α to β conformational transitions in proteins*. *J. Biol. Chem.* **277**, 10150–10155 (2002).
311. Bendit, E. G. A quantitative X-ray diffraction study of the alpha-beta transformation in wool keratin. *Textile Res. J.* **30**, 547–555 (1960).
312. Monterroso, B. *et al.* Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor slmA. *EMBO Rep.* **20**, e45946 (2019).
313. Azaldegui, C. A., Vecchiarelli, A. G. & Biteen, J. S. The emergence of phase separation as an organizing principle in bacteria. *Biophys. J.* (2020).
314. Tomares, D. T., Whitlock, S., Mann, M., DiBernardo, E. & Childers, W. S. Repurposing peptide nanomaterials as synthetic biomolecular condensates in bacteria. *ACS Syn. Bio.* (2022).
315. Jin, X. *et al.* Membraneless organelles formed by liquid-liquid phase separation increase bacterial fitness. *Sci. Adv.* **7**, eabh2929 (2021).
316. Basile, W., Salvatore, M., Bassot, C. & Elofsson, A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput. Biol.* **15**, e1007186 (2019).
317. Tabaka, M., Kalwarczyk, T., Szymanski, J., Hou, S. & Holyst, R. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. *Front. Phys.* **2** (2014).
318. Schramm, F. D., Schroeder, K. & Jonas, K. Protein aggregation in bacteria. *FEMS Microbiol. Rev.* **44**, 54–72 (2020).
319. Yogurtcu, O. N. & Johnson, M. E. Cytosolic proteins can exploit membrane localization to trigger functional assembly. *PLOS Comp. Biol.* **14**, e1006031 (2018).
320. Kramer, D. A., Piper, H. K. & Chen, B. WASP family proteins: Molecular mechanisms and implications in human disease. *Eur. J. Cell Biol.* **101**, 151244 (2022).
321. Case, L. B., Zhang, X., Ditlev, J. A. & Rosen, M. K. Stoichiometry controls activity of phase-separated clusters of actin signaling proteins. *Science* **363**, 1093–1097 (2019).
322. Yuan, B. *et al.* Wiskott-Aldrich syndrome protein forms nuclear condensates and regulates alternative splicing. *Nat. Commun.* **13**, 3646 (2022).

323. Tillu, V. A. *et al.* Cavin1 intrinsically disordered domains are essential for fuzzy electrostatic interactions and caveola formation. *Nat. Commun.* **12**, 931 (2021).
324. Zhao, Y. G. & Zhang, H. Phase separation in membrane biology: The interplay between membrane-bound organelles and membraneless condensates. *Dev. Cell* (2020).
325. Ditlev, J. A. Membrane-associated phase separation: organization and function emerge from a two-dimensional milieu. *Journal of Mol. Cell Biology* (2021).
326. Su, W.-C. *et al.* Kinetic control of shape deformations and membrane phase separation inside giant vesicles. *Nat. Chem.* (2023).
327. Jiang, Y., Lookman, T. & Saxena, A. Phase separation and shape deformation of two-phase membranes. *Phys. Rev. E. Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* **61**, R57–60 (2000).
328. Lang, V. *et al.* Fyn membrane localization is necessary to induce the constitutive tyrosine phosphorylation of Sam68 in the nucleus of T lymphocytes. *J. Immunol.* **162**, 7224–32 (1999).
329. Michaelson, D. *et al.* Postprenylation CAAX processing is required for proper localization of Ras but not Rho GTPases. *Mol. Biol. Cell* **16**, 1606–16 (2005).
330. Anderson, P. & Kedersha, N. RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat. Rev. Mol. Cell Biol.* **10**, 430–436 (2009).
331. Narlikar, G. J. Phase-separation in chromatin organization. *J. Biosci.* **45** (2020).
332. Edun, D. N., Flanagan, M. R. & Serrano, A. L. Does liquid–liquid phase separation drive peptide folding? *Chem. Sci.* **12**, 2474–2479 (2021).
333. O’Flynn, B. G. & Mittag, T. The role of liquid-liquid phase separation in regulating enzyme activity. *Curr. Opin. Cell Biol.* **69**, 70–79 (2021).
334. Yeates, T. O. Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu. Rev. Biophys.* **46**, 23–42 (2017).
335. Li, M. *et al.* Controlling synthetic membraneless organelles by a red-light-dependent singlet oxygen-generating protein. *Nat. Commun.* **13**, 3197 (2022).
336. Qian, Z.-G., Huang, S.-C. & Xia, X.-X. Synthetic protein condensates for cellular and metabolic engineering. *Nat. Chem. Bio.* **18**, 1330–1340 (2022).
337. Shapiro, D. M., Ney, M., Eghtesadi, S. A. & Chilkoti, A. Protein phase separation arising from intrinsic disorder: First-principles to bespoke applications. *J. Phys. Chem. B* **125**, 6740–6759 (2021).
338. Naudin, E. A. *et al.* From peptides to proteins: coiled-coil tetramers to single-chain 4-helix bundles. *Chem. Sci.* **13**, 11330–11340 (2022).
339. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

340. Polizzi, N. F. *et al.* De novo design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017).
341. Kimura, N., Mochizuki, K., Umezawa, K., Hecht, M. H. & Arai, R. Hyperstable de novo protein with a dimeric bisecting topology. *ACS Syn. Bio.* **9**, 254–259 (2020).
342. Anil, B., Craig-Schapiro, R. & Raleigh, D. P. Design of a hyperstable protein by rational consideration of unfolded state interactions. *J. Am. Chem. Soc.* **128**, 3144–5 (2006).
343. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* **535**, 136–9 (2016).
344. Gainza, P. *et al.* Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **17**, 184–192 (2020).
345. Gainza, P. *et al.* De novo design of protein interactions with learned surface fingerprints. *Nature* **617**, 176–184 (2023).
346. Bennett, N. R. *et al.* Improving de novo protein binder design with deep learning. *Nat. Commun.* **14**, 2625 (2023).
347. Krapp, L. F., Abriata, L. A., Cortés Rodriguez, F. & Dal Peraro, M. PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nat. Commun.* **14**, 2175 (2023).

Chapter 8

Appendix

HERD #	Helical region 1 <i>gabcdef gabcdef EIAAIKW EIAAIKE G</i>	Linker	Helical region 2 <i>gabcdef gabcdef EIQWQLK EIQWQLK EIQWQLK G</i>
0	G EIAAIKE EIAAIKE EIAAIKW EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQWQLK EIQWQLK EIQEQLK G
1.1	G EIAAIKE EIAAIKW EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQWQLK EIQEQLK G
2.1	G EAAAAIKE EAAAAIKW EAAAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQAQ EIQWQAQ EIQEQAQ G
2.2	G IKE EAAAAIKW EAAAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EIQWQAQ EIQEQAQ G
2.3	G EISSIKE EISSIKW EISSIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQWQLK EIQEQLK G
2.4	G EISSIKE EISSIKW EASSIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQWQLK EIQEQAQ G
2.5	G EAAAAIKW EAAAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQWQAQ EIQEQAQ G
2.6	G EASSIKE EASSIKE EASSIKW EASSIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQAQ EIQEQAQ EIQWQAQ EIQEQAQ G
2.7	G EISSIKE EASSIKW EASSIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQWQAQ EIQEQAQ G
2.8	G EASSIKE EASSIKW EASSIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQAQ EIQWQAQ EIQEQAQ G
3.1	G IKE EAAAAIKW EAAAAIKE G	ASPEPQPKPSGDPQSKQTPEASPRP QPEPSGKPKQSEQTTPKASPEPQPKPS	G QAK EIQWQAQ EIQEQAQ G
3.2	G IKE EAAAAIKW EAAAAIKE G	ASESQKPS	G QAK EIQWQAQ EIQEQAQ G
3.3	G IKE EAAAAIKW EAAAAIKE G	ASEPKQSDPKGDPKQSEKQKSESR	G QAK EIQWQAQ EIQEQAQ G
3.4	G IKE EAAAAIKW EAAAAIKE G	ASPTSYPAPSWAPQYLQTPGPSVSQ	G QAK EIQWQAQ EIQEQAQ G
Ctrl1	G EAAAAAKE EAAAAAKW EAAAAAKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQAQ EIQWQAQ EIQEQAQ G
Ctrl2	G EAAAAAKE EAAAAAKW EAAAAAKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EAQEQAQ EAQWQAQ EAQEQAQ G
Ctrl3	G KEI EAAAAIKW EAAAKIE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK IEQWQAQ EEQIQAK G
Ctrl4	G PKE EAAAAPKW EAAAPKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EPQWQAQ EPQEQAK G
Ctrl5	G GKE EAAAAGKW EAAAGKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EGQWQAQ EGQEQAK G
Ctrl6	G IKE EAPAIKW EAPAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EIQQQAQ EIQEPQAK G
Ctrl7	G IKE EAGAIKW EAGAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EIQQQAQ EIQEGQAK G

Table 8.1: Sequences of the designed HERD proteins used in Chapter 3.

HERD #	Helical region 1 <i>gabcdef gabcdef gabcdef gabcdef</i>	Linker	Helical region 2 <i>gabcdef gabcdef gabcdef gabcdef</i>
0	G EIAAIKE EIAAIKE EIAAIKW EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQWQLK EIQEQLK G
2.2	G IKE EAAAIKW EAAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EIQWQAK EIQEQAQ G
4.1	G IKQ QAAAIKW QAAAIKQ G	ASPQPQPKPSGQPQSKQTPQPSRSQ	G QAK QIQWQAK QIQQQAK G
TC-2.2	G IKE EAAAIKW EAAAIKE G	ASPEPQPKPSGCCQTPPEPSRSQ	G QAK EIQWQAK EIQEQAQ G
TC-4.1	G IKQ QAAAIKW QAAAIKQ G	ASPEPQPKPSGCCQTPPEPSRSQ	G QAK QIQWQAK QIQQQAK G

Table 8.2: Sequences of the designed HERD proteins used in Chapter 4.

HERD #	Helical region 1 <i>gabcdef gabcdef gabcdef gabcdef</i>	Linker	Helical region 2 <i>gabcdef gabcdef gabcdef gabcdef</i>
2.2	G IKE EAAAIKW EAAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EIQWQAK EIQEQAQ G
2.2-E	G IKE EAAAIKW EAAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EIQWQAK EIQEQAQ GE

Table 8.3: Sequences of the designed HERD proteins used in Chapter 5.

HERD #	Helical region 1 <i>gabcdef gabcdef gabcdef gabcdef</i>	Linker	Helical region 2 <i>gabcdef gabcdef gabcdef gabcdef</i>
0	G EIAAIKE EIAAIKE EIAAIKW EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQWQLK EIQEQLK G
2.2	G IKE EAAAIKW EAAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G QAK EIQWQAK EIQEQAQ G
5.1	G EIAAIKE EIAAIKE EIAAIKW EIAAIKE EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQWQLK EIQEQLK EIQEQLK
5.2	G EIAAIKE EIAAIKE EIAAIKE EIAAIKW EIAAIKE EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQEQLK EIQWQLK EIQEQLK EIQEQLK G
5.3	G EIAAIKE EIAAIKE EIAAIKE EIAAIKW EIAAIKE EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQEQLK EIQWQLK EIQEQLK EIQEQLK G
5.4	G EIAAIKE EIAAIKE EIAAIKW EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQWQLK EIQEQLK G
5.5	G EIAAIKE EIAAIKE EIAAIKW EIAAIKE G	ASESQKPS	G EIQEQLK EIQEQLK EIQWQLK EIQEQLK G
5.6	G EIAAIKE EIAAIKE EIAAIKE EIAAIKW EIAAIKE EIAAIKE G	ASPEPQPKPSGDPQSKQTPEPSRSQ	G EIQEQLK EIQEQLK EIQEQLK EIQWQLK EIQEQLK EIQEQLK G
5.7	G EIAAIKE EIAAIKE EIAAIKE EIAAIKW EIAAIKE EIAAIKE G	ASESQKPS	G EIQEQLK EIQEQLK EIQEQLK EIQWQLK EIQEQLK EIQEQLK G

Table 8.4: Sequences of the designed HERD proteins used in Chapter 6.

Name	Sequence
mEmerald	MVSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGEGDATYKGLTLKFCITTKGLPVPWPTLVTLLTYGVQCFARYPDHMKQHDFFKSAMPE GYVQERTIFFKDDGNKYTRAEVKFEGDTLVNRIELKIGIDFKEDGNILGHKLEYNYNHSHKVIYITADKQKNGIKVNFKTRHNIEDGVSQVLADH YQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMLLLEFVTAAGITLGMDELYK
mCherry	MVSKGEEDNMAIIKEFMRFKVHMEGVSNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDLILSPQFMVSGSKAYVKHPADIPDYLKLS FPEGFKWERVMNFEDGGVVTVDSSSLQDGEFIYKVKLRGTNFPSDGPMQKKTGMWEASSERMYPEDGALKGEIKQRLKLDKGGHYDAEV KTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIIVEQYERAEGRHSTGGMDELYK
Alcohol dehydrogenase	MKGFAMLSIGKVGWIEKEKPAPGFFDAIVRPLAVAPCTSDIHTVFEGAI GERHNMILGHEAVGEVVEVGEVKDFKPGDRVVVPAITPDWR TSEVQRYHQHSGGMLAGWKFSNVKDGVEFFHVNDADMNLHLPKIPLAAVMIIPDMMTTGFHGAELADIELGATVAVLIGIGPVGLMA VAGAKLRGAGRI IAVGSRPVCVDAAKYYGATDIVNYKDGPIESQIMNLTGKGVDAIIAGGNADIMATAVKIKVPGGTIANVNYFGEVEV LPVPRLEWGGCGMAHKTIKGGLCPGGRLRMRERLIDLVFYKRVDP SKLVTHVFRGFDNIEKAFMLMKDKPKDCLKPVVILA
Glucose-6-phosphatase	MAEQVALSRTQVCGILREELFQGD AFHQSDTHIF IIMGASGDLAKKIIYPTIWWLFRDGLLPENTFIVGYARSRLTVADIRKQSEPPFFKAT PEEKLKLEDDFFARNSYVAGQYDDAASYQRNLNSHMNALHLGSQANRLFYALPPTVYEAVTKNIHESCMSQIGWNRIIVEKPFGRDLQSSDR LSNHISSLFREDQIYRIDHYLGKEMVQNLMLVLRFANRIFGP IWNRDNIACVILTFKEPFGTEGRGGYDFEFGIIRDVMQNHLQLMLCLVAM EKPASTNSDDVRDEKVKVLKCISEVQANNVVLGQYVGNPDGEGEATKGYLDDPTVPRGSTTATFAAVVLYVENERWDGVPFILRCGKALNE RKAEVRLQFHDVAGDIFHQQCKRNELVIRVQNEAVYTKMNTKPGMFFNPESELDLTYGNRYKNVKLPDAYERLILDFVFCGSQMHFVRS DELREAWRIFTPLLHQIELEKPKPIYIYGSRGPTEADELMKRVGFYEGTYKWNPHKL
Tryptophanase	MENFKLLPEFRIRVIEPVKRTTRAYREEAIKSGMNFLLDSEDFIDLITDSGTGAVTQSMQAMMRGDEAYSGRSYYALAESVKNIIF GYQYTIPTHQGRGAEQIYIPVLIKKREQEKGLDRSKMVAFSNYFFDTQGHSQINGCTVRNVIKEAFDTGVRYDFKGNFDLEGLERGIEE VGPNNVPIVATITSNAGGQPVSLANLKAMYSIAKKYDIPVVMDSARFAENAYFIKQREAEYKDWITIEQITRETYKYADMLAMSAKKDAM VPMGGLLCKMKDDSFDDVYTECRTL CVVQEGFPTYGGLEGGAMERLAVGLYDGMNLDWLAIRIAQVYLLVDGLEEIGVVCQAGGHAAFVDA GKLLPHIPADQFPAQALACELYKVAGIRAVEIGSFLLRDPKTKGQLPCPAELLRLTIIPRATYQTTHMDFIIIEAFKHVKENAAAIKGLLFTT YEPKVLRHFTAKLKEV
Flavin containing monoxygenase	MATRIAILGAGPSGMAQLRAFQSAQEKGAEIPELVCFEKQADWGGQWNYTWRTGLDENGEVHSSMYRYLWSNGPKCELEFADYTFDEHFG KPIASYPPREVLWDYIKGRVEKAGVRKYIRFNTAVRHVEFNEDSQFTVTVDHTTDTIYSEEDYVVCCTGHFSTPYVPEFEGFEKFGGR ILHAHDFRDALEFKDKTLLVGVSSYSAEDIGSQCYKYGAKKLI SCYRTAPMGYKWPENWDERPDLVRVDTENAYFADGSSEKVDAILCTG YIHFFPFLNDDLRLLVTNNRLWPLNLYKGVVWEDNPKFFYIGMQDQWYFNMFDQAQAWYARDVIMGRPLP SKKEEMKADSMAWREKELLLVT AEEMYTYQGDYIQNLIDMTDYP SFDIPATNKTFLEWKHHKKENIMTFRDHSYRSLMTGTMAPKHHTPWIDALDDSLAEAYLSDKSEIPVAKE AGS
Fyn tag	MGCVQCKDKEATKLTTEERDGSLNQT
CAAX tag	KGKKKKKKSKTKCVIM

Table 8.5: Additional protein sequences used in this thesis.